DESIGN AND ANALYSIS OF CASE-CROSSOVER STUDY WITH MULTIPLE EVENTS

B. Zhang[1,2] , J. A. Hanley[1,2] , R. Platt[2,3] , S. Suissa[1,2]

[1] *Royal Victoria Hospital, McGill University Health Centre, Montreal, Canada*
[2] *McGill University, Montreal, Canada.*
[3] *Montreal Children's Hospital, McGill University Health Centre, Montreal, Canada*

Email: *bin.zhang@webmail.mcgill.ca*

In the conventional data analysis of most medical research, we assume that all observations are mutually independent. However, this assumption is questionable in many applications where the data can be grouped into clusters with responses in the same cluster tending to be more alike than responses in the other different clusters. In epidemiology, the "first event" approach is often used to investigate the risk of drug utilization; i.e., all the subsequent events are ignored. This can be wasteful. Our concern is how the estimate in terms of bias and precision of the odds ratio would differ if an alternative approach that used "multiple events" is used. We address this question in the context of a case-crossover study design.

The estimates from three different statistical methods were compared; these are based on 3-level of data analysis units (overall, subject-level and event-level): the Mantel-Haenszel 2×2 table estimator; the conditional logistic regression model considering matching; and the generalized estimation equations technique involving different working correlation structures as well as matching factors.

A simulation study with various combinations of the design parameters (sample size, correlation coefficient, hazard ratio and intensities of exposure and outcome) was conducted. The mean squared error (MSE) is employed to evaluate the performance of these three different methods when the data are correlated. We compare these three different methods with data on the study of the association between the use of benzodiazepine and repeated motor vehicle crashes (MVCs).

The alternative approach using "multiple events" produces practically identical point estimates of the odds ratios as those from the "first event" approach; however, the estimates from the former are more efficient (lower standard error, i.e., smaller confident interval). Furthermore, if multiple levels of clusters are occurred in the research data, the data analysis needs to be conducted at the finest level of cluster in order to obtain an unbiased point estimate of the odds ratio.

The major contribution of this study is to provide guidelines indicating under what circumstances the multiple events should be chosen to produce better estimates, or whether the use of the “first event” approach is sufficient to research our goal in epidemiological studies.