J Clin Epidemiol Vol. 51, No. 3, pp. 233–236, 1998

0895-4356/98/$19.00
PII S0895-4356(97)00283-7

On Miettinen's Multivariate Confounder Score

*David Strauss**
Department of Statistics, University of California, Riverside, California

ABSTRACT. Consider a regression model for the effect of a treatment on an outcome variable in the presence of potential confounders. It is common to test for main effects and interactions of the treatment variable; if the confounding variables are discrete, one might then compare the treatment groups within strata formed by suitable covariate patterns. An alternative, due to Miettinen, is to stratify according to a "multivariate confounder score" and test for treatment effects within strata. This test has been shown to be flawed, and the method appears largely to have fallen into disuse. Here we show that such stratification nevertheless provides a sound and often useful graphical comparison of treatment and control groups. J CLIN EPIDEMIOL 51;3:233–236, 1998. © 1998 Elsevier Science Inc.

KEY WORDS. Logistic regression, multivariate confounder score, mental retardation, tube feeding, mortality

INTRODUCTION

One frequently deals with regression models for an outcome variable Y on a categorical "treatment" or "exposure" variable X and potential confounder variables $Z_1, \ldots, Z_k$. The model may be an ordinary multiple regression, with a continuous dependent variable Y, or it may, for example, be a logistic regression with a binary Y-variable such as lived/died. Two issues usually of interest are; (1) Does X have an effect on Y (or, at least, is it associated with Y) when the Zs are taken into account? and (2) If so, how if at all does the pattern of association of X and Y vary according to the values of the Zs?

The standard approach to (1) uses regression-based tests of the main effects of X and examines how much these change when Z is added to the model. If such effects are present, and if the Zs are categorical, one proceeds to (2) by stratification on those Zs judged to have significant interactions with X. As is well known, however, if there are several such Zs the result may be too many strata containing too few observations. This is especially likely occur when both X and many of the Zs are correlated with an underlying variable such as "risk."

Miettinen [1] suggested a different approach. The regression model is used to predict a Y-value after adjustment for treatment effects, called a multivariate confounder score, for each observation. Observations are partitioned into homogenous strata on the basis of their scores, and the comparison of treatment groups is carried out within strata. The method has intuitive appeal, but it was subsequently shown that the resulting significance test is flawed, and the approach seems to have fallen into disuse [2].

The purpose of this article is to show that despite its shortcomings for task (1), the method can provide a sound and often revealing graphical comparison of treatment groups [task (2)]. It will be especially helpful when the confounders all represent aspects of a single unobserved variable such as "susceptibility" or "risk." This is illustrated with an example that compares two treatment groups across a spectrum of risk.

MULTIVARIATE CONFOUNDER SCORE

Consider the following model for the expected value of the outcome variable Y on a binary treatment indicator variable X and categorical potential confounder variable $\mathbf{Z} = (Z_1, \ldots Z_k)$:

$$E(Y_i) = c(z_i) + I[X_i = 1]\, t(z_i). \quad (1)$$

Here $c(z_i)$ is the expected value of Y_i for a "control" unit with covariates z_i, $I[.]$ is an indicator variable, and $t(z_i)$ is the additional effect of the treatment. If t is a constant then the effect of the treatment is additive; otherwise Eq. (1) indicates interaction between treatment and covariates.

The multivariate confounder score for the ith individual may be defined as

$$S = c(z_i). \quad (2)$$

This is obtained from Eq. (1) by setting X to zero. The mean of Y_i is either $S + t(z_i)$ or S, depending on whether the ith observation is a treatment or a control unit. In this sense, S may be interpreted as the expected Y-score corresponding

Address for correspondence: David Strauss, Ph.D., Department of Statistics, University of California, Riverside, CA 92521.
Accepted for publication on 4 November 1997.

to z_i (after subtraction of the treatment effect for units receiving the treatment).

Note that, formally, S is a function only of the covariates. Given an infinite population, the procedure now would be to stratify the observations into subgroups within which S is constant. Let A be such a stratum; it consists of two subsets, a treatment group A_1 for which X = 1 and a control group A_0. Using the notation $\{V|A\}$ to denote the average value of V over the subset A, we have from Eq. (1) that

$$\{EY|A_0\} = \{c(z)|A_0\},$$

$$\{EY|A_1\} = \{c(z)|A_1\} + \{t(z)|A_1\},$$

and thus

$$\{EY|A_1\} - \{EY|A_0\} = \{t(z)|A_1\}, \quad (3)$$

since the function c is constant in stratum A. This means that the confounding effects due to Z have been removed: the difference in means between treatment and control groups within A is equal to the average treatment effect in A. If interactions are present, so that t is not constant, this will vary between the strata.

In practice the set of observations is finite and such a stratification would be too fine, in the sense that most strata would lack either treatment or control observations. Instead, one would stratify into a reasonable number of subgroups within which the S values are fairly homogeneous. Further it would generally be necessary to work with a model such as the generalized linear form

$$g\{EY\} = \alpha X + \Sigma \beta_j Z_j, \quad (4)$$

for a suitable link function g. Here Z_1 is identically 1. This leads to estimated confounder scores

$$S = \Sigma b_j Z_j.$$

This is the form of confounder score originally proposed by Miettinen [1]. Note that it would not be appropriate to work with fitted values from the model

$$EY = \Sigma \gamma_j Z_j,$$

which excludes X, instead of Eq. (4). To do so would assign a portion of the effect of X to the confounders.

Miettinen [1] suggested that the sample be partitioned into roughly homogeneous strata on the basis of the confounder scores, and that differences in Ys for X = 0 and X = 1 be tested within strata in the usual ways. Pike *et al.* [2] showed, however, that in general the nominal significance level in such tests underestimates that the actual significance level. Their simulations indicated that a nominal 5% test may in practice have a 20% or more chance of rejecting a true null hypothesis. Essentially, the problem is as follows: Miettinen's test is roughly equivalent to a comparison of the residual sum of squares associated with Eq. (1) with the (larger) residual sum of squares in a simple regression of Y on S. But the latter is in general inflated because S is not exactly the regression function of Y on the Zs when X is excluded. Because of this bias, Breslow and Day [3] and other sources recommend the standard regression approach over the use of a confounder score.

When there is a fairly large number (10 or more, say) of covariates that are all correlated with an unobserved variable such as "risk," the number of significant treatment-covariate interactions may be large. In this case stratification by all the indicated covariates patterns may result in strata that are both sparse and too numerous. An additional problem in interpretation is that there will be no natural ordering of the strata with respect to overall risk. In this case, once the preliminary significance testing has been carried out the multivariate confounder score could be used to generate an alternative stratification. This may result in a useful descriptive comparison of treatment and control groups within a convenient number of strata that are homogeneous with respect to risk. As shown in Eq. (3), apart from the effects of sampling error in the regression estimates and the use of "coarse" strata, such comparisons are free of biases related to the confounder variables.

Miettinen [1] suggested the use of five strata, each containing equal numbers of observations. The optimal number of strata will depend on the application, however. In particular, (1) the more data are available and the better the model's fit, the more strata can be used; (2) there is no need to insist on equal number of cases. For example, if Y is a binary lived/died variable and if deaths are relatively uncommon then variances within strata will be approximately equal if boundaries are drawn so that the numbers of deaths (either actual or expected from the model) are the same for each stratum.

In the case of categorical covariates there will usually be many covariate patterns aggregated into a single risk stratum. In some situations, even within strata the treatment units will systematically tend to be associated with the higher-risk (or lower-risk) covariate patterns than the control units. This may induce a bias, though usually a small one, essentially because the stratification is not sufficiently fine. If needed, the bias is easily corrected with the use of direct standardization [4]. Either the treatment or the control group may be taken as the standard population, the rate for the other group being then standardized to it, or if appropriate both rates may be standardized to an external population.

EXAMPLE

There has been increasing use of gastrostomy procedures and subsequent tube feeding for individuals with serious mental retardation and related disabilities. In the most severe cases tube feeding is accepted to be necessary for the patient's survival. The procedure carries some risk, however, and for less severely disabled individuals the risks and benefits have not been fully assessed [5]. Eyman *et al.* [6], working

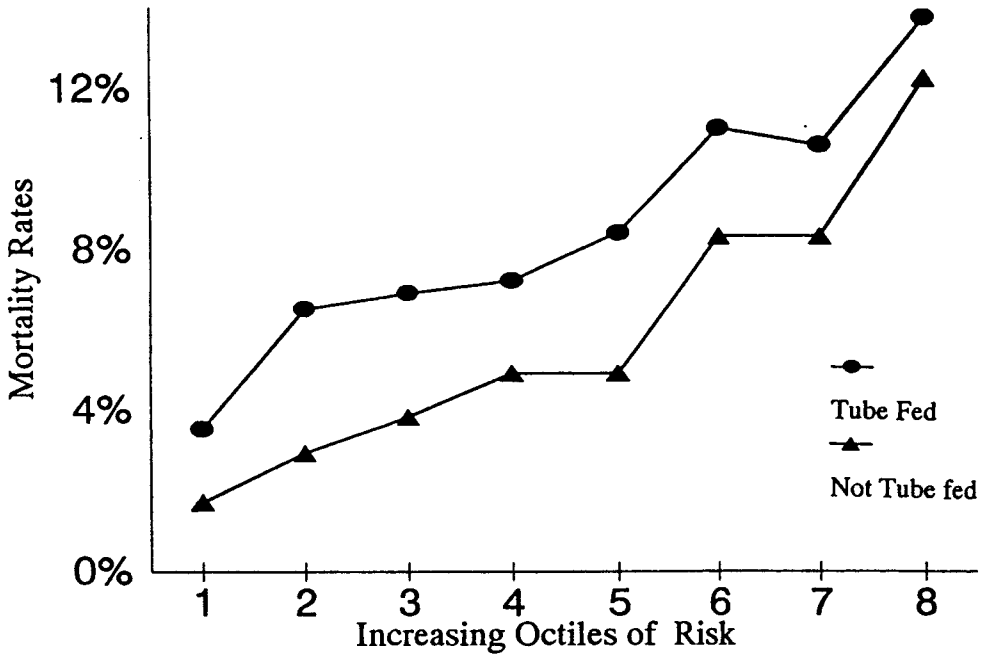

FIGURE 1. Mortality rates for person-years with (*dots*) and without (*triangles*) tube feeding (within each risk octile).

with a sample of all severely disabled children with mental retardation in California who receive state assistance, found that the mortality rate of tube-fed individuals was several times higher than that of those not tube-fed. In that study, however, only limited attempts were made to control for potential confounding variables. Doubtless much of the difference in mortality rates was due to a higher rate of serious medical conditions in the tube-fed group.

Using the same data source, Strauss *et al.* [7] compared mortality rates for those with and without tube feeding. In this study an attempt was made to control for all potential confounding variables for which data was available. These included known risk factors such as immobility and type of residence (e.g., large state institutions versus small group homes), together with presence or absence of major medical problems (heart disease, upper respiratory infection, pneumonia, esophageal conditions, etc.). The unit of analysis was a person-year (the period between a subject's consecutive birthdays). For the set of 13,688 person-years, a logistic regression model was developed for the binary lived/died dependent variable Y on the "exposure" variable X—presence or absence of tube feeding that year—together with a set of covariates $\{Z_i\}$. These included the risk factors, together with the child's age and the calendar year. This form of logistic regression has been frequently employed in the Framingham study [8] and elsewhere.

The final model showed significant main effects of tube feeding (X) and of some 10 of the covariates. Many of the latter had significant interactions with X: the tube feeding-mortality association was generally weaker for subjects with serious medical conditions than for those without. This is consistent with the earlier suggestion that tube-fed subjects tend to have more recorded and unrecorded medical problems. A parametric interpretation would be quite complicated in view of the large number of interactions. Further, as noted previously, the alternative of stratification by all the confounding variables would result in a large number of unordered strata containing rather few observations. For this reason, we preferred a graphical approach based on the multivariate confounder scores.

The confounder scores were used to partition the person-years into eight fairly homogeneous "risk-octiles." Cut points were chosen so that equal numbers of deaths occurred in each group. Within each stratum, the mortality rate for the not-tube-fed group was simply the raw percentage of deaths. For the tube-fed group, rates were directly standardized. The not-tube-fed group was chosen as standard population here because it was substantially the larger of the two. Mortality rates for person-years with and without tube feeding within each risk octile are plotted in Figure 1.

The spacings between the eight groups have been made equal in Figure 1. This is arbitrary, and other choices would have been possible. The issue will generally not be critical, as typically one would be interested in whether the relative risk of treatment and control tends to increase or decrease with risk rather than, for example, the linearity of such a trend.

The figure shows that the relative risk associated with tube feeding is about 2.0 in the lowest risk groups, but decreases almost to unity as the overall risk level increases. (Note, however, that the excess risk—the difference of the two rates—is only slightly larger in the lower risk groups than in the higher ones.) As noted previously, it would not be proper to carry out an overall significance test or to compute confidence intervals etc. for the data of the figure.

