Response to Invited Commentary

Pischon et al. Respond to “Variable Selection versus Shrinkage in Control of Confounders”

Tobias Pischon, Matthias B. Schulze, Dagmar Drogan, and Heiner Boeing

From the Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke, Nuthetal, Germany.

Received for publication October 24, 2007; accepted for publication November 7, 2007.

We read with great interest Dr. Greenland’s invited commentary (1) about variable selection to control for confounding in observational studies. We agree with Dr. Greenland that the identification of confounders should be based primarily on background knowledge and not on significance testing. However, our proposed method (2) is not meant primarily as a variable selection procedure. Currently, relative risk estimates are commonly presented from nested models with increasing complexity of covariate use (3). This is not caused predominantly by the uncertainty of selecting the proper model, but rather by the interest to quantify the relative effect of adjustment for specific covariates on risk estimates. For example, relative risks from a multivariate model for a specific nutrient might be compared with those from a model with additional adjustment for other nutrients or foods to evaluate the relative importance of confounding by dietary characteristics (3). Another potential application is the situation where it may be more likely that a covariate reflects a potential mediator rather than a confounder. Although such type of analysis requires several assumptions and careful interpretation, the comparison of models without and with adjustment for potentially intermediate variables might be informative to quantify the change in a beta coefficient when these covariates are taken into account (4).

As reviewed by Dr. Greenland, inclusion of all measured potential confounders as covariates may be a desirable approach to avoid the complexities of variable selection, and shrinkage methods have been proposed to obtain unbiased results when following this strategy in the case of collinearity or sparse data (1). However, the example that Dr. Greenland presents in table 1 (1) illustrates that the additional inclusion of 85 food items using the shrinkage method to 35 nutrients within a conditional logistic regression model had little effect on the point estimates for most nutrients, while the confidence intervals for all nutrients presented largely increased. The rationale of selecting 35 nutrients out of hundreds of potential nutrients and non-nutritive food constituents remains unknown, but it seems clear from this example that the pure availability of food and nutrient intake values should not guide covariate selection. Rather, such selection should be driven by a priori hypotheses and should be limited to plausible covariates, as it is the case for other epidemiologic questions. In addition, when using the prior-data method, we find that the degree of shrinkage seems somewhat arbitrary. Finally, shrinkage adds complexity to the regression models, thus further complicating the biologic interpretation of the coefficients provided in table 1. Therefore, despite advances in methodology, the decisions involved in the definition of the confounder structure, as well as the statistical penalty for added covariates in terms of loss of precision, are problems epidemiologists still face in the current millennium.

Note: With great sadness we report the loss of our colleague Dr. Kurt Hoffmann. He died shortly after acceptance of our article by the Journal, without having the opportunity to read Dr. Greenland’s comment.

ACKNOWLEDGMENTS

Conflict of interest: none declared.
REFERENCES