Improving Population Attributable Fraction Methods: Examining Smoking-attributable Mortality for 87 Geographic Regions in Canada

Peter Tanuseputro1, Douglas G. Manuel1,2, Susan E. Schultz1, Helen Johansen3, and Cameron A. Mustard2,4

1 Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada.
2 Department of Public Health Sciences, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.
3 Statistics Canada, Ottawa, Ontario, Canada.
4 Institute for Work & Health, Toronto, Ontario, Canada.

Received for publication February 16, 2004; accepted for publication October 26, 2004.

Smoking-attributable mortality (SAM) is the number of deaths in a population caused by smoking. In this study, the authors examined and empirically quantified the effects of methodological problems in the estimation of SAM through population attributable fraction methods. In addition to exploring common concerns regarding generalizability and residual confounding in relative risks, the authors considered errors in measuring estimates of risk exposure prevalence and mortality in target populations and estimates of relative risks from etiologic studies. They also considered errors resulting from combining these three sources of data. By modifying SAM estimates calculated using smoking prevalence obtained from the 2000–2001 Canadian Community Health Survey, a population-based survey of 131,535 Canadian households, the authors observed the following effects of potential errors on estimated national SAM (and the range of effects on 87 regional SAMs): 1) using a slightly biased, mismatched definition of former smoking: 5.3% (range, 1.8% to 11.6%); 2) using age-collapsed prevalence and relative risks: 6.9% (range, 1.1% to 15.5%) and 21.5% (range, 7.9% to 21.0%), respectively; 3) using relative risks derived from the same cohort but with a shorter follow-up period: 8.7% (range, 4.5% to 11.8%); 4) using relative risks for all diseases with age-collapsed prevalence: 49.7% (range, 24.1% to 82.2%); and 5) using prevalence estimates unadjusted for exposure-outcome lag: −14.5% (range, −20.8% to 42.6%) to −1.4% (range, −0.8% to −2.7%), depending on the method of adjustment. Applications of the SAM estimation method should consider these sources of potential error.

bias (epidemiology); effect modifiers (epidemiology); epidemiologic methods; mortality; prevalence; risk; smoking

Abbreviations: AFp, attributable fraction in the population; CCHS, Canadian Community Health Survey; CPS II, Cancer Prevention Study II; SAM, smoking-attributable mortality; SAMMEC, Smoking-attributable Mortality, Morbidity, and Economic Costs.

The World Health Organization reported in 2000 that tobacco use was responsible for the fourth largest global attributable and avoidable burden of disability-adjusted life years (behind underweight, unsafe sex, and suboptimal blood pressure) (1). In developed countries, tobacco ranks first in this respect. Estimates of the number of deaths caused by tobacco use in a population are often derived using population attributable fraction (AFp) methods. AFp, as applied to mortality, estimates the fraction of deaths that can be ascribed to a particular exposure. The product of AFp and the number of deaths in the population yields an attributable mortality count, which for smoking is commonly termed smoking-attributable mortality (SAM).

The general principle of AFp was first discussed by Levin in 1953 (2) and can be described as $AF_p = (I_t - I_u)/I_t$, where $I_t$ is the incidence rate of the outcome in the target population (defined as the population for which AFp is being derived) and $I_u$ is the incidence rate of the outcome in the unexposed population. Since incidence rates for

