Multivariable Analysis: A Primer for Readers of Medical Research

Mitchell H. Katz, MD

Many clinical readers, especially those uncomfortable with mathematics, treat published multivariable models as a black box, accepting the author's explanation of the results. However, multivariable analysis can be understood without undue concern for the underlying mathematics. This paper reviews the basics of multivariable analysis, including what multivariable models are, why they are used, what types exist, what assumptions underlie them, how they should be interpreted, and how they can be evaluated. A deeper understanding of multivariable models enables readers to decide for themselves how much weight to give to the results of published analyses.

Ann Intern Med. 2003;138:644-650. www.annals.org
For author affiliation, see end of text.

Most published medical research uses multivariable analysis. Unfortunately, many readers, especially those uncomfortable with mathematics, treat multivariable models as a black box, accepting the author's explanation of the results without independently assessing whether the models are correctly constructed or interpreted. However, multivariable models can be understood without undue concern for the underlying mathematics. I review the basics of multivariable analysis, including why multivariable models are used, what types exist, what assumptions underlie them, how they should be interpreted, and how they can be evaluated.

What Is Multivariable Analysis?

Multivariable analysis is a statistical tool for determining the unique contributions of various factors to a single event or outcome. For example, numerous factors are associated with the development of coronary heart disease, including smoking, obesity, sedentary lifestyle, diabetes, elevated cholesterol level, and hypertension. These factors are called *risk factors*, *independent variables*, or *explanatory variables*. Multivariable analysis allows us to determine the independent contribution of each of these risk factors to the development of coronary heart disease (called the *outcome*, the *dependent variable*, or the *response variable*).

Why Is Multivariable Analysis Needed?

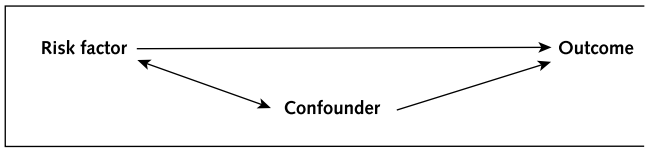
In many clinical situations, experimental manipulation of study groups would be unfeasible, unethical, or impractical. In these circumstances, multivariable analysis can be used to assess the association between multiple risk factors and outcomes. For example, we cannot test whether smoking increases the likelihood of coronary heart disease by randomly assigning persons to groups who smoke and groups who do not smoke. Although bivariate analysis of longitudinal data demonstrates that smokers are more likely than nonsmokers to develop coronary heart disease, this is weak evidence of a causal association. Perhaps the only reason smokers are more likely to develop coronary heart disease is that they are more likely to be male, live in poverty, and have a sedentary lifestyle. In other words, the relationship between smoking and coronary artery disease may be *confounded* by these other variables.

Confounding occurs when the apparent association between a risk factor and an outcome is affected by the relationship of a third variable to the risk factor and to the outcome; the third variable is a *confounder*. For a variable to be a confounder, the variable must be associated with the risk factor and causally related to the outcome (**Figure 1**). Male sex, poverty, and sedentary lifestyle could be confounders because they are associated with both smoking and coronary heart disease. With multivariable analysis, we can demonstrate that even after adjusting for male sex, poverty, and sedentary lifestyle, smoking has an independent relationship with coronary artery disease (**Figure 2**).

A study of the association between periodontal disease and coronary heart disease illustrates how multivariable analysis can be used to identify confounders (2). Bivariate analysis demonstrates that persons with periodontitis have a markedly increased rate of coronary heart disease (relative hazard, 2.66 [95% CI, 2.34 to 3.03]). If this relationship were independent and causal, then interventions that would reduce periodontitis would decrease the occurrence of coronary heart disease. However, periodontitis is also associated with several factors known to be related to coronary heart disease, including older age, male sex, poverty, smoking, increased body mass index, and hypertension, raising the question of whether the association between periodontitis is due to confounding by these factors (**Figure 3**). With multivariable adjustment for these variables, sampling design, and sampling weights, the association between periodontitis and coronary heart disease weakens substantially: the relative hazard decreases to 1.21; the 95% CIs for the relative hazard (0.98 to 1.50) crosses 1.0; and the association between periodontitis and coronary artery disease is no longer statistically significant.

