

## Empirical Bayes Adjustments for Multiple Results in Hypothesis-generating or Surveillance Studies<sup>1</sup>

Kyle Steenland,<sup>2</sup> Isabelle Bray, Sander Greenland, and Paolo Boffetta

National Institute for Occupational Safety and Health, Cincinnati, Ohio 45226 [K. S.]; School of Math and Statistics, University of Plymouth, Drake Circus, Plymouth PL48AA, United Kingdom [I. B.]; Department of Epidemiology, University of California Los Angeles School of Public Health, Los Angeles, California 90095-1772 [S. G.]; and IARC, 69372 Lyon, France [P. B.]

### Abstract

**Traditional methods of adjustment for multiple comparisons (e.g., Bonferroni adjustments) have fallen into disuse in epidemiological studies. However, alternative kinds of adjustment for data with multiple comparisons may sometimes be advisable. When a large number of comparisons are made, and when there is a high cost to investigating false positive leads, empirical or semi-Bayes adjustments may help in the selection of the most promising leads. Here we offer an example of such adjustments in a large surveillance data set of occupation and cancer in Nordic countries, in which we used empirical Bayes (EB) adjustments to evaluate standardized incidence ratios (SIRs) for cancer and occupation among craftsmen and laborers. For men, there were 642 SIRs, of which 138 (21%) had a  $P < 0.05$  (13% positive with  $SIR > 1.0$  and 8% negative with  $SIR \leq 1.0$ ) when testing the null hypothesis of no cancer/occupation association; some of these were probably due to confounding by nonoccupational risk factors (e.g., smoking). After EB adjustments, there were 95 (15%) SIRs with  $P < 0.05$  (10% positive and 5% negative). For women, there were 373 SIRs, of which 37 (10%) had  $P < 0.05$  before adjustment (6% positive and 4% negative) and 13 (3%) had  $P < 0.05$  after adjustment (2% positive and 1% negative). Several known associations were confirmed after EB adjustment (e.g., pleural cancer among plumbers, original SIR 3.2 (95% confidence interval, 2.5–4.1), adjusted SIR 2.0 (95% confidence interval, 1.6–2.4). EB can produce more accurate estimates of relative risk by shrinking imprecise outliers toward the mean, which may reduce the number of false positives otherwise flagged for further investigation. For example, liver cancer among chimney sweepers was reduced from an original SIR of 2.2 (range, 1.1–4.4) to an adjusted SIR of 1.1 (range, 0.9–1.4). A**

**potentially important future application for EB is studies of gene-environment-disease interactions, in which hundreds of polymorphisms may be evaluated with dozens of environmental risk factors in large cohort studies, producing thousands of associations.**

### Introduction

Traditional methods of adjustment for multiple comparisons (e.g., the Bonferroni adjustment) have fallen into disuse. It has been argued that such multiple comparison adjustments are unnecessary and in fact ill-advised because they assume a global null hypothesis, which is neither plausible nor of interest, and because they may lead investigators to ignore unexpected but important findings (1). When faced with a large number of comparisons, many epidemiologists currently do no adjustment at all but instead use whatever *a priori* knowledge exists, as well as common sense and biological plausibility, to evaluate what findings are important in their data.

However, Greenland and Robins (2) and Greenland and Poole (3) have argued that in some circumstances, EB<sup>3</sup> or semi-Bayes adjustments can be useful as an alternative to traditional multiple comparison adjustments. EB or semi-Bayes adjustments are useful under circumstances when (a) a large number of comparisons are made; (b) the comparisons can be grouped into sets within which all comparisons can be considered similar or “exchangeable;” (c) random error is present and presumably accounts for much of the observed variation in the parameters estimated to evaluate the comparisons (e.g., relative risks, rate ratios, and regression coefficients); and (d) investigators must choose which comparisons to investigate further, and there is a significant cost to such additional investigations. Even if all these conditions are not fully met, EB or semi-Bayes adjustments may be useful. The methods used in this study also require that the estimated parameters have an approximately normal distribution, although other distributional assumptions are possible. For the sake of simplicity, we will refer hereafter to these parameters generically as relative risks, which are generally used after a log transformation to approximate normality.

A typical current example is a large occupational surveillance data set, where many relative risks are evaluated with few or no *a priori* beliefs about which ones might be real (i.e., causal). Another example is the analysis of many gene-environment interactions, in which relative risks for several environmental exposures are analyzed in conjunction with multiple genetic polymorphisms, some of which might confer susceptibility (4). Data on gene-environment interactions to date have involved relatively small data sets and expensive laboratory techniques to identify polymorphisms. Technology is evolving

Received 10/7/99; revised 6/6/00; accepted 6/19/00.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> I. B.’s work on this study was done while working at the IARC under a Special Training Award.

<sup>2</sup> To whom requests for reprints should be addressed, at National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, OH 45226.

<sup>3</sup> The abbreviations used are: EB, empirical Bayes; SIR, standardized incidence ratio.

that will provide the investigator with hundreds of polymorphisms simultaneously at low cost, with little information about which ones might be of *a priori* interest. Furthermore, this information will soon be available for large cohorts in which blood samples have been collected at baseline. Such studies will generate thousands or even tens of thousands of relative risks for different environmental agents combined with different genotypes. Another important application in which such methods can have dramatic impact is in adjustment for multiple confounders. Semi-Bayes adjustments of confounder coefficients can allow more thorough and realistic control for confounder effects, especially in settings in which the confounders are highly correlated with the study exposure (5).

The basic idea of EB adjustments for multiple associations, such as relative risks, is that the observed spread or variation of the estimated relative risks around their geometric mean is larger than the variation of the true but unknown relative risks. EB adjustments attempt to estimate this extra variation from the data at hand and then use this estimate to adjust the observed relative risks. Typically, this adjustment serves to pull or shrink outlying relative risks in toward their geometric mean, more so if the estimate to be adjusted has a large individual variance. This shrinkage attempts to anticipate “regression to the mean,” in which outlier observations tend to shrink toward (*i.e.*, become closer to) the mean upon obtaining new data. A consequence of this shrinkage is that the overall variance of the EB-adjusted estimates is smaller than that of the unadjusted estimates. The variance of each estimated relative risk is also reestimated. Furthermore, although the individual EB-adjusted estimates are not statistically unbiased, the average squared error of the adjusted estimates will be less than the average squared error of the original estimates. EB estimators are part of a class of “shrinkage” estimators with a long history in the statistical literature (6–10). This class includes estimators based on hierarchical, multilevel, and mixed (random) coefficient models; see the article by Greenland (11) for a nontechnical review of this class.

