

## Practice of Epidemiology

### Estimation of the Population Attributable Fraction for Mortality in a Cohort Study Using a Piecewise Constant Hazards Model

Maarit A. Laaksonen\*, Paul Knekt, Tommi Härkänen, Esa Virtala, and Hannu Oja

\* Correspondence to: Maarit Laaksonen, National Institute for Health and Welfare, PL 30, 00271 Helsinki, Finland (e-mail: maarit.laaksonen@thl.fi).

Initially submitted March 19, 2009; accepted for publication December 21, 2009.

Quantification of the impact of exposure to modifiable risk factors on a particular outcome at the population level is a fundamental public health issue. In cohort studies, the population attributable fraction (PAF) is used to assess the proportion of the outcome that is attributable to exposure to certain risk factors in a given population during a certain time interval. This is done by combining information about the prevalence of the risk factor in the population with estimates of the strength of the association between the risk factor and the outcome. In case of mortality, the PAF demonstrates what proportion of mortality can be delayed during the given follow-up time. However, literature on carrying out model-based estimation of PAF and its variance in cohort studies while properly taking follow-up time into account is still scarce. In this article, the authors present formulas for estimation of PAF, its variance, and its confidence interval using the piecewise constant hazards model and apply a SAS macro created for the estimation of PAF (SAS Institute Inc., Cary, North Carolina) to estimate the mortality attributable to some common risk factors.

attributable risk; cohort studies; mortality; proportional hazards models; risk factors

Abbreviations: CI, confidence interval; PAF, population attributable fraction.

In public health research and in the planning of interventions, it is essential to be able to quantify the impact of exposure to modifiable risk factors on the outcome of interest at the population level. In epidemiologic studies, the strength of an association between exposure to a risk factor and the occurrence of a particular outcome is often assessed using the relative risk or the odds ratio. However, these measures do not describe the importance of the risk factor at the population level, as the prevalence of the risk factor is not taken into account. The population attributable fraction (PAF) is an integrated measure, taking into account both the strength of the association between the risk factor and the outcome and the prevalence of the risk factor in the population, which assesses the proportion of a specific outcome in a population that is attributable to exposure to 1 or several risk factors.

Since its introduction in 1953 (1), the PAF has been increasingly dealt with in methodological research, but relatively few practical applications of this measure have been

presented. A variety of methods for estimating PAF, adjusted for potentially confounding factors, have been proposed and applied in case-control, cross-sectional, and cohort studies (2). In a cohort study design, PAF assesses the prognostic impact of exposure to certain risk factors on the occurrence of an outcome in a given population during a certain follow-up period. In case of mortality, PAF demonstrates how much mortality can be delayed during a certain follow-up period. However, demonstrations of model-based PAF measures in which follow-up time has been properly taken into account are scarce. To date, definitions of PAF have relied primarily on instantaneous hazard ratios, calculated at a certain point in time (3–5). This may be due to the popularity of the semiparametric Cox proportional hazards model (6), which enables the elimination of the underlying baseline hazard from an instantaneous hazard ratio but not from a hazard ratio extended over time. A PAF measure extended over time in which the baseline hazard is estimated using the Breslow estimator (7) has been proposed (4, 8). As far as

we know, however, calculation of the variance of these PAF estimates has so far been done through bootstrapping (8), and an analytic variance estimate of the PAF based on the delta method is still missing. Instead of deriving this complicated variance estimate, a parametric piecewise constant hazards model for the estimation of PAF and its variance may be applied. In this approach, the follow-up time is divided into fixed time intervals and the piecewise constant baseline hazards are estimated for each interval. The shorter the intervals, the smaller the bias in the hazard rate estimation.

In this paper, we derive formulas for the model-adjusted PAF estimate with its standard error and confidence interval based on the piecewise constant hazards model. We also demonstrate the performance of this model in a data example estimating the mortality attributable to selected common risk factors and evaluate the reliability of these estimates.

## CALCULATION OF PAF FOR TOTAL MORTALITY IN A COHORT STUDY USING A PIECEWISE CONSTANT HAZARDS MODEL

### Definition of PAF

PAF estimates the proportion of the occurrence of an outcome that could be reduced if it were possible to change some risk factor values  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im})^T$  to their chosen target values  $\mathbf{x}_i^* = (\mathbf{x}_{i1}^*, \dots, \mathbf{x}_{im}^*)^T$ . In this notation,  $\mathbf{x}_i$  is the vector of all risk factors of the  $i$ th individual (modifiable, nonmodifiable, and confounding factors) used in the model; thus, only the modifiable risk factors whose effect we wish to measure will have a different value in  $\mathbf{x}_i^*$ , while the rest of the factors will retain their values. Let  $R(\mathbf{x}_i, \boldsymbol{\beta})$  denote the model-based probability of the outcome occurrence for the  $i$ th individual with risk factor values  $\mathbf{x}_i$  and regression coefficients  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m)^T$  corresponding to them. The expected outcome incidence  $I$  in a study population of  $n$  individuals given the risk factor values  $\mathbf{x}_i$  for individual  $i$ ,  $i = 1, \dots, n$ , can then be calculated as

$$I(\mathbf{x}; \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n R(\mathbf{x}_i, \boldsymbol{\beta}) \quad (1)$$

and given the target values  $\mathbf{x}_i^*$  as

$$I(\mathbf{x}^*; \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n R(\mathbf{x}_i^*, \boldsymbol{\beta}). \quad (2)$$

The excess outcome incidence due to risk factors is then given by

$$\text{PAF}(\boldsymbol{\beta}) = \frac{I(\mathbf{x}; \boldsymbol{\beta}) - I(\mathbf{x}^*; \boldsymbol{\beta})}{I(\mathbf{x}; \boldsymbol{\beta})} = 1 - \frac{I(\mathbf{x}^*; \boldsymbol{\beta})}{I(\mathbf{x}; \boldsymbol{\beta})}. \quad (3)$$

The greater the prevalence of the risk factor  $\mathbf{x}$ —that is, the greater the number of persons who have a certain, presumably harmful level of the risk factor—the more its modification to the target level ( $\mathbf{x} \rightarrow \mathbf{x}^*$ ) reduces the outcome

incidence ( $I(\mathbf{x}; \boldsymbol{\beta}) \rightarrow I(\mathbf{x}^*; \boldsymbol{\beta})$ ) at the population level, thus making the PAF larger.

### Application of PAF in a cohort study using a piecewise constant hazards model

Suppose that at baseline ( $t = 0$ ) we have a study population of  $n$  individuals who are free of the outcome of interest. Each person's  $m$  risk factor values  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im})^T$  are measured. The study population is subsequently followed for a given period of time, with the length of follow-up for each individual ( $T_i$ ) being determined as the time from baseline to the date of the outcome of interest or censoring due to the end of follow-up, whichever comes first (9). These time-to-event data are then used to analyze the effects of the risk factors  $\mathbf{x}_i$  on the outcome occurrence. In this study, the outcome of interest is total mortality.

The expected mortality at a chosen interval ( $t, t + \Delta t$ ) in the whole study population of  $n$  individuals, given the risk factor values  $\mathbf{x}_i$ , can be calculated as in equation 1:

$$\begin{aligned} I(\mathbf{x}; \boldsymbol{\beta})_{t,t+\Delta t} &= \frac{1}{n} \sum_{i=1}^n P(t < T_i \leq t + \Delta t | \mathbf{x}_i) \\ &= \frac{1}{n} \sum_{i=1}^n [S(t; \mathbf{x}_i) - S(t + \Delta t; \mathbf{x}_i)], \end{aligned} \quad (4)$$

where  $S(t; \mathbf{x}_i) = P(T_i > t | \mathbf{x}_i)$  is the survival function up to time  $t$ . Calculation of the expected mortality given the target values follows similarly through replacement of  $\mathbf{x}_i$  by  $\mathbf{x}_i^*$  in equation 4. The PAF for mortality at interval ( $t, t + \Delta t$ ) can then be calculated as in equation 3:

$$\text{PAF}(\boldsymbol{\beta})_{t,t+\Delta t} = 1 - \frac{\frac{1}{n} \sum_{i=1}^n [S(t; \mathbf{x}_i^*) - S(t + \Delta t; \mathbf{x}_i^*)]}{\frac{1}{n} \sum_{i=1}^n [S(t; \mathbf{x}_i) - S(t + \Delta t; \mathbf{x}_i)]}, \quad (5)$$

where  $S(t; \mathbf{x}_i) = \exp[-\int_0^t \lambda(s; \mathbf{x}_i) ds]$ , and  $\lambda(t; \mathbf{x}_i)$  is the hazard function at time  $t$  for the  $i$ th individual with risk factors  $\mathbf{x}_i^T = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im})$ . The PAF for mortality from baseline ( $t = 0$ ) by time  $t + \Delta t$  is a special case of  $\text{PAF}_{t,t+\Delta t}$ , where  $S(t; \mathbf{x}_i)$  and  $S(t; \mathbf{x}_i^*)$  are reduced to 1.

In this study, estimation of the PAF for mortality is carried out using a parametric piecewise constant hazards model. In this approach, the follow-up time is partitioned into  $J$  intervals  $(0 = a_1, a_2], (a_2, a_3], \dots, (a_{j-1}, a_j], \dots, (a_{J-1}, a_J]$ , and the hazard is allowed to depend on time by letting the baseline hazard change from one interval to another (10). Virtually any baseline hazard can be well approximated by choosing closely spaced cutpoints for the intervals. The effect of age can be taken into account in the model by dividing the range of individual birth dates into  $C$  birth cohorts  $(v_0, v_1], \dots, (v_C, v_{C-1}], \dots, (v_{C-1}, v_C]$  and by further stratifying the baseline hazard by birth cohort (9). The hazard function at time  $t$  for the  $i$ th individual given the birth cohort  $c_i$  and risk factors  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im})^T$  can then be expressed as

$$\lambda(t; c_i, \mathbf{x}_i) = \sum_{j=1}^J 1\{a_{j-1} < t \leq a_j\} \lambda_{0jc_i} \exp(\mathbf{x}_i^T \boldsymbol{\beta}),$$

where  $\lambda_{0jc_i}$  is the baseline hazard in the  $j$ th interval and birth cohort  $c_i$ . The survival function is then given by

$$\begin{aligned} S(t; c_i, \mathbf{x}_i) &= \exp\left(-\sum_{j=1}^J \lambda_j(c_i, \mathbf{x}_i) \delta_j(t)\right) \\ &= \exp\left(-\sum_{j=1}^J \exp(\boldsymbol{\alpha}_{jc_i} + \mathbf{x}_i^T \boldsymbol{\beta}) \delta_j(t)\right) \\ &= \exp\left[-\sum_{j=1}^J \exp(\mathbf{z}_{ij} \boldsymbol{\gamma}) \delta_j(t)\right], \end{aligned} \quad (6)$$

where  $\lambda_j(c_i, \mathbf{x}_i)$  is the hazard function in the  $j$ th interval,  $\log \lambda_{0jc_i} = \boldsymbol{\alpha}_{jc_i} = \boldsymbol{\alpha}_{1j} + \boldsymbol{\alpha}_{2c} + \boldsymbol{\alpha}_{1j2c}$ ,  $\mathbf{z}_{ij}$  is the design matrix corresponding to the regression coefficients  $\boldsymbol{\gamma} = \boldsymbol{\alpha}_{11}, \dots, \boldsymbol{\alpha}_{JC}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m)^T$ , and  $\delta_j(t)$  defines the length of follow-up in the  $j$ th interval:

$$\delta_j(t) = \begin{cases} 0, & t \leq a_{j-1} \\ t - a_{j-1}, & a_{j-1} < t \leq a_j \\ a_j - a_{j-1}, & t > a_j. \end{cases}$$

The PAF at interval  $(t, t + \Delta t)$  can then be calculated as in equation 5:

$$\begin{aligned} \text{PAF}(\boldsymbol{\gamma})_{t,t+\Delta t} &= \\ &1 - \frac{\frac{1}{n} \sum_{i=1}^n \left\{ \exp\left[-\sum_{j=1}^J \exp(\mathbf{z}_{ij}^* \boldsymbol{\gamma}) \delta_j(t)\right] - \exp\left[-\sum_{j=1}^J \exp(\mathbf{z}_{ij}^* \boldsymbol{\gamma}) \delta_j(t+\Delta t)\right] \right\}}{\frac{1}{n} \sum_{i=1}^n \left\{ \exp\left[-\sum_{j=1}^J \exp(\mathbf{z}_{ij} \boldsymbol{\gamma}) \delta_j(t)\right] - \exp\left[-\sum_{j=1}^J \exp(\mathbf{z}_{ij} \boldsymbol{\gamma}) \delta_j(t+\Delta t)\right] \right\}}. \end{aligned} \quad (7)$$

## Estimation of PAF

In order to estimate the PAF, we first need to estimate the model parameters  $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\alpha}}_{11}, \dots, \hat{\boldsymbol{\alpha}}_{JC}, \hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_m)^T$ . In this section of the paper, the persons who are used in the estimation of parameters are not necessarily the same as those in the other sections, but we retain the same notation:  $i = 1, \dots, n$ . In some applications, parameters might be estimated in 1 population, the PAF being calculated in another (standard) population. The maximum likelihood estimation of  $\boldsymbol{\gamma}$  is demonstrated in Appendix 1.