Although one can theoretically distinguish independent associations from confounding, a variable may have both an independent effect on outcome and be a confounder of another variable's relationship to outcome. For example, poverty is a confounder of the relationship between smoking and coronary artery disease (poor people are more likely to smoke and to develop coronary artery disease), but poverty also has an independent effect on

Figure 1. **Relationship among risk factor, confounder, and outcome.**



Reprinted with permission from Cambridge University Press, Cambridge, United Kingdom (1).

development of coronary artery disease (after adjustment for smoking, cholesterol level, and other known risk factors, poor persons are more likely to develop coronary artery disease).

Multivariable analysis is not the only statistical method for eliminating confounding. *Stratified analysis* can also assess the effect of a risk factor on an outcome while holding other variables constant, thereby eliminating confounding. For example, the effect of periodontitis on coronary heart disease can be examined separately for men and women, which removes the effect of sex on the relationship between these diseases. If periodontitis is no longer significantly associated with coronary heart disease when men and women are looked at separately, then sex was confounding the relationship between the two. If periodontitis is still associated with coronary heart disease when men and women are assessed separately, then the effect of periodontitis on coronary heart disease is independent of sex.

Stratification works well when there are only two or three confounders. However, when there are many potential confounders, stratifying for all of them will create literally hundreds of groups in which the investigators would need to determine the relationship between periodontitis and coronary heart disease. Because the sample sizes would be small, the estimates of risk would be unstable.

Whether investigators use multivariable analysis or stratification, it is important to remember that they can only adjust for measured variables. Results may still be confounded by known and unknown unmeasured factors.

What Types of Multivariable Analysis Are Commonly Used in Clinical Research?

The three types of multivariable analysis that are commonly used in clinical research are multiple linear regression, multiple logistic regression, and proportional hazards (Cox) regression (**Table**). Linear regression is used with interval (also called *continuous*) outcomes (such as blood pressure). With interval variables, equally sized differences on all parts of the scale are equal. Blood pressure is an interval variable because the difference between a blood pressure of 140 and 143 mm Hg (3 mm Hg) is the same as the difference between a blood pressure of 150 and 153 mm Hg (3 mm Hg). Logistic regression is used with dichotomous outcomes (yes or no; for example, death). Proportional hazards regression is used when the outcome is the length of time to reach a discrete event (such as time from baseline visit to death).

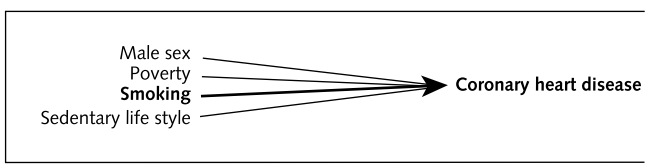
How Is the Effect of an Individual Variable on Outcome Assessed in a Multivariable Analysis?

The regression coefficient for each variable must be estimated by fitting the model to the data and adjusting for all other variables in the model. With logistic regression and proportional hazards regression, the coefficients have a special meaning. The antilogarithm of the coefficient equals the odds ratio (for logistic regression) and the relative hazard (for proportional hazards regression). The hazard is the probability that a person experiences an outcome in a short time interval, given that the person has survived to the beginning of the interval. When the outcome is uncommon (<15%), the odds ratio and relative hazard are reasonable estimates of the relative risk. For example, if the odds ratio or relative hazard for the association between smoking and fatal heart attacks is 3.0 (assuming that fatal heart attacks occurred in <15% of patients), then smoking roughly triples the risk for a fatal heart attack. If the odds ratio or relative hazard for the association between estrogen use and development of a pathologic fracture is 0.33, then persons who take estrogen have roughly a third of the risk for fracture as persons who do not take estrogen.

When the outcome is common, the odds ratio remains a useful measure of association, but it does not approximate the relative risk. For example, a randomized trial of persons with bronchopulmonary aspergillosis showed better response to itraconazole (13 of 28 patients) than to placebo (5 of 27 patients) (3). The odds ratio is 4.7 [(13 × 27)/(15 × 5)], but the relative risk is only 3.0 [(13/28)/(5/32)].

