

BRIEF REPORT

Reporting of Multivariable Methods in the Medical Literature

Jeanette M. Tetrault, MD, Maor Sauler, MD, Carolyn K. Wells, MPH, and John Concato, MD, MS, MPH

■ ABSTRACT

Background: Multivariable models are frequently used in the medical literature, but many clinicians have limited training in these analytic methods. Our objective was to assess the prevalence of multivariable methods in medical literature, quantify reporting of methodological criteria applicable to most methods, and determine if assumptions specific to logistic regression or proportional hazards analysis were evaluated.

Methods: We examined all original articles in *Annals of Internal Medicine*, *British Medical Journal*, *Journal of the American Medical Association*, *Lancet*, and *New England Journal of Medicine*, from January through June 2006. Articles reporting multivariable methods underwent a comprehensive review; reporting of methodological criteria was based on each article's primary analysis.

Results: Among 452 articles, 272 (60%) used multivariable analysis; logistic regression (89 [33%] of 272) and proportional hazards (76 [28%] of 272) were most prominent. Reporting of methodological criteria, when applicable, ranged from 5% (12/265) for assessing influential observations to 84% (222/265) for description of variable coding. Discussion of interpreting odds ratios occurred in 13% (12/89) of articles reporting logistic regression as the primary method and discussion of the proportional hazards assumption occurred in 21% (16/76) of articles using Cox proportional hazards as the primary method.

Conclusions: More complete reporting of multivariable analysis in the medical literature can improve understanding, interpretation, and perhaps application of these methods.

Key Words: multivariable analysis, statistical models, regression analysis

■ INTRODUCTION

Clinicians are expected to interpret the results of studies found in the medical literature, and multivariable statistical techniques are often used to assess complex associations.^{1–3} Although generally accepted methodological criteria exist for the application of multivariable analysis,^{1,2} these criteria may not always be applied, or at least not reported. For clinicians who encounter such analyses, medical training offers little instruction in multivariable methods.^{4,5} Accordingly, if authors do not conduct and report the application of methodological criteria appropriately, the results of a study may be misinterpreted or perhaps be incorrect.¹

The objectives of this review were to (1) assess the frequency of multivariable methods reported in the medical literature, (2) quantify reporting of methodological criteria applicable to most multivariable models, and (3) determine if assumptions specific to logistic regression or proportional hazards analysis (Cox regression) were reported.

■ METHODS

We manually reviewed all abstracts of original research articles published in the *Annals of Internal Medicine*, *British Medical Journal*, *Journal of the American Medical Association*, *Lancet*, and *New England Journal of Medicine*, from January 2006 through June 2006. Articles underwent complete review if a multivariable method was mentioned (within the abstract) as a statistical analysis or if information suggestive of multivariable modeling techniques was mentioned in the results. Data were then extracted onto a standardized form regarding the types of analytic methods used, reporting of common methodological criteria, and confirmation of model assumptions.

If more than one multivariable method was reported, all methods used were noted, but results were extracted based on the method reported in the abstract. If more than one method was found in the abstract, results were extracted based on the method used to evaluate the primary research question. Finally, if the primary method used was still uncertain, data were extracted based on the method that received the most emphasis or were presented first in the methods section.

From the Veterans Affairs Clinical Epidemiology Research Center, West Haven; and Yale University School of Medicine, New Haven, CT.

Presented at the Society of General Internal Medicine, 30th Annual (2007) Meeting, Toronto, Canada, April 26, 2007.

Reprints: John Concato, MD, MS, MPH, Clinical Epidemiology Research Center, 950 Campbell Ave, Bldg 35A, Mail Code 151B, West Haven, CT 06516. E-mail: john.concato@yale.edu.

TABLE 1. Frequency of Multivariable Methods (n = 272 articles)

| Method | n | % |
|--|----|----|
| Logistic regression | 89 | 33 |
| Proportional hazards analysis (Cox regression) | 76 | 28 |
| Mixed models | 20 | 7 |
| Linear regression | 19 | 7 |
| Analysis of covariance | 17 | 6 |
| Poisson regression | 12 | 4 |
| Generalized estimating equations | 9 | 3 |
| Unspecified method | 7 | 3 |
| Other* | 23 | 8 |

*Includes Weibull regression, accelerated failure time models, and others.

