Methodological Issues in Estimating Smoking-attributable Mortality in the United States

Ann M. Malarcher, Jane Schulman, Leonardo A. Epstein, Michael J. Thun, Paul Mowery, Ben Pierce, Luis Escobedo, and Gary A. Giovino

The authors explored two methodological issues in the estimation of smoking-attributable mortality for the United States. First, age-specific and age-adjusted relative risk, attributable fraction, and smoking-attributable mortality estimates obtained using data from the American Cancer Society’s second Cancer Prevention Study (CPS II), a cohort study of 1.2 million participants (1982–1988), were compared with those obtained using a combination of data from the National Mortality Follow-back Survey (NMFS), a representative sample of US decedents in which information was collected from informants (1986), and the National Health Interview Survey (NHIS), a nationally representative household survey (1987). Second, the potential for residual confounding of the disease-specific age-adjusted smoking-attributable mortality estimates was addressed with a model-based approach. The estimated smoking-attributable mortality based on the CPS II for the four most common smoking-related diseases—lung cancer, chronic obstructive pulmonary disease, coronary heart disease, and cerebrovascular disease—was 19% larger than the estimated smoking-attributable mortality based on the NMFS/NHIS, yet the two data sources yielded essentially the same smoking-attributable mortality estimate for lung cancer alone. Further adjustment of smoking-attributable mortality for disease-appropriate confounding factors (education, alcohol intake, hypertension status, and diabetes status) indicated little residual confounding once age was taken into account. Am J Epidemiol 2000;152:573–584.

Received for publication March 29, 1999, and accepted for publication October 11, 1999.

Abbreviations: CPS II, Cancer Prevention Study II; ICD-9, International Classification of Diseases, Ninth Revision; NHIS, National Health Interview Survey; NMFS, National Mortality Follow-back Survey.

1 Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, GA.
2 Battelle Memorial Institute, Centers for Public Health Research and Evaluation, Baltimore, MD.
3 American Cancer Society, Atlanta, GA.

Reprint requests to Dr. Ann M. Malarcher, Office on Smoking and Health, Centers for Disease Control and Prevention, 4770 Buford Highway NE, Mailstop K-58, Atlanta, GA 30341 (e-mail: aym8@cdc.gov).
different data sources for estimating age-adjusted relative risks, attributable fractions, and smoking-attributable mortality. In particular, we compared age-specific and age-adjusted relative risks estimated using data from the CPS II with those estimated using a combination of data from the NMFS and the NHIS. In contrast to Sterling et al., we directly standardized all age-specific relative risks to the same reference population. Second, we assessed the potential residual confounding of the disease-specific age-adjusted attributable fractions and smoking-attributable mortality estimates. Many researchers have been interested in estimating adjusted attributable fractions by using either case-control data or cohort data (7–16). In our application, we adapted the method of Bruzzi et al. (17) for use with case-control data to estimate the adjusted attributable fraction when combining information from one cross-sectional national survey and one cohort study.

MATERIALS AND METHODS

Data sources

We used four data sources to estimate smoking prevalences, relative risks, attributable fractions, and smoking-attributable mortality: US mortality data from the National Center for Health Statistics and data from the CPS II, the NMFS, and the NHIS. From the National Center for Health Statistics’ US mortality data, as reported by the states from death certificates (18), we obtained the total number of deaths for each smoking-related disease. The American Cancer Society’s CPS II is a cohort study of approximately 1.2 million participants that began in 1982 (19). American Cancer Society volunteers identified potential participants—persons whom they would be able to maintain contact with over the course of the study, such as neighbors and friends. CPS II participants reside in all 50 states, the District of Columbia, Puerto Rico, and Guam. Smoking status and age were determined at baseline from a self-administered questionnaire on medical history, current health status, and lifestyle factors. The volunteers reported vital status for participants every year. For participants who died during follow-up, the underlying cause of death (as indicated by the International Classification of Diseases, Ninth Revision (ICD-9) (20)) was obtained from the death certificate. After 6 years of follow-up, death certificates had been obtained for 94.1 percent of subjects known to have died. The 1986 NMFS used a stratified sample of death certificates in the Current Mortality Sample (21), which was a 10 percent sample of death certificates of persons aged 25 years or older. To ensure sufficient sample sizes for subpopulations, certain age and racial/ethnic groups and causes of death were oversampled or selected with certainty. In all, 18,733 death certificates were sampled, of which 2,274 were sampled with certainty. Proxy information on decedents’ health behaviors was collected from informants (usually next of kin) and from facility abstract questionnaires (i.e., nursing homes). Information was obtained using a combination of self-administered questionnaires and telephone or field interviews. The response rate for the 1986 NMFS was 88.6 percent.

The NHIS is a national, multistage cluster survey that has been conducted continuously since 1957 (22). The target population is the civilian, noninstitutionalized population of the United States aged 18 years or older. Questions about tobacco smoking frequently appear on supplements to the NHIS. The 1987 NHIS supplement included tobacco questions as part of a special section on cancer control. Although proxy responses to the core NHIS are permitted, supplemental interviews must be completed by the sampled respondents. A total of 47,240 interviews were completed in 1987.

Comparison of CPS II and NMFS/NHIS estimates

We restricted our comparison of age-adjusted estimates obtained from CPS II data and the NMFS/NHIS data to the four most common smoking-related diseases: lung cancer (ICD-9 code 162), chronic obstructive pulmonary disease (ICD-9 codes 490–492 and 496), coronary heart disease (ICD-9 codes 410–414), and cerebrovascular disease (ICD-9 codes 430–438). We used four age groups (35–59, 60–69, 70–79, and ≥80 years) when we analyzed lung cancer, coronary heart disease, and cerebrovascular disease; we used only two age groups (50–69 and ≥70 years) when we analyzed chronic obstructive pulmonary disease, because there are so few deaths below age 50 years for this cause of death. All estimates were calculated separately by gender. Because 93 percent of the CPS II respondents are White, we limited our study to Whites. Subjects were classified as current smokers (persons who reported that they smoked now), former smokers (persons who reported that they had ever smoked but did not smoke now), or never smokers (persons who reported that they had never smoked).