Strauss *et al.* [7] concluded that the advantage of tube

feeding for the less debilitated subjects was still not established, and that a randomized prospective study may in some cases be ethically justified. Our main point here, however, is that the pattern revealed in Figure 1 would be difficult to obtain without use of the multivariate confounder score.

DISCUSSION

We have seen that although treatment effects should not be tested using groups defined form multivariate confounder scores, such groups may provide a useful descriptive comparison. Formal inference may be based on the regression model in the standard way, while graphs such as Figure 1 may provide a helpful visual summary. The method is likely to be useful when most covariates are related to an unobserved variable, such as risk. This situation is not uncommon. For example, in studies of mortality for people with mental retardation living in different residential settings, confounder variables such as age, gender, and various measures of disability were all associated with mortality risk [5,9,10].

The method invites comparison with the use of propensity scores [11]. This stratifies observations into subgroups within which the chance of receiving the treatment, $P(X = 1|z)$, is approximately constant. It is known that within strata the distribution of covariates z is then roughly the same among the treatment and control subgroups [11]. Such "balancing" of the zs removes most of the bias when an outcome variable Y is compared across treatment and control groups. Formally, it is in fact easy to show that the method satisfies the same condition in Eq. (3) as does the multivariate confounder score. A potential disadvantage of the propensity score is that when strata are defined on the basis of the $P(X = 1|z)$, strata with large values of P will consist mostly of treatment observations and those with small values will contain mostly controls. This may cause numerical instability.

As pointed out by a reviewer, if the model contains interactions between the treatment variable X and the covariates then the two sets of confounder scores obtained by assigning X to 0 or to 1 will not be monotonically related. This means that the grouping of observations into quantiles may differ according to whether we adjust for the effect of the treatment or the control group. It would, of course, be possible to construct a graph for each case. If the two showed markedly different patterns, however, the conclusions to be drawn may not be so clear.

One limitation of the method should be noted: generally the strata lack a biological interpretation. Instead, each stratum contains observations at similar risk but with possibly very different covariate patterns. When strata can be grouped sensibly on biological characteristics, they may be relatively risk-homogeneous. We have seen, however, that when there are a large number of confounders, all related to "risk," a biologically meaningful grouping is not always possible. In such cases the multivariate confounder score may be a useful tool.

The author thanks Dr. Robert Shavelle for numerous valuable discussions. He also thanks the editor, Professors Norman Breslow and David Hosmer, and two anonymous reviewers for helpful comments on earlier drafts of the manuscript. The work was supported by Grant No. HD21056 from the National Institute of Child Health and Human Development.

References

1. Miettinen O. Stratification by a multivariate confounder score. **Am J Epidemiol** 1976; 104: 609–620.
2. Pike MC, Anderson J, Day NE. Some insight into Miettinen's multivariate confounder score approach to case-control studies. **J Epidemiol Community Health** 1979; 33: 104–106. 1979.
3. Breslow NE, Day NE. **Statistical Methods in Cancer Research: The Analysis of Case-Control Studies, Vol. 1**. International Agency for Cancer Research: Lyon, France; 1980; 101.
4. Fleiss JL. **Statistical Methods for Rates and Proportions**. 2nd Ed. Wiley: New York; 1981.
5. Sulkes S. Life expectancy of profoundly handicapped people. **N Engl J Med** 1990; 324: 420. (Letter to Editor)
6. Eyman RK, Grossman HJ, Chaney RH, Call TL. Life expectancy of profoundly handicapped people. **N Engl J Med** 1990; 323: 584–589.
7. Strauss DJ, Kastner TA, Ashwal S, White JF. Tubefeeding and mortality in severely disabled children with mental retardation. **Pediatrics** 1997; 99: 358–362.
8. Cupples LA, D'Agostino RB, Anderson K, Kannel WB. Comparison of baseline and repeated measure covariate techniques in the Framingham heart study. **Stat Med** 1988; 7: 205–218.
9. Strauss DJ, Kastner TA. Comparative mortality of people with mental retardation in institutions and the community. **Am J Mental Retardation** 1996; 101: 26–40.
10. Strauss D, Eyman RK, Grossman HJ. The prediction of mortality in children with severe mental retardation: The effect of placement. **Am J Public Health** 1996; 86: 1422–1429.
11. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. **J Am Stat Assoc** 1984; 79: 516–524.