Published in final edited form as:

Am J Public Health. 2013 June ; 103(6): e1. doi:10.2105/AJPH.2013.301234.

Achieving Consensus on Terminology Describing Multivariable Analyses

Alexander C. Tsai, MD, PhD

Center for Global Health and the Chester M. Pierce, MD Division of Global Psychiatry, Massachusetts General Hospital, Boston. He is also with Harvard Medical School, Boston, MA.

In their recent article, Hidalgo and Goodman¹ call our attention to the need for consistent and distinctive use of the terms “multivariable” and “multivariate.” They introduced a point of confusion, however, with their suggestion that the terms “linear, logistic, multivariate, or proportional hazards” be employed to indicate continuous, dichotomous, repeated measures, or time-to-event outcomes, respectively. I find their suggestion confusing because it suggests the absence of an overlap between “linear,” “logistic,” and “multivariate.” Yet a regression model fit to repeated-measures data may assume a normal *or* logistic distribution (or any of a number of other distