269

When Does Factoring Increase The Accuracy Of Rate Estimation? G. Maldonado,* S. Greenland (University of Minnesota, Minneapolis, MN 55455 and University of California, Los Angeles, CA 90024).

A common epidemiologic strategy for modeling a continuous or ordered discrete variable is to transform it into a series of dichotomous indicator variables (i.e., to factor the variable). Factoring tends to increase estimation variance, but also tends to decrease bias, and thus may increase total accuracy (as measured by mean squared error). However, the conditions under which factoring increases accuracy are not well understood. The authors conducted a simulation study to examine the impact of factoring on the accuracy of rate estimation. Factored and unfactored Poisson regression models were fit to follow-up study data sets that were randomly generated from 18,750 population model forms that ranged from sub-additive to supra-multiplicative. In the situations we examined, factoring sometimes substantially improved accuracy relative to fitting the corresponding unfactored model, but sometimes substantially decreased accuracy, and sometimes made little difference. The difference in accuracy between factored and unfactored models depended in a complicated fashion on the difference between the true and fitted model forms, the strength of exposure and covariate effects in the population, and the sample size. Factoring always decreased accuracy when effects were weak, especially at a smaller sample size. Factoring always provided the largest increase in accuracy when one of two variables in the model had a weak effect and the other variable had a strong effect. As expected from theory, factoring always performed better as the sample size increased.

270

An Empirical Evaluation of Hospital versus Random Control Selection for Case-Control Studies using an HMO Population. S.K. Van Den Eeden* (Division of Research, Kaiser Permanente, Oakland, CA 94611).

In case-control studies, the sampling framework for the controls is determined by the population from which the cases arose. The choice of the population itself is often dictated by either the difficulty in assembling cases or by the availability of certain data. Two common methods are random selection from a defined, usually geographic, population and random selection from a hospitalized population. The author sought to compare selected baseline characteristics from the same base cohort (n=206,974) who underwent a multiphasic health checkup (MHC) between July 1964 and August 1973. Only subjects free of cancer at the time of the checkup were included in this study. The MHC is a voluntary general health check-up. As part of the multiphasic health checkup, members provided information on demographic characteristics, tobacco and alcohol use, and individual and family medical history. In addition, laboratory data were obtained which included a complete blood count, roentgenographies of the lung, and pulmonary function. An arbitrary study period, 1976-1979, from which to select idealized controls was defined. One group was 20,000 individuals randomly selected from the population who were still alive and a member (called the random group), while the other included 20,000 individuals randomly selected from the living members who were hospitalized. Compared to the hospitalized group, the random group reported less often being remarried or separated, divorced or widowed ($p<0.001$), more likely to be a current smoker ($p<0.001$), more likely to drink less ($p=0.008$), to have a lower mean diastolic ($p<0.001$) and systolic ($p<0.001$) blood pressure and have a lower body mass index ($p<0.001$).

271

Toenail Selenium Level and Lung Cancer among Men and Women in a High Seleniferous Region of the USA. J. Ma,* M.J. Stampfer, J.S. Morris, W.C. Willett. (Harvard University, Boston, MA 02115).

In animal studies and ecologic studies, higher selenium levels have been associated with decreased risks of cancer at various sites, but studies among individuals have been inconsistent. To address these issues, the authors conducted a case-control study in a region with high variability of selenium status. Toenail clippings provide a stable time-integrated measure of selenium status. These were collected in southwestern South Dakota and northeastern Wyoming from patients (141 men and 65 women) with newly diagnosed lung cancer and from randomly selected controls (217 men and 351 women), and analyzed for selenium content. The mean selenium level in toenails was lower among cancer cases than controls: for men 1.03 vs 1.14 ppm, $P = 0.0007$; for women 1.09 vs 1.11 ppm, $P = 0.6$ respectively. After adjusting for age, county of residency, and smoking, the relative risks of lung cancer in the three higher quartiles, compared with the lowest quartile, for men were 0.5, 0.3, and 0.5, Ptrend: 0.03. The Odds Ratio [OR] for each 0.1 ppm increasing in selenium level was 0.9 (95% Confidence Interval [CI]: 0.8,1.0). For women, however, a significant trend in the opposite direction was observed, with OR's of 1.3, 1.3, 3.2, Ptrend: 0.04. The OR for each 0.1 ppm increasing in selenium level was 1.1 (95% CI: 1.0,1.3). After restricting the analyses to current and former smokers, the results for men did not change, while the positive association among women was attenuated and nonsignificant. These findings suggest that high selenium may be associated with a lower lung cancer risk in men but not in women.

272

Employment Status and Effect of Coal Mine Dust Exposure. P.K. Henneberger,* M. D. Attfield (Division of Respiratory Disease Studies, NIOSH, Morgantown, WV 26505).

The purpose of this study was to determine if coal miners who quit the profession suffer greater declines in pulmonary function associated with coal mine dust exposure than miners who continue working or are laid off. The subjects were 1829 white male coal miners who were participants in the National Study of Coal Workers' Pneumoconiosis and had been tested at two surveys an average of 15.1 years apart. At the initial survey, all subjects were active coal miners with an average of 10.8 years mining. By the final survey, 379 (21%) had quit for a variety of reasons (e.g. sickness, injury, another job) and the remaining 1450 were either still mining (938) or laid off (512). Least squares linear regression was used to model the average annual change in the forced expiratory volume in one second (FEV1) from the initial to the final testing. The change in FEV1 (milliliters/year, or ml/yr) associated with the mean inter-survey coal mine dust exposure (milligrams/meter3, or mg/m3) was negative for those who quit (-5.9 ml/yr per mg/m3, standard error (se)=3.5, $p=0.09$) and slightly positive for those who kept on mining or were laid off (+2.6 ml/yr per mg/m3, se=1.8, $p=0.15$). The difference in effect was -8.5 ml/yr per mg/m3 (se=3.90, $p=0.03$). This finding suggests that miners who quit have a greater adverse response to coal mine dust exposure.

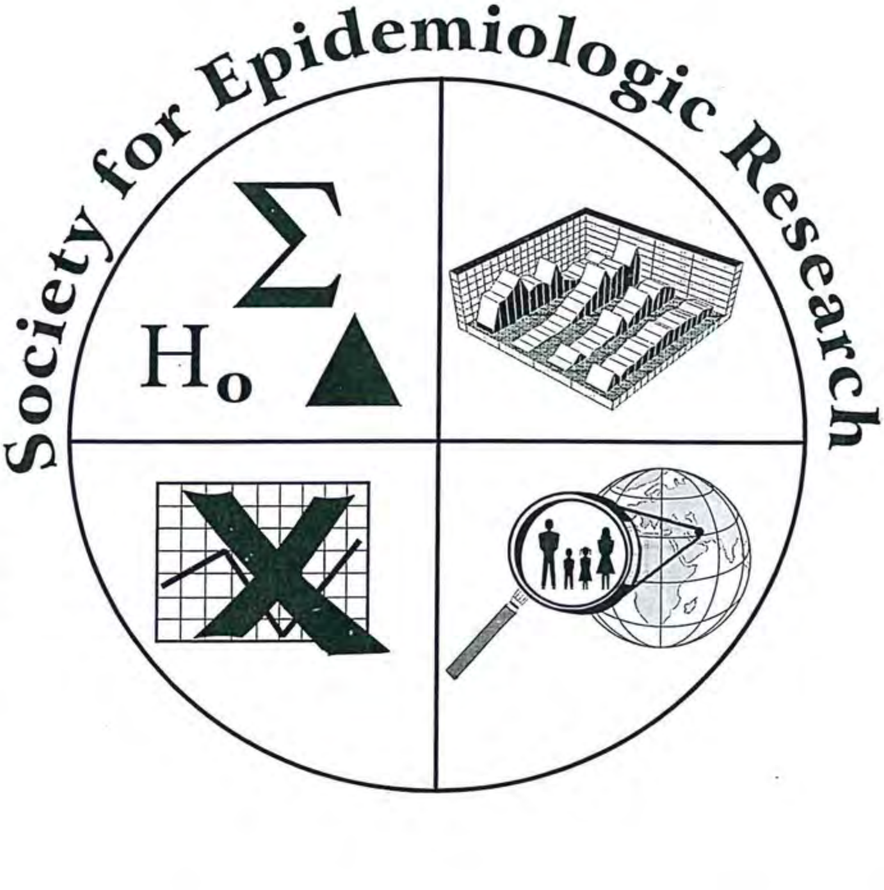
ISSN 0002-9262
Printed in the U.S.A.

American Journal of EPIDEMIOLOGY

Volume 141
Number 11
June 1, 1995

Published by The Johns Hopkins University
School of Hygiene and Public Health

Sponsored by the Society for Epidemiologic Research



ABSTRACTS OF THE 28TH ANNUAL MEETING SNOWBIRD, UTAH, JUNE 21–24, 1995

Founded 1920 by W. H. Welch and W. H. Howell as the American Journal of Hygiene