<table>
<thead>
<tr>
<th>Baseline conditions for SAM* analyses</th>
<th>Methods used to explore the effect of possible sources of error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposure measurement</strong></td>
<td></td>
</tr>
<tr>
<td>1. Prevalences of current, former, and never smoking were obtained from the 2000–2001 CCHS† and were used to calculate SAM using AFp* methods. Current smokers were defined as those reported to have been smoking cigarettes daily or occasionally at the time of the survey. Former smokers were defined in the CCHS as persons reported to have smoked at least 100 cigarettes in their lifetime.</td>
<td>To determine the effect of using a slightly different definition of smoking on SAM, former smokers were redefined as those who had smoked at least one cigarette in their lifetime (a question also asked in the CCHS, and as originally defined in Statistics Canada’s derived variable specifications (34)).</td>
</tr>
<tr>
<td>2. Respondents who did not answer the questions required for classification into a smoking category (did not know, refused, not stated) were excluded from the smoking prevalence estimates.</td>
<td>The possibility of respondent bias was examined by treating all nonrespondents as smokers or as nonsmokers.</td>
</tr>
<tr>
<td><strong>Exposure categorization</strong></td>
<td></td>
</tr>
<tr>
<td>3. Prevalence estimates were categorized into sex-specific 5-year age groups (ages ≥70 years were combined).</td>
<td>To examine the effect of using non-age-specific prevalence estimates (when prevalence customarily varies over different age groups), we used two alternate specifications. First, non-age-specific prevalence estimates for adults aged ≥35 years were used in place of the baseline 5-year age-group estimates. Second, the method of calculating SAM used in the Centers for Disease Control and Prevention’s SAMMEC† computer program (5) has been widely employed (13, 38–41, 46–48). SAMMEC and the 1989 Surgeon General’s report (4) use prevalence data from two age groups, 35–64 years and ≥65 years, while using 5-year age groups (except for ages ≥85 years) for mortality (5). The effect of using these age groupings for prevalence and mortality estimates was also observed.</td>
</tr>
<tr>
<td><strong>Risk and outcome categorization</strong></td>
<td></td>
</tr>
<tr>
<td>4. RR* estimates from CPS II† consisted of a select number of diseases associated with smoking. Disease-group-specific and 5-year age-group-specific (except for ages ≥70 years) mortality counts (1995–1997) were used in the calculation of baseline SAM. Conservatively, deaths due to perinatal conditions, burns, environmental tobacco smoke, and all other causes for persons under 35 years of age were excluded.</td>
<td>SAM and other attributable mortality estimates are occasionally calculated using non-disease-specific, all-cause RR and mortality estimates. To observe the effect of doing so, RR and mortality estimates for all causes from CPS II (1982–1986) and vital statistics (1995–1997) were alternatively used to calculate SAM for the 87 regions (Web appendix table A) (4). This was done in combination with age-specific and non-age-specific prevalence estimates.</td>
</tr>
<tr>
<td>5. RR estimates were sex-specific but not age-specific, except for coronary heart disease and cerebrovascular disease (age groups 35–64 years and ≥65 years).</td>
<td>While the baseline RR estimates were stratified into two age groupings for only two disease groups, 5-year age-group-specific RR estimates (1982–1988) were used for lung cancer, chronic obstructive pulmonary disease, cerebrovascular disease, and coronary heart disease (Web appendix table C) (29).</td>
</tr>
<tr>
<td><strong>Lag time/latency period</strong></td>
<td></td>
</tr>
<tr>
<td>6. Original SAM estimates were not adjusted to take into account the lag time that exists between past prevalence of current smoking and its present effect on mortality. SAM estimates were originally obtained using prevalence estimates from the 2000–2001 CCHS.</td>
<td>The effect of using smoking prevalence estimates calculated from the indirect method was observed (Appendix). The indirect method infers the smoking prevalence that would be necessary to cause the current excess lung cancer mortality in the population as compared with the lung cancer mortality in a nonsmoking population. The use of the indirect method’s prevalence estimates was done while maintaining the diseases/disease groupings of the original AFp method (outlined in Web appendix table A) and without the halving of excess risks (Appendix), the two other proposed changes in the indirect method.</td>
</tr>
</tbody>
</table>

Table continues

unexposed populations are usually unavailable, AFp is commonly derived in terms of the prevalence of risk exposure (Pp) in the population and the relative risk (RR) of the outcome for those exposed:

\[ AF_p = \frac{P_p (RR - 1)}{1 + P_p (RR - 1)} \]

Most AFp and SAM estimates, including those presented in the Surgeon General’s report on smoking in the United States (3, 4) and those calculated using the Centers for Disease Control and Prevention’s Smoking-attributable Mortality, Morbidity, and Economic Costs (SAMMEC) computer program (5), are generated using variations of the latter equation.

A competing method for estimating SAM uses the approach of Peto et al. (6). This “indirect method” infers the prevalence of smoking by observing the excess rate of lung cancer mortality (primarily influenced by smoking) in the
### TABLE 1. Continued

<table>
<thead>
<tr>
<th>Baseline conditions for SAM analyses (possible sources of variation and error)</th>
<th>Methods used to explore the effect of possible sources of error</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Prevalence estimates from the 2000–2001 CCHS were unadjusted for lag time.</td>
<td>CCHS current smoking prevalence estimates were corrected backward in time using smoking data from other past surveys. Two sets of lag times were selected. First, for smoking AFs associated with neoplasms and selected respiratory diseases, this was done by comparing the national CCHS current smoking estimate with the estimate from the 1985 General Social Survey (i.e., a 15- or 16-year lag time to the 2000–2001 CCHS) (42). For cardiovascular disease-associated smoking AFs, data from the 1991 General Social Survey were used (42). Second, AFs associated with neoplasms and selected respiratory diseases were adjusted using the 1991 General Social Survey, while cardiovascular disease-associated smoking AFs were adjusted using the 1996–1997 National Population Health Survey (42).</td>
</tr>
<tr>
<td>8. Disease-specific RR estimates were originally obtained from CPS II (Web appendix table A) (4).</td>
<td>In light of some criticisms regarding the generalizability of CPS II RR estimates to external populations (6, 11, 13), SAM estimates were also generated using the results of a meta-analysis that combined tobacco-associated RR estimates from 10 different studies (Web appendix table D) (Eric Single, University of Toronto, personal communication, 2003) (30).</td>
</tr>
<tr>
<td>9. The CPS II RR estimates used included data from the first 4 years of follow-up (1982–1986) of the cohort and were used in older versions of the SAMMEEC program (49) and in many studies (13, 39, 40).</td>
<td>SAM estimates were also calculated using 6-year follow-up (1982–1988) RR estimates from the CPS II cohort, as opposed to the baseline 4-year follow-up (1982–1986), to examine potential biases resulting from length of follow-up in prospective studies (Web appendix table A) (4, 33, 50).</td>
</tr>
<tr>
<td>Alternate method: Peto et al.’s indirect method for calculating SAM</td>
<td>The effects of individual and different combinations of changes made by the indirect method were explored. First, the effect of substituting the indirect method’s smoking prevalence estimates alone was observed. Second, SAM estimates were calculated using the indirect method (using indirect prevalence estimates and RRs and outcome with the specified disease groupings) but without the halving of excess risks. Third, all three changes were implemented together to obtain the final SAM estimates for the indirect method. Fourth, as a crude method of adjusting for confounding, the effect of halving the excess risks for all disease groupings was then observed alone. Finally, SAM from the indirect method was also calculated using disease groupings outlined in the 1989 Surgeon General’s report (4) (Web appendix table A).</td>
</tr>
<tr>
<td>10. There are three distinct changes in Peto et al.’s (6) indirect method (Appendix). The first substitutes prevalence estimates inferred from excess lung cancer mortality rates for observed prevalence estimates. The second changes the disease and disease groupings considered attributable to smoking. The third halves excess risks (RR - 1) in all SAM calculations for all diseases other than lung cancer (Appendix).</td>
<td></td>
</tr>
</tbody>
</table>