We used standard methods to estimate age-, gender-, and cause-specific relative risks and attributable fractions, age-adjusted relative risks, and age-adjusted attributable fractions and smoking-attributable mortality (see Appendix A). Briefly, age-specific relative risks for current or former smokers compared with never smokers for each disease/gender group were directly age-adjusted by taking a weighted sum of the age-specific relative risks (using the proportion of the 1986 US population in each age group as the weights). Age-specific attributable fractions for each disease/gender group were estimated by combining age-specific relative risks and smoking prevalence estimates for never, former, and current smokers. Age-adjusted attributable fractions were calculated as a weighted sum of the age-specific attributable fractions (using the proportion of deaths in the United States in 1986 in each age group as the weights). For each disease/gender group, we estimated age-adjusted smoking-attributable mortality by multiplying the age-specific attributable fractions by the number of deaths within the age group and then summing across age groups.

Comparison of age-adjusted and fully adjusted estimates

To examine the potential for residual confounding of the age-adjusted estimates, we compared age-adjusted estimates...
and estimates adjusted for multiple confounding factors (age, education, alcohol use, hypertension status, and diabetes status). As before, we restricted the analysis to the four most common smoking-related diseases: lung cancer, chronic obstructive pulmonary disease, coronary heart disease, and cerebrovascular disease. The results are presented separately by gender, and the analysis was limited to Whites. Subjects were classified as current, former, or never smokers.

We varied the set of potential confounders by disease. Age (<59, 60–64, 65–69, 70–75, 76–79, 80–84, and ≥85 years) and education (<12 years, 12 years, and >12 years) were included for all diseases. Alcohol use (nondrinking, and <1, 1, and ≥2 drinks per day) was also included for chronic obstructive pulmonary disease, coronary heart disease, and cerebrovascular disease. History of hypertension (yes/no) and diabetes mellitus (yes/no) was also included for coronary heart disease and cerebrovascular disease.

The method used for adjusting the attributable fraction was an adaptation of that used by Bruzzi et al. (17). Details on the method are presented in Appendix B. Briefly, the adjusted attributable fraction (AF) is given by the formula

\[ AF_c = 1 - \sum_j \frac{\rho_j}{R_j}, \]

where \( \rho_j \) is the proportion of deaths in the \( j \)th cell defined by exposure and confounder status (e.g., smoking \( \times \) age) and \( R_j \) is the relative risk for smokers compared with nonsmokers, adjusted for confounder(s) \( C \) (e.g., age). We used data from the NMFS to estimate the values of \( \rho_j \), accounting for the design of the survey. The \( R_j \) values were estimated using Poisson regression with the CPS II data. The purpose of the model in this instance was to pool information from different cells to estimate adjusted relative risks while preserving, to the extent possible, the structure of the data. Our goal was to define disease- and gender-specific models with predicted cell counts similar to the observed cell counts. The final model for each disease/gender group included all main effects and all significant two- and three-factor interactions. The final models also included all relevant two-factor interaction terms needed to maintain a hierarchical model structure (23).

We calculated the confidence intervals for the adjusted attributable fraction and smoking-attributable mortality estimates using bootstrap methods (24), because traditional methods for estimating standard errors are based on asymptotic results, the validity of which may not hold with sparse data and with models containing many parameters. To generate bootstrap estimates for the adjusted attributable fraction, we combined bootstrap estimates of the relative risks computed from the CPS II data with bootstrap estimates of the proportions computed from the NMFS data. Details on the bootstrap calculation are presented in Appendix C.

RESULTS

Comparison of CPS II and NMFS/NHIS estimates

Patterns in the age-specific relative risks estimated using the CPS II data were similar to those estimated using a combination of data from the NMFS and the NHIS in that 1) relative risks for current smokers versus never smokers were larger than relative risks for former smokers versus never smokers; 2) relative risks tended to decrease with increasing age; and 3) relative risks for lung cancer and chronic obstructive pulmonary disease were generally larger than the corresponding relative risks for coronary heart disease and cerebrovascular disease (tables 1 and 2).

Among men, the age-specific relative risks from the CPS II were generally (up to 3.2 times) larger than the corresponding NMFS/NHIS relative risks, except for lung cancer, for which NMFS/NHIS estimates were 1.4–3.0 times larger than CPS II estimates among persons aged 35–59 years and persons aged 80 years or older (table 1). For coronary heart disease and cerebrovascular disease among men aged 60 years or older, the relative risks from the CPS II indicated excess risk among former smokers, whereas the NMFS/NHIS relative risks for former smokers indicated a protective effect. The relative risk estimates from the NMFS/NHIS data were considerably more variable than the CPS II estimates. Among women, the differences between the age-specific relative risks estimated from the CPS II and those estimated from the NMFS/NHIS were smaller (table 2). The pattern for lung cancer was similar to that of men, and the NMFS/NHIS relative risk estimates were up to 3.7 times larger than CPS II estimates among women aged 35–59 years and women aged 70 years or older.

For both genders, age-adjusted relative risks for lung cancer were larger when estimated using the NMFS/NHIS data than when estimated using the CPS II data, and the age-adjusted relative risks for cerebrovascular disease were slightly higher in the CPS II than in the NMFS/NHIS (tables 1 and 2). The age-adjusted relative risks for chronic obstructive pulmonary disease and coronary heart disease based on the CPS II were larger than those based on the NMFS/NHIS for men but not for women.

The total age-adjusted smoking-attributable mortality for the four diseases combined was 36 percent higher for men and 8 percent lower for women when it was estimated using CPS II data rather than NMFS/NHIS data (table 3). For men, age-adjusted attributable fraction and smoking-attributable mortality estimates based on CPS II data were consistently larger for all four diseases than they were for estimates based on NMFS/NHIS data. For women, however, estimates based on the two data sources were quite comparable for chronic obstructive pulmonary disease, coronary heart disease, and cerebrovascular disease, whereas the age-adjusted attributable fraction and smoking-attributable mortality for lung cancer based on CPS II data were approximately 11 percent lower than the age-adjusted attributable fraction and smoking-attributable mortality based on NMFS/NHIS data.