In this study, the SAS procedures LIFEREG and TPHREG (SAS Institute Inc., Cary, North Carolina) are used to compute the maximum likelihood estimates  $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\alpha}}_{11}, \dots, \hat{\boldsymbol{\alpha}}_{JC}, \hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_m)^T$  and their estimated covariance matrix  $\widehat{\text{cov}}(\hat{\boldsymbol{\gamma}})$  (11). The point estimate of  $\text{PAF}(\boldsymbol{\gamma})_{t,t+\Delta t}$  can then be obtained by replacing the unknown parameter values  $\boldsymbol{\gamma}$  in equation 6 by their point estimates  $\hat{\boldsymbol{\gamma}}$ . The variance estimate of  $\text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}$  can be obtained using the delta method, according to which

$$\widehat{\text{var}}[\text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}] = \left[ \frac{\partial \text{PAF}(\boldsymbol{\gamma})_{t,t+\Delta t}}{\partial \boldsymbol{\gamma}} \bigg|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} \right]^T \widehat{\text{cov}}(\hat{\boldsymbol{\gamma}}) \left[ \frac{\partial \text{PAF}(\boldsymbol{\gamma})_{t,t+\Delta t}}{\partial \boldsymbol{\gamma}} \bigg|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} \right]. \quad (10)$$

The vector of derivatives of  $\text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}$  with respect to  $\boldsymbol{\gamma}$  is presented in Appendix 2. The approximate 95% confidence interval of PAF is then obtained by

$$\text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t} \pm 1.96 \sqrt{\widehat{\text{var}}[\text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}]}.$$

This normal approximation for the sampling distribution of  $\text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}$  is not accurate when the sample size is small and the distribution of  $\text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}$  is skewed, and therefore some symmetrizing transformation of PAF, such as the complementary logarithmic transformation, should be used:

$$\begin{aligned} \log[1 - \text{PAF}(\boldsymbol{\gamma})_{t,t+\Delta t}] &= \log\left[1 - \left(1 - \frac{I(\mathbf{x}^*; \boldsymbol{\gamma})_{t,t+\Delta t}}{I(\mathbf{x}; \boldsymbol{\gamma})_{t,t+\Delta t}}\right)\right] \\ &= \log I(\mathbf{x}^*; \boldsymbol{\gamma})_{t,t+\Delta t} - \log I(\mathbf{x}; \boldsymbol{\gamma})_{t,t+\Delta t}. \end{aligned}$$

The 95% confidence interval of  $\log[1 - \text{PAF}(\boldsymbol{\gamma})_{t,t+\Delta t}]$  is transformed back to the original scale by

$$\begin{aligned} 1 - \exp\left\{\log[1 - \text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}] \pm 1.96 \right. \\ \left. \times \sqrt{\widehat{\text{var}}[\log[1 - \text{PAF}(\hat{\boldsymbol{\gamma}})_{t,t+\Delta t}]]}\right\}. \end{aligned}$$

SAS code for calculating PAF for total mortality with the piecewise constant hazards model, based on the formulas provided here and in Appendix 2, is given in Appendix 3.

## DATA EXAMPLE

### Population and methods

The present study is based on data from the Mini-Finland Health Survey, which was carried out in 1978–1980 (12). The Mini-Finland Health Survey consisted of 8,000 persons from 40 geographic areas in Finland and was a representative sample of the Finnish population of adults aged 30 years or more. For the present analysis, the data included a total of 6,267 men and women who participated in the study, were 30–69 years of age, and had been born between 1890 and 1949. As part of the baseline examination, the participants took part in an interview and completed a self-administrated questionnaire that yielded information on health-related lifestyle choices, such as smoking, alcohol consumption, and physical exercise. Body weight and height were measured and body mass index (weight (kg)/height (m)<sup>2</sup>) was calculated. These potential risk factors were all categorized (Table 1). The subjects were systematically followed for

**Table 1.** Estimated Age- and Sex-adjusted Relative Risk of Death in Categories of Potential Risk Factors According to Cox's Model, Mini-Finland Health Survey, 1978–1994

Variable and Category	No. of Deaths	Total No. of Subjects	Adjusted for Age and Sex		Adjusted for All Variables		
			RR	95% CI	RR	95% CI	
Nonmodifiable variables							
Sex							
Male	683	2,980	1		1		
Female	423	3,305	0.43*	0.38, 0.49	0.47*	0.41, 0.55	
Age group, years							
30–39	70	1,862	1		1		
40–49	156	1,635	2.58*	1.95, 3.42	2.74*	2.06, 3.65	
50–59	314	1,590	6.08*	4.69, 7.88	6.35*	4.87, 8.28	
60–69	566	1,198	18.29*	14.26, 23.46	19.59*	15.11, 25.38	
Modifiable variables							
Smoking							
Never smoker	439	3,336	1		1		
Ex-smoker	277	1,336	1.23*	1.04, 1.46	1.28	1.08, 1.53	
Current smoker, cigarettes/day							
1–19	222	990	1.90*	1.59, 2.26	2.00*	1.64, 2.36	
≥20	166	618	2.75*	2.25, 3.36	2.73*	2.22, 3.37	
Alcohol consumption, g/week							
0	541	2,592	1		1		
1–99	401	2,875	0.78*	0.68, 0.90	0.72*	0.63, 0.84	
≥100	160	812	1.13	0.93, 1.38	0.85	0.70, 1.05	
Body mass index <sup>a</sup>							
≤21.4	123	825	1		1		
21.5–29.9	745	4,529	0.64*	0.53, 0.78	0.69*	0.57, 0.84	
≥30.0	236	928	0.91	0.73, 1.13	0.91	0.76, 1.19	
Physical exercise							
Little or none	514	2,113	1		1		
Occasional or regular	588	4,165	0.62*	0.55, 0.69	0.70*	0.62, 0.79	

Abbreviations: CI, confidence interval; RR, relative risk.

\*  $P < 0.05$ .<sup>a</sup> Weight (kg)/height (m)<sup>2</sup>.

mortality since the baseline examination, using individual mortality information obtained from a nationwide registry maintained by Statistics Finland. During a 17-year follow-up period, 683 men and 423 women died.

