Reporting of Multivariable Methods in the Medical Literature

Maor Sauler MD, Jeanette M. Tetrault, MD, John Concato, MD, MPH

**Background:** Multivariable models are frequently used in the medical literature, allowing for researchers to assess the complex relationship of multiple independent variables with an outcome variable. (For example, studies might report “after controlling for variables x and y, factor z is shown to significantly increase mortality.”) The use of these tools is not often taught in medical school, however, and many physicians are unfamiliar with them. In addition, despite generally accepted methodological criteria for the use and discussion of multivariable models, corresponding methodological standards are not often reported in published reports. This scenario can lead to misinterpretation of the data or misleading inferences when clinicians read medical journals.

**Specific Aims:** Our goals were to determine: 1) the frequency of use of multivariable methods in the medical literature; and 2) how often appropriate methodological criteria and assumptions specific to logistic regression or proportional hazards analysis were reported.

**Hypotheses:** Multivariable models will be found commonly in the medical literature, yet there will be underreporting of appropriate criteria for these statistical methods.

**Methods:** A review was conducted of original articles published in the Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, Lancet, and the New England Journal of Medicine from January 2006 through June 2006. If multivariable models were applied for analysis, the article was reviewed for adequate reporting of methodological criteria: 1) reporting the coding of independent and dependent variables; 2) determining how these variables were selected; 3) providing information to determine the number of events per variable; 4) reporting tests for interactions; 5) determining whether model “fit” was mentioned; 6) determining if independent variables were tested for co-linearity; and 7) describing how “outlier” observations were handled. Finally, we determined if any mention was made of the proportional hazards assumption when evaluating Cox Regression, or frequency of the outcome event (related to an underlying assumption) with logistic regression. My role included a 10% re-review, for quality assurance, of the 452 articles reviewed initially; discussions regarding the focus of the manuscript; and participation in writing the manuscript.

**Results:** Among 452 articles, 272 (60%) used multivariable analysis, with logistic regression and proportional hazards being the most common methods. A wide range was found regarding reporting methodological criteria: from 5% (12/265) for “outliers” to 84% (222/265) for coding of variables, as applicable. Discussion of the proportional hazards assumption occurred in only 21% (16/76) of articles using Cox proportional hazards, and frequency of the outcome event was mentioned in 13% (12/89) of articles reporting logistic regression.

**Conclusions:** Multivariable analysis is commonly encountered as a statistical method in the medical literature. Despite generally accepted methodological criteria for the use of multivariable methods, however, a high rate of underreporting exists. As a consequence, research data and their meaning can be misinterpreted by clinicians.