An Appraisal of Multivariable Logistic Models in the Pulmonary and Critical Care Literature*

Marc Moss, MD; D. Andrew Wellman, MD; and George A. Cotsonis, MA

Objective: Multivariable modeling techniques are appearing in today's medical literature with increasing frequency. Improper reporting of these statistical models can potentially make the results of a study inaccurate, misleading, or difficult to interpret. We performed a manual literature search of five international pulmonary and critical care journals to determine the accuracy in the reporting of logistic regression modeling strategies.

Design: We examined all of the published manuscripts for 12 potential limitations in the reporting of important statistical methodologies over a 6-month period from July 1, 2000, until December 31, 2000.

Results: Of the 81 articles that included multivariable logistic regression analyses, only 65% (53 analyses) properly reported the coding classification of pertinent independent variables that were included in the final model. An odds ratio and confidence interval were reported for the independent variables included in the final model for 79% (64 analyses) and 74% (60 analyses), respectively. Only 12% (10 articles) referenced whether interaction terms or effect modifications were examined, 1% (1 article) reported testing for collinearity, and only 16% (13 articles) included a goodness-of-fit analysis of the logistic model. The type of statistical package was reported in 69% (56 articles). Finally, approximately 39% of the articles (22 of 57) may have overfit the logistic regression model, leading to potentially unreliable regression coefficients and odds ratios.

Conclusions: Our results indicate that the reporting of multivariable logistic regression analyses in the pulmonary and critical care literature is often incomplete, therefore making it difficult for the reader to accurately interpret the manuscript. We recommend the implementation of adequate guidelines that will lead to overall improvements in the reporting and possibly to the conducting of multivariable analyses in the pulmonary medicine and critical care medicine literature.

(CHEST 2003; 123:923–928)

Key words: critical care medicine; logistic models; pulmonary medicine; study design

Abbreviation: ROC = receiver operating characteristic

Although most investigators of basic science and clinical medicine receive little or no training in statistics and epidemiology, multivariable regression-modeling techniques are now frequently used in medical research.¹ These statistical models are used primarily to identify either the effect of an individual variable on a specific outcome while adjusting for differences in other factors (descriptive modeling) or to estimate the likelihood of an outcome for a given patient from a set of observations that are particular to that patient (predictive modeling).² All multivariable modeling strategies have specific assumptions and several limitations that, when violated, may impact on the accuracy of the results and the ability of the reader to properly interpret the study. Logistic
regression analysis is one type of multivariable modeling procedure that is used to assess the relationship between two or more continuous or categoric, explanatory (independent) variables and a single dichotomous response or outcome variable, such as hospital mortality (alive or deceased). Logistic regression is commonly used in the medical literature for several reasons. There is a high degree of flexibility associated with logistic regression models, including that the continuous explanatory variables do not need to be normally distributed and that other explanatory variables can be dichotomous variables such as gender (male or female). In addition, the results of logistic regression models are easy to interpret because the coefficients can be easily transformed into odds ratios, which are an accepted measure of association in medical research.

In 1993, Concato and colleagues reported a review of 60 randomly selected articles from the Lancet and the New England Journal of Medicine that utilized multivariable statistical methods. In 73% of the articles, they identified violations and omissions of methodological guidelines that would make the reported results potentially inaccurate, misleading, or difficult to interpret. Based on their observations, the authors concluded that there was a need for improvement in the reporting and perhaps in the conducting of multivariable analysis in the general medical literature.

It has been almost a decade since this important observational study was published. The methodological and statistical review policies of many major medical journals have improved over the last decade. However, deficiencies in the reporting of multivariable regression models still exist in some subspecialty journals, including the obstetrics and gynecology literature. It is presently unknown whether similar deficits remain in the pulmonary medicine and critical care medicine literature. Therefore, we used preexisting guidelines for assessing multivariable models and investigated the quality of reporting multivariable logistic regression analyses in five major pulmonary and critical care journals.