To avoid problems in estimation, we selected the birth cohorts and follow-up intervals so that there would be at least 1 mortality case for each cohort in each interval. The length of follow-up for each person was determined as the time from baseline to either death or censoring, whichever came first. The strength of the association between the selected potential risk factors (sex, age, smoking, alcohol consumption, body mass index, and physical exercise) and death was estimated in terms of the relative risk, using Cox's

model. The variability of the regression parameter estimates obtained in the piecewise constant hazards model using different follow-up intervals and birth cohorts was examined and compared with the estimates obtained by means of the Cox model. The PAF and its variance for the number of deaths attributable to selected risk factors were estimated with the piecewise constant hazards model using the SAS code given in Appendix 3.

## RESULTS

Statistically significant differences in risk of death between categories of all potential risk factors (sex, age,

**Table 2.** Comparison of Regression Parameter Estimates Obtained in a Piecewise Constant Hazards Model Using Different Follow-up Intervals and Birth Cohorts With Estimates Obtained in a Stratified Cox Model, Mini-Finland Health Survey, 1978–1994

Variable and Category <sup>a</sup>	Piecewise Constant Hazards Model <sup>b</sup>							Cox Model (Model 8)
	Model 1	Model 2	Model 3	Model 4	Model 5 <sup>c</sup>	Model 6	Model 7	
No. of follow-up intervals	10	10	4	4	4	2	2	4
No. of birth cohorts	25	10	25	10	5	10	5	5
Regression parameter								
Female sex	–0.756	–0.742	–0.753	–0.741	–0.744 (0.078) <sup>d</sup>	–0.727	–0.732	–0.746 (0.078)
Smoking								
Ex-smoker	0.281	0.277	0.276	0.275	0.269 (0.089)	0.274	0.269	0.266 (0.088)
Current smoker, cigarettes/day								
1–19	0.732	0.710	0.729	0.707	0.693 (0.093)	0.699	0.686	0.689 (0.093)
≥20	1.062	1.024	1.054	1.024	1.023 (0.107)	1.011	1.013	1.025 (0.107)
Alcohol consumption, g/week								
1–99	–0.306	–0.310	–0.302	–0.307	–0.320 (0.073)	–0.300	–0.313	–0.319 (0.073)
≥100	–0.174	–0.148	–0.169	–0.146	–0.169 (0.104)	–0.133	–0.158	–0.174 (0.103)
Body mass index <sup>e</sup>								
21.5–29.9	–0.383	–0.379	–0.382	–0.376	–0.348 (0.100)	–0.362	–0.338	–0.345 (0.100)
≥30	–0.081	–0.074	–0.081	–0.074	–0.041 (0.115)	–0.064	–0.035	–0.034 (0.115)
Occasional or regular physical exercise	–0.363	–0.368	–0.362	–0.365	–0.367 (0.062)	–0.358	–0.361	–0.368 (0.062)

<sup>a</sup> Category of the variable which was compared with the reference category (lowest) (see Table 1).

<sup>b</sup> Piecewise constant hazards model with different lengths of follow-up and birth cohorts.

<sup>c</sup> The results from model 5 were the most comparable with those of the Cox model.

<sup>d</sup> Numbers in parentheses, standard error.

<sup>e</sup> Weight (kg)/height (m)<sup>2</sup>.

smoking, alcohol consumption, body mass index, and physical exercise) were found (Table 1). Smoking and physical exercise, in addition to age, showed the strongest associations with mortality. No statistically significant interactions between the variables in the model were found, and thus we used a model with main effects only to study their simultaneous effects on the risk of death. No notable differences in the relative risks were found between the sex- and age-adjusted models including only 1 potential risk factor and the simultaneous model including all of them.

No notable differences were found in the regression parameters obtained for the closely spaced partition of both the follow-up interval and the birth cohort to 2-year intervals and the wider partition to 10-year intervals in these data, where the distribution of deaths during follow-up within each interval was assumed to be exponential (Table 2). Regression parameter estimates obtained using the piecewise constant hazards model and estimates obtained using the Cox model were very similar, thus supporting an adequate approximation of the piecewise constant hazards model to Cox's model.

We then used the piecewise constant hazards model to estimate PAFs and their confidence intervals for the potential risk factors in each of the 4 5-year follow-up intervals and in the whole 20-year interval using 10-year birth cohorts (Table 3). Of the 4 risk factors, smoking seemed to have the greatest impact on risk of death, reducing it by 14% if the current smokers had never started smoking (95% confidence interval (CI): 11, 17). Reduction of alcohol consumption seemed to have the smallest impact, the PAF being 3% (95% CI: 2, 5) during the 20-year follow-up period. In addition to changes in these 2 risk factors, an increase in physical exercise and a better body mass index would altogether have led to a 30% reduction in mortality risk during the 20 years of the study. A decreasing tendency in the PAF estimates during follow-up was demonstrated through estimation of PAF for the 5-year follow-up intervals. This tendency was more notable in the oldest 10-year age group as compared with the youngest 10-year age group (Figure 1).

Finally, the cumulative PAF estimate obtained from the full model including all 4 risk factors and the entire

**Table 3.** Comparison of PAF Estimates Calculated Using Complementary Logarithmic Transformation for Each 5-Year Time Interval ( $PAF_{t,t+\Delta t}$ ) and the Entire 20-Year Follow-up Period ( $PAF_{0,\Delta t}$ ) Using 10-Year Birth Cohorts, Mini-Finland Health Survey, 1978–1994

Variable <sup>a</sup>	$PAF_{0,5}$	95% CI	$PAF_{5,10}$	95% CI	$PAF_{10,15}$	95% CI	$PAF_{15,20}$	95% CI	$PAF_{0,20}$	95% CI
Smoking <sup>b</sup>	0.20	0.16, 0.24	0.18	0.14, 0.21	0.13	0.10, 0.16	0.10	0.07, 0.13	0.14	0.11, 0.17
Alcohol consumption <sup>c</sup>	0.04	0.02, 0.07	0.04	0.02, 0.06	0.03	0.01, 0.04	0.02	0.01, 0.04	0.03	0.02, 0.05
Body mass index <sup>d</sup>	0.10	0.06, 0.14	0.09	0.06, 0.12	0.07	0.04, 0.09	0.05	0.03, 0.07	0.07	0.05, 0.10
Physical exercise	0.18	0.14, 0.23	0.16	0.12, 0.20	0.13	0.10, 0.16	0.09	0.07, 0.12	0.13	0.10, 0.17
All 4 variables	0.39	0.34, 0.44	0.35	0.31, 0.40	0.29	0.24, 0.33	0.22	0.18, 0.27	0.30	0.26, 0.34

Abbreviations: CI, confidence interval; PAF, population attributable fraction.

<sup>a</sup> The age- and sex-adjusted variable for which the PAF was calculated by estimating the reduction in mortality (proportion) during the given time interval if all subjects had belonged to the target category with the lowest risk of death (see Table 1), unless otherwise noted.