In semi-Bayes adjustments, the investigator specifies or chooses an *a priori* value for the extra variation. If the investigator cannot specify what this extra variation might be, then EB adjustments may be used. EB adjustments may be inherently more satisfying to epidemiologists than semi-Bayes adjustments because all parameters are estimated from the data without any *a priori* specification by the investigator. However, when accurate prior information regarding the parameters is available (*e.g.*, a reasonable range in which they are likely to fall), semi-Bayes estimates can outperform EB estimates (12).

Forecasting examples and simulation studies (in which the true values of all parameters are known, and observed values with random error are generated) have been used to show that these types of adjustments can provide more accurate point estimates and narrower confidence intervals than the original estimates (9, 10, 12–15), as predicted theoretically (6–8, 10). Such accuracy improvements have also appeared in real examples (4, 12, 16, 17).

Here we provide an example of the use of EB adjustments in a large surveillance data set of occupation and cancer in the Nordic countries. The statistics are not very complicated in our example. We also provide an S-plus program for the interested reader in the “Appendix.” A more general regression program in SAS, PROC GLIMMIX, can also be easily adapted to carry out these analyses (18).

## Materials and Methods

**General Perspective.** Following the presentation by Greenland and Poole (3), the variance of the true log relative risks ( $\text{Var}_{\text{true}}$ ) can be estimated approximately by

$$\text{Var}_{\text{true}} \cong \text{Var}_{\text{obs}} - \text{Var}_{\text{mean}} \quad (1)$$

where  $\text{Var}_{\text{obs}}$  is the observed sample variance of the log relative risk estimates, and  $\text{Var}_{\text{mean}}$  is the mean of the estimated variances of each estimate (*e.g.*, each observed log relative risk has an estimated variance, and  $\text{Var}_{\text{mean}}$  is the mean of these estimated variances). In practice,  $\text{Var}_{\text{obs}}$  and  $\text{Var}_{\text{mean}}$  are weighted statistics. The weights themselves depend on the estimated  $\text{Var}_{\text{true}}$  and therefore must be derived iteratively (see below).

Two comments can be made about the formula above. It follows from the above equation that the observed variance of the estimates can be decomposed into the variance of the true log relative risks and the average variance of the individual estimates, *i.e.*, into the variance of the true values plus a component due to random error. Furthermore, because  $\text{Var}_{\text{true}}$  must be positive,  $\text{Var}_{\text{obs}}$  must be greater than  $\text{Var}_{\text{mean}}$  for the procedure to work. When the estimated variances fail to satisfy this inequality, the approximate EB methods used here must be abandoned in favor of semi-Bayes methods, in which the  $\text{Var}_{\text{true}}$  is specified by the investigator. For example, Greenland and Poole (3) in one example specify that the  $\text{Var}_{\text{true}}$  is 0.25 ( $\text{SD}_{\text{true}} = 0.5$ ), which (assuming normality) implies that 95% of the true relative risks are within a 7-fold range of each other (based on the width of the 95% prior probability interval, *e.g.*,  $\exp(2(1.96)0.5) = 7.1$ ). Caution must be exercised in choosing such a range, of course; the range should reflect what has been found in other studies of the same type.

There are two important restrictions to the types of data for which one can use these kinds of Bayesian adjustments. First, as mentioned above, the original observations to be adjusted must be able to be considered as arising from an “ensemble” or population of true relative risks that have an approximately log normal distribution around some single unknown geometric mean relative risk. For example, if the mean of the log relative risks is 0, some true log relative risks may be greater than 0 because exposure causes disease, and some may be less than 0 because exposure protects against disease, whereas a large number will be clustered around 0 because exposure has little or no effect on disease. However, if some exposures are known to cause disease whereas others are not, it would not be appropriate to treat them as coming from a single distribution. This means that some exposure-disease relative risks should not be grouped with others. For example, one should not include relative risks for diet items and lung cancer with relative risks for smoking and lung cancer in a Bayesian adjustment because the two sets of log relative risks would not be expected to have the same mean. This example is rather extreme; lung cancer relative risks are very high for smoking-related variables. In many practical situations, there will be few established exposure-disease associations, and even for them, relative risks may be expected to be only modestly elevated, so that investigators will not need to create separate ensembles within their data.

A second restriction is that quantitative exposures must be scaled so that the log relative risks (or regression coefficients in linear regression) from an increment of one unit of exposure are comparable, *i.e.*, can be expected to have the same mean. This restriction, for example, would be important in a study of neurological function in which a battery of different tests had been done, *e.g.*, 10 nerve conduction tests, 10 tests of postural sway, and 10 tests of tremor. All these subgroups of tests would

be measured on a different scale (meters per second, centimeters of sway, and frequency of tremor). Assume that 100 regressions are done in which exposed are compared to nonexposed while adjusting for covariates like age and sex. The 100 resulting regression coefficients for exposure will estimate the change in neurological function for the exposed *versus* nonexposed, but the different neurological functions will have been measured on different scales, and the coefficients (and their variances) therefore will not be comparable. These coefficients cannot all be assumed to come from some common distribution, as required by the EB method used here. Nevertheless, they can be transformed to meet this requirement or analyzed with more general EB regression methods that relax this requirement (12, 13).

Finally, if the relative risks of interest have prior associations (in that new information about one would change our expectation for another), these associations should be taken into account. The simplest way to do so is to group or regress the relative risks on factors that explain the associations (12). Similarly, if the relative risk estimates are statistically associated, these associations can and should be taken into account, as in matrix-weighted and penalized likelihood methods (12, 15). We consider here only the simpler case in which the associations among the relative risks or among their estimates are absent or negligible.