Materials and Methods

We manually searched all of the original research articles published in five prominent pulmonary and critical care journals over a 6-month period (July 1 through December 31, 2000). The five journals (American Journal of Respiratory and Critical Care Medicine, CHEST, Critical Care Medicine, The European Respiratory Journal, and Thorax) were chosen for the following three reasons: their large circulation size; their international distribution; and the inclusion of original research articles. Any article stating that either a multivariable logistic regression analysis was performed in the “Methods” section or reporting the results of a multivariable logistic regression analysis in the “Results” section was selected for further examination and review.

Various guidelines have been recommended that can improve the appropriate execution, interpretation, and reporting of multivariable logistic regression methods. Using some of the criteria established by Concato et al., we examined published manuscripts for 12 potential limitations in their reporting of multivariable logistic regression analysis, including the following: (1) inclusion of odds ratio and confidence intervals for variables in the final model; (2) listing the type of statistical software package that was used to perform the analysis; (3) classification or coding of the independent variables included in the final model; (4) testing for interaction terms; (5) collinearity of independent variables; (6) goodness of fit; (7) potential overfitting of the model; (8) conformity of linear gradients; and (9) explanation of how the explanatory variables that were used in the model were initially selected. The 10th potential limitation was reserved only for those articles that used pair-matched case-control data. We determined whether the authors reported that a conditional logistic regression model was used. Finally, if a logistic regression analysis was used to create a predictive model, we determined (11) whether a receiver operating characteristic (ROC) curve was reported for the final model and (12) whether a separate validation set was used.

Study Criteria

The following criteria were used to categorize each article in a systematized manner for each of the potential limitations in reporting style:

1. Inclusion of Odds Ratio and Confidence Intervals for the Final Model. This criterion was fulfilled if the odds ratio and 95% confidence interval for at least the primary explanatory (independent) variable of interest were included anywhere in the text of the manuscript or in a table. If the article included more than one multivariable logistic regression model, only one of the models had to include an odds ratio and confidence interval to be classified as being properly recorded. In addition, we recorded whether the article reported that the stepwise regression was performed in a forward or backward manner.

2. List Statistical Package. Articles also were examined to determine whether the authors included the type of statistical program that was used in the analyses. Citing the statistical program used to perform the multivariable analysis should be analogous to a basic science researcher indicating the particular experimental protocol used for physiologic measurements. The type of program could be mentioned anywhere in the text of the manuscript or included as a reference in order to fulfill this criterion.

3. Coding of Variables. The proper reporting of the coding of variables is important because the effect of an independent variable on the outcome variable depends on the corresponding units of measurement and the manner in which the variable was coded. For example, the regression coefficient for the impact of age on mortality will be different if age is coded in 1-year increments, in 10-year increments, or dichotomously as < 65 years vs ≥ 65 years. Articles were considered to properly report the coding of variables if the method of coding for all of the variables that remained in the final statistical model could easily be determined (such as gender being a dichotomous variable) or were referenced anywhere in the article. If the odds ratio could not be calculated accurately because the coding of all of the variables in the final model was not designated or able to be
determined, the article was considered to have improperly reported the coding of variables.

4. Interaction or Effect Modification. An interaction occurs between independent variables if the impact of one variable on the outcome event depends on the level of another variable. Articles were considered to properly report their testing for interaction terms if the concepts of interaction or effect modification were mentioned anywhere in the manuscript or if a single independent variable in any multivariable logistic regression model was reported as a cross between two single variables.

5. Collinearity. If two of the independent variables are highly correlated with one another, then collinearity occurs and may create highly unstable estimated regression coefficients. Articles were considered to properly report testing for collinearity if a test for collinearity or the concept of collinearity was discussed anywhere in the article.

6. Goodness of Fit. Indexes of goodness-of-fit evaluate how effectively the calculated model fits the actual data for estimating the outcome variable. Articles were considered to properly report goodness-of-fit testing if a test for goodness of fit, such as the Hosmer-Lemeshow statistic, was used in the article. If the article did not mention any goodness-of-fit techniques (eg, jackknife, bootstrap, Hosmer-Lemeshow, deviance, or reporting of a percentage for the dependent variable that was correctly identified by the model), the article was classified as having improperly reported a goodness-of-fit test.