<sup>b</sup> The category with the lowest risk of death (i.e., never smokers) was used as the target category, but the risk of death among ex-smokers remained unchanged.

<sup>c</sup> The category with the lowest risk of death (i.e., moderate alcohol consumption (1–99 g/week)) was used as the target category, but the risk of death among nondrinkers remained unchanged.

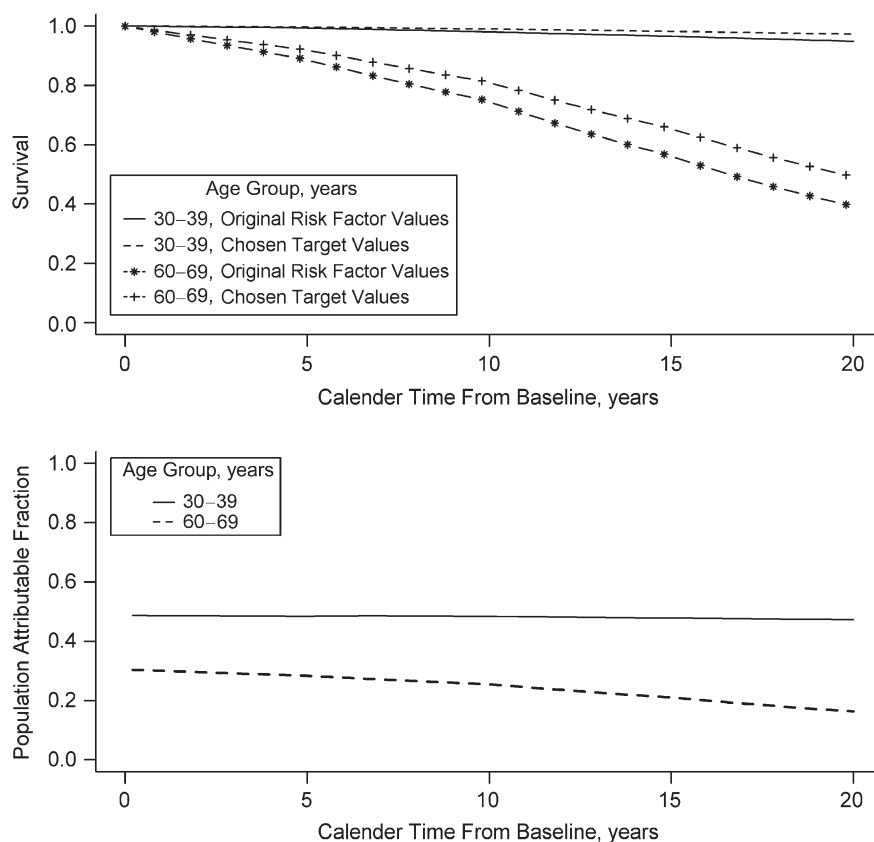
<sup>d</sup> Weight (kg)/height (m)<sup>2</sup>.

follow-up period, using the analytic piecewise constant hazards method introduced in this paper (Table 3), was compared with the PAF estimate obtained using bootstrap estimation (13), so we could study the usability and accuracy of this method. The bootstrap with 2,000 samples yielded the same point estimate of PAF (29.9%) as the analytic method, whereas there was some variation in the estimates of the 95% confidence intervals for PAF (analytic

method—95% CI: 25.5, 34.0; percentile points 2.5 and 97.5 of the bootstrap distribution—95% CI: 25.7, 35.3).

## DISCUSSION

Model-based estimation of the PAF and its standard error in a cohort study that properly takes follow-up time into account has received very little attention.

**Figure 1.** Estimates of the survival curve and the population attributable fraction by duration of follow-up for the youngest (30–39 years) and oldest (60–69 years) 10-year age groups, Mini-Finland Health Survey, 1978–1994.



The PAF in a cohort study is defined as the expected excess incidence, during a certain follow-up time, due to certain risk factors in comparison with their chosen target values. The expected outcome incidences are calculated by estimating the change in the survival function during that time. In this study, the survival function was estimated using a piecewise constant hazards model, in which a linear model for the logarithm of the hazard is assumed. Follow-up time was defined as time since the baseline examination, and we took the effect of age into account by stratifying the baseline hazards by birth cohort. The new method for estimating PAF and its standard error on the basis of these assumptions was found to be very flexible in that both categorical and continuous variables and their interactions could be included in the model. In addition, judicious choice of the cutpoints in the piecewise constant hazards model allows us to well approximate almost any baseline hazard for large data sets. We demonstrated this method by estimating the numbers of deaths attributable to certain well-known risk factors using data from the Mini-Finland Health Survey. In this application, the piecewise constant hazards model and the Cox model gave similar relative risk estimates.

Thus far, the variance of PAF has been estimated using methods based on resampling, such as bootstrapping (8). This paper, in which we have presented an analytic method based on the piecewise constant hazards model, offers a fast method for estimation of the variance of PAF. Alternative models may also be used (14, 15). Furthermore, the use of complementary logarithmic transformation for estimation of the confidence interval of PAF guarantees that it remains in its natural range from  $-\infty$  to 1 (15); therefore, the proposed analytic method may also be applied for protective factors, in which case negative PAF estimates and their confidence intervals would be obtained. When it was compared with the PAF estimates and confidence intervals obtained using the bootstrap method, however, we noted that the analytic piecewise constant hazards method with complementary logarithmic transformation produced somewhat lower estimates for the upper limit of the 95% confidence interval of PAF. This may indicate that complementary logarithmic transformation is not an optimal method for correcting the skewed sampling distribution, and thus further studies on comparison of the confidence intervals obtained using different methods in a cohort study design are needed.