For both genders combined, the CPS II smoking-attributable mortality estimates were larger than the NMFS/NHIS estimates for three of the four diseases: They were 13 percent higher for chronic obstructive pulmonary disease, 54 percent higher for coronary heart disease, and 31 percent higher for cerebrovascular disease. The smoking-attributable mortality estimates for lung cancer differed by only 2.6 percent. Overall (for both genders and all four diseases combined), the
estimated smoking-attributable mortality based on CPS II data was 19 percent larger than that based on NMFS/NHIS data.

Comparison of age-adjusted and fully adjusted estimates from the CPS II data

The age-adjusted attributable fractions were within 10 percent of the fully adjusted attributable fractions for six of the eight disease- and gender-specific comparisons (tables 3 and 4); four of the eight age-adjusted attributable fractions were within 5 percent of the fully adjusted attributable fractions. The most meaningful difference was for males who died from cerebrovascular disease, for whom the age-adjusted attributable fraction was 60 percent larger than the fully adjusted attributable fraction (0.16 vs. 0.10).

For all four diseases and both genders combined, the age-adjusted smoking-attributable mortality (236,942) was 2.5 percent smaller than the fully adjusted smoking-attributable mortality (243,047). For men, for all four diseases combined, the age-adjusted smoking-attributable mortality almost equaled the fully adjusted smoking-attributable mortality (167,468 vs. 167,504). For women, for all four diseases combined, the age-adjusted smoking-attributable mortality was 8.0 percent smaller than the fully adjusted smoking-attributable mortality (69,474 vs. 75,543).

For both genders combined, the fully adjusted smoking-attributable mortality was almost equal to the age-adjusted smoking-attributable mortality for lung cancer (94,670 vs. 93,387) and chronic obstructive pulmonary disease (53,587 vs. 52,821). The fully adjusted smoking-attributable mortality was 10.0 percent larger than the age-adjusted smoking-attributable mortality for coronary heart disease (82,666 vs. 75,092) and 22.5 percent smaller than that for cerebrovascular disease (12,124 vs. 15,642).

DISCUSSION

The number of deaths attributable to cigarette smoking in the United States has been estimated by the Surgeon General (1), by Sterling and colleagues (4, 5), and by Peto et al. (25).
TABLE 2. Age-specific and age-adjusted relative risks of mortality for current and former smokers versus never smokers (US White women)*

<table>
<thead>
<tr>
<th>Cause of death and age (years)</th>
<th>NMFS†/NHIS†</th>
<th>CPS II†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current smokers vs. never smokers</td>
<td>Former smokers vs. never smokers</td>
</tr>
<tr>
<td></td>
<td>RR† 95% CI†</td>
<td>RR 95% CI</td>
</tr>
<tr>
<td>Lung cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35–59</td>
<td>32.13, 7.56, 136.55</td>
<td>12.77, 2.71, 60.05</td>
</tr>
<tr>
<td>60–69</td>
<td>11.22, 4.64, 27.13</td>
<td>5.65, 2.13, 15.01</td>
</tr>
<tr>
<td>70–79</td>
<td>21.70, 9.50, 49.54</td>
<td>7.50, 3.09, 18.21</td>
</tr>
<tr>
<td>≥80</td>
<td>27.19, 11.14, 66.36</td>
<td>3.68, 1.28, 10.53</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–69</td>
<td>15.22, 3.17, 73.12</td>
<td>13.32, 2.62, 67.60</td>
</tr>
<tr>
<td>≥70</td>
<td>8.80, 4.96, 15.61</td>
<td>3.75, 2.03, 6.91</td>
</tr>
<tr>
<td>Age-adjusted</td>
<td>12.51, 4.47, 35.02</td>
<td>8.46, 2.91, 24.59</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35–59</td>
<td>4.27, 2.75, 6.62</td>
<td>1.68, 0.84, 3.35</td>
</tr>
<tr>
<td>60–69</td>
<td>1.89, 1.26, 2.85</td>
<td>1.28, 0.79, 2.08</td>
</tr>
<tr>
<td>70–79</td>
<td>1.50, 1.05, 2.14</td>
<td>1.29, 0.94, 1.78</td>
</tr>
<tr>
<td>≥80</td>
<td>1.62, 1.01, 2.61</td>
<td>0.74, 0.50, 1.09</td>
</tr>
<tr>
<td>Age-adjusted</td>
<td>2.91, 2.21, 3.83</td>
<td>1.44, 0.95, 2.19</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35–59</td>
<td>3.95, 1.70, 9.18</td>
<td>1.12, 0.26, 4.78</td>
</tr>
<tr>
<td>60–69</td>
<td>2.43, 1.17, 5.26</td>
<td>1.00, 0.35, 2.84</td>
</tr>
<tr>
<td>70–79</td>
<td>2.29, 1.34, 3.91</td>
<td>0.94, 0.50, 1.76</td>
</tr>
<tr>
<td>≥80</td>
<td>1.60, 0.86, 3.00</td>
<td>0.69, 0.41, 1.17</td>
</tr>
<tr>
<td>Age-adjusted</td>
<td>3.11, 1.85, 5.24</td>
<td>1.03, 0.43, 2.47</td>
</tr>
</tbody>
</table>

* Data were obtained from the National Mortality Follow-back Survey (1986), the National Health Interview Survey (1987), and the American Cancer Society’s second Cancer Prevention Study (1982–1988).
† NMFS, National Mortality Follow-back Survey; NHIS, National Health Interview Survey; CPS II, Cancer Prevention Study II; RR, relative risk; CI, confidence interval.

In contrast to the Surgeon General and Sterling and colleagues, Peto et al. estimated smoking-attributable mortality indirectly using vital statistics data. Sterling et al. (4) found that the disease-specific, age-adjusted relative risks calculated from NMFS/NHIS data were considerably smaller than those presented in the Surgeon General’s report (1). We found that the age-adjusted relative risks estimated from the two data sources for coronary heart disease and cerebrovascular disease did not differ substantially. The age-adjusted relative risks for lung cancer estimated using NMFS/NHIS data were 1.3–2.0 times larger than the relative risks estimated using the CPS II data (for both current and former smokers). This pattern applied to African Americans as well (data not shown). We also observed that the age-adjusted relative risks for chronic obstructive pulmonary disease based on the CPS II data were approximately 1.7 times larger than those based on the NMFS/NHIS data for men but were about the same for women.