* Corresponding results are shown in tables 3 and 4.
† SAM, smoking-attributable mortality; CCHS, Canadian Community Health Survey; AFs, population attributable fraction; SAMMEEC, Smoking-attributable Mortality, Morbidity, and Economic Costs; RR, relative risk; CPS II, Cancer Prevention Study II.

target population, as compared with an unexposed reference population. In AFp calculations, the indirect method substitutes observed current exposure prevalence estimates with prevalence that is considered necessary for causing the current lung cancer mortality burden. This method provides prevalence and SAM estimates for populations without reliable exposure data (data on lung cancer mortality are often available when data on smoking prevalence are not). In the estimation of AFp and SAM, use of the indirect method’s prevalence estimates also avoids the potential error resulting from the lag time between population changes in smoking prevalence and the resulting change in disease outcome. For most smoking-related outcomes, the current burden of disease is largely influenced by the past smoking exposure in the population. Lag-time error occurs when AFp is calculated with the current prevalence of smoking to estimate the current burden of illness in a population with changing smoking exposure (7–9).

While attributable fraction methods in general (9) and SAM in particular are helpful tools for public health planning (10), evaluation of the impact of different assumptions in their application has been infrequent. Evaluation of SAM estimates has typically been limited to criticism regarding the estimates of relative risks used. For example, following the 1989 Surgeon General’s report (4), one set of criticisms pointed out that the relative risk estimates used, obtained from the American Cancer Society’s Cancer Prevention Study II (CPS II), were adjusted for age but not for possible confounding factors such as alcohol consumption or dietary factors (11–13). When smoking relative risk estimates were adjusted for potential confounding, the effects on total SAM were found to range from...
a 26 percent decrease (11) to a 1 percent decrease (12) to a 2.5 percent increase (13). Another set of criticisms remarked that the CPS II relative risk estimates were not generalizable to the entire population (11, 13). Using an alternate set of relative risk estimates from a more nationally representative US survey yielded SAM estimates 39.5 percent and 16.2 percent lower than those in the Surgeon General’s 1989 report (11, 13). Other issues relating to AF, estimation that have been examined include interpretation of multiple competing risks (14–17), the extrapolation of AF, findings to new populations (18, 19), the theoretical effect of nondifferential exposure and outcome misclassification on the attributable fraction (20, 21), and the use of broad definitions of exposure (22–24).

Beyond these concerns, however, there are a number of issues related to the estimation of AF, that have not been empirically examined. Included are three sources of potential error that result from measuring and analytically combining prevalence and outcome estimates from target populations with relative risks from etiologic studies. The first arises from differences in risk factor exposure measurement in the target population (the source of prevalence estimates) and the etiologic study used (the source of relative risk estimates). The second results from differences in the categorization of risk factor exposure, relative risks associated with the exposure, and the outcome associated with the risk. These include differences in the grouping of diseases attributable to smoking and differences in the age-specificity of the estimates. The third source of potential error relates to temporal issues that arise from uncertainty concerning the lag time between population changes in risk factor exposure and the subsequent changes in disease burden.

In this study, we empirically examined the impact of these potential errors on SAM estimates using data from 87 geographic regions in Canada. The large number of regions examined demonstrates the potential magnitude/importance of these errors in different population settings. We explicitly contrasted conventional AF, estimation methods with those presented by Peto et al. (6), partly to examine potential errors arising from lag-time assumptions. SAM is commonly derived and occasionally validated at the national/local level (3, 11–13, 25, 26); to our knowledge, SAM has never been validated for such a large number of regions. We expect that these findings will aid in improving the accuracy of AF, and attributable mortality estimation techniques for a wide range of exposures and outcomes.