There are certain issues related to the cohort study design to be noted in the interpretation of PAF. First, according to the traditional definition of PAF, the expected excess risk is defined as the proportion of the outcome which could be avoided if the risk factor were eliminated. Since in the case of mortality the outcome can only be delayed, it is useful to calculate PAF estimates as a function of time in order to demonstrate the effect of a potential intervention in the long run. Second, in the definition of PAF, an immediate reduction in risk is assumed to follow from the change in risk factor. Often, however, a certain amount of time is needed before the effect of the change can be seen. To be able to evaluate the length of this delay, we would need a randomized clinical trial in which the effect of changing certain risk factor values to their target values

would be followed and compared with the effect of not changing them. Third, whenever the effects of several risk factors on the outcome are evaluated simultaneously, part of the effect is due to the interaction of these factors. Therefore, to be able to evaluate the relative importance of a certain risk factor in different risk factor combinations, the joint effect of the risk factors should be partitioned to the individual risk factors so that the separate PAF estimates for the different risk factors sum to the total PAF estimate (16, 17). Fourth, a decreasing tendency in the PAF estimates during follow-up, demonstrated through estimation of PAF for 5-year follow-up intervals, requires some clarification. It is a well-known phenomenon in cohort studies that the predictive value of the risk factors measured at baseline diminishes with a longer duration of follow-up. Repeated measurements of the risk factors during follow-up would be needed to estimate the effect of this phenomenon and thus study the accuracy of the multiplicativity assumption of the piecewise constant hazards model. If data on such time-varying covariates were available, however, they could also be included in the formulas presented in this study. The decrease in PAF estimates during follow-up may also be partly due to the effect of age, since risk factors are not strong predictors for older persons. Ultimately, however, the tendency of PAF estimates to decrease is related to the inevitability of death; if the follow-up time were extended enough, eventually everyone would (of course) die, and the PAF estimates would approach zero.

There are also certain issues directly related to the piecewise constant hazards model used in the present study. The choice of the cutpoints in this model depends on the form of the hazard. In the present study, the relatively slowly changing hazard was well approximated by relatively wide intervals. When the hazard varies more rapidly, however, more closely spaced cutpoints will be needed to well approximate the hazard. This leads to the issue of the sufficiency of data, especially in the case of stratified baseline hazards, since at least 1 case per birth cohort within each interval is required to estimate the levels of the baseline hazard rate. This may limit the choice of cutpoints and thus the approximation of the hazard, especially in the case of smaller data sets. In the case of a more rapidly varying hazard, a flexible choice of intervals of varying lengths, instead of the fixed cutpoints presented in this paper, might also be useful.

In conclusion, a comparison of model-based PAF estimates with estimates obtained from an intervention study including repeated measurements would enhance our knowledge of the appropriateness of the underlying assumptions of the piecewise constant hazards model used to estimate PAF in this study. Development of a strategy for an optimal choice of cutpoints in a piecewise constant hazards model would also be of interest. Studying the performance of a piecewise constant hazards model with data with more rapidly varying hazards would improve our knowledge of the applicability of this model. It would also be of interest to extend the PAF formulas presented here, applicable under the assumption that the outcome is total mortality, to the situation where the outcome is a certain disease and censoring due to death is taken into account.

## ACKNOWLEDGMENTS

Author affiliations: National Institute for Health and Welfare, Helsinki, Finland (Maarit A. Laaksonen, Paul Knekt, Tommi Härkänen, Esa Virtala); and Tampere School of Public Health, University of Tampere, Tampere, Finland (Hannu Oja).

The first author received financial support from the University of Tampere Doctoral Programs in Public Health.

Conflict of interest: none declared.

## REFERENCES

- Levin ML. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum*. 1953;9(3):531–541.
- Benichou J. A review of adjusted estimators of attributable risk. *Stat Methods Med Res*. 2001;10(3):195–216.
- Bruzzi P, Green SB, Byar DP, et al. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol*. 1985;122(5):904–914.
- Chen YQ, Hu C, Wang Y. Attributable risk function in the proportional hazards model for censored time-to-event. *Bio-statistics*. 2006;7(4):515–529.
- Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control*. 2007;18(5):571–579.
- Cox DR. Regression models and life tables (with discussion). *J R Stat Soc Series B Stat Methodol*. 1972;34(2):187–220.
- Burr D. On inconsistency of Breslow's estimator as an estimator of the hazard rate in the Cox model. *Biometrics*. 1994;50(4):1142–1145.
- Samuelsen SO, Eide GE. Attributable fractions with survival data. *Stat Med*. 2008;27(9):1447–1467.
- Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am J Epidemiol*. 1997;145(1):72–80.
- Friedman M. Piecewise constant hazards models for survival data with covariates. *Ann Stat*. 1982;10(1):101–113.
- SAS Institute Inc. *SAS/STAT User's Guide, Version 9.1*. Cary, NC: SAS Institute Inc; 2007.
- Aromaa A, Heliövaara M, Impivaara O, et al. Aims, methods and study population. Part 1. In: Aromaa A, Heliövaara M, Impivaara O, et al, eds. *The Execution of the Mini-Finland Health Survey* [in Finnish with English summary]. Helsinki, Finland: Social Insurance Institution; 1989.
- Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat*. 1979;7(1):1–26.
- Laird N, Olivier D. Covariance analysis of censored survival data using log-linear analysis techniques. *J Am Stat Assoc*. 1981;76(374):231–240.
- Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics*. 1993;49(3):865–872.
- Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. *J Clin Epidemiol*. 1995;48(5):645–665.
- Rabe C, Lehnert-Batar A, Gefeller O. Generalized approaches to partitioning the attributable risk of interacting risk factors can remedy existing pitfalls. *J Clin Epidemiol*. 2007;60(5):461–468.
- Bickel PJ, Doksum KA. *Mathematical Statistics: Basic Ideas and Selected Topics*. 2nd ed. Upper Saddle River, NJ: Prentice Hall, Inc; 2001:132,434.

## APPENDIX 1

### Maximum Likelihood Estimation of the Model Parameters Needed for Calculation of the Population Attributable Fraction Using a Piecewise Constant Hazards Model

The maximum likelihood estimates  $\hat{\gamma} = (\hat{\alpha}_{11}, \dots, \hat{\alpha}_{JC}, \hat{\beta}_1, \dots, \hat{\beta}_m)^T$  of  $\gamma$  can be obtained by maximizing the overall likelihood function, which is given by

$$L(\gamma) = \prod_{i=1}^n \prod_{j=1}^J \{ \exp(d_{ij} \mathbf{z}_{ij} \gamma) \exp[-\exp(\mathbf{z}_{ij} \gamma) \delta_j(t_i)] \}$$

$$= \exp \left[ \left( \sum_{i=1}^n \sum_{j=1}^J d_{ij} \mathbf{z}_{ij} \right) \gamma \right] \exp \left[ - \sum_{i=1}^n \sum_{j=1}^J \exp(\mathbf{z}_{ij} \gamma) \delta_j(t_i) \right]$$

or the logarithm of the likelihood function, which is given by

$$l(\gamma) = \log(L(\gamma)) = \left( \sum_{i=1}^n \sum_{j=1}^J d_{ij} \mathbf{z}_{ij} \right) \gamma - \sum_{i=1}^n \sum_{j=1}^J \delta_j(t_i) \exp(\mathbf{z}_{ij} \gamma),$$

where

$$d_{ij} = \begin{cases} 1, & 0 < \delta_j(t_i) < a_j - a_{j-1} \\ 0, & \text{otherwise} \end{cases}.$$