With interval-independent variables, the coefficient and the resulting odds ratio or relative hazard can be misunderstood. For example, an observational study reported that the odds ratio for the effect of low-density lipoprotein cholesterol on coronary artery calcification was 1.01 (CI, 1.00 to 1.02) (4). This may seem like a trivial effect until you notice that the odds ratio of 1.01 is for each increase of 1 mg/dL of low-density lipoprotein cholesterol. An increase of 40 mg/dL of cholesterol would produce an odds ratio of $1.49\ (1.01)^{40}$. This example demonstrates that the

Figure 2. **Multivariable association between four risk factors and coronary artery disease.**



The thicker arrow indicates that smoking is associated with coronary heart disease, even after adjustment for male sex, poverty, and sedentary lifestyle.

Table. **Types of Multivariable Analysis**

Type	Use	Special Feature
Multiple linear regression	Interval outcome	Variable coefficients have a linear relation with outcome
Logistic regression analysis	Dichotomous outcome	Model constrains probability of outcome to 0 to 1
Proportional hazards regression	Length of time to a discrete event	Useful for longitudinal studies in which persons may be lost to follow-up

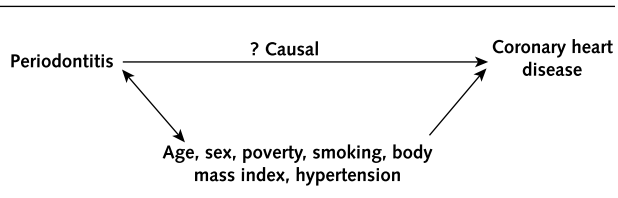
size of the coefficient of an interval variable is entirely dependent on the units being used.

With interval-independent variables, readers must also assess whether the model accurately captures the relationship between the variable and the outcome. Multivariable models assume that increases (or decreases) in an interval-independent variable will be associated with increases (or decreases) in the outcome variable. However, what if this is not the case? For example, there is a *J*-shaped relationship between alcohol consumption and mortality; nondrinkers and heavy drinkers have higher rates of mortality than do moderate drinkers. Readers would miss this important effect if alcohol consumption were entered into a logistic regression model (as an untransformed interval variable) to predict mortality. The analysis might instead indicate that there is no association between the two (the lower and higher levels of alcohol consumption could cancel each other out).

If independent variables are left in their interval form, a discrete increase (or decrease) anywhere along the scale must have an equal effect on the outcome (5). Thus, if an increase of systolic blood pressure from 150 to 160 mm Hg is associated with a 50% increase in stroke, an increase of blood pressure from 100 to 110 mm Hg must also be associated with a 50% increase in stroke. However, for many clinical variables, including blood pressure, increases (or decreases) may be relevant only after a particular threshold.

Because researchers usually do not report whether they have assessed the relationship of an interval-independent variable to the outcome over the range of values for the independent variable, it can be difficult for readers to independently judge this aspect of the analysis (5). For this reason, readers may prefer articles that use multiple dichotomous variables (also called *dummy* variables), since these allow readers to assess the effect of the risk factor on outcome for a broad range of values of the risk factor. For example, the relationship between weekly alcohol consumption and death was modeled in one study by using multiple dichotomous variables: 0 drinks (relative risk, 1.0 [referent]), 1 to 7 drinks (relative risk, 0.82), 8 to 21 drinks (relative risk, 0.82), 22 to 35 drinks (relative risk, 1.00), and more than 35 drinks (relative risk, 1.10) (6).

Figure 3. **Potential confounders of the relationship between periodontitis and coronary heart disease.**



What Assumptions Underlie Multivariable Models?

Multivariable models are mathematical expressions. We choose particular models because we believe that the data will follow the form of that model. If the model does not fit the data, our understanding of the data will be distorted.

The underlying assumption of multiple linear regression is that, as the independent variables increase (or decrease), the mean value of the outcome increases (or decreases) in a linear fashion. For example, a linear combination of age and body mass index is a good predictor of bone density in postmenopausal women. Although the relationship between the independent variable and the outcome must be linear, nonlinear relationships can be modeled by transforming the variables so that the independent variables have a linear relationship to the outcome. Logarithmic and spline transformations are often used to model nonlinear relationships.

Logistic regression models the probability of an outcome and how that probability changes with a change in the predictor variables. The basic assumption is that each one-unit increase in a predictor multiplies the odds of the outcome by a certain factor (the odds ratio of the predictor) and that the effect of several variables is the multiplicative product of their individual effects. The logistic function produces a probability of outcome bounded by 0 and 1.