7. Overfitting of the Model. Risk estimates may be unreliable if the model includes too many independent variables and too few outcome events. Although controversial, results having <10 outcome events per explanatory (independent) variable may result in inaccurate risk estimates. We attempted to determine the number of outcome events and the number of explanatory variables included in the final regression model. In those articles that reported this information, we categorized the articles with a ratio of <10:1 (10 outcome events for each single explanatory variable in the final model) as possibly having an overfit model.

8. Nonconformity to a Linear Gradient. When a linear regression coefficient is estimated for a variable, the implication is that, regardless of the value of the variable, each specific unit of change of this variable should have the same effect on the outcome. For explanatory variables that are ordinal or continuous, several gradients of effect on the outcome may occur for different ranges of the explanatory variable. If the value of the regression coefficient is reported as an average effect over the range of the explanatory variable, this may be inaccurate as the variable may have different effects at different ranges. Therefore, in the multivariable logistic model that include continuous or ordinal explanatory variables, we examined the article to assess whether the authors had assessed for conformity to a linear gradient for these variables. Unless the manuscript reported determining the impact of each explanatory variable separately in zones of ranked data or mentioned that conformity to a linear gradient was addressed, the manuscript was coded as being improperly reported.

9. Selection Process of Explanatory Variable Initially Included in the Multivariable Model. The explanatory variables that are included in the initial multivariable logistic regression analysis should have a specific reason for being selected. We classified an article as properly addressing this concern if the article included any explanation as to why the explanatory variables included in the final analysis were selected, such as significance in a prior univariate analysis, a previously reported association in the medical literature, or some degree of evidence of the biological plausibility of an association.

10. Criteria for Conditional Regression Analyses. For those articles that used pair-matched, case-control data, we required the authors to report in any part of the manuscript that a conditional (as opposed to unconditional) logistic regression analysis was used.

11. and 12. Criteria for Prediction Models. In some manuscripts, multivariable logistic regression analyses are used to develop predictive models. We developed additional criteria for these specific articles. In regard to goodness of fit, the authors were required to report some measure of model fit such as an ROC curve to receive credit for properly reporting their results. In addition, we classified articles as to whether a specific validation patient cohort also was used to test the validity of the initial model (eg, performing a test analysis on a subsample of patients followed by a subsequent validation analysis on the remaining patients or repeating the analysis on an independent sample of patients).

All of the articles were distributed and examined by two of the three authors to determine the accuracy of logistic regression analysis reporting. If there was disagreement between the initial reviews, then the third person, who had not previously examined the article, rendered the final decision.

RESULTS

Over the 6-month period, a total of 964 original articles in five pulmonary-critical care journals were manually searched to determine whether each individual manuscript included a multivariable logistic regression analysis. A total of 81 articles was identified that used multivariable logistic regression models in the analysis of the data, representing 8.4% of all published original research articles (81 articles). These articles were distributed among the following five journals: American Journal of Respiratory and Critical Care Medicine, 25 articles; CHEST, 19 articles; Critical Care Medicine, 17 articles; The European Respiratory Journal, 12 articles; and Thorax, 8 articles.

The overall results are summarized in Table 1. Odds ratios for the independent variables included in the final multivariable logistic model were reported in 79% (64 articles). Confidence intervals for these odds ratios were included in 74% of the articles (60). Only 22 of the articles (27%) specified whether the multivariable regression analysis was performed in a forward or backward manner. Of these 22 articles, 12 used a backward regression analysis and 10 used a forward process. The type of statistical package was reported in 69% (56 articles). The most commonly used programs were SPSS (SPSS; Chicago, IL) [39 articles; 39%], SAS (SAS Institute; Cary, NC) [18 articles; 32%], and STATA (STATA; College Station, TX) [9 articles; 16%].

Only 65% of the 81 articles (53 articles) properly included a coding classification for those independent variables that had been included in the final
model. Only 12% (10 articles) mentioned whether interaction terms were examined in their statistical models. Testing for collinearity was identified in 1% (1 article). Goodness-of-fit analyses were included in 16% (13 articles). In addition, only three articles (4%) specifically mentioned in the manuscript how missing data points were handled.