**Statistical Methods.** Again following Greenland and Poole (3), consider a group of exchangeable SIR estimates, an “ensemble” of SIRs, each with an estimated variance, and let  $i$  denote individual members of this ensemble. Taking logarithms, we derive a weighted average, as shown below.

$$\ln \text{SIR}_{\text{mean}} = \sum (\ln \text{SIR}_i * w_i) / \sum w_i \quad (2)$$

The weights have the form

$$w_i = 1 / (S_i^2 + \hat{\text{V}}_{\text{true}}) \quad (3)$$

where  $S_i^2$  is the variance of each  $\ln \text{SIR}_i$ , and  $\hat{\text{V}}_{\text{true}}$  is the estimated  $\text{Var}_{\text{true}}$ . Note that because  $\hat{\text{V}}_{\text{true}}$  is the result of these calculations in an EB analysis, we cannot know it at the beginning ( $\text{Var}_{\text{true}}$  is specified at the start of a semi-Bayes analysis). Thus EB analyses require the use of iteration, where an initial guess of  $\text{Var}_{\text{true}}$  is used, and then this initial guess is refined iteratively. Now let

$$D = \ln \text{SIR}_i - \ln \text{SIR}_{\text{mean}} \quad (4)$$

and

$$\hat{\text{V}}_{\text{obs}} = \sum (w_i * D_i^2) / \sum w_i \quad (5)$$

so  $\hat{\text{V}}_{\text{obs}}$  is our estimate of  $\text{Var}_{\text{obs}}$ . Then derive the estimate for  $\text{Var}_{\text{mean}}$  as described below.

$$\hat{\text{V}}_{\text{mean}} = \sum (w_i * S_i^2) / \sum w_i \quad (6)$$

We can now estimate  $\text{Var}_{\text{true}}$  by  $\hat{\text{V}}_{\text{true}} = \hat{\text{V}}_{\text{obs}} - \hat{\text{V}}_{\text{mean}}$  or by  $\max(\hat{\text{V}}_{\text{obs}} - \hat{\text{V}}_{\text{mean}}, \tau^2)$ , where  $\tau^2$  is a user-specified minimum plausible value for  $\text{Var}_{\text{true}}$ . Finally, we can derive our EB estimate of each true SIR as a weighted average of the original estimate and the mean of the estimates as described below.

$$EB_i = ((\hat{\text{V}}_{\text{true}} * \ln \text{SIR}_i) + (S_i^2 * \ln \text{SIR}_{\text{mean}})) / (\hat{\text{V}}_{\text{true}} + S_i^2) \quad (7)$$

Note that if  $\hat{\text{V}}_{\text{true}}$  is large, this gives more weight to the original estimate. On the other hand, if the variance of the individual estimate  $S_i^2$  is large, this gives more weight to the overall mean of the estimates.

Details of the iterative methods necessary to do the above

calculations are provided in the appendix to Greenland and Poole (3), supplemented by an improved formula for the variance of the adjusted estimates found in Greenland (15); these two articles provide the basis for the S-plus program in the “Appendix.” Note that this program is quite general and can account for associations among the relative risks by grouping of some subensembles within the general ensemble of interest (but assumes the same estimated  $\text{Var}_{\text{true}}$  across subensembles). It also allows for correlations among the relative risk estimates. When there is no subgrouping of ensembles (subensembles), and there is no correlation among the estimated relative risks, the matrices in the S-plus program reduce to vectors, and the computations are considerably simplified.

**Data Used for Example.** Data for our example were derived from a record-linkage study of cancer by occupational group in the Nordic countries (19). In this study, occupation was recorded at the time of the 1970 or 1971 census for the population aged 25–64 years of Sweden, Denmark, Norway, and Finland (approximately 10.1 million people). Follow-up was conducted through 1987–1991 for cancer incidence, with the exact date varying by country. The four countries had nationwide cancer registration during this period. Indirectly standardized SIRs for each sex, using the whole population as the referent, were calculated for 35 cancers and 53 occupational groups after stratification of the data by age, calendar time, and country. Exact confidence intervals were calculated for SIRs with less than 100 observed cases by assuming a Poisson distribution for the observed cases, whereas an approximation (based on a normal approximation to the Poisson distribution) was used for SIRs with more than 100 cases.

Following the suggestion in Greenland and Poole (3) to exclude estimates based on very small numbers, we eliminated all SIRs for which there were five or fewer observed cases. We furthermore eliminated all SIRs for the “unknown” cancer site and also eliminated four occupational categories, *i.e.*, those not economically active, those active but not otherwise classifiable, those in the armed forces, and those in the public safety/protection group (a heterogeneous group). This left 2357 cancer/occupation combinations.

The published results across all occupational groups exhibit considerable confounding of occupation-specific results by nonoccupational variables related to socioeconomic status. For example, women in higher socioeconomic classes (in profession/managerial occupations) had more breast cancer, likely reflecting different patterns in reproductive behavior (Fig. 1) rather than different exposures to occupational breast carcinogens. Similarly, men and women in higher socioeconomic classes had less smoking-related cancers such as lung, larynx, and bladder cancer. These patterns are well known. We did two things to the data to limit confounding by nonoccupational variables related to social class and to create an ensemble of exchangeable SIRs for EB adjustments.

First, we restricted our analyses to manual laborers and craftsmen, a relatively homogenous socioeconomic group. We were particularly interested in this group because we felt it was more likely to be exposed to occupational carcinogens. There were 23 distinct occupations in this group; miners (men only), seamen (men only), drivers, textile workers, leather workers, foundry workers, mechanic/metalworkers, plumbers (men only), welders (men only), electricians, woodworkers, painters, construction, bricklayers (men only), printers, chemical workers, food manufacture workers, beverage manufacture workers, tobacco manufacture workers, glass/ceramic workers, packers/warehouse workers, engine operators, and chimney sweeps

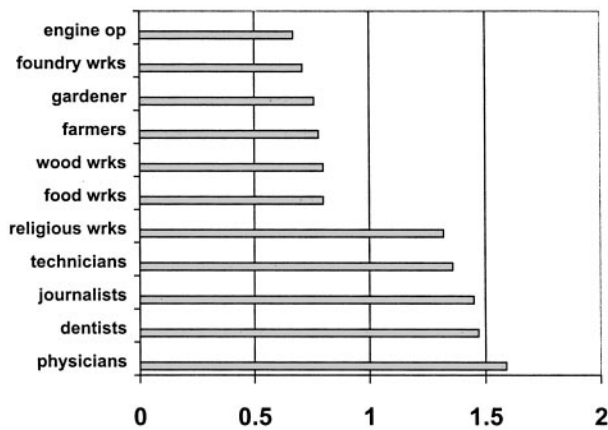


Fig. 1. Highest and lowest SIRs for female breast cancer.

Table 1 Summary of cancer and occupation SIRs for manual laborers/craftsmen with  $P < 0.05$ , adjusted for social class, before and after EB adjustment

	Positive, <sup>a</sup> before EB Adjustment	Negative, <sup>a</sup> before EB adjustment	Positive, after EB adjustment	Negative, after EB adjustment
Total, 1015	107 (11%)	68 (7%)	71 (7%)	37 (4%)
Men, 642	84 (13%)	54 (8%)	62 (10%)	33 (5%)
Women, 373	23 (6%)	14 (4%)	9 (2%)	4 (1%)

<sup>a</sup> Positive, SIR > 1.0; negative, SIR ≤ 1.0.

(men only). There were 1015 occupation-specific, sex-specific, cancer-specific SIRs within this group.

Second, for men and women separately, we made an adjustment for social class by scaling the observed 23 occupation-specific SIRs for each cancer in the manual laborer/craftsmen group so that each overall (across all occupations) sex-specific, cancer-specific SIR would be 1.0. We did this because the reported cancer-specific SIRs in the manual laborer/craftsmen group were calculated in reference to the total Nordic population (across all socioeconomic groups) and hence would reflect the influence of nonoccupational variables (e.g., smoking) that differ across socioeconomic groups. For example, if the overall lung cancer SIR for male manual laborers/craftsmen was 1.5, all 23 occupation-specific lung cancer SIRs in this group were multiplied by  $1/1.5$ , so that the overall lung cancer SIR after this adjustment in this group became 1.0. In effect, this adjustment was equivalent to adjusting the expected number of group-specific cancers without affecting the observed number. The variance was not correspondingly adjusted. The adjustment made the overall SIR (across all cancers and all occupations and both sexes) for manual laborers/craftsmen equal to 1.0, creating an ensemble of interchangeable SIRs for our EB adjustments.

## Results

Table 1 shows the data for manual laborers/craftsmen, adjusted for social class, before and after EB adjustment. Before adjustment, both men and women (especially men) show more than the expected 2.5% of observations with  $P < 0.05$  in each tail, with more excesses in the upper tail. Fig. 2, *a* and *b*, shows the distribution of  $P$  values for the data in Table 1 before EB

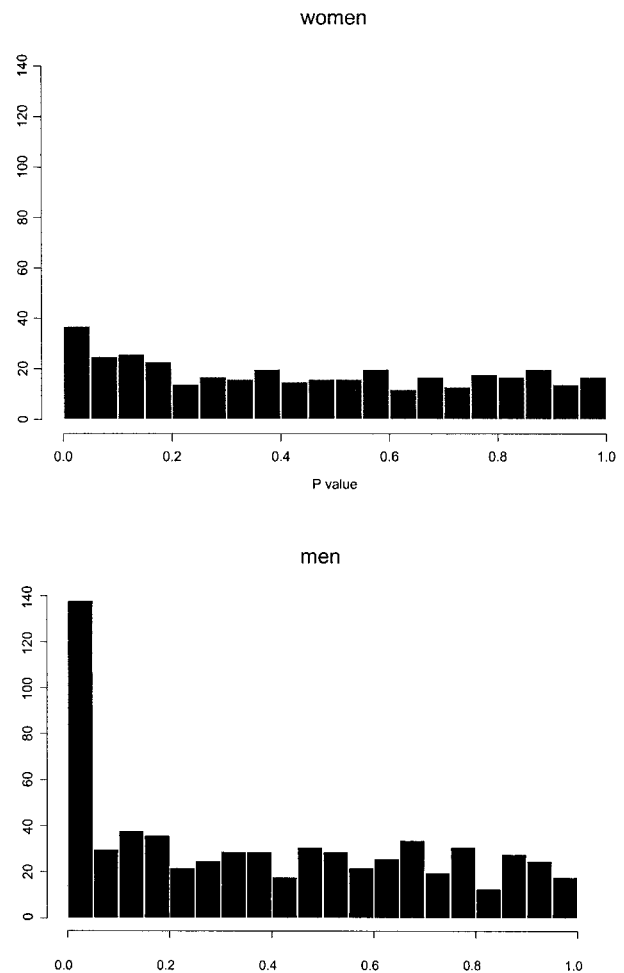


Fig. 2. *a*,  $P$  values for SIRs for female manual laborers/craftsmen. *b*,  $P$  values for SIRs for male manual laborers/craftsmen.

adjustment, for females and males. The observed positive (SIR > 1.0) findings with low  $P$  values are presumably due to a combination of true associations due to occupational carcinogens, residual confounding by nonoccupational variables related to occupation within the manual laborer/craftsmen group, and random error. After EB adjustment, there are fewer observations with low  $P$  values, with women showing approximately the expected numbers of very high and very low SIRs, but men still showing more than the expected number, especially for high SIRs. Fig. 3, *a* and *b*, shows the distribution of  $P$  values for females and males after EB adjustments. We think it is likely that the remaining lifestyle differences (residual confounding) among occupations account for some of the elevated relative risks among men with low  $P$  values, but that some others are probably due to the effect of occupational carcinogens. The adjusted  $P$  values for women are skewed toward 1; however, this can result from overadjustment by the EB procedure due to underestimation of  $V_{\text{true}}$  and may be avoided by using semi-Bayes adjustments instead (see below).

For men (642 relative risks), the EB estimate of overall (weighted) mean of the lnSIRs was 0.020, corresponding to a mean SIR of 1.02. The unweighted sample variance of the lnSIRs was 0.056, whereas the weighted variance estimated by

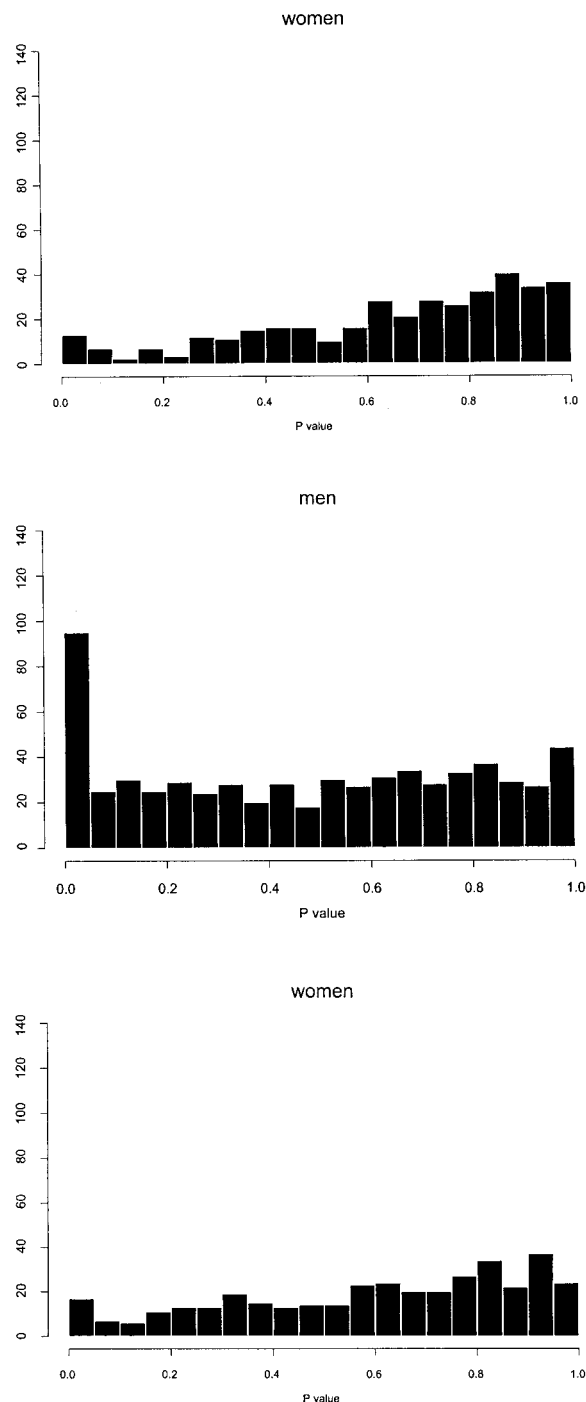


Fig. 3. *a*, *P* values for SIRs for females after EB adjustment. *b*, *P* values for SIRs for males after EB adjustment. *c*, *P* values for SIRs for females after semi-Bayes adjustment (assuming that true variance of lnSIRs for females was equal to that of males, *i.e.*, 0.019).

EB ( $\hat{V}_{\text{obs}}$ ) was 0.032. The weighted average of the individual variances estimated by EB ( $\hat{V}_{\text{mean}}$ ) was 0.012. The EB estimated variance of the true lnSIRs ( $\hat{V}_{\text{true}}$ ), approximately the difference between  $\hat{V}_{\text{obs}}$  and  $\hat{V}_{\text{mean}}$ , was thus 0.019. The large size of  $\hat{V}_{\text{obs}}$  compared with  $\hat{V}_{\text{true}}$  resulted in considerable shrinkage of the new EB-estimated lnSIRs toward the mean of

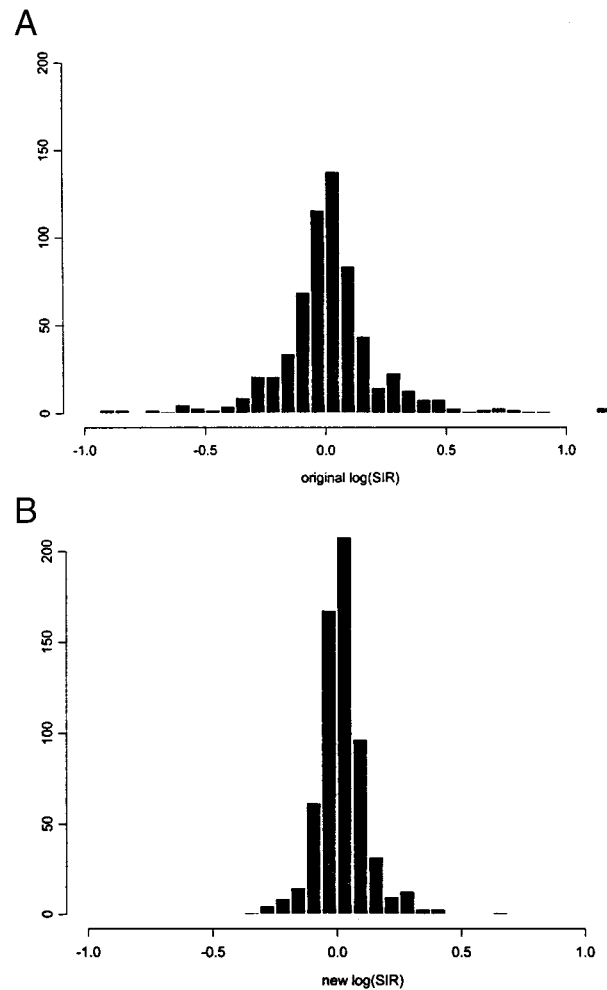


Fig. 4. *a*, lnSIRs for male manual laborers/craftsmen before EB adjustment. *b*, lnSIRs for male manual laborers/craftsmen after EB adjustment.

1.02 (Fig. 4, *a* and *b*). The mean estimate and  $\hat{V}_{\text{true}}$  need to be evaluated against background knowledge by translating them to an estimated 95% prior interval for the relative risks. For men, this interval is  $\exp(0.020 \pm 1.96(0.019)^{1/2})$ , or 0.78–1.28.

For women (373 relative risks), the EB-estimated overall mean of the lnSIRs was 0.015 (SIR, 1.02). The unweighted sample variance of the lnSIRs was 0.066,  $\hat{V}_{\text{obs}}$  was 0.031, and  $\hat{V}_{\text{mean}}$  was 0.022, resulting in a  $\hat{V}_{\text{true}}$  of 0.009. This again resulted in considerable shrinkage of the lnSIRs toward their mean (Fig. 5, *a* and *b*). The estimated 95% prior interval for the relative risks is  $\exp(0.015 \pm 1.96(0.009)^{1/2})$  or 0.84–1.19, which may be too narrow, consistent with the impression from Fig. 3*a* that  $V_{\text{true}}$  is underestimated here.