Differences in methodology may explain why our findings concerning the age-adjusted relative risks differed from Sterling et al.’s (4). We directly adjusted the NMFS/NHIS data for age, whereas Sterling et al. (4) used indirect adjustment. The method of age adjustment did not appear to affect the magnitude of the age-adjusted relative risks from the NMFS/NHIS data for current and former smoking with regard to coronary heart disease, cerebrovascular disease, and chronic obstructive pulmonary disease (or with regard to lung cancer for former smoking). However, our age-adjusted relative risks from the NMFS/NHIS data for current smoking and lung cancer were 2–3 times larger than Sterling et al.’s (4) (our relative risks were also larger for African Americans (data not shown)). We also used the same reference population for direct standardization of both the NMFS/NHIS and the CPS II relative risks, whereas different populations were used for standardization in Sterling et al.’s (4) comparison of relative risks from the NMFS/NHIS and the Surgeon General’s report.

We noted that the NMFS/NHIS data produced several age-specific relative risk estimates in former smokers that were less than 1.0 for coronary heart disease and cerebrovascular disease, particularly among older men. Sterling et al. (4) also reported this pattern. This pattern is probably
related to 1) limited sample sizes in specific cells (e.g., the confidence intervals for cerebrovascular disease for men included 1.0) and 2) the fact that information about deceased’s smoking habits was obtained from proxy respondents for the NMFS (several studies have shown that proxy respondents tend to underreport smoking history). There is no biologically plausible explanation for smoking’s being protective against coronary heart disease or cerebrovascular disease, or indeed against lung cancer or chronic obstructive pulmonary disease.

Similar to other cohort studies (19, 30), we observed lower age-adjusted relative risks for lung cancer among female smokers than among male smokers in the CPS II and NMFS/NHIS data. These differences largely reflect gender differences in patterns of smoking across birth cohorts. Women began smoking in the 1930s and 1940s, approximately 20–30 years after men, which resulted in a lower cumulative smoking exposure. In addition to starting smoking at a later age, women in the CPS II reported smoking fewer cigarettes per day and were less likely to inhale

### TABLE 3. Age-adjusted attributable fractions (AFs) and smoking-attributable mortality (SAM) among US White men and women aged ≥35 years*

<table>
<thead>
<tr>
<th></th>
<th>White men</th>
<th>White women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NMFS†/NHIS†</td>
<td>CPS II†</td>
</tr>
<tr>
<td>Lung cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.83 (0.77, 0.90)†</td>
<td>0.91 (0.89, 0.92)</td>
</tr>
<tr>
<td>SAM</td>
<td>61,952 (57,029, 66,876)</td>
<td>67,497 (66,160, 68,834)</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.72 (0.62, 0.82)</td>
<td>0.85 (0.82, 0.88)</td>
</tr>
<tr>
<td>SAM</td>
<td>29,993 (25,800, 34,186)</td>
<td>35,256 (33,819, 36,693)</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.11 (0.04, 0.17)</td>
<td>0.23 (0.20, 0.26)</td>
</tr>
<tr>
<td>SAM</td>
<td>27,215 (10,751, 43,679)</td>
<td>56,852 (49,292, 64,412)</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.08 (0.00, 0.21)</td>
<td>0.16 (0.09, 0.22)</td>
</tr>
<tr>
<td>SAM</td>
<td>3,936 (0.10, 300)</td>
<td>7,863 (4,464, 11,262)</td>
</tr>
<tr>
<td>Total SAM</td>
<td>123,098 (104,298, 141,895)</td>
<td>167,468 (158,949, 175,966)</td>
</tr>
</tbody>
</table>

* Data were obtained from the National Mortality Follow-back Survey (1986), the National Health Interview Survey (1987), and the American Cancer Society’s second Cancer Prevention Study (1982–1988).
† NMFS, National Mortality Follow-back Survey; NHIS, National Health Interview Survey; CPS II, Cancer Prevention Study II.
‡ Numbers in parentheses, 95% confidence interval.

### TABLE 4. Fully adjusted* attributable fractions (AFs) and smoking-attributable mortality (SAM), by gender and disease, in the American Cancer Society’s second Cancer Prevention Study (1982–1988)

<table>
<thead>
<tr>
<th></th>
<th>White men</th>
<th>White women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.89 (0.87, 0.91)†</td>
<td>0.77 (0.71, 0.80)</td>
</tr>
<tr>
<td>SAM</td>
<td>66,541 (64,423, 67,984)</td>
<td>28,129 (25,994, 29,247)</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.88 (0.83, 0.92)</td>
<td>0.68 (0.61, 0.76)</td>
</tr>
<tr>
<td>SAM</td>
<td>36,342 (34,483, 38,048)</td>
<td>17,245 (15,259, 19,235)</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.24 (0.21, 0.27)</td>
<td>0.10 (0.06, 0.13)</td>
</tr>
<tr>
<td>SAM</td>
<td>59,683 (52,844, 66,789)</td>
<td>22,983 (19,927, 26,778)</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.10 (0.02, 0.19)</td>
<td>0.09 (0.05, 0.12)</td>
</tr>
<tr>
<td>SAM</td>
<td>4,938 (1,050, 9,719)</td>
<td>7,186 (3,631, 9,832)</td>
</tr>
</tbody>
</table>

* Fully adjusted attributable fractions were based on models that included all main effects and significant two- and three-factor interactions. The main effects for each disease were education and age for lung cancer; education, alcohol use, and age for chronic obstructive pulmonary disease; and education, age, alcohol use, hypertension status, and diabetes status for both coronary heart disease and cerebrovascular disease.
† Numbers in parentheses, 95% confidence interval.
Smoking-attributable Mortality in the United States

Am J Epidemiol Vol. 152, No. 6, 2000

deeply than men (19). Lung cancer risk increases with amount smoked, duration of smoking, and depth of inhalation in both men and women (30). However, the CPS II relative risks for lung cancer were still higher among men than among women within strata of 5-year duration of smoking (30–50 years) and cigarettes per day (20 vs. 40) (19). These differences may have been due to remaining gender differences in duration and amount within these broad strata. In contrast to cohort studies, some recent case-control studies have found a comparable risk or even a higher risk for female smokers than for male smokers (32–36). Possible explanations include a lower risk of lung cancer among nonsmoking women (perhaps due to occupational exposures that increase the nonsmoker risk in men), differences in the reporting of exposure information, differences in the dynamics of cigarette smoking, increased female susceptibility to carcinogens in tobacco smoke, and increased exposure levels among women (32, 36, 37).