Proportional hazards models assume that the ratio of the hazard functions for persons with and without a given risk factor is the same over the entire study period. This is known as the *proportionality assumption* (1, 7–9). Take, for example, a study that compares surgery to watchful waiting in patients with carotid artery stenosis. To fulfill the proportional hazards assumption, the ratio between the hazard of death with watchful waiting and the hazard of death with surgery should be constant over the course of the study. If the hazard of death is greater with surgery at the

beginning of the study, then the hazard of death should also be greater in the later follow-up period. It would be a violation of the proportionality assumption if the hazard of death were higher with surgery at the beginning of the study (as is often the case with surgical interventions because of perioperative mortality) but lower with surgery later in the study (because persons who survived after surgery had a longer life expectancy as a result of the beneficial effects of carotid endarterectomy).

If the hazard of death associated with surgery were higher at the beginning of the study and lower later in the study, the model could indicate that the risk for death (relative hazard of 1) did not differ between surgery and watchful waiting. The reason is that the higher short-term mortality associated with surgery would be averaged with the lower long-term mortality to produce a null finding. However, a finding of "no difference" would be a misleading way of characterizing that surgery has a higher short-term but lower long-term mortality.

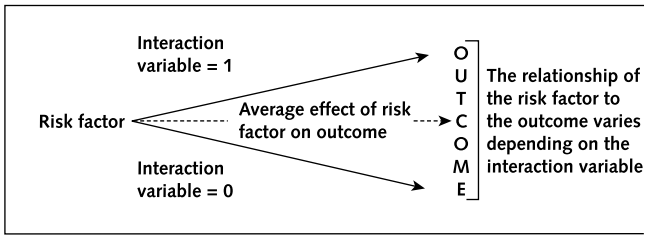
When the data do not support the proportionality assumption, proportional hazards analysis can still be performed by using *time-varying covariates*. Time-varying covariates, also called *time-dependent covariates*, are independent variables whose values change over time (1). With time-varying covariates, the model can correctly account for hazard ratios that vary over the course of the study.

A major advantage of proportional hazards analysis is that it includes persons with varying lengths of follow-up. Length of follow-up often varies in longitudinal studies for several reasons, including persons being lost to follow-up; persons developing a condition that precludes their evaluation for the study's outcome of interest; and persons being enrolled at different times (10). A person who does not experience the outcome of interest by the end of the study is considered *censored*.

In proportional hazards analyses, it is assumed that censored persons have had the same course (as if they had not been censored) as persons who were not censored. In other words, the losses occur randomly, independent of outcome. This assumption allows the follow-up time of censored persons to be included into the analysis. However, losses sometimes occur because of a systematic bias. This is the case when persons lost to follow-up are more likely to have experienced the outcome of interest than persons not lost to follow-up. For example, a randomized, controlled trial of zidovudine in HIV-infected persons reported that the condition of patients lost to follow-up was more likely to be deteriorating at the time these patients left the study than patients who stayed in the study (11).

Multiple linear regression, logistic regression, and proportional hazards models all assume that observations are independent of one another. In other words, these models cannot incorporate the same outcome occurring more than once in the same person. Although some outcomes (such as cancer) rarely occur in one patient more than once during the follow-up period, other outcomes may occur repeatedly

Figure 4. **An interaction effect.**



The effect of the risk factor on outcome (*solid lines*) differs depending on the value of the interaction variable. The dotted line is the average of the two effects. Adapted with permission from Cambridge University Press, Cambridge, United Kingdom (1).

to the same patient. For example, a patient may have multiple urinary tract infections during a follow-up period (12). In these situations, researchers may use generalized estimating equations that adjust for the correlation between repeated observations of the same patients (13). Generalized estimating equations are also used to assess outcomes that may occur in more than one body part. For example, generalized estimating equations were used to assess determinants of osteoarthritis developing in either the left or right knee (14).

Does the Model Fit the Data?

Residual analysis is the best way to assess whether a model fits the data. *Residuals* are the differences between the observed and the estimated values (1, 15). They can be thought of as the error in estimation. Large residuals suggest that the model does not fit the data. It may be that certain variables should be transformed (discussed earlier) or the correct variables are not included in the model (discussed later). Unfortunately, journals rarely print residual plots; readers must assume that the investigators reviewed them.