The log-likelihood function will be maximized where the score function  $S(\gamma)$ , first derivative of the log-likelihood function with respect to  $\gamma$ , equals zero:

$$S(\gamma) = \frac{\partial l(\gamma)}{\partial \gamma} = \left( \sum_{i=1}^n \sum_{j=1}^J d_{ij} \mathbf{z}_{ij} \right) - \sum_{i=1}^n \sum_{j=1}^J \delta_j(t_i) \exp(\mathbf{z}_{ij} \gamma) \mathbf{z}_{ij} = 0.$$

The asymptotic variance for the estimates  $\hat{\gamma}$  can be obtained using the inverse of the Fisher information matrix  $I(\gamma)$ , the second derivative of the negative log-likelihood function:

$$I(\gamma) = -\frac{\partial^2 l(\gamma)}{\partial \gamma^2} = \sum_{i=1}^n \sum_{j=1}^J \delta_j(t_i) \exp(\mathbf{z}_{ij} \gamma) \mathbf{z}_{ij} \mathbf{z}_{ij}^T.$$

Since the score function cannot be solved in closed form, however, maximum likelihood estimation with iterative methods, such as Newton-Raphson or Fisher scoring (18), can be used to obtain the parameter estimates  $\hat{\gamma} = (\hat{\alpha}_{11}, \dots, \hat{\alpha}_{JC}, \hat{\beta}_1, \dots, \hat{\beta}_m)^T$  and their estimated covariance matrix  $\widehat{\text{cov}}(\hat{\gamma})$ . These 2 methods are available in SAS (11).



## APPENDIX 2

### Derivatives of $\text{PAF}(\gamma)_{t,t+\Delta t}$ and $\log[1 - \text{PAF}(\gamma)_{t,t+\Delta t}]$

The population attributable fraction (PAF) at interval  $(t, t + \Delta t)$  is given by

$$\text{PAF}(\gamma)_{t,t+\Delta t} = 1 - \frac{I(\mathbf{x}^*, c; \gamma)_{t,t+\Delta t}}{I(\mathbf{x}, c; \gamma)_{t,t+\Delta t}}$$

$$= 1 - \frac{\frac{1}{n} \sum_{i=1}^n \left\{ \exp \left[ - \sum_{j=1}^J \exp(\mathbf{z}_{ij}^* \gamma) \delta_j(t) \right] - \exp \left[ - \sum_{j=1}^J \exp(\mathbf{z}_{ij}^* \gamma) \delta_j(t + \Delta t) \right] \right\}}{\frac{1}{n} \sum_{i=1}^n \left\{ \exp \left[ - \sum_{j=1}^J \exp(\mathbf{z}_{ij} \gamma) \delta_j(t) \right] - \exp \left[ - \sum_{j=1}^J \exp(\mathbf{z}_{ij} \gamma) \delta_j(t + \Delta t) \right] \right\}}.$$

The components of the  $1 \times (C + J + C \times J + m)$  vector of derivatives of  $\text{PAF}_{t,t+\Delta t}(\gamma)$  and  $\log[1 - \text{PAF}(\gamma)_{t,t+\Delta t}]$  with respect to  $\gamma$  are

$$\frac{\partial \text{PAF}(\gamma)_{t,t+\Delta t}}{\partial \gamma_r} = \frac{\frac{\partial I(\mathbf{x}, c; \gamma)_{t,t+\Delta t}}{\partial \gamma_r} \times I(\mathbf{x}^*, c; \gamma)_{t,t+\Delta t} - \frac{\partial I(\mathbf{x}^*, c; \gamma)_{t,t+\Delta t}}{\partial \gamma_r} \times I(\mathbf{x}, c; \gamma)_{t,t+\Delta t}}{I(\mathbf{x}, c; \gamma)_{t,t+\Delta t}^2},$$

and

$$\begin{aligned} & \frac{\partial \left\{ \log \left[ 1 - \text{PAF}(\gamma)_{t,t+\Delta t} \right] \right\}}{\partial \gamma_r} \\ &= \frac{\partial \left\{ \log I(\mathbf{x}^*, c; \gamma)_{t,t+\Delta t} - \log I(\mathbf{x}, c; \gamma)_{t,t+\Delta t} \right\}}{\partial \gamma_r} \\ &= \frac{\frac{\partial I(\mathbf{x}^*, c; \gamma)_{t,t+\Delta t}}{\partial \gamma_r}}{I(\mathbf{x}^*, c; \gamma)} - \frac{\frac{\partial I(\mathbf{x}, c; \gamma)_{t,t+\Delta t}}{\partial \gamma_r}}{I(\mathbf{x}, c; \gamma)}, \end{aligned} \quad (\text{A1})$$

where  $r = 1, \dots, (m + J + C + J \times C)$ , and

$$\begin{aligned} \frac{\partial I(\mathbf{x}, c; \gamma)_{t,t+\Delta t}}{\partial \gamma_r} &= \frac{1}{n} \sum_{i=1}^n \left\{ \exp \left[ - \sum_{j=1}^J \exp(\mathbf{z}_{ij} \gamma) \delta_j(t + \Delta t) \right] \right. \\ &\quad \times \left( \sum_{j=1}^J \exp(\mathbf{z}_{ij} \gamma) \delta_j(t + \Delta t) \mathbf{z}_{ijr} \right) \\ &\quad \left. - \exp \left[ - \sum_{j=1}^J \exp(\mathbf{z}_{ij} \gamma) \delta_j(t) \right] \right. \\ &\quad \times \left( \sum_{j=1}^J \exp(\mathbf{z}_{ij} \gamma) \delta_j(t) \mathbf{z}_{ijr} \right) \left. \right\}; \end{aligned} \quad (\text{A2})$$

and

$$\frac{\partial I(\mathbf{x}^*, c; \gamma)_{t,t+\Delta t}}{\partial \gamma_r}$$

follows similarly by replacing  $\mathbf{x}_i$  with  $\mathbf{x}_i^*$ .

## APPENDIX 3

### Sample SAS Code for Calculating the Population Attributable Fraction for Total Mortality With the Piecewise Constant Hazards Model

The SAS program (SAS Institute Inc., Cary, North Carolina) for estimation of the population attributable fraction (PAF) and its 95% confidence interval requires the SAS procedures LIFEREG and IML and the following inputs:

DES\_BASE = design matrix  $((n \times J) \times (C + J + C \times J))$  for the baseline hazard parameters, which indicates to which categories of the baseline hazard variables (follow-up time intervals, birth cohorts, and their interactions) each individual belongs in each follow-up time interval.

DES\_COVAR = design matrix  $((n \times J) \times m)$  for observed covariates, which indicates which values of the risk factor each individual has.