MATERIALS AND METHODS

Data sources

We estimated AF, by combining smoking prevalence estimates from a population health survey with relative risks from an etiologic study. To generate SAM estimates, we then multiplied AF,’s with death counts from vital statistics. We obtained geographic region-, sex-, and age-group-specific prevalence estimates of smoking from cycle 1.1 (2000–2001) of the Canadian Community Health Survey (CCHS), conducted by Statistics Canada. This survey provides cross-sectional estimates of health determinants, health status, and health-system utilization at a subprovincial level (health region or combination of health regions). The target population of the CCHS includes household residents over 12 years of age in all provinces and territories, with the principal exclusion of populations on Indian reserves, Canadian military bases, and some remote areas. There was one randomly selected respondent per household, although planned oversampling of youths resulted in a second member of certain households being interviewed. The 14-month CCHS 1.1 data collection cycle began in September 2000, and the total sample size was 131,535 household respondents, representing a response rate of 84.7 percent.

Annual average geographic region-, sex-, age-group-, and disease-specific death counts were obtained from the Canadian mortality database for 1995–1997, the three most recent years for which data were available from Statistics Canada (27). Disease-specific relative risk estimates for current smokers and former smokers, as compared with never smokers, were obtained from CPS II (4, 28), an ongoing prospective study of 1,185,106 adults (at baseline) over the age of 30 years living in the United States (Web appendix tables A, B, and C) (29). A second set of relative risks from a meta-analysis of 10 different studies was also used for sensitivity analyses (Web appendix table D) (Eric Single, University of Toronto, personal communication, 2003) (30).

Because of concerns about statistical power, prevalence and mortality data for health regions with populations of less than 50,000 were aggregated within each province. We also combined the three northern territories (Nunavut, Yukon, and the Northwest Territories) with Zone 2 in Nova Scotia and Nord-du-Québec in Quebec, the only two health regions within their respective provinces with populations of less than 50,000. This process reduced the number of distinct geographic regions in the analyses from 136 to 87.

Calculation of SAM

Smoking AF, represents the proportion of deaths due to a given disease that has been caused by smoking in a population. Baseline AF, was estimated using the general AF, equation, modified to take into account multiple levels of exposure to smoking (Appendix) (5). While exposure estimates for AF, are often obtained from population health surveys, the indirect method uses the excess lung cancer mortality rate in the target population, as compared with the lung cancer mortality rate of nonsmokers in a reference population, to infer a hypothesized percentage of current smokers in the population. SAM is the sum of the product of AF, and the number of deaths in the population for all diseases considered. Details on both methods can be found in the Appendix.

Sensitivity analyses

Table 1 presents background information on the potential errors we examined and describes the methods we used to estimate their effects. Briefly, we examined the influence of potential error related to risk factor exposure measurement in using alternate definitions of smoking exposure and in
treating nonrespondents in different ways (table 1). We examined the effect of errors related to exposure, outcome, and risk categorization through several different specifications, including varying age aggregations of smoking prevalence and relative risk and varying disease groupings for relative risk and mortality (table 1). We quantified the effects of lag time by examining the effect of 1) substituting prevalence estimates obtained by means of the indirect method for prevalence estimates obtained from the CCHS and 2) adjusting current CCHS smoking prevalence estimates to reflect past prevalence (using historic surveys as a guide). We also examined the effect of using alternate relative risks from a different etiologic study and the effect of the indirect method’s crude adjustment for confounding (halving of excess risks) (table 1). In addition to differences in prevalence estimation and in the halving of excess risks (Appendix), we examined the effect of the indirect method’s use of an alternate set of CPS II (1984–1988) relative risks, which considered all medical causes of death to be partially attributable to smoking (Web appendix table B).

**RESULTS**

**Local SAM estimates**

Table 2 shows estimates of SAM calculated using the set of baseline assumptions outlined in table 1. Under these assumptions, an estimated 40,770 annual deaths, or 273