Are the Correct Variables in the Model?

At a minimum, each multivariable model should include the risk factor or factors and potential confounders. However, deciding which potential confounders to include is neither standard nor straightforward. Therefore, readers need to pay close attention to which variables are included in and excluded from the model.

Ideally, models should include all variables that have been hypothesized on theoretical grounds or that have been shown in previous research to be confounders of the relationship being studied. For example, a study assessing whether microalbuminuria predicted cardiovascular death included age, sex, smoking status, hypertension, dyslipidemia, diabetes, abdominal obesity, and creatinine levels in the proportional hazards model (16).

Although researchers should err on the side of including potentially important variables in the analysis, it is important to exclude extraneous ones. For example, seat belt

use should not be included in a model predicting HIV prevalence, even though it may well be associated with safer-sex practices. The reason is that seat belt use is not on the causal pathway between behavior and HIV infection. When the data set includes highly similar variables, only one should be chosen. For example, a study of neonatal mortality showed that birth weight and gestational age were too closely related to one another for both to be included in the model (17). The investigators excluded gestational age because more data were missing for this variable than for birth weight.

In deciding how many variables to include in a model, it is important to distinguish between two purposes for models: explanatory and predictive (18, 19). In an explanatory model, the goal is to correctly characterize the relationship of each predictor to the outcome variable. For that purpose, the identities of the variables in the model are critical, and the analyst must take great care in choosing which variables to include and in what mathematical form. Predictive models aim to calculate a probability that an event will occur, and as long as the model performs well in different settings, the number and identity of the variables in the model are not important. In other words, the accuracy of a predictive model's output is more important than the details of its inputs.

The number of variables in a model are often reduced by using automatic variable selection algorithms. These algorithms allow the computer to choose the variables to be included in the model, based on criteria specified by the investigator. Variable selection methods include *forward stepwise selection*, *backward deletion*, and *best subset*. For forward stepwise selection, the variable with the strongest association with the outcome is entered first, followed by the next strongest, until all variables that are related to the outcome (at a significance level specified by the investigator) are entered into the model. Any variable that has been entered into the model but that is no longer significant when the other variables have been added to the model will be sequentially deleted. For backward deletion, all variables are entered into the model and are sequentially deleted starting with the variable having the weakest association with the outcome and continuing until the only variables left in the model are those related to the outcome (at a significance level specified by the investigator). For best subset, the subset of variables that maximizes the specifications chosen by the investigator are entered into the model.

Although automatic variable selection techniques often produce models with a smaller number of independent variables, they have important limitations (20, 21). Readers may not be able to tell whether all important confounders have been included in the model. For example, a survey of oncologists' views about euthanasia (22) reported that "Predictors of support for physician-assisted suicide and euthanasia were identified by using stepwise logistic regression analysis . . . [with] selection criteria for entry into the model . . . set at an alpha level of 0.005." The authors found that five factors were associated with oncologists' being less likely to support euthanasia: reluctance to increase the morphine dose for a hypothetical patient, sufficient time, religiosity, being Catholic, and not being a surgical oncologist. Age and year of graduation from medical school were not included in the model because they did not meet the criteria of being associated with supporting euthanasia at a P value less than 0.05. This omission may be important because older physicians are more likely to be religious and less likely to support euthanasia. The authors state that attitudes toward euthanasia did not differ by age and year of graduation, but observational studies often have variables that are only weakly associated with outcome when tested singly but are strongly associated with outcome when tested jointly. To eliminate this possibility, the investigators should have rerun the analysis with all variables in the model (their sample size was large enough).

Another problem with automatic variable selection techniques is that the variables that are retained in the model are not necessarily clinically more important than the variables that are excluded. If two variables are significantly associated with one another, the model will probably choose the one with the better statistical characteristics. For these reasons and the fact that automatic methods have a high probability of generating spurious findings, statisticians strongly discourage their use for any purpose other than as an exploratory tool (23).

How Should Interactions Between Independent Variables Be Interpreted?