DES\_STAR\_COVAR = design matrix  $((n \times J) \times m)$  for modified covariates, which indicates which values of the risk factor each individual has after the hypothetical change of the risk factors of interest.

EST\_BASE = vector  $((C + J + C \times J) \times 1)$  of parameter estimates for the baseline hazard variables obtained from the LIFEREG analysis.

EST\_COVAR = vector  $(m \times 1)$  of parameter estimates for the risk factors obtained from the LIFEREG analysis.

COVB\_ALL = covariance matrix  $((C + J + C \times J + m) \times (C + J + C \times J + m))$  of the parameter estimates for the baseline hazard variables and the risk factors obtained from the LIFEREG analysis.

Note that the parameter estimates obtained from the LIFEREG analysis are negative.

To estimate PAF for a chosen time interval ( $t, t + \Delta t$ ), the user must define the exposure at different time intervals up to time  $t$  (DELTA\_1) and time  $t + \Delta t$  (DELTA\_2). For example, to estimate PAF for the time interval (0, 20) when the follow-up period is divided into 4 5-year intervals, the user must define:

DELTA\_1 = {0, 0, 0, 0};

DELTA\_2 = {5, 5, 5, 5};

Then, the following SAS code can be applied to obtain the point estimate of PAF for total mortality and its lower and upper 95% confidence limits (IPAF\_CL\_l and IPAF\_CL\_u):

```
start_COLSUM_(inmatrix, outmatrix, groupsize);
  if missing(groupsize) then groupsize = nrow(inmatrix);
  ncolumns = ncol(inmatrix);
  outmatrix = btran(btran(inmatrix, groupsize,
                          ncolumns) [+ , ], 1, ncolumns);
finish_COLSUM_;

start_PAF_;

/* POINT ESTIMATE OF PAF */;

/* Number of follow-up time intervals */;
PERIODCOUNT = nrow (DELTA_1);

/* Hazard of death for observed and modified ('star') covariate values */;
lambda      = exp (DES_BASE * (-EST_BASE) +
                  DES_COVAR * (-EST_COVAR));
lambda_star = exp (DES_BASE * (-EST_BASE) +
                  DES_STAR_COVAR * (-EST_COVAR));

/* Exposure at different follow-up time intervals until time t */;
lambda_times_delta_t1 = lambda # DELTA_1;
lambda_star_times_delta_t1 = lambda_star # DELTA_1;

/* Exposure at different follow-up time intervals until time */;
lambda_times_delta_t2 = lambda # DELTA_2;
lambda_star_times_delta_t2 = lambda_star # DELTA_2;

/* Individual total exposure until time t: */;
/* sum of exposure at different follow-up time intervals */;
run_COLSUM_(lambda_times_delta_t1,
            sum_lambda_times_delta_t1, PERIODCOUNT);
run_COLSUM_(lambda_star_times_delta_t1,
            sum_lambda_star_times_delta_t1, PERIODCOUNT);

/* Individual total exposure until time : */;
/* sum of exposure at different follow-up time intervals */;
run_COLSUM_(lambda_times_delta_t2,
            sum_lambda_times_delta_t2, PERIODCOUNT);
run_COLSUM_(lambda_star_times_delta_t2,
            sum_lambda_star_times_delta_t2, PERIODCOUNT);

/* Survival until time t (see formula 6) */;
S_t1      = exp (-sum_lambda_times_delta_t1);
S_t1_star = exp (-sum_lambda_star_times_delta_t1);

/* Survival until time (see formula 6) */;
S_t2      = exp (-sum_lambda_times_delta_t2);
S_t2_star = exp (-sum_lambda_star_times_delta_t2);

/* Point estimate of PAF (see formula 7) */;

I      = (S_t1 - S_t2) [:,];
I_star = (S_t1_star - S_t2_star) [:,];
PAF     = 1 - (I_star/I);
```

```

%* CONFIDENCE INTERVAL OF PAF USING COMPLEMENTARY LOGARITHMIC TRANSFORMATION *;

%* Logarithmic transformation of PAF: log(1-PAF) *;
lPAF = log(1-PAF);

%* Design matrix including baseline hazard parameters *;
%* and observed (Z) or modified (Z*) covariates *;
DES_ALL      = DES_BASE || DES_COVAR;
DES_ALL_STAR = DES_BASE || DES_STAR_COVAR;

%* Derivative of I w.r.t.  $\gamma$  (see formula A2) *;
lambda_times_delta_t1_times_Z = lambda_times_delta_t1 # DES_ALL;
run_COLSUM_(lambda_times_delta_t1_times_Z,
             sum_lambda_times_delta_t1_times_Z, PERIODCOUNT);
lambda_star_times_delta_t1_times_Z_star = lambda_star_times_delta_t1 # DES_ALL_STAR;
run_COLSUM_(lambda_star_times_delta_t1_times_Z_star,
             sum_lambda_star_times_delta_t1_times_Z_star,
             PERIODCOUNT);
lambda_times_delta_t2_times_Z = lambda_times_delta_t2 # DES_ALL;
run_COLSUM_(lambda_times_delta_t2_times_Z,
             sum_lambda_times_delta_t2_times_Z,
             PERIODCOUNT);

lambda_star_times_delta_t2_times_Z_star = lambda_star_times_delta_t2 # DES_ALL_STAR;
run_COLSUM_(lambda_star_times_delta_t2_times_Z_star,
             sum_lambda_star_times_delta_t2_times_Z_star, PERIODCOUNT);

dI_gamma      = t((S_t2 #
                  sum_lambda_times_delta_t2_times_Z)
                  - (S_t1 #
                     sum_lambda_times_delta_t1_times_Z)
                  )[:,]);

dI_star_gamma = t((S_t2_star #
                  sum_lambda_star_times_delta_t2_times_Z_star)
                  - (S_t1_star #
                     sum_lambda_star_times_delta_t1_times_Z_star)
                  )[:,]);

%* Derivative of log(1-PAF) w.r.t  $\gamma$  (see formula A1) *;
dlPAF_gamma = (dI_star_gamma / I_star) - (dI_gamma / I);

%* Standard error of log(1-PAF) *;
se_lPAF = sqrt((t(dlPAF_gamma)*COVB_ALL) * dlPAF_gamma);

%* 95% confidence limits for inverse log(1-PAF) *;
lPAF_CL_l = 1 - exp(lPAF + PROBIT(0.975) *se_lPAF);
lPAF_CL_u = 1 - exp(lPAF - PROBIT(0.975) *se_lPAF);

print PAF lPAF_CL_l lPAF_CL_u;
finish_PAF_;

```