<table>
<thead>
<tr>
<th>Population count (no.)</th>
<th>Smoking prevalence (%)</th>
<th>Mortality rate (per 100,000)</th>
<th>All-cause mortality count (no.)</th>
<th>Smoking-attributable mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current smoking</td>
<td>Former smoking</td>
<td>Never smoking</td>
<td>Lung cancer mortality</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region with highest value</td>
<td>578,899</td>
<td>43.1</td>
<td>52.8</td>
<td>46.6</td>
</tr>
<tr>
<td>75th percentile</td>
<td>104,774</td>
<td>29.9</td>
<td>43.9</td>
<td>35.3</td>
</tr>
<tr>
<td>Median</td>
<td>52,108</td>
<td>27.6</td>
<td>41.2</td>
<td>31.8</td>
</tr>
<tr>
<td>25th percentile</td>
<td>33,317</td>
<td>25.0</td>
<td>37.7</td>
<td>27.2</td>
</tr>
<tr>
<td>Region with lowest value</td>
<td>24,023</td>
<td>12.9</td>
<td>28.6</td>
<td>17.2</td>
</tr>
<tr>
<td>Range</td>
<td>1,185,919</td>
<td>42.4</td>
<td>26.5</td>
<td>32.7</td>
</tr>
<tr>
<td>Canada</td>
<td>7,160,595</td>
<td>27.0</td>
<td>38.7</td>
<td>34.3</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region with highest value</td>
<td>666,786</td>
<td>32.9</td>
<td>41.1</td>
<td>70.6</td>
</tr>
<tr>
<td>75th percentile</td>
<td>110,004</td>
<td>26.8</td>
<td>31.7</td>
<td>50.4</td>
</tr>
<tr>
<td>Median</td>
<td>55,433</td>
<td>23.9</td>
<td>29.5</td>
<td>46.8</td>
</tr>
<tr>
<td>25th percentile</td>
<td>36,400</td>
<td>20.8</td>
<td>27.2</td>
<td>43.3</td>
</tr>
<tr>
<td>Region with lowest value</td>
<td>24,415</td>
<td>7.5</td>
<td>18.0</td>
<td>34.0</td>
</tr>
<tr>
<td>Range</td>
<td>1,258,980</td>
<td>39.0</td>
<td>23.1</td>
<td>41.6</td>
</tr>
<tr>
<td>Canada</td>
<td>7,747,004</td>
<td>22.4</td>
<td>27.8</td>
<td>49.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region with highest value</td>
<td>1,243,685</td>
<td>37.9</td>
<td>46.4</td>
<td>58.7</td>
</tr>
<tr>
<td>75th percentile</td>
<td>222,117</td>
<td>28.1</td>
<td>37.2</td>
<td>42.7</td>
</tr>
<tr>
<td>Median</td>
<td>107,542</td>
<td>25.8</td>
<td>34.9</td>
<td>39.1</td>
</tr>
<tr>
<td>25th percentile</td>
<td>69,744</td>
<td>23.2</td>
<td>32.5</td>
<td>36.5</td>
</tr>
<tr>
<td>Region with lowest value</td>
<td>50,324</td>
<td>12.7</td>
<td>23.3</td>
<td>26.1</td>
</tr>
<tr>
<td>Range</td>
<td>2,444,871</td>
<td>38.7</td>
<td>23.1</td>
<td>35.9</td>
</tr>
<tr>
<td>Canada</td>
<td>14,907,599</td>
<td>24.6</td>
<td>33.1</td>
<td>42.3</td>
</tr>
</tbody>
</table>

* Health regions with a population of less than 50,000 were aggregated within each province. This decreased the number of geographic regions in the analyses from 136 to 87.

† Rates were age- and sex-standardized to the 1991 Canadian population.

‡ Calculated using the set of baseline assumptions outlined in table 1.
deaths per 100,000 population (aged ≥35 years) per year, could be attributed to smoking (20 percent of all deaths). Canadian men incur approximately twice the SAM as Canadian women (table 2). Age- and sex-standardized rates of SAM (SAM per 100,000 population) varied widely across geographic regions: For example, the two most populated regions—the Montréal-Centre region and the Toronto Public Health Unit—had estimated standardized SAM rates of 332 per 100,000 population and 213 per 100,000 population, respectively (Web appendix table E). Geographically, there was generally a gradient of increasing SAM rates from west to east (31). This follows the west-to-east gradient in life expectancy (32).

**Prevalence of smoking, mortality rate, and SAM across geographic regions**

Examining some of the necessary components for the calculation of SAM, table 2 and Web appendix table E show

---

**TABLE 3. Effect of modifying conditions that influence the estimation of smoking-attributable mortality through the population attributable fraction method for 87 Canadian geographic regions, 2000–2001**

