273

PARTICIPATION IN EPIDEMIOLOGIC STUDIES: SURVEY OF PRACTICE.

L M Morton, J Cahill, *P Hartge (Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Rockville, MD, 20852)

Low participation ("response rates") in epidemiologic studies can threaten validity. Many commentators note apparent declines in participation in recent years. Some research organizations have suggested guidelines for consistent reporting of participation, but the degree of application of these guidelines in epidemiologic research is unknown. We therefore abstracted original analytic epidemiology research articles published during 2003 in 7 general epidemiology and public health journals (AJE, AJPH, Annals of Epid., CEBP, Epid., IJE, J. Clin. Epi.), and 3 major medical journals (JAMA, Lancet, NEJM). Here we present preliminary results from the first 65 articles we reviewed. Among 24 questionnaire- and/or interview-based case-control and cross-sectional studies, 5 (21%) adequately reported participation (including the number eligible, contacted, and participated), whereas 12 (50%) provided some and 7 (29%) provided no participation information. Of the 17 studies that provided at least some information, the participation proportion ranged from 15% to 99% (median 71%). In these 2003 reports, studies conducted more recently were more likely to include biologic specimens. Specimens were collected in 28% (5/18) of studies begun in 1985 or earlier, 27% (3/11) of studies begun during 1986–90, 35% (6/17) of studies begun during 1991–95, and 47% (9/19) of studies begun in 1996 or later. Very few studies reported information on participation separately for the biologic specimen study component. Additional analyses will consider apparent changes in the participation proportion in recent studies (≥1996) compared with earlier studies (≤1985), and the relation between participation and various study characteristics, including design, inclusion of biologic specimens, and location. Our preliminary results suggest that stronger adherence to reporting guidelines in epidemiologic studies is needed.

274

MAINTAINING QUALITY CONTROL IN A LARGE SCALE RANDOMIZED CONTROLLED TRIAL IN THE AMAZON REGION OF PERU.

R Larocque and *T W Gyokos (McGill University, Montreal, QC Canada)

A randomized controlled trial was undertaken from April 2003 to July 2004 to investigate the effectiveness of antenatal single-dose mebendazole plus iron supplements versus single-dose placebo plus iron supplements in reducing low birthweight in a hookworm-endemic area of Peru. A total of 1042 pregnant women were recruited from 12 health centers in the Iquitos region from April 2003 to November 2003. The size of the field team during the peak recruitment phase reached 12 people. The trial included an interviewer-administered questionnaire, laboratory measures (fecal samples and blood samples for malaria), hemoglobin measures with digitalized HemoCue® portable machines, as well as weight measures with digital scales. Once included in the study the women were followed for an average of 7 months. They were seen by research personnel at least once a month. In order to maintain the quality of the data for such a large prospectively recruited and rurally dispersed study population, there needed to be constant supervision of data collection activities during fieldwork and in the laboratory. The task of the field supervisor was to ensure 100% systematic quality control of all field data collection activities. One principal element in ensuring the level of quality control was constant training. In addition to training, the following daily quality control issues will be discussed: 1) Minimizing loss to follow-up, 2) Maximizing compliance with ongoing iron supplementation, 3) Minimizing reporting errors, 4) Maximizing accuracy of questionnaire responses, 5) Calibration of measurement instruments and 6) Role of external quality control assessors. Meeting the daily challenge of quality control not only ensures quality of the data but also stimulates positive teamwork and builds local research capacity.

275

MODELING SMOKING HISTORY USING AN OVERALL INDICATOR OF EXPOSURE.

*K Leffondré, M Abrahamowicz and J Siemiatycki (University of Montreal, Montreal, Quebec, H2W 1V1)

Multi-dimensionality of smoking history presents challenges for the analyses of epidemiologic studies. Although all aspects of exposure (smoking status, duration, intensity, time since cessation, and age at initiation) need to be carefully adjusted for each other to prevent confounding, such adjustments are difficult because of multicollinearity. To avoid these problems, a new single aggregate measure of smoking that simultaneously accounts for intensity, duration, and time since cessation has been proposed by Hoffmann and Bergmann (American Journal of Epidemiology 2003;158:393). This overall indicator of smoking implies that the effects of duration and time since cessation depend (i) on the presumed half-life of the carcinogen and (ii) a lag time parameter that both have to be fixed a priori, or estimated by maximizing the fit. Our objective was to assess the advantages and limitations of this indicator, compared to conventional modeling that represents smoking status, cigarette-years, and time since cessation as separate variables. To this aim, different multiple logistic regression models were fitted using data on 640 lung cancer cases and 915 controls from a case-control study undertaken in Montreal in 1979–1985. We found that for well-selected values of half-life and lag parameters, the new overall indicator yielded a better fit than the conventional representation of smoking, especially when ex-smokers were included in the analyses. This suggests that the indicator may better represent time since cessation. When investigating the effects of intensity and duration on the estimated risks of lung cancer for different profiles of smokers, we found that the overall indicator was more responsive to increasing intensity than duration. Our results suggest that using the new indicator for (i) testing or [illegible] of smoking, or (iii) investigating interactions with other factors, is an attrac[illegible] alternative to conventional modeling.

276

SIMULATIONS FOR THE COMPARISON AND VALIDATION OF IMPUTATION METHODS FOR MISSING DIAGNOSES OF LIFETIME DISORDERS.

*M. Abrahamowicz, M-P. Sylvestre, A Schwartzman, R Tamblyn, C Senneville, L Serbin, D Stack, S Hodgins, J Ledingham (McGill University, Montreal, Qc, H3A 1A2)

Some epidemiologic studies assess putative risk factors for a disease for which the 'true' diagnosis is known for only a subset of subjects. We used simulations to evaluate and validate methods for imputing missing diagnosis when markers of possible disease are available. We focus on potential bias in the estimated impact of predictors on the risks of developing disorders of interest and the accuracy of statistical inference about these associations. A life course of a cohort was simulated. For each cohort member, we randomly generated characteristics and the events corresponding to the development of the disease. Two scenarios are considered: (a) the "true diagnosis" is missing at random, i.e. the probability of having a missing disease status does not depend on its presence or on the subject's characteristics; (b) the risk of missing diagnosis increases for subjects who did develop the disease and for those with high risk factor values. We generate 1,000 random samples for each scenario. The analysis involves 3 stages: (1) using subjects with known true disease status, estimate the multivariable regression equation that yields the expected probability of a subject having the disease; (2) impute the disease status for subjects for whom it was missing. Two alternative methods based on the estimated probability of the disease (P) are employed to impute missing diagnoses. The *deterministic* method assigns the disease if and only if P ≥0.5. The *probabilistic* method randomly assigns the presence or absence of the disease with a probability of disease P. (3) estimate the association between the risk of the disease and the putative risk factors. The *probabilistic* method of diagnosis imputation performs uniformly better than the *deterministic* method. When the true diagnosis is missing at random, the estimated associations are practically unbiased. Non-randomly missing data introduce some bias.

Material may be protected by copyright law (Title 17, U.S. Code)

Material may be protected by copyright law (Title 17, U.S. Code)