Sufficient data

Eugene H. Blackstone, MD

From the Department of Thoracic and Cardiovascular Surgery, Heart and Vascular Institute, and Department of Quantitative Health Sciences, Research Institute, Cleveland Clinic, Cleveland, Ohio.

This work was funded by the Cleveland Clinic Department of Thoracic and Cardiovascular Surgery.

Disclosures: Author has nothing to disclose with regard to commercial support.

Received for publication July 18, 2016; accepted for publication July 18, 2016; available ahead of print Aug 26, 2016.

Address for reprints: Eugene H. Blackstone, MD, Department of Thoracic and Cardiovascular Surgery, Cleveland Clinic, 9500 Euclid Ave/Desk JJ-4, Cleveland, OH 44195 (E-mail: blackse@ccf.org).

J Thorac Cardiovasc Surg 2016;152:1235-6

0022-5223/\$36.00

Copyright © 2016 by The American Association for Thoracic Surgery

<http://dx.doi.org/10.1016/j.jtcvs.2016.07.046>

Statistical reviewers often indicate they believe that investigators' analyses of binary (yes/no) endpoints (death, stroke, reoperation) are not supported by sufficient data (underpowered). Thus, we offer the following guidance and explanations about "sufficient data," taken with slight modification from Chapter 6 of the Kirklin/Barratt-Boyes book *Cardiac Surgery*, Fourth Edition.¹

A common misconception is that the larger the study group, the larger the amount of information available for analysis of an event. In studies of binary events, however, the effective sample size for analysis is the number of events that have occurred, not the total size of the study group. Thus, a study of 2000 patients experiencing 10 events has an effective sample size of 10, not 2000.

Ability to detect differences in outcome events is coupled with effective sample size. A statistical quantification of the ability to detect a difference is the *power* of a study. Binary outcomes always have low power compared with a continuous outcome such as forced expiratory volume in 1 second after lung transplantation or left ventricular end-systolic volume after repair of ischemic mitral regurgitation. This is why randomized trials with clinical events endpoints must be huge compared with trials that use surrogated continuous endpoints.²⁻⁴

To make matters worse, many variables in a data set represent subgroups of patients, and events in some of them may be few in number. If a single patient in a small subgroup experiences an event, multivariable analysis of that binary endpoint may identify that subgroup as one at high risk when, in fact, the variable only identifies a specific patient, not a common denominator of risk. The purpose of multivariable analyses, particularly of observational data, is to identify general risk factors associated with an event, not individual patients experiencing events (something already known)!

Thus, more than 1 event needs to be associated with every variable considered in the analysis. The question is, "how many events are sufficient?" The rule of thumb in multivariable analysis is that the ratio of events to risk factors identified should be about 10 to 1,⁵ but the number varies from 5

Eugene H. Blackstone, MD

Central Message

For outcome events, the effective sample size is number of events, not total study size. Thus, "sufficient data" means a sufficient number of events associated with individual risk factors.

to as many as 30, representing liberal versus conservative thought.⁶⁻⁹

We have explored this matter by bootstrap methods.^{10,11} Quite simply, we randomly sample a data set, pulling out patients to build another data set of equal size. This is done "with replacement," meaning that the original data set is left intact, so a given patient can be selected again, by chance. This is known as a bootstrap sample. This new data set is analyzed and the results stored. This process is repeated hundreds to thousands of times. The results are then aggregated across all these analyses by tallying the number of times a variable has been selected. Variables appearing in, say, half the analyses, are said to be "risk factors." For a typical size surgical study and a *P*-value criterion of .05 to retain variables in each analysis, we have found that the rule of 10 is pretty good, and if anything, conservative.

Thus, we think that a rule of 30 events per factor is probably too conservative, 10 about right (sufficient data), and less than 5 too liberal (insufficient data). This means that if a variable representing a subgroup is associated with fewer than 5 of the events of interest, that variable must be excluded from consideration in a multivariable analysis. Bear in mind, however, that variables may be highly correlated and subgroups may overlap; thus, in the course of analysis, the number of events in a subgroup may effectively fall below 5, which is insufficient data.

Thus, there is both an upper limit of risk factors that can be identified by a multivariable analysis of a binary endpoint, which is related to the total number of events experienced, and a lower limit of events to allow a particular variable to be considered in the analysis. Sufficient data

implies having enough events available to test all relevant risk factors for their association with the event.

References

1. Kouchoukos NT, Blackstone EH, Hanley FL, Kirklin J. *Cardiac Surgery*. 4th ed. Philadelphia: Elsevier; 2012:265.
2. Hsieh FY. Sample size tables for logistic regression. *Stat Med*. 1989;8:795-802.
3. Perrault LP, Moskowitz AJ, Kron IL, Acker MA, Miller MA, Horvath KA, et al. Optimal surgical management of severe ischemic mitral regurgitation: to repair or to replace? *J Thorac Cardiovasc Surg*. 2012;143:1396-403.
4. Smith PK, Michler RE, Woo YJ, Alexander JH, Puskas JD, Parides MK, et al. Design, rationale, and initiation of the Surgical Interventions for Moderate Ischemic Mitral Regurgitation Trial: a report from the Cardiothoracic Surgical Trials Network. *J Thorac Cardiovasc Surg*. 2012;143:111-7. 117.e1.
5. McDonald CJ, Hui SL. The analysis of humongous databases: problems and promises. *Stat Med*. 1991;10:511-8.
6. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. New York: Springer; 2015.
7. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. 2007;165:710-8.
8. Courvoisier DS, Combescure C, Agoritsas T, Gayet-Ageron A, Perneger TV. Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. *J Clin Epidemiol*. 2011;64:993-1000.
9. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol*. 2003;158:280-7.
10. Breiman L. Bagging predictors. *Machine Learning*. 1996;24:123-40.
11. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med*. 1992;11:2093-109.