Overall (for both genders and for all four diseases combined), we found that the age-adjusted smoking-attributable mortality estimate using CPS II data was 19 percent larger than the estimate using NMFS/NHIS data. Sterling et al. (4), however, found the CPS II-based estimate from the Surgeon General’s report to be approximately 40 percent larger than the NMFS/NHIS-based estimate. The lower percentage we calculated may reflect the fact that the NMFS/NHIS estimates in our analysis were not uniformly smaller than the CPS II estimates. For example, the CPS II-based estimated smoking-attributable mortality for the four diseases combined was 36 percent larger for men but 8 percent smaller for women. In addition, the two data sources yielded essentially the same smoking-attributable mortality estimates for lung cancer, which accounts for approximately 40 percent of the total smoking-attributable mortality for all four diseases combined.

Two reasons why our results concerning smoking-attributable mortality differ from Sterling et al.’s (4) are that 1) we used similar methods to calculate the relative risks, attributable fractions, and smoking-attributable mortalities from the NMFS/NHIS and the CPS II and 2) we substituted the point estimate of 1.0 for the age-, sex-, and cause-specific relative risks that were less than 1.0 (thereby assuming that smoking did not cause or prevent deaths for these groups). We made this substitution because the confidence intervals for all of the relative risks that were less than 1.0 included 1.0 (partly due to small sample sizes) and because the literature does not support a protective relation between cigarette smoking and these four diseases (29). Sterling et al. (4) calculated smoking-attributable mortality from the NMFS/NHIS data using age-specific relative risks, and they compared these with smoking-attributable mortality estimates from the Surgeon General’s report that were calculated from age-adjusted relative risks. Sterling et al. (4) used relative risks that were less than 1.0 in their smoking-attributable mortality calculations from the NMFS/NHIS data, thereby estimating that cigarette smoking prevented 7,200 deaths from coronary heart disease and cerebrovascular disease among persons aged 65 years or older.

Further adjustment of the attributable fractions and smoking-attributable mortality estimates from the CPS II for disease-appropriate confounders (education, alcohol intake, hypertension status, and diabetes status) indicated little residual confounding once age was taken into account. For example, for both genders combined, the age-adjusted estimate of smoking-attributable mortality was only 2.5 percent smaller than the fully adjusted estimate. This result contrasts with that of Sterling et al. (4), who found that the age-adjusted smoking-attributable mortality estimate (using the NMFS/NHIS data) for these same four diseases combined was 26 percent larger than the estimate adjusted for age, alcohol intake, and income.

We noted that adjusting the attributable fraction for covariates other than age was more important for women than for men and more important for coronary heart disease and cerebrovascular disease than for lung cancer and chronic obstructive pulmonary disease. Sterling et al. (4) also found that there is more residual confounding for coronary heart disease and cerebrovascular disease than for lung cancer and chronic obstructive pulmonary disease. Several methodological differences between our work and Sterling et al.’s (4) are likely to account for the disparate findings regarding adjustment for multiple confounders. First, we used a model-based approach to control for confounding, while Sterling et al. (4) used a stratified analysis. We believe the model-based approach produced more robust estimates in comparison with the stratified analysis, where the sample sizes for estimating relative risks may have been small for certain cells. Second, as discussed above, we substituted 1.0 in calculations when the estimated relative risks were less than 1.0. Much of the change that Sterling et al. (4) noted between the age-adjusted and the fully adjusted smoking-attributable mortalities was due to the apparent protective effect of smoking in older men for coronary heart disease and cerebrovascular disease (they calculated 7,200 deaths prevented in the age-adjusted model and 47,600 deaths prevented in the alcohol-, income-, and age-adjusted model). Third, we examined the issue of confounding within the CPS II, whereas Sterling et al. (4) used the NMFS/NHIS. Therefore, it is possible that the nature of the confounding of the relation between smoking and mortality varied by data set.

It is not surprising that the results from the NMFS/NHIS data and the CPS II data differ because of the differences in study populations and in data collection. The NMFS was based on a stratified sample of death certificates, and information was obtained from proxy respondents through mailed self-administered questionnaires. The CPS II was a longitudinal study of a cohort of friends and relatives of American Cancer Society volunteers which obtained self-reported data. Proxy respondents are more likely to misreport smoking history (26–28, 38). It has been debated whether it is inappropriate to use relative risks from the CPS II to estimate smoking-attributable mortality in the United States, because the CPS II was not a representative sample of the target population. However, Rothman has argued that the utility of the relative risks should be determined not by whether the study population is a representative subgroup of the target population but by understanding the biologic relation between smoking and disease (39).
Epidemiologists and public health leaders have used the attributable fraction and its corresponding number of attributable deaths to educate public health policy-makers on the potential benefits of intervention with regard to cigarette smoking and several other underlying causes of disease (40, 41). Policy-makers, in turn, use these estimates to help determine current priorities and strategies for disease control (41). Therefore, it is important to calculate the most valid and accurate smoking-attributable mortality estimates for the United States and the world as possible. In our analysis, we assumed that the stratum-specific relative risks were generalizable to the US population, because we stratified the data according to disease-specific biologically relevant covariates. With such an assumption, we estimate that approximately 240,000 annual deaths from lung cancer, chronic obstructive pulmonary disease, coronary heart disease, and cerebrovascular disease among US Whites can be attributed to smoking. In 1990–1994, the total annual average smoking-attributable mortality for the entire US population—for all smoking-related cancers, cardiovascular diseases, and respiratory diseases, lung cancer deaths due to environmental tobacco smoke, and deaths due to smoking-related fires—was approximately 430,000 (3).