For females, we also conducted a semi-Bayes adjustment in which we assumed that  $V_{\text{true}}$  for females was 0.019, the same as males, rather than the 0.009 estimated via EB. *P* values for the semi-Bayes-estimated SIRs for women are shown in Fig. 3*c* and tend to have a more uniform profile than those in Fig. 3*a*, although there is still some clumping toward higher *P* values. The corresponding log SIRs are shown in Fig. 5*c* and can be seen to have somewhat more spread than the EB-adjusted SIRs. In this analysis, there were 13 positive (SIR > 1.0) associations

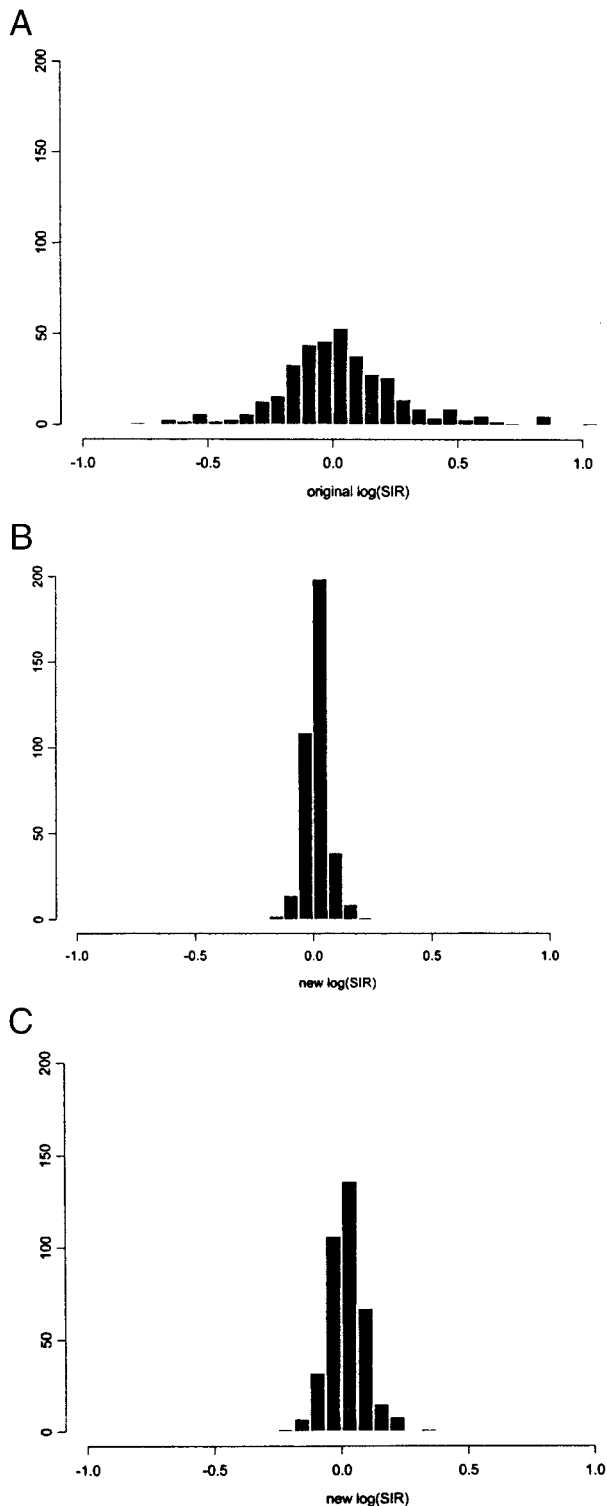


Fig. 5. *a*, lnSIRs for female manual laborers/craftsmen before EB adjustment. *b*, lnSIRs for female manual laborers/craftsmen after EB adjustment. *c*, lnSIRs for female manual laborers/craftsmen after semi-Bayes adjustments (assuming that the true variance of lnSIRs for females was equal to that of males, *i.e.*, 0.019).

Table 2 Selected positive (SIR > 1.0) findings suspected *a priori*, and supported by EB adjustment (males)

Occupation/cancer	Social class-adjusted SIR (95% CI) <sup>a</sup>	Social class- and EB-adjusted SIR (95% CI)
Plumber/pleura	3.21 (2.52–4.09)	1.96 (1.62–2.36)
Woodworkers/nasal	1.62 (1.39–1.90)	1.46 (1.27–1.67)
Miners/lung	1.37 (1.27–1.48)	1.34 (1.24–1.45)

<sup>a</sup> CI, confidence interval.

Table 3 Selected positive (SIR > 1.0) findings not suspected *a priori*, and not supported by EB adjustment (males)

Occupation/cancer	Social class-adjusted SIR (95% CI) <sup>a</sup>	Social class- and EB-adjusted SIR (95% CI)
Chimney sweep/liver	2.17 (1.06–4.43)	1.12 (0.87–1.45)
Printer/breast	2.08 (1.10–3.93)	1.14 (0.88–1.47)
Beverage worker/oral	2.25 (1.10–4.60)	1.13 (0.87–1.46)

<sup>a</sup> CI, confidence interval.

Table 4 Selected positive (SIR > 1.0) findings not suspected *a priori* and supported by EB adjustment

Occupation/cancer	Sex	Social class-adjusted SIR (95% CI) <sup>a</sup>	Social class- and EB-adjusted SIR (95% CI)
Mechanic/breast	M	1.41 (1.09–1.83)	1.21 (1.01–1.47)
Construction/lip	M	1.48 (1.35–1.63)	1.42 (1.30–1.56)
Welder/NHL <sup>b</sup>	M	1.23 (1.06–1.42)	1.17 (1.03–1.34)
Printer/breast	F	1.24 (1.13–1.37)	1.19 (1.09–1.30)
Machine operator/lung	F	2.36 (1.70–3.27)	1.23 (1.03–1.48)

<sup>a</sup> CI, confidence interval.

<sup>b</sup> NHL, non-Hodgkin's lymphoma.

with  $P < 0.05$  (3%) and 5 negative (SIR  $\leq 1.0$ ) associations with  $P < 0.05$  (1%), compared with 9 such positive and 5 such negative associations using EB.

