445

USING CAUSAL DIAGRAMS OPTIONALLY AUGMENTED WITH FUNCTIONAL MAPPING TO UNDERSTAND COVARIATE BALANCE IN PROPENSITY SCORE MODELS. *O A Arah (Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90095-1772)

Propensity score models are increasingly being used to estimate the causal effects of exposures in observational studies. A central concern in propensity score models is achieving covariate balance. Usually, this covariate balance is expected to be achieved for those variables included in the propensity score model. This study shows how to use traditional causal diagrams as well as those augmented with the functional mapping encoded in the propensity score to explain the attainment and consequences of covariate balance in propensity score models. The researcher uses functional mapping augmentation to replace solid directed edges between the exposure and other covariates with broken directed edges interconnecting the propensity score with its selected covariates and the exposure. The graphical rules of causal diagrams combined with the conditional independence properties of the propensity score can then be applied to check for the attainment and consequences of covariate balance. Using both traditional and augmented causal diagrams, this study demonstrates how outcome-only predictors are always balanced within strata of the exposure whether or not they are included in the propensity score model. Other types of variable such as confounders and exposure-only predictors must be included in the propensity score model in order to achieve covariate balance between the exposed and unexposed. Nonetheless, only imbalances in covariates which are on confounding paths lead to biased estimates of the causal effect of the exposure. All these findings are illustrated using Monte Carlo simulations.

446

FORMALIZING VARIABLE SELECTION IN PROPENSITY SCORE METHODS. *O A Arah (Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90095-1772)

The propensity score, defined as the probability that a study participant would have been exposed given their observed background characteristics, is a robust method for estimating unbiased exposure effect. Although it is becoming one of the most applied estimation methods for causal analysis in observational studies, the propensity score still suffers from confusions about its variable selection. This study uses statistical formalization and simulations to demonstrate that the current definition of the propensity score as true exposure-assignment modeling can be misleading for the effect estimation of point exposures using matching, stratification, or covariate adjustment. The most accurate and precise exposure effect estimate is obtained when the propensity score model selects confounding variables plus those related only to the outcome, but not those related only to the exposure. This 'best' model also yields the largest matched sample. Including sufficient confounders yields an unbiased exposure effect estimate. Additional inclusion of outcome-only or exposure-only predictors respectively creases or increases the variance of the exposure effect estimate without improving the point estimate. Non-collapsible effect measures such as the odds ratio require the inclusion of outcome-only predictors after adjusting for confounders. For the commoner point exposure effect estimation, the propensity score can be viewed as a robust functional mapping that mimics the outcome-model space. Propensity score models can then receive similar causal considerations for variable selection as estimation techniques aimed at outcome modeling.

447

ESTIMATION OF SPATIAL RISK DISTRIBUTION IN CASE-CONTROL STUDIES USING MULTINOMIAL MODEL. *L B Nucci, A C C N Mafra, R Cordeiro (State University of Campinas – Unicamp, Campinas, São Paulo 13083-970, Brazil)

Case-control studies make a large contribution to epidemiological methodology, providing an efficient way of estimating population relative risk. Even with outstanding developments in geographic information systems and spatial analysis of data in the last 20 years, there still seems to be little use of the spatial dimension in case-control studies, although important studies have appeared in this area, particularly in the field of environmental epidemiology. This study aims to describe a multinomial model as a method to estimate risk factors in case-control studies where the cases are subclassified, including the geographic position of housing from cases and controls. The polytomous logistic model supports the calculation of specific risk for each sub-class of the response variable, but a semi-parametric generalized additive model (GAM) is more appropriate to include geographic coordinates as unidimensional splines and the co-variables. As an illustration we present an application of these models in a population based case-control study of dengue, with cases classified as severe or light. The analyses have found areas of increased risk as well as co-variables that influence the infection. The multinomial approach allows searching, in a unique analysis, the association among co-variables and one or some sub-classes of the cases, opening the possibility of individualized risk and protection factors identification for each sub-class studied which is of great epidemiologic interest, as well as GAM is able to estimate spatial distribution of the risk according to their severity.

448

GENERALIZED ROC CRITERION FOR MULTIPLE BIOMARKERS WITH LIMITS OF DETECTION. *N J Perkins, E F Schisterman, A Vexler (NIH, NICHD/DESPR, Rockville, MD 20852)

Authors have shown that a best linear combination (BLC) of multiple biomarkers following a multivariate normal distribution can be formed parametrically that maximizes the area under the receiver operating characteristic curve (AUC). However, biomarkers are often measured with limits of detection (LOD). Common solutions, of omitting or naïvely replacing observations missing below the LOD, lead to negatively biased estimates of the AUC. Maximum likelihood methods have been developed for estimating the parameters of a bivariate normal distribution of two biomarkers with LODs. We generalize the bivariate likelihood for left censored normal random variables to p biomarkers with LODs, develop the point estimator and confidence interval for AUC and demonstrate asymptotic unbiasedness and nominal coverage probability, respectively. Simulations demonstrate that this method yields relative bias ranging from five to less than one percent of the AUC for various levels of correlation and missingness, bias similar in magnitude to estimates based on complete data. For motivation, we apply the methods to measurements of three different polychlorinated biphenyls in women with and without endometriosis to exemplify the discriminatory ability of a BLC of biomarkers each measured with LOD.