An interaction occurs when the effect of a risk factor on an outcome is changed by the value of a third variable. As shown in **Figure 4**, the effect of a risk factor on outcome differs depending on the value of the interaction variable. Because the value of the third variable changes the effect of the risk on an outcome, interaction is often called *effect modification*.

Interactions are different from confounding; with interactions, the relationship between the risk factor and the outcome is not due to a third variable; rather, the relationship varies depending on the value of the third variable. For example, a randomized, controlled trial (24) showed that the drug alendronate decreased the risk for fractures among patients with osteoporosis (relative hazard, 0.64 [CI, 0.50 to 0.82]) but not among those with a higher baseline bone mineral density (relative hazard, 1.08 [CI, 0.87 to 1.35]). The interaction was demonstrated statistically by entering a variable for treatment group, a variable for bone mineral density, and a product term consisting of treatment group multiplied by the bone mineral density into a proportional hazards analysis. The product term was statistically associated with outcome ($P = 0.01$), reflecting the fact that the effect of alendronate varied by baseline bone mineral density.

Although the search for interactions can be clinically

meaningful, as was the case in this study, readers should be skeptical of interaction terms especially if, unlike in this study, the researchers did not specify the hypothesis a priori. The reason is that when investigators search for interactions, they are essentially performing subgroup analyses. The more interactions searched for, the more subgroups tested, and the greater the possibility that the relationship between the dependent variable and the outcome will differ because of chance in one or more of the different subgroups.

How Well Does the Model Predict Outcome?

To assess the power of a linear regression model to predict outcome, most investigators report the adjusted R^2. The value of R^2 ranges from 0 to 1; multiplied by 100, R^2 can be thought of as the percentage of the variance in the outcome accounted for by the independent variables. Because R^2 increases in value as additional variables are included in the model, adjusted R^2 charges a penalty for every additional variable included. In a model with an R^2 close to 1, the dependent variables together accurately predict outcome.

For logistic regression models, investigators often use the Hosmer–Lemeshow goodness-of-fit test (25). This statistic compares the estimated-to-observed likelihood of outcome for groups of persons. In a well-fitting model, the estimated likelihood will be similar to the observed likelihood. Readers should be aware that the Hosmer–Lemeshow goodness-of-fit test and other available goodness-of-fit tests (26) have substantial limitations.

Goodness-of-fit tests are rarely reported with proportional hazards regression. Instead, some investigators compare estimated-to-observed time to outcome in tabular form (27). In a well-fitting model, the estimated and observed times to outcome for different groups of persons will be similar.

Although goodness-of-fit statistics are an adequate measure of how well an explanatory model accounts for the outcome, predictive models require a more quantitative measure of their ability to predict outcome. This is commonly done with a logistic regression model by computing the sensitivity, specificity, and accuracy of a model's predictions at a particular cut-point (for example, assuming that all persons with a predicted probability of ischemia of $\geq$15% actually have ischemia). The area under the receiver-operating characteristic (ROC) curves allows assessment of the predictive value of a logistic regression model over various cut-offs of probability of outcome (28, 29).

Is the Model Reliable?

Readers should note whether the investigator has shown that a multivariable model is reliable before accepting it at face value. The reliability of a model depends on its purpose. If the model is explanatory, reliability means that a different set of data would probably yield a model with the same variables and similar coefficients. A reliable predictive model predicts outcomes equally well for settings or for data other than those for which it was developed.

An important threat to reliability is insufficient sample size. As a rule of thumb, to have confidence in the results, there should be at least 20 persons for each independent variable eligible to be included in a linear regression model and at least 10 outcomes for each independent variable eligible to be included in a logistic regression or proportional hazards model (30–32). Sample size requirements for logistic regression and proportional hazards regression are expressed as outcomes per variable (rather than persons per variable). The required sample size is based on the less frequent state of the dichotomous outcome. If only 6 persons in a study develop cancer, the model will have difficulty predicting how three variables independently predict cancer development, even if 994 persons did not develop cancer.

Wide CIs are the result of insufficient sample size. For example, a logistic regression analysis was used to assess the association between cigarette smoking during adolescence and development of a panic disorder in early adulthood (33). The investigators adjusted for eight potential confounders and found that smoking during adolescence was strongly associated with development of a panic disorder in young adulthood (odds ratio, 15.6). However, 95% CIs for the odds ratio varied greatly from 2.31 to 105.14. Although the sample size was large ($n = 688$), a panic disorder developed in only 7 persons.