Table 2 gives some results for relative risks for males that were suspected *a priori* (known from the literature) to be elevated; although reduced somewhat due to the EB adjustment, they remained unambiguously positive. Table 3 gives some results for male relative risks that were not previously suspected; these were reduced considerably by EB. The original findings for these relative risks are likely to have been false positives. Finally, Table 4 presents a few relative risks that remained slightly elevated and still had  $P < 0.05$  after EB adjustment. Lip cancer among construction workers may be due to solar radiation, whereas the others are largely unanticipated, although one can hypothesize that blood neoplasms might be increased among welders due to high EMF exposure.

## Discussion

It is not likely that occupational cancer surveillance of this sort will uncover many new true associations between cancer and occupation because (*a*) many strong carcinogens present in specific occupations have been identified previously, and (*b*) it is likely that only a minority of workers in a specific occupation have high exposures to specific previously unidentified carcinogens, decreasing the likelihood of their identification. This may be why relatively few dramatic positive findings “survive” the EB adjustments.

We have not tried to systematically analyze the results for

this data set; instead, we selected only a few associations to highlight as examples. Readers are referred to the original publication for a detailed discussion of results (19). Our main purpose here has been to illustrate the methods rather than try to identify all associations worth further investigation in this population. Furthermore, our adjustment for socioeconomic class was necessarily crude, undoubtedly leaving residual confounding by nonoccupational risk factors.

One important result of using Bayes adjustments in this type of study is the weeding out of false positives. Of course, some true associations may also be weeded out in the process. The usefulness of these kinds of adjustments therefore depends on the cost associated with further investigation of false positives. These costs will be specific to each type of study, and few generalizations can be made. If the result of a false positive in a large number of cancer-gene-environment associations is the launching of a case-control study to try to replicate an original finding, the cost may be high, and the EB adjustment may therefore be worthwhile. For presentation of results, investigators might wish to present both the original unadjusted findings and the Bayesian-adjusted findings. Assuming a decision rule for further investigation is based on an elevated SIR with  $P < 0.05$ , based on unadjusted results investigators would further investigate 84 of 642 (13%) SIRs among male laborers/craftsmen. With an EB adjustment, investigators would further investigate 62 SIRs (10%). With a Bonferroni adjustment ( $P < 0.000778$  or  $0.05/642$ ), investigators would pursue only 18 SIRs (3%). Thus, the EB adjustment results in a decision rule that is between no adjustment and a Bonferroni adjustment with regard to the number SIRs selected for further investigation. EB adjustments result in a selection of parameter estimates that are more accurate than those selected by the classical decision rules; therefore, these adjustments should result in a more accurate prediction of future results than either classical rule.

It is worth noting that the percentage of unambiguous associations remaining in Table 3 after EB adjustment is highest for positive associations in males, in accordance with the *a priori* belief that true positive occupational associations are most likely to occur among males (true associations because

men are more likely than females to have had higher and longer exposures to occupational carcinogens, and positive associations because occupational agents are unlikely to protect against cancer). This may be an indication that the adjustment has been relatively successful in weeding out false positives but retaining true positives.

We have used an example from a field, occupation and cancer, in which there is a general body of *a priori* knowledge (although not necessarily directly applicable to most of the occupation/cancer associations studied in the Nordic population). Thus, in our example, EB adjustments are only a supplement to attempts to weed out false positives based on *a priori* considerations (*e.g.*, knowledge of likely exposures within occupations). Future cancer-gene-environment studies, on the other hand, are likely to produce a very large number of associations about which there is little or no *a priori* knowledge whatsoever; EB or semi-Bayes adjustments may prove essential in making sense of such data.

### Acknowledgments

We thank Aage Andersen for comments regarding the Nordic occupation/cancer surveillance study and anonymous journal reviewers for helpful comments on the manuscript. Giles Ferro provided valuable computer support.

### Appendix

**S-Plus Program to Perform EB Analysis of Epidemiological Data with Correlations.** The program (Table 5) requires a vector of original estimates (bhat), a corresponding variance-covariance matrix (vhat), a matrix to describe prior knowledge (z), and the number of iterations to be carried out (niter). Let  $n$  be the number of log relative risks to be estimated. For our simple example, we do not incorporate prior knowledge in the form of grouping into subensembles or other prior knowledge constraining the original estimates. Therefore,  $z$  is simply a column vector of 1s with length  $n$ , and each estimate will be pulled toward the same mean ( $\ln \text{SIR}_{\text{mean}}$ ). However, more generally, suppose we wish to define  $g$  groups (subensembles), each with a different mean but the same  $\text{Var}_{\text{true}}$ . Then  $z$  comprises  $g$  columns of dummy variables indicating the group membership of each of the  $n$  log relative risks. More generally still,  $z$  may include other ancillary information upon which the original estimates can be regressed. In the two-stage modeling example presented by Witte *et al.* (17),  $z$  had 87 rows corresponding to the food constituents in 87 dietary items, for which the original estimates had been derived in the first stage.

Table 5 S-Plus program

```
function (bhat, vhat, z, niter)
{
### Define the number of estimates (n) and the number of parameters in the prior model (p)
  n <- length (bhat)
  p <- ncol (z)
### Set initial value for Vhattrue (vart) to 0
  vart <- 0
### Generate vector to store values of Vhattrue from each iteration
  v <- vector (length = niter + 1)
  v [1] <- 0
### Generate column of 1s for use in calculations
  I <- matrix (1, n, 1)
### Start iterative loop
  for (i in 1:niter) {
### Define weights (W) and calculate their sum (w)
    W <- solve (vhat + vart * diag(n))
    w <- t(I) %*% W %*% I
### Calculate lnSIRmean (pi) and a vector of prior means (mu)
    pi <- solve (t(z) %*% W %*% z) %*% t(z) %*% W %*% bhat
    mu <- z %*% pi
### Calculate D, Vhatobs (varo) and Vhatmean (varm)
    D <- bhat - mu
    varo <- t(D) %*% W %*% D/w
    varm <- (t(I) %*% W %*% vhat %*% I)/w
```