Overfitting of the regression model could be assessed in 57 of the manuscripts. Using the criteria of maintaining < 10 outcome events per each explanatory (independent) variable, 39% of the articles (22 articles) included final logistic regression models that were suspicious for overfitting. The criterion for nonconformity to a linear gradient did not apply to 42 of the articles in which the analyses used only binary independent variables. For the remaining 39 articles, only 6 (15%) included any indication of assessing for nonconformity to a linear gradient in the text of the manuscript. Finally, the reason that the specific explanatory variables were chosen for the multivariable logistic regression analysis was reported in only 23% of the 81 articles (19 articles). The most common reason for the inclusion of a specific variable was a significant association in a univariate analysis (95%; 18 of 19 articles). One manuscript (5%) reported that the selection of explanatory variables was due to reasons of biological plausibility.

The majority of articles (94%; 76 of 81 articles) used multivariable logistic regression modeling in a descriptive manner in order to determine the effect of an individual variable on a specific outcome while adjusting for differences in other factors. The remaining articles (6%; 5 articles) were primarily concerned with creating a predictive modeling strategy to estimate the likelihood of an outcome for a given patient from a set of observations particular to that patient. Goodness-of-fit tests and ROC curves were reported in 80% (four of five articles). However, a validation set was included in only three of the articles (60%). Five other articles were pair-matched case-control studies. Four of the five articles properly indicated that a conditional multivariable logistic regression analysis had been performed.

**Discussion**

The results of this study raise concerns about the proper reporting of multivariable logistic regression analyses in today’s pulmonary medicine and critical care medicine literature. In our review of the pulmonary and critical care medicine literature, we identified several important deficiencies in the manner in which multivariable logistic regression models were expressed. Previous investigations have reported the use of inefficient statistical analyses in univariate models in the general medical literature and multivariable logistic regression analyses in the obstetrics and gynecology literature. The inaccurate reporting of multivariable models makes a proper and comprehensive interpretation of a study’s results difficult, if not impossible.

For example, the apparent effect of an independent variable will depend on the corresponding units of measure and the coding of that variable. If the values of the regression coefficients are reported without identifying the units of coding for the independent variable, then readers will be unable to interpret the actual magnitude of the risk estimates. The importance of testing for interaction terms is evident in a 1999 trial examining the effect of body position on the incidence of nosocomial pneumonia in mechanically ventilated patients. If interaction was not considered in the logistic regression model, then the regression coefficient for the independent variable (eg, the body position of an intubated patient that had been coded as semi-recumbent or horizontal) represents the impact of this variable on the outcome event (such as the development of ventilator-associated pneumonia) for all levels of another variable (eg, the use of enteral nutrition coded as present or absent). If body position and enteral nutrition have a significant interaction, the impact of body position depends on the presence or absence of enteral feedings. Without attention to
interaction, the coefficient for body position will report a misleading quantitative estimate of the impact of the body position of an intubated patient on the development of ventilator-associated pneumonia. Finally, goodness-of-fit indexes evaluate how effectively the calculated model fits the actual data for estimating the outcome variable, and, therefore, the validity of all of the results and conclusions strongly depends on assessment of the adequacy of the regression model.

Most clinicians and researchers depend on journals, via the editorial and peer review processes, to ensure that statistical methods in published articles are being used and interpreted appropriately. Goodman and colleagues conducted a cross-sectional survey of medical journals to determine the general policies of the statistical review process. Approximately one third of the 114 journals that responded to the survey require statistical review for all accepted manuscripts. The statistical review policies differed between journals according to the size of their circulation. Eighty-two percent of journals with a circulation of > 25,000 maintained a statistical consultant on staff compared to only 31% for those journals with a circulation of < 4,100. In addition, the managing editors of these sampled journals estimated that a formal statistical review resulted in an important change in approximately 50% of the manuscripts. Goodman and colleagues further evaluated the peer review process by examining the quality of manuscripts before and after editing at the Annals of Internal Medicine. The quality of multivariable reporting was rated as one of the most deficient factors at the time of submission. However, after the manuscript was edited and revised, the clarity in the reporting of multivariable analyses was markedly improved.