One other caution about sample size is as follows: Even if a study has a large enough number of events per independent variable, the estimates of the association between a risk factor and an outcome may still be inaccurate if the risk factor is rare. For example, if only 10 of 800 persons in a study are injection drug users, the model cannot accurately estimate the relationship between being an injection drug user and the outcome.

In addition, a sufficient sample size is no guarantee that results from a model can be reproduced with new data. Although some decrement in performance is acceptable when a model is rerun with new data, a reliable model will perform well with new data. Unfortunately, investigators cannot always collect additional data. In these situations, investigators may report one of three alternative methods for assessing the reliability of a model: *split-group*, *jackknife*, or *bootstrap* (1). With split-group validation, investigators divide the data set into two parts; the model is developed on the first data set and then validated on the second data set. With a jackknife procedure, the investigators sequentially delete persons from a data set and repeatedly recomputes the model with each person missing once. With a bootstrap procedure, the investigators take random samples of persons from a data set with replacement of the previously selected persons so that they are eligible to be resampled. The result is that a person may be chosen more

than once. Although none of these methods can be considered definitive, if they closely approximate the original model, readers can have greater confidence in the results.

Even if a prediction model is reliable, it may not be useful in clinical practice for several reasons (18, 19). It may require clinicians to have certain laboratory results that may not be available, or it may have been developed and validated on patients different from those seen in clinical practice. For example, if an analysis included only men between 20 and 60 years of age, we would not assume that the results would be applicable to a 70-year-old woman. Similarly, if less than 5% of the sample was younger than 30 years of age, the model's predictions for 22-year-old men may not be very robust. These principles can be seen clearly by translating the results of a multivariable model into tabular form; this allows readers to see the probabilities and CIs of outcome for different groups of persons (18).

Conclusion

By understanding the concepts that underlie multivariable analysis, clinical readers can better evaluate the goals and results of published studies.

From San Francisco Department of Public Health, San Francisco, California.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Mitchell H. Katz, MD, 101 Grove Street, Room 308, San Francisco, CA 94102; e-mail, Mitch.Katz@sfdph.org.

References

1. **Katz MH.** Multivariable Analysis: A Practical Guide for Clinicians. Cambridge: Cambridge Univ Pr; 1999.

2. **Hujoel PP, Drangsholt M, Spiekerman C, DeRouen TA.** Periodontal disease and coronary heart disease risk. JAMA. 2000;284:1406-10. [PMID: 10989403]

3. **Stevens DA, Schwartz HJ, Lee JY, Moskovitz BL, Jerome DC, Catanzaro A, et al.** A randomized trial of itraconazole in allergic bronchopulmonary aspergillosis. N Engl J Med. 2000;342:756-62. [PMID: 10717010]

4. **O'Malley PG, Jones DL, Feuerstein IM, Taylor AJ.** Lack of correlation between psychological factors and subclinical coronary artery disease. N Engl J Med. 2000;343:1298-304. [PMID: 11058674]

5. **Concato J, Feinstein AR, Holford TR.** The risk of determining risk with multivariable models. Ann Intern Med. 1993;118:201-10. [PMID: 8417638]

6. **Grønbaek M, Becker U, Johansen D, Gottschau A, Schnohr P, Hein HO, et al.** Type of alcohol consumed and mortality from all causes, coronary heart disease, and cancer. Ann Intern Med. 2000;133:411-9. [PMID: 10975958]

7. **Kahn HA, Sempos CT.** Statistical Methods in Epidemiology. New York: Oxford Univ Pr; 1989:193-8.

8. **Lawless JF.** Statistical Models and Methods for Lifetime Data. New York: Wiley; 1982:394-5.

9. **Kalbfleish JD, Prentice RL.** The Statistical Analysis of Failure Time Data. New York: Wiley; 1980:89-98.

10. **Kelsey JL, Whittemore AS, Evans AS, Thompson WD.** Methods in Observational Epidemiology. New York: Oxford Univ Pr; 1996:130-4.