If current smoking patterns continue, an estimated 5 million US residents who were under age 18 years in 1995 will die prematurely from smoking-related illnesses (42). Unfortunately, the recent increase in adolescent smoking will probably have an undesirable effect on smoking-attributable mortality for this birth cohort (43). Annual medical costs attributable to cigarette smoking have been estimated at $53.3 billion (44), and annual indirect costs associated with morbidity and premature mortality from cigarette smoking have been estimated at $6.9 billion and $40.3 billion, respectively (45). The human and economic costs of smoking will continue to accumulate until effective public health efforts to prevent initiation, promote cessation, and protect nonsmokers from the adverse effects of environmental tobacco smoke are established at every level of society.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Sander Greenland for serving as a consultant to this project and Cathy Day-Lally for her help and expertise with the CPS II data.

REFERENCES

APPENDIX A

**Point and Interval Estimation for Relative Risks, Attributable Fractions, and Smoking-Attributable Mortality**

This appendix outlines the methods we used to compare age-specific and age-adjusted relative risks, smoking-attributable fractions, and smoking-attributable mortality as calculated from the National Mortality Follow-back Survey (NMFS)/National Health Interview Survey (NHIS) data with those calculated from the Cancer Prevention Study II (CPS II) data. In this application, the data are stratified by gender and are limited to Whites. First, we examine age- and cause-specific relative risks, smoking-attributable fractions, and death rates; then the age- and cause-specific relative risks from the two data sources are directly age-adjusted to the 1986 US population. Age-adjusted attributable fractions for each cause are obtained as a weighted sum of the age-specific estimates (using the proportion of deaths in each age group as the weights).

**Point and confidence interval estimation for relative risks calculated from the CPS II data**

We estimate age-specific relative risks directly from the CPS II data, which are in the form of deaths and person-years. As usual, confidence intervals for the relative risks are obtained by finding confidence intervals for the (natural) log relative risks and exponentiating.

**Point and confidence interval estimation for relative risks calculated from the NMFS/NHIS data**

Smoking prevalences within each age group, stratified by gender, are obtained using the 1987 NHIS data as $p_j$ ($i = 1, 2, \ldots$ total number of age groups, $j = 0, 1, 2$ smoking categories). The corresponding smoking prevalences within each age group, stratified by gender and disease, among US decedents are $z_{ij}$ and are obtained from the 1986 NMFS. $p_j$ and $z_{ij}$ represent age-specific probabilities, i.e., $\sum p_j = 1$ and $\sum z_{ij} = 1$. We estimated these prevalences and their associated variances by using SUDAAN (version 6.34; Research Triangle Institute, Research Triangle Park, North Carolina), which accounted for the survey designs.

Let $D_i$ be the total number of US deaths for each disease/gender group in age group $i$ and $N_i$ the number in the US population of each gender in age group $i$. Then, the age-specific death rate for smoking category $j$ is $I_j = \frac{z_{ij}D_i}{(p_jN_i)}$, and the relative risk is a function of the smoking prevalence estimates $p_j$ and $z_{ij}$; that is, $R_j = \frac{I_j}{I_0} = \frac{(z_{ij}p_j)}{(p_0z_{ij})}$. 

Am J Epidemiol Vol. 152, No. 6, 2000
A confidence interval for \( R_i \) is obtained by finding a confidence interval for the log relative risk and exponentiating. The natural log of the relative risk is \( \ln(R_i) = \ln(z_{ij}) + \ln(p_{ij}) - \ln(p_i) - \ln(z_{i0}) \). Using a Taylor series approximation, the variance of the log relative risk is

\[
\text{Var}(\ln(R_i)) = \frac{1}{z_{ij}} \text{Var}(z_{ij}) + \frac{1}{z_{i0}} \text{Var}(z_{i0}) + \frac{1}{p_{ij}} \text{Var}(p_{ij}) + \frac{1}{p_i} \text{Var}(p_i) + 2 \left[ \frac{1}{d_i} + \frac{1}{n_i} \right],
\]

assuming that the age-specific prevalences \( z_{ij} \) and \( p_{ij} \) follow a multinomial distribution within each age group. The \( d_i \) and \( n_i \) are the numbers of deceased and living persons, respectively, in age group \( i \) from the NMFS and NHIS surveys, so \( \sum_i d_i = d \) and \( \sum_i n_i = n \) are the total numbers of respondents in each survey.

### Directly age-adjusted relative risks

From both the NMFS data and the CPS II data, we obtain the age-adjusted log relative risks \( \ln(R_i) \) by taking weighted sums of the age-specific log relative risks. The weights are based on the proportion of the US population in each age group. The variance of \( \ln(R_i) \) is \( \text{Var}(\ln(R_i)) = \sum_i w_i^2 \text{Var}(\ln(R_i)) \), where \( w_i \) represents the proportion of the population in age group \( i \) and \( \text{Var}(\ln(R_i)) \) is estimated as appropriate for either the CPS II or the NMFS (see above paragraphs). To obtain a confidence interval for \( R_i \), we assume that \( \ln(R_i) \) follows a normal distribution.

### Age-specific attributable fractions

The age-specific attributable fraction (AF) is

\[
\text{AF}_i = \frac{p_{ij}(R_{ij} - 1) + p_{i1}(R_{i1} - 1)}{p_{ij}(R_{ij} - 1) + p_{i1}(R_{i1} - 1) + 1},
\]

where \( p_{ij} \) represents the joint smoking prevalence in age group \( i \) and smoking category \( j \), so that \( p_i \) is the proportion of the US population in age group \( i \). The logit of \( \text{AF}_i \) is \( \text{logit}(\text{AF}_i) = \ln \left[ \frac{2 \sum_j p_{ij}(R_{ij} - 1)}{p_i} \right] \). The variance of the logit \( (\text{AF}_i) \) is

\[
\text{Var}(\text{logit}(\text{AF}_i)) = \frac{\sum_j (R_{ij} - 1)^2 \text{Var}(p_{ij}) + \sum_j p_{ij}^2 \text{Var}(R_{ij}) - 2(R_{ij} - 1)(R_{ij} - 1)p_{ij}p_{j2} - 2(p_{i1})(p_{i2})(R_{ij})(R_{i2})}{n \left[ \sum_j p_{ij}^2(R_{ij} - 1) \right]^2}
\]

\[+ \frac{1}{p_i} \left[ \frac{\text{Var}(p_i)}{p_i} - \frac{2}{n(1 - p_i)} \right],\]

where \( b = n_{i0} \) for CPS II and \( b = z_{i0} D_i \) for the NMFS.