<table>
<thead>
<tr>
<th>Baseline condition</th>
<th>Modified condition</th>
<th>Percentage change in regional SAM estimates</th>
<th>Total SAM (no.)</th>
<th>Change in total SAM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM estimate using all baseline conditions</td>
<td></td>
<td></td>
<td>40,770</td>
<td></td>
</tr>
<tr>
<td><strong>Exposure measurement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Former smokers are defined as having smoked at least 100 cigarettes in a lifetime.</td>
<td>Former smokers are defined as having smoked at least one cigarette in a lifetime.</td>
<td>1.8 3.7 4.8 6.4 11.6 5.2</td>
<td>42,928</td>
<td>5.3</td>
</tr>
<tr>
<td>2. Nonrespondents are not included in numerator or denominator of prevalence.</td>
<td>Nonrespondents are treated as never smokers.</td>
<td>−0.9 −0.3 0.0 0.0 0.0 −0.1</td>
<td>40,717</td>
<td>−0.1</td>
</tr>
<tr>
<td><strong>Exposure categorization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Use of 5-year age-group-specific prevalence estimates</td>
<td>Use age-combined prevalence estimates (except for CHD† and CVD†).</td>
<td>1.1 5.0 6.5 8.2 15.5 6.9</td>
<td>43,579</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td>Use two age groups for prevalence estimates (35–64 years and ≥65 years) instead of 5-year age groupings.</td>
<td>−2.9 1.1 1.9 3.1 9.1 2.2</td>
<td>41,636</td>
<td>2.1</td>
</tr>
<tr>
<td><strong>Risk and outcome categorization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Cause- and sex-specific RR† and mortality estimates; prevalence by geographic region, 5-year age group, and sex</td>
<td>All-cause, sex-specific RR and mortality estimates; prevalence by geographic region and sex</td>
<td>24.1 43.1 48.7 57.0 82.2 49.7</td>
<td>61,050</td>
<td>49.7</td>
</tr>
<tr>
<td></td>
<td>All-cause, sex-specific RR and mortality estimates; prevalence by geographic region, 5-year age group, and sex</td>
<td>16.5 23.7 28.8 33.0 43.9 28.6</td>
<td>52,489</td>
<td>28.8</td>
</tr>
</tbody>
</table>

Table continues
the variation in population counts, prevalences of smoking, and lung cancer and all-cause mortality rates for the 87 Canadian geographic regions used in the analyses. There was more than a twofold difference between the geographic region with the highest value and the geographic region with the lowest value for most of the characteristics examined. The interquartile ranges for the prevalences of current, former, and never smoking were 4.9 percent (23.2–28.1 percent), 4.7 percent (32.5–37.2 percent), and 6.2 percent (36.5–42.7 percent), respectively. The ranges in age- and sex-standardized lung cancer and all-cause mortality rates were strikingly large, although the interquartile ranges were noticeably smaller.

**Sources of potential error and variation in SAM**

Table 3 shows changes in local and national SAM resulting from applying the different methods listed in table 1, which explore possible sources of variation and error.

**Effects of modifying smoking exposure measurement.**

Slightly altering the definition of exposure in the CCHS...
<table>
<thead>
<tr>
<th>Modification of the indirect method</th>
<th>Use RRs(\text{\textdagger}) aggregating original 22 disease groupings from the Surgeon General's report (6) into five disease groupings that include all medical causes of death (Web appendix table B)</th>
<th>Percentage change (\dagger) in regional SAM (\ddagger) estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Halve excess risks ((RR - 1)) in (AF_p) (\dagger) calculations for all diseases except lung cancer (Appendix)</td>
<td>Minimum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-20.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-29.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-29.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2.4</td>
</tr>
</tbody>
</table>

* See text and table 1 for complete description of methods.
\(\dagger\) Change is compared with the SAM estimates obtained from using all of the baseline conditions for the \(AF_p\) method, as outlined in tables 1 and 3 (national SAM estimate of 40,770).
\(\ddagger\) SAM, smoking-attributable mortality; RR, relative risk; \(AF_p\), population attributable fraction.

The authors' findings show the effects of using different combinations of methods in calculating smoking-attributable mortality in Canadian geographic regions. The modifications of Peto et al.'s (6) indirect method were applied to estimate the percentage change in regional SAM. The table highlights the effects of using different combinations of methods, such as using RRs, halving excess risks, and adjusting for lag time. The results indicate that changes in the indirect method can significantly impact the calculation of smoking-attributable mortality, with regional SAM estimates varying from 20.8% to 42.6%.

The authors note that the changes in SAM are compared with the SAM estimates obtained from using all of the baseline conditions for the \(AF_p\) method. The table includes estimates for different scenarios, such as using age-group-specific relative risk and outcome estimates, adjusting for lag time, and modifying exposure categorization. The changes in SAM range from 1.1% to 40.2%, with the largest change observed when using different disease groupings and adjusting for lag time.

The authors conclude that the indirect method can be modified to improve the accuracy of smoking-attributable mortality estimates, with the choice of method having a significant impact on the results. The modifications lead to changes in the national SAM estimate of 40,770, with the largest change observed when using different disease groupings and adjusting for lag time.

These findings are important for public health practitioners and policymakers as they provide insights into the impact of different methods on smoking-attributable mortality estimates, which can inform decisions on tobacco control measures.
decreased SAM by approximately one quarter nationally and by 21.4–29.6 percent for the various geographic regions. The three modifications of the indirect method combined increased SAM for half of the regions (43 of 87), leading to a final SAM national estimate of 41,224, which is close to (1.1 percent higher than) the baseline estimate obtained with the AFₚ technique. Web appendix table F presents a detailed, disease-specific comparison of SAM estimates obtained by means of the two methods.