11. **Volberding PA, Lagakos SW, Koch MA, Pettinelli C, Myers MW, Booth DK, et al.** Zidovudine in asymptomatic human immunodeficiency virus infection. A controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. The AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases. N Engl J Med. 1990;322:941-9. [PMID: 1969115]

12. **Hooton TM, Scholes D, Hughes JP, Winter C, Roberts PL, Stapleton AE, et al.** A prospective study of risk factors for symptomatic urinary tract infection in young women. N Engl J Med. 1996;335:468-74. [PMID: 8672152]

13. **Liang KY, Zeger SL.** Longitudinal data analysis using generalized linear models. Biometrika. 1986;73:13-22.

14. **McAlindon TE, Felson DT, Zhang Y, Hannan MT, Aliabadi P, Weissman B, et al.** Relation of dietary intake and serum levels of vitamin D to progression of osteoarthritis of the knee among participants in the Framingham Study. Ann Intern Med. 1996;125:353-9. [PMID: 8702085]

15. **Glantz SA, Slinker BK.** Primer of Applied Regression and Analysis of Variance. New York: McGraw-Hill; 1990:110-80.

16. **Gerstein HC, Mann JF, Yi Q, Zinman B, Dinneen SF, Hoogwerf B, et al.** Albuminuria and risk of cardiovascular events, death, and heart failure in diabetic and nondiabetic individuals. JAMA. 2001;286:421-6. [PMID: 11466120]

17. **Phibbs CS, Bronstein JM, Buxton E, Phibbs RH.** The effects of patient volume and level of care at the hospital of birth on neonatal mortality. JAMA. 1996;276:1054-9. [PMID: 8847767]

18. **Braitman LE, Davidoff F.** Predicting clinical states in individual patients. Ann Intern Med. 1996;125:406-12. [PMID: 8702092]

19. **Wasson JH, Sox HC.** Clinical prediction rules. Have they come of age? [Editorial] JAMA. 1996;275:641-2. [PMID: 8594248]

20. **Greenland S.** Modeling and variable selection in epidemiologic analysis. Am J Public Health. 1989;79:340-9. [PMID: 2916724]

21. **Steyerberg EW, Eijkemans MJ, Habbema JD.** Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999;52:935-42. [PMID: 10513756]

22. **Emanuel EJ, Fairclough D, Clarridge BC, Blum D, Bruera E, Penley WC, et al.** Attitudes and practices of U.S. oncologists regarding euthanasia and physician-assisted suicide. Ann Intern Med. 2000;133:527-32. [PMID: 11015165]

23. **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;15:361-87. [PMID: 8668867]

24. **Cummings SR, Black DM, Thompson DE, Applegate WB, Barrett-Connor E, Musliner TA, et al.** Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. JAMA. 1998;280:2077-82. [PMID: 9875874]

25. **Hosmer DW, Lemeshow S.** Applied Logistic Regression. New York: Wiley; 1989:187-215.

26. **Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S.** A comparison of goodness-of-fit tests for the logistic regression model. Stat Med. 1997;16:965-80. [PMID: 9160492]

27. **Colford JM Jr, Tager IB, Hirozawa AM, Lemp GF, Aragon T, Petersen C.** Cryptosporidiosis among patients infected with human immunodeficiency virus. Factors related to symptomatic infection and survival. Am J Epidemiol. 1996;144:807-16. [PMID: 8890659]

28. **Hanley JA, McNeil BJ.** The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36. [PMID: 7063747]

29. **Hsiao JK, Bartko JJ, Potter WZ.** Diagnosing diagnoses. Receiver operating characteristic methods and psychiatry. Arch Gen Psychiatry. 1989;46:664-7. [PMID: 2735814]

30. **Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR.** A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49:1373-9. [PMID: 8970487]

31. **Peduzzi P, Concato J, Feinstein AR, Holford TR.** Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48:1503-10. [PMID: 8543964]

32. **Harrell FE Jr, Lee KL, Matchar DB, Reichert TA.** Regression models for prognostic prediction: advantages, problems, and suggested solutions. Cancer Treat Rep. 1985;69:1071-77. [PMID: 4042087]

33. **Johnson JG, Cohen P, Pine DS, Klein DF, Kasen S, Brook JS.** Association between cigarette smoking and anxiety disorders during adolescence and early adulthood. JAMA. 2000;284:2348-51. [PMID: 11066185]