Given the large number of papers submitted for publication and the limitations on journal space, editors must make important decisions concerning what type of information about regression models should be included, excluded, or made available on request. Campillo suggested that clear-cut publication criteria for generalized regression models should be made available for contributors to medical journals. Lang and Secic have recommended that an article with a multivariable logistic regression analysis should include a table with the coefficient (β), SE, Wald χ² test value, p value, odds ratio, and 95% confidence interval for each independent variable. In addition, the coding of each independent variable and a statement concerning whether the model was validated should be included. Finally, they recommended that the important issues of interaction and collinearity should be addressed in these manuscripts (Table 2). One of the major pulmonary and critical care journals has changed its "Instructions for Contributors" in order to limit the length of each article and to utilize the online reporting of the medical literature. Authors will now be required to limit the "Methods" section to 500 words; however, an extended account of methods, when appropriate, will be included online in the journal’s Web repository. Therefore, an additional venue for reporting multivariable statistical analyses in the future may be online, thereby optimally utilizing a Web repository site.

There are some potential modifiers that may influence the interpretation of our results. It was not possible to determine whether the authors improperly performed the statistical analyses or simply did not accurately report their methods in their articles. However, an improvement in the required reporting of multivariable logistic regression results would clearly facilitate the reader’s ability to interpret the data appropriately. In addition, our review examined articles from 2000, and it is possible that the standards for reporting statistical procedures have improved over the last few years. Finally, manuscripts with other forms of multivariable regression analyses, such as linear or proportional hazards regression, were not studied due to there being relatively few articles that used these techniques.

Accurate and understandable results are required to properly communicate medical research. Several deficiencies in the reporting of multivariable logistic regression analyses in the pulmonary and critical care medicine literature were identified. We recommend that the editorial boards of these pulmonary and critical care journals implement sufficient guidelines for the proper statistical reporting of multivariate analyses that will enable the reader to accurately interpret an article’s results (Table 2). In addition, it may be reasonable for the editorial boards to con-

---

Table 2—Potential Requirements for the Reporting of Multivariable Logistic Regression Analyses in the Pulmonary and Critical Care Medicine Literature

<table>
<thead>
<tr>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summarize the logistic regression equation in a table including</td>
</tr>
<tr>
<td>• No. of observations</td>
</tr>
<tr>
<td>• Coefficient of the explanatory variable and the associated SE</td>
</tr>
<tr>
<td>• Odds ratio</td>
</tr>
<tr>
<td>• 95% confidence interval of the odds ratio</td>
</tr>
<tr>
<td>• p value</td>
</tr>
<tr>
<td>Name the statistical package used in the analysis</td>
</tr>
<tr>
<td>Identify the variable used in the analysis and include descriptive statistics</td>
</tr>
<tr>
<td>Specify whether the explanatory variables were assessed for collinearity</td>
</tr>
<tr>
<td>Specify whether the explanatory variables were tested for interaction or effect modification</td>
</tr>
<tr>
<td>When possible, state whether the model was validated</td>
</tr>
</tbody>
</table>

---

www.chestjournal.org

CHEST / 123 / 3 / MARCH, 2003

927
consider increasing the frequency of involving statistical consultants to review manuscripts to a similar degree as journals use copy editors to correct the grammar and format.

It has been nearly a decade since Cancato et al first identified problems in the reporting of multivariable statistical analyses in the medical literature. Presently, the pulmonary and critical care medicine literature has not fully addressed these issues or implemented acceptable guidelines or criteria for the reporting of multivariable analyses. Hopefully, this article will promote improvements in the reporting of multivariable analyses and will allow readers to more accurately interpret study results in future articles.

**REFERENCES**

14. Tobin MJ. Authors, authors, authors: follow instructions or expect delay. Am J Respir Crit Care Med 2000; 162:1193–1194