DISCUSSION

We examined the effects of often-neglected sources of potential error in the calculation of SAM. The results show that a range of factors—including exposure measurement; source of relative risks; length of follow-up in prospective studies used to determine relative risks; categorization of exposure, risk, and outcome; and lag-time effects—may have important effects on local and national SAM estimates. While the interquartile ranges for regional changes in SAM were comparable to the changes observed at the national level, some regions experienced outlying effects. This large variation indicates the complexity of the interaction between variation in regional estimates and sources of potential error. It emphasizes the need to carefully examine assumptions about the use of exposure, risk, and outcome estimates in etiologic studies and target populations for SAM and other attributable mortality calculations.

Exposure measurement

When relative risk estimates are mathematically combined with prevalences of exposure in AFₚ estimation methods, the same definitions of exposure in the target and etiologic populations should be used. The use of slightly differing definitions of smoking led to changes of up to 11.6 percent in local area SAM estimates. The CPS II relative risks used a definition of at least one cigarette per day for at least 1 year for former smokers (33). For former smoking prevalence in the target population, even the use of the alternate CCHS definition of having smoked at least 100 cigarettes in a lifetime (34) would probably have led to overestimation of the AFₚ in the Canadian population, since CPS II required greater exposure, thereby leading to higher relative risk estimates. The manner in which smoking status, especially former smoking, is defined varies considerably between etiologic studies and health surveys (35), and this issue should be carefully examined in the estimation of AFₚ.

Exposure, risk, and outcome categorization

Care should also be taken when deciding on how to categorize exposure groups. SAM estimates differed considerably when age-specific prevalence estimates or age-specific relative risk estimates were used (T3-3 and T3-5). The use of stratum-specific estimates is important, since exposure prevalence and disease effect (relative risk and mortality) often vary across age groups, leading to the possibility of effect modification and bias when these components are combined in attributable mortality calculations. For Canada, the prevalence of smoking and the smoking AFₚ decrease with increasing age. The use of age-pooled prevalence estimates then underestimates smoking AFₚ for younger age groups and overestimates it for older age groups. Since the number of deaths is consistently greater for older age groups, the degree of overestimation in SAM (the product of AFₚ and death count) for older persons will be greater in absolute terms than the degree of underestimation for younger persons. This explains the 6.9 percent inflation in national SAM (T3-3). Because these trends exist within the age groups 35–64 years and ≥65 years, the use of these two collapsed age groupings for prevalence estimates (as replicated from SAMMEC) still led to overestimation of SAM.

This bias can similarly be observed when relative risk is modified by age and non-age-specific relative risks are used. Further supporting the use of age-specific estimates, non-age-specific relative risks such as those calculated in CPS II are often age-standardized to a particular standard population, which may not be representative of the target population. When stable age-group-specific prevalence/relative risk estimates are available and age-group-specific mortality data (more readily available) are being used, age-group-specific AFₚ and SAM values should be calculated.

Furthermore, the use of relative risks and deaths for all diseases substantially increased SAM estimates (T3-4 and use of the indirect method’s disease groupings). Investigators in previous studies have argued that calculating SAM using a select number of disease-specific relative risks and outcomes may result in underestimation of the true burden on mortality (36, 37). However, other investigators have argued that since unadjusted relative risk estimates are prone to the observation of spurious associations in the presence of confounding variables, inclusion of diseases with weak evidence of a causal relation with smoking would lead to overestimation of SAM (36).

Lag-time issues

The indirect method’s prevalence estimation technique may be useful for populations for which prevalence estimates are unavailable (6). Furthermore, the indirect method’s use of current lung cancer mortality rates in place of current smoking prevalence estimates avoids error from the lag time between population changes in the prevalence of exposure and the resulting change in mortality. Adjustment for this error is rarely carried out (4, 11–13, 30, 38–41). Lag-time error is largest when there is a long latency period between exposure and disease (true of smoking and many of its associated diseases) coinciding with rapidly changing prevalence of exposure in the population. The prevalence of current smoking has decreased by more than 10 percent in Canada since 1985 (42), but the full resulting effect on mortality has not yet been observed. Using the indirect method to obtain prevalence estimates that were based on current mortality rates would thus be expected to lead to a greater SAM estimate than would be obtained from using CCHS smoking prevalence estimates. As expected, the indirect method estimated SAM that was...
14.5 percent greater than estimates based on current smoking prevalence estimates.

Indirect prevalence estimates can be used independently of other changes made with the indirect method. One must exercise interpretative caution, however, since applying the indirect prevalence estimates for other diseases (done in the indirect method) would apply to them the same lag time as lung cancer in the calculation of SAM. As an alternative, current SAM estimates can be adjusted for changes in exposure through the use of past smoking prevalence estimates, when available. This method has the advantage of being able to apply different latency periods for different diseases. Doing so led to an increase in estimated SAM that was smaller than that observed from using prevalence data from the indirect method. This difference may result from the indirect method’s adjustment for underreporting of smoking in health surveys (4, 43–45) and from the application of the relatively long lag time of lung cancer to other diseases considered. Since current regional estimates of lung cancer mortality were available while past regional prevalence estimates were not, only the indirect method allowed for bidirectional adjustment for lag time in different regions (the direction of trend in the prevalence of smoking may differ among regions).