### Age-adjusted attributable fractions and smoking-attributable mortality

We obtained cause-specific age-adjusted attributable fractions \( (AF_A) \) as a weighted sum of the age-specific estimates by using the proportion of deaths in each age group as the weights, such that \( AF_A = \sum_i D_i \text{AF}_i \text{Var}(\text{AF}_i) = \sum_i \left[ \frac{D_i}{D} \text{AF}_i (1 - \text{AF}_i) \right]^2 \text{Var}(\text{logit}(\text{AF}_i)) \) using \( \text{Var}(\text{logit}(\text{AF}_i)) \) from the previous section. An approximate confidence interval for \( AF_A \) is then \( AF_A \pm 1.96 \sqrt{\text{Var}(AF_A)} \). Similarly, smoking-attributable mortality (SAM) \( = \sum D_i \text{AF}_i \text{Var}(\text{SAM}) = \sum_i \left[ D_iAF_i (1 - AF_i) \right]^2 \text{Var}(\text{logit}(\text{AF}_i)) \), and an approximate confidence interval for smoking-attributable mortality is \( \text{SAM} \pm 1.96 \sqrt{\text{Var}(\text{SAM})} \).
APPENDIX B

Estimation of the Adjusted Smoking-Attributable Fraction

Bruzzi et al. (17) have described a method for estimating the adjusted attributable fraction when using case-control data. This appendix describes the adaptation of their method for application to the CPS II cohort data. We first introduce the notation to present the adapted method in general terms. We then describe a Poisson regression model for the number of deaths in the CPS II and close with a derivation of a formula for the adjusted attributable fraction in terms of parameters in the Poisson regression model.

Notation

The derivation of a formula for the adjusted attributable fraction from cohort data requires the following notation. Let E denote the exposure factor with levels $e = 0, 1, \ldots, L_E$, and let C denote the (cross-classification of) confounding factor(s) with levels $c = 0, 1, \ldots, L_C$. That is, c indexes the combination of levels of the confounding factors. Let $j = (e,c)$ denote the index for the stratum in the cross-classification of exposure by confounding factors. If C is a cross-classification of several confounding factors, then the index c is multidimensional. More concretely, let there be $F$ factors, $C_p, f = 1, \ldots, F$, with factor 1 the exposure (that is, $C_1 = E$) and factors $C_p, f = 2, \ldots, F$, the potential confounders, and let $L_p$ be the number of levels of factor $C_p$. Then index j is an F-tuple $(j_1, \ldots, j_F)$, and the set of indices J is the Cartesian set $L_1 \times \ldots \times L_F$.

The data from the CPS II cohort study consist of counts of deaths $d_j$ and person-years $n_j$ in stratum $j$, $j \in J$. Usually, these data are modeled conditionally on the $n_j$, such that the $d_j$ are independent and Poisson distributed, $d_j \sim \text{Poisson} (\lambda_j n_j)$, $j \in J$. With these assumptions, the expectation of $d_j$ is $E(d_j | \lambda_j n_j) = \lambda_j n_j$, which clarifies the interpretation of $\lambda_j$ as a rate.

Poisson regression model

The search for a parsimonious description of the CPS II data may be attempted with the fit of a log-linear model for the Poisson rates $\lambda_j$; that is, $\log E(d_j | \lambda_j n_j) = \log n_j + \log \lambda_j$, and

$$\log \lambda_j = \text{constant} + \text{main effects} + \text{interactions}. \quad (B1)$$

The adjusted attributable fraction

The adjusted attributable fraction may be expressed in terms of rate ratios, which in turn depend on parameters in a Poisson regression model. To derive an expression for the adjusted rate ratios, one compares the death rate of individuals in an elevated risk stratum (e.g., $e$) and level $c$ of the confounding factors with the death rate of individuals at the baseline level of the risk factor ($e = 0$), keeping level $c$ fixed. To conduct this comparison, with each $j = (e,c)$ associate $j^* = (0,c)$, which has the same level of the confounder but shifts the exposure to the baseline level.

To follow steps similar to the original derivation of Bruzzi et al. (17), let $N_j$ be the population number of person-years in stratum $j$ and let $D_j$ be the population number of deaths in that stratum. Now define the following:

- $\lambda_j = D_j/N_j = \text{death rate for cell } j \text{ in the population}.$
- $\lambda_0 = D_0/N_0 = \text{population death rate at the strata at the baseline exposure level, }$ where $D_0 = \sum_d D_{0,c} = \text{combined number of deaths in the population for strata with exposure at the baseline level and } N_0 = \text{combined number of person-years in the population for strata with exposure at the baseline level}.$
- $R_j = \lambda_j/\lambda_0 = \text{population death rate of stratum } j \text{ relative to stratum } 0.$
- $R^*_j = \lambda_j/\lambda_0 = \text{adjusted rate ratio; that is, the rate ratio of stratum } j \text{ relative to stratum } j^*.$

The attributable fraction $AFC = (\lambda_C N - \lambda_C N)/\lambda_C N = 1 - \lambda_C/\lambda_T$ may be expressed in terms of the adjusted relative risks $\tilde{R}_j$; that is, $AFC = 1 - \sum_j \rho_j/\tilde{R}_j$, where $\rho_j = D_j/D$. The last formula is used for the estimation of $AFC$. The rate ratios $\tilde{R}_j$ may be obtained directly from the Poisson regression via $\tilde{R}_j = \lambda_j/\lambda_{j^*}$, or alternatively they can be computed from the parameters in the log-linear parameterization shown in equation B1.