Table 5 Continued

```

### Calculate new estimate of Vartrue, set to zero if negative and
### store in vector v
  vart <- (n/(n - p)) * varo - varm
  vart <- max(vart, 0)
  v[i + 1] <- vart
### Terminate iterations when Vhattrue has converged (at 10-8) and calculate
new estimates of lnSIR (newbeta) and var(lnSIR) (newvar)
  if (Mod(v[i + 1] - v) < 1e-007) {
    vart <- v[i + 1]
    vstar <- (((n - p - 2) * vhat)/(n - p))
    tstar <- vart * diag(n) + vhat - vstar
    newbeta <- W %%% ((tstar %%% bhat) + (vstar %%% mu))
### The definition of newvar is taken from Ref. 15
    H <- z %%% (solve (t(z) %%% W %%% z)) %%% t(z) %%% W
    bstar <- ((n - p - 2) * W %%% vhat)/(n - p)
    E <- bstar %%% D * sqrt (varm/diag(vhat))
    newvar <- diag(vhat - (t(diag(n) - H) %%% vhat %%% bstar) +
      (2 * E %%% t (E))/(n - p - 2))

    break
  }
}
}

```

In an example such as that presented in this article, for which we assume no correlations between the original estimates (and no prior knowledge is incorporated), the above program simplifies. Vhat becomes a vector of variances instead of a variance-covariance matrix, and z is no longer necessary.

```

function (bhat, vhat, niter)
{
### Define the number of estimates (n)
  n <- length (bhat)
### Set initial value for Vhattrue (vart) to 0
  vart <- 0
### Generate vector to store values of Vhattrue from each iteration
  v <- vector (length = niter + 1)
  v [1] <- 0
### Start iterative loop
  for (i in 1:niter) {
### Define a vector of weights (w) and calculate lnSIRmean (pi)
### vhat is a vector of the variances of the original estimates
    w <- 1/(vhat + vart)
    pi <- sum (w * bhat)/sum(w)
### Calculate D, Vhatobs (varo) and Vhatmean (varm)
    D <- bhat - pi
    varo <- sum(w * D * D)/sum(w)
    varm <- sum(w * vhat)/sum(w)
### Calculate new estimate of Vartrue, set to zero if negative and
### store in vector v
    vart <- max((varo - varm), 0)
    v[i + 1] <- vart
### Terminate iterations when Vhattrue has converged and calculate new
estimates of lnSIR (newbeta) and var(lnSIR) (newvar)
    if (Mod(v[i + 1] - v) < 1e-007) {
      vart <- v[i + 1]
      newbeta <- w*((vart * bhat) + (vhat * pi))
      E <- w * vhat * D * sqrt(varm/vhat)
      newvar <- vhat * (1 - vhat * w) + (2 * E * t(E))/n

      break
    }
  }
}

```

## References

1. Rothman, K. No adjustments are needed for multiple comparisons. *Epidemiology*, *1*: 43–46, 1990.
2. Greenland, S., and Robins, J. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. *Epidemiology*, *2*: 244–251, 1991.
3. Greenland, S., and Poole, C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch. Environ. Health*, *48*: 9–16, 1994.
4. Aragaki, C., Greenland, S., Pobst-Hensch, N., and Haile, R. Hierarchical modeling of gene-environment interactions: estimating NAT2 genotype-specific dietary effects on adenomatous polyps. *Cancer Epidemiol. Biomark. Prev.*, *6*: 307–314, 1997.
5. Greenland, S. When should epidemiologic regressions use random coefficients? *Biometrics*, in press, 2000.
6. Robbins, H. An empirical Bayes approach to statistics. In: J. Neyman (ed.), *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 131–148. Berkeley, CA: University of California Press, 1955.
7. Good, I. On estimation of small frequencies in contingency tables. *J. R. Stat. Soc. B*, *18*: 113–124, 1956.
8. James, W., and Stein, C. Estimation with quadratic loss. In: J. Neyman (ed.), *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 361–379. Berkeley, CA: University of California Press, 1961.
9. Good, I. *The Estimation of Probabilities*. Cambridge, MA: MIT Press, 1962.
10. Efron, B., and Morris, C. Stein's estimation rule and its competitors: an empirical Bayes approach. *J. Am. Stat. Assoc.*, *68*: 117–130, 1973.
11. Greenland, S. Principles of multilevel modeling. *Int. J. Epidemiol.*, *29*: 158–167, 2000.
12. Greenland, S. Methods for epidemiologic analyses of multiple-exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. *Stat. Med.*, *12*: 717–736, 1993.
13. Morris, C. Parametric empirical Bayes theory and applications (with discussion). *J. Am. Stat. Assoc.*, *78*: 47–65, 1983.
14. Witte, J., and Greenland, S. Simulation study of hierarchical regression. *Stat. Med.*, *15*: 1161–1170, 1996.
15. Greenland, S. Second-stage least squares *versus* penalized quasi-likelihood for fitting hierarchical models in epidemiologic analyses. *Stat. Med.*, *16*: 515–526, 1997.
16. Efron, B., and Morris, C. Data analysis using Stein's estimator and its generalization. *J. Am. Stat. Assoc.*, *70*: 311–319, 1975.
17. Witte, J., Greenland, S., Haile, R., and Bird, C. Hierarchical regression analyses applied to a study of multiple dietary exposures and breast cancer. *Epidemiology*, *5*: 612–621, 1994.
18. Witte, J., Greenland, S., Kim, L., and Arab, L. Multilevel modeling in epidemiology with GLIMMIX. *Epidemiology*, in press, 2000.
19. Andersen, A., Barlow, L., Engeland, A., Kjaerheim, K., Lynge, E., and Pukkala, E. Occupational cancer in the Nordic countries. *Scand. J. Work Environ. Health*, *25* (Suppl. 2): 1–116, 1999.

# Cancer Epidemiology, Biomarkers & Prevention

AACR American Association  
for Cancer Research

## Empirical Bayes Adjustments for Multiple Results in Hypothesis-generating or Surveillance Studies

Kyle Steenland, Isabelle Bray, Sander Greenland, et al.

*Cancer Epidemiol Biomarkers Prev* 2000;9:895-903.

**Updated version** Access the most recent version of this article at:  
<http://cebp.aacrjournals.org/content/9/9/895>

**Cited articles** This article cites 13 articles, 1 of which you can access for free at:  
<http://cebp.aacrjournals.org/content/9/9/895.full#ref-list-1>

**Citing articles** This article has been cited by 5 HighWire-hosted articles. Access the articles at:  
<http://cebp.aacrjournals.org/content/9/9/895.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cebp.aacrjournals.org/content/9/9/895>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.