Addressing confounding

When SAM is calculated using unadjusted relative risks, the indirect method’s halving of excess risks is a crude yet conservative method of adjusting for confounding. While it has been argued that the halving of excess risk does not lead to a guarantee of conservatism (25), our results show that the magnitude of the reduction in SAM estimates for all geographic regions (from 21.4 percent to 29.6 percent) was similar to or greater than the magnitudes of reductions observed in previous attempts to control for confounding (11–13).

Conclusions

Concerns regarding relative risk generalizability should be extended in $A\text{F}_p$ calculations for consideration of how relative risks from an etiologic study and prevalences of exposure and outcome (i.e., death) from a target population are defined, categorized, and analytically combined. For the reasons discussed above, in the calculation of SAM estimates we recommend (when possible) the use of closely matching definitions of exposure in the study and target populations; the use of age- and sex-specific prevalence, mortality, and relative risk estimates; and (to be conservative) the use of a select number of causally linked, disease-specific relative risk estimates. Consideration of lag-time effects will also be important if there is a long latency period between exposure and disease occurrence combined with changing prevalence in the population over time. Attention to the often-neglected issues examined in this paper will lead to $A\text{F}_p$ estimates with greater validity or will, at least, inform authors and readers about the magnitude of potential biases when adjustment is not performed.

ACKNOWLEDGMENTS

Support for this project was received from the Canadian Population Health Institute.
Dr. Douglas G. Manuel is a Career Scientist with the Ontario Ministry of Health and Long-Term Care.

REFERENCES


**APPENDIX**

**Population attributable fraction method**

Using prevalence data from the Canadian Community Health Survey and relative risk (RR) estimates from a separate study, the population attributable fraction (AFp) for smoking is

\[ AF_p = \frac{[p_0 + p_1 \times RR_1 + p_2 \times RR_2] - 1}{[p_0 + p_1 \times RR_1 + p_2 \times RR_2]} \]

where \( p_0, p_1, \) and \( p_2 \) represent the percentage prevalence of never, current, and former smoking, while \( RR_1 \) and \( RR_2 \)
represent the relative risk of death due to a given disease for current and former smokers, respectively, with never smokers serving as the reference group.

The calculation of smoking attributable mortality (SAM) involves taking the product of smoking AF, and mortality count (\(D\)). For each geographic region, the SAM estimate for each disease group, \(k\), for which relative risk estimates (RR1’s and RR2’s) were obtained (Web appendix tables A–D) is

\[
SAM_k = \sum_{j=1}^{2} \sum_{i=1}^{8} AF_{pji} \times D_{ji},
\]

where \(i = 1, 2, \ldots 8\) represent 5-year age groups (except for \(\geq 70\)) and \(j = 1, 2\) represents males and females. The total SAM (total number of deaths) for each geographic region is simply the sum of SAM from all \(k\) disease groups (i.e., \(\sum SAM_k\)).

### Indirect method

The indirect method of Peto et al. (6) can be broken down into three steps:

1. For each of the 87 distinct geographic regions and for each sex \(j\), we calculated a hypothesized prevalence of current smoking, \(F\), for six age groups (35–59, 60–64, 65–69, 70–74, 75–79, and \(\geq 80\) years), by assuming that the lung cancer mortality rate, \(I_i\), in each geographic region/age group is composed of the lung cancer mortality of smokers, \(I_e\), and nonsmokers, \(I_u\) (solve for \(F\)):

\[
I_i = FI_e + (1 - F)(I_u).
\]

The lung cancer mortality rate in the population, \(I_i\), is obtained from the Canadian mortality database for each geographic region, while \(I_e\) and \(I_u\) are obtained from Cancer Prevention Study II data. Former smokers are ignored.

2. The calculated prevalence of current smoking (\(F\)) is then used to calculate the smoking AF, using AF techniques. For lung cancer, \(AF_{pji} = F_{ji}(RR_l - 1)/[F_{ji}(RR_l - 1) + 1]\), where \(RR_l = \) relative risk of lung cancer mortality for current smokers (Web appendix table B). For other disease groups, \(k\), the excess risk associated with current smoking (excess risk = RR - 1) is halved to compensate for possible confounding resulting from the use of unadjusted relative risk estimates from Cancer Prevention Study II. Thus, for other disease groups (Web appendix table B),

\[
AF_{pji} = F_{ji}(RR_k - 1)/[F_{ji}(RR_k - 1) + 2].
\]

Note that the indirect method takes into account somewhat different diseases and disease groupings from those presented and used in the Surgeon General’s reports (Web appendix table A) (4).

3. SAM counts for all disease groups (\(SAM_k\)) are calculated from equation 2 of the AF, method (see above), using the six new age groupings (\(i\)).