APPENDIX C

Confidence Intervals for the Adjusted Attributable Fractions

The stratification of the CPS II data by exposure and by levels of potential confounders yields tables in which some strata have low counts. Furthermore, the decision to include all main effects and certain interactions in the models limits the reduction of the number of parameters in the models. The combination of sparse data and models with many parameters warrants concern that the Wald standard errors reported in the SAS statistical packages (versions 6.08, 6.10, and 6.11; SAS Institute, Inc., Cary, North Carolina) may be inadequate. To avoid the use of Wald standard errors to derive the standard error for the attributable fraction, we made recourse to the bootstrap by approximating the distribution of $\hat{AFC}$, the adjusted estimate of $AFC$. 
A broad description of this application of the bootstrap follows. One obtains a large number of independent bootstrap estimates of the AFs, that is, $\hat{AF}^{(b)}$, $b = 1, \ldots, B$, by independently generating $B$ samples with replacement from the original sample.

There are two sources of data used in this analysis, the CPS II and the NMFS. Therefore, to generate a bootstrap estimate $\hat{AF}^{(b)}$, we combined bootstrap estimates of the rate ratios $\hat{R}^{(b)}_j$, $j \in J$, computed using the CPS II data, with bootstrap estimates of the prevalences $\hat{\rho}^{(b)}_j$, $j \in J$, computed using the NMFS data, by means of $\hat{AF}^{(b)} = 1 - \sum_j (\hat{\rho}^{(b)}_j / \hat{R}^{(b)}_j)$, $b = 1, \ldots, B$.

Below, we first describe the generation of $\hat{R}^{(b)}_j$ and then the generation of $\hat{\rho}^{(b)}_j$. The generation of $\hat{R}^{(b)}_j$ requires resampling from the CPS II, whereas the generation of $\hat{\rho}^{(b)}_j$ requires resampling from the NMFS.

### Bootstrap estimates $\hat{R}^{(b)}_j$

The method of generating bootstrap estimates $\hat{R}^{(b)}_j$ for each disease/gender combination has four steps: 1) restructuring the data; 2) drawing the sample; 3) stratifying the sample; and 4) calculating the rate ratio.

In step 1, the CPS II data are restructured because the data are stratified counts and individual-specific covariate values are unavailable. Therefore, we compiled the CPS II data into separate tables for each disease/gender combination. We further stratified each one of these tables by factors of interest (smoking status, age, education, etc.). Let $j \in J$ index the stratum in a given disease × gender table and let $J$ denote the index set for that table. Then the number of deaths ($d_j$) and the person-years ($h_j$) are calculated for every $j \in J$. Therefore, the tables contain counts of deaths and person-years stratified by factor levels. In this fashion, we built a list of $d = \sum d_j$ records (one record for each death) with individual-specific factor levels. These expanded data sets were amenable to bootstrap resampling.

For a given disease/gender combination, in the $b$th bootstrap iteration $d$ records were randomly and independently selected with replacement from an expanded table (step 2). We used the selected records to build a bootstrap stratification table having the same factors as the original stratification, namely with index set $J$, where the data associated with stratum $j$ are the number of selected records (deaths) with that combination of factors. We then combined the deaths in stratum $j$ of the bootstrap table with the person-years in that stratum in the original CPS II stratification table (step 3).

The final step was to fit a Poisson regression model to the bootstrap table defined by disease and gender. We used the estimated coefficients from the model to develop bootstrap estimates of $\hat{R}^{(b)}_j$, $j \in J$ (step 4).

### Bootstrap estimates $\hat{\rho}^{(b)}_j$

For a given disease/gender combination, the method of generating separate bootstrap estimates of $\hat{\rho}^{(b)}_j$, $b = 1, \ldots, B$, has five steps: 1) generating the sample; 2) developing weights to adjust for nonresponse; 3) developing poststratification weights; 4) applying the weights and stratifying the sample; and 5) calculating $\hat{\rho}^{(b)}_j$.

Unlike the CPS II, the NMFS data set contains individual-specific data records, and therefore bootstrap sampling was immediate. Thus, for each bootstrap iteration, the first step was to take, with replacement, a separate sample of records from each of the original 18 NMFS sampling strata (step 1). Each of these samples was of the same size as the corresponding original NMFS sampling stratum. For example, there were 540 records in the first stratum of the original NMFS; i.e., the first stratum is defined to include deaths among American Indians, Eskimos, and Aleuts, regardless of the cause. Therefore, the bootstrap samples also had 540 records in the first stratum. It is important to note that deaths in three of the 18 original sampling strata were selected with certainty for inclusion in the NMFS. These records were also selected with certainty for each of the bootstrap samples.

We weighted each record in the NMFS so that a nationally representative estimate of $\hat{\rho}^{(b)}_j$ could be obtained. Therefore, the next step was to develop weights for the bootstrap sample drawn in the first step. Three weighting factors were required: sampling weights, nonresponse weights, and poststratification weights. The sampling weights, the inverse of the probabilities of selection, are fixed by design and therefore were held constant in the bootstrap. In contrast, the values for the weighting factors used to adjust for nonresponse and poststratification depended on the distribution of deaths among the factor levels. Therefore, these weighting factors differed from bootstrap sample to bootstrap sample.

The adjustment for nonresponse in the original NMFS was calculated for subsets, defined by age intervals, within each of the 18 sampling strata. We conducted this adjustment for each bootstrap sample (step 2). That is, we stratified each bootstrap sample into the same subsets as the original NMFS sample and calculated the response rate within each of these subsets. The reciprocals of these response rates were used for the nonresponse adjustment. With this approach, all deaths in a particular subset have the same nonresponse adjustment.

The next step was to develop poststratification weights (step 3). We used the original NMFS poststrata, which were defined by race, sex, and age. We calculated poststratification weights for each stratum in a bootstrap sample by dividing an estimate of the number of US deaths for that stratum (estimated from the original NMFS sample) by the weighted sum of the deaths in the poststratum for that sample.

After calculation and application of the final weight for each death (the final weight for each death being the product of three weights: the inverse of the probability of selection, the reciprocal of the response rate, and the poststratification weight), step 4 was to create eight stratification tables defined by disease and gender from each bootstrap sample (step 4). We further stratified each one of these by other factors of interest (smoking status, age, education, etc.). The final step was to calculate $\hat{\rho}^{(b)}_j$, $j \in J$, $b = 1, \ldots, B$, separately for each one of the eight tables (step 5).