We evaluated the adequacy of reporting of methodological criteria common to most multivariable models¹: reporting the coding scheme for independent and dependent variables (to interpret coefficients), providing information to calculate the number of events per variable for models with discrete outcome events (to avoid overfitting of the model^{6–8}), reporting of tests for interactions (or mention of a lack thereof), describing the process of variable selection, such as backward or forward selection (to identify the strategy used), whether the model was validated (eg, with an assessment of model “fit”), whether independent variables were tested for collinearity, and whether a method for evaluating outliers was considered (even if data were “left as is”).

Finally, for logistic regression, we assessed whether potential problems regarding interpreting odds ratios were mentioned (such that an odds ratio for each independent variable approximates a relative risk only if the outcome being assessed is uncommon). Similarly, for proportional hazards models, we evaluated reporting of the proportional hazards assumption, which involves a relatively constant “hazard” of the outcome for the compared groups over time.

All data were extracted onto a standardized form, and data were double entered into an Excel spreadsheet. For the methodological criteria, categories of no mention versus the combination of “mentioned, with detail” and “mentioned, without detail” were analyzed. A 10% random re-review was performed for data quality assurance by two authors (J.M.T. and M.S.). Descriptive statistics regarding frequencies of methods and criteria were evaluated in SAS version 9.1 (SAS Institute Inc, Cary, NC).

■ RESULTS

A total of 452 abstracts of original research articles (listed at www.cerc.med.va.gov) were reviewed, with 26% (n = 119) from *British Medical Journal*, 23% (n = 105) from *Journal of the American Medical Association*, 22% (n = 100) from *New England Journal of Medicine*, 20% (n = 90) from *Lancet*, and 8% (n = 38) from *Annals of Internal Medicine* (published semimonthly). Multivariable meth-

ods were reported in 60% (n = 272) of the articles, including 28% (n = 77) using more than one multivariable method; 2% (n = 9) reported a multivariable analysis for bivariate (“unadjusted”) purposes only. For the elements included on the data extraction form, the average percent agreement was 96.9%, and the average κ statistic was 0.91, indicating “almost perfect” agreement.⁹

As shown in Table 1, logistic regression (33%, n = 89/272) and proportional hazards analysis (28%, n = 76/272) were the most frequently reported methods. Other less common methods were found in 8% (n = 23) of the studies, including Weibull regression¹⁰ and accelerated failure time models.^{11,12} Three percent of the studies (n = 7) used an unspecified multivariable modeling technique that precluded further review; 1 study reported the use of more than 5 different methods.

Table 2 shows the adequacy of reporting of the 265 articles with clearly stated multivariable methods. The coding scheme for variables was described in 84% (n = 222) of the studies evaluated. The numbers of events per variable were described in 79% (152/192) of studies with categorical outcome events, and 45% (n = 118) of the studies reported data on testing for interaction terms. The model selection process was described in 15% of studies (n = 41); assessing model “fit” or other mechanisms of model validation were described in 10% of studies (n = 27), including techniques such as bootstrapping¹³ (n = 4) and the Hosmer-Lemeshow test¹⁴ (n = 13). The text described issues relating to collinearity in 9% (n = 24) of the studies, and a method for dealing with outliers was found in 5% (n = 12). Of note, 5% (n = 13) of the studies failed to report any of the criteria; only 1 study met all of the criteria.

We also looked for discussion of assumptions specific to logistic regression and proportional hazards analysis. Among studies reporting data from logistic regression models, 13% (12/89) discussed the interpretation of odds ratios; 21% (16/76) of studies using proportional hazards analysis discussed the proportional hazards assumption. As examples of analytic strategies, 1 study

TABLE 2. Adequacy of Reporting of Methodological Criteria (n = 265*)

| Criterion | n | % |
|---|-----|----|
| Coding of variables | 222 | 84 |
| No. events per variable (n = 192 [†]) | 152 | 79 |
| Tests for interactions | 118 | 45 |
| Model variable selection | 41 | 15 |
| Model validation | 27 | 10 |
| Testing for collinearity | 24 | 9 |
| Method for evaluating outliers | 12 | 5 |
| Studies meeting no criteria | 13 | 5 |
| Studies meeting all criteria | 1 | <1 |

*Seven studies (among 272) did not have clearly stated multivariable methods.

[†]Studies with categorical outcome events.

reported using Poisson regression rather than logistic regression because the outcome event was common, and another study reported use of logistic regression because the proportional hazards assumption was not met. Examples of how the proportional hazards assumption was tested were use of log-log plots and use of the Schoenfeld residual test.¹⁵

■ DISCUSSION

In a review of prominent medical journals, we found that multivariable methods of data analysis were used frequently, with logistic regression and proportional hazards analysis the most commonly reported methods. “Any mention” of methodological criteria applicable to most multivariable models varied widely, suggesting an opportunity for improved reporting (and possibly conduct) of these methods. In addition, model assumptions specific to logistic regression and proportional hazards were infrequently discussed.

The results of our review can be considered in the context of a prior review¹ finding an 18% prevalence of 4 common multivariable methods in *Lancet* and *New England Journal of Medicine* as of 1989. In the current review (as of 2006), we found that 60% of studies used multivariable methods. Across the 2 time periods, logistic regression and proportional hazards analysis remained the most frequently used methods. The range of “adequate” reporting of general criteria in the current analysis ranged from 5% (method for evaluating outliers) to 84% (coding of variables). Among the criteria evaluated in both reviews, most were met more frequently in the current review. For example, testing for interactions was evident in 45% of articles in the current review, versus 27% of articles in the earlier review; only reporting of model selection process occurred less frequently in the current versus former review (15% vs 86%, respectively). The frequency of reporting of the proportional hazards assumption remained approximately the same (and was “low”) in both reviews.

Reviews of multivariable analysis have been published in the medical specialty¹⁶ and obstetrics-gynecology literature,¹⁷ with similar findings of incomplete reporting. Other research has focused on the statistical review process at the level of the journal or authors. In a masked before-and-after study,¹⁸ the peer review and editing process at the *Annals of Internal Medicine* were found to improve the quality of reporting of multivariable methods. In another study¹⁹ of 114 journals responding to a survey, only one third of journals required statistical review for all accepted manuscripts. Finally, a study²⁰ of 704 authors submitting to either of 2 general medical journals found that 73% received input from a methodologist (most often a biostatistician or epidemiologist);

papers without methodological input were more likely to be rejected without review.

As a potential limitation of our work, the manual search, although extensive, may not have identified all possible studies using multivariable methods. In addition, the 10% random review indicated minimal interobserver variability, but some studies may still have been misclassified. As a strength of our study, we “gave credit” if any information was presented for the various methodological criteria. For example, mentioning the number of outcome events per independent variable was considered awareness of the issue; we did not apply a threshold value (eg, 10 events per variable^{7,8}) for “appropriateness.” Similarly, a figure showing the relationship of 2 factors with an outcome variable was considered evidence of assessing interactions, even if not mentioned in the text.

This review suggests the continued need for more complete reporting of multivariable methods in the medical literature. Inadequate reporting of these methods increases the potential for unclear or misinterpreted results. Efforts to standardize reporting of multivariable methods would improve the quality of publications and assist clinicians in reaching appropriate conclusions.

■ ACKNOWLEDGMENTS

The authors thank the staff at the Clinical Epidemiology Research Center, and especially Richard Feinn, for assistance with this article.

■ REFERENCES

1. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med.* 1993;118:201–210.
2. Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Intern Med.* 2003;138:644–650.
3. Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* 1988;80:1198–1202.
4. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *Am J Med.* 1981;71:991–998.
5. Windish DM, Diener-West M. A clinician-educator’s roadmap to choosing and interpreting statistical tests. *J Gen Intern Med.* 2006;21:656–660.
6. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* 1996;49:1373–1379.
7. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J Clin Epidemiol.* 1995;48:1495–1501.

8. Peduzzi P, Concato J, Feinstein AR, et al. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48:1503–1510.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
10. O'Quigley J, Flandre P. Predictive capability of proportional hazards regression. *Proc Natl Acad Sci U S A*. 1994;91:2310–2314.
11. Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat Med*. 1992;11:1871–1879.
12. Walker S, Mallick BK. A Bayesian semiparametric accelerated failure time model. *Biometrics*. 1999;55:477–483.
13. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
14. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: Wiley; 1989.
15. Grambsch RM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81:515–526.
16. Moss M, Wellman DA, Cotsonis GA. An appraisal of multivariable logistic models in the pulmonary and critical care literature. *Chest*. 2003;123:923–928.
17. Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology literature. *Obstet Gynecol*. 1999;93:1014–1020.
18. Goodman SN, Berlin J, Fletcher SW, et al. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Ann Intern Med*. 1994;121:11–21.
19. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector? *J Gen Intern Med*. 1998;13:753–756.
20. Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. *JAMA*. 2002;287:2817–2820.