

EDITORIAL

Case-Control Study Design: Spurious Associations Between Exposure and Outcome

The paper in this issue by Deubner et al (*J Occup Environ Med.* 2007; 49:953–959.) illustrates an important point in study design: correlations among time-related variables may produce apparent associations between outcome variables and covariates that have no meaningful relationship to outcome. Although confounding is well recognized as a threat to the validity of epidemiologic studies, investigators sometimes do not adequately assess the study design as a source of confounding. This paper provides an important contribution to our understanding of how correlations among time-related covariates in a commonly used case-control study design may introduce spurious associations between exposure and outcome. The data set was derived from an important study of lung cancer risk related to beryllium exposure.¹

The paper suggests that the source of confounding in this instance was subtle. It was not recognized either by the authors of the original research from which the data set was derived or by the reviews of that study. The paper offers no remedy in the design of case control studies, and no universal remedy is obvious. What the authors do offer is the recommendation that confounding be evaluated by performing simulations to test the effects of the study design on the measures of association. Their approach was facilitated

by having access to the original cohort from which the cases and controls were selected so that they could resample subjects at random from the cohort. This procedure, replicated 10 times, made it clear that the results were not dependent on any particular set of cases or controls or their exposures but were due to the study design itself. The authors then methodically unraveled exactly what it was about the study design that produced the association.

The factors that they identified should serve as warnings to researchers. First, when matching criteria require that age of censor be different (either higher or lower) in cases and controls, the researcher should take note that the study design is creating an association between a time-related variable and outcome. This should be followed by recognition that any other time-related variable that is associated with age at censor may now potentially be associated with the outcome simply because of confounding. It is common in occupational studies to use several time-related variables as surrogates for exposure (such as duration of exposure) or in the calculation of exposure estimates (such as cumulative exposure, cumulative exposure during defined periods, and duration of exposure above a threshold), and it should be a concern that associations among such exposure variables and the outcome may be confounded by the matching procedure.

The second warning is that age at censor is associated with age at hire. Age at hire is commonly a determinant of exposure because younger workers often end up in more arduous, less skilled jobs. Recognition of

this association signals that matching is a likely confounder because the matching factor (age at hire) is likely to be associated with both exposure and outcome.

The third warning is not intuitively obvious: lagging exposures to account for latency can dramatically exaggerate the difference in exposure between cases and controls because it censors a great deal of exposure information, and this censoring may be different between cases and controls because of the matching alone. Simulation is an efficient way to recognize this problem.

The fourth warning is old news: transforming the probability distribution (in this case, a logarithmic transformation was used) changes its mean and standard deviation, which may give very different results from the untransformed data when groups are compared.

Matching cases and controls can create confounding by the matching factor, which must be controlled by keeping the matching factor in the analysis. What is unusual about the matching used in the original study of beryllium workers is that the cases and controls were not chosen to be equal on the matching factor (age of censor); they were chosen to be unequal. Thus, matching of this type not only created confounding, but it also could not be controlled in the analysis. Simulation offers a way to examine the influence of the study design when the standard approaches for controlling are confounding and are not possible.

What does the Deubner study suggest that researchers should do? Most importantly, they should be alert to the possibility that study

In the interest of full disclosure, David H. Garabrant traveled to Washington, DC, to meet With David Deubner to discuss the topic of beryllium epidemiology. Brush Wellman, Inc paid the travel expenses for the meeting. Dr Garabrant has received no other compensation from Brush Wellman and has no relationship with the company.

design decisions may introduce confounding, particularly when they are matching on time-related factors. This is emphasized when the measures of exposure under study are also time related, as they usually are. Secondly, the Deubner study suggests that when matching on time-related factors is used, simulations offer a way to evaluate whether the study design also produced a spurious association between exposure and outcome.

There is no reason to believe that the problems Deubner et al elucidated are unique to their data set. Furthermore, there is no assurance that these issues, under a different

study design or data, could not obscure a true association and produce negative results. The Deubner paper suggests that confounding as a result of study design may be an important problem that should be evaluated rigorously by assessing confounding and by simulations designed to show the effects of study design in the absence of differences in exposure.

The dilemma is that there are good reasons to match on time-related variables to satisfy assumptions under proportional hazards models, yet doing so may introduce confounding. The Deubner study provides evidence that investiga-

tors should perform empirical simulations to evaluate whether their design decisions have led to valid results.

David H. Garabrant, MD, MPH
*Professor of Occupational Medicine
and Epidemiology
University of Michigan School of
Public Health
Ann Arbor, Michigan*

Reference

1. Sanderson WT, Ward EM, Steenland K, Petersen MR. Lung cancer case-control study of beryllium workers. *Am J Ind Med.* 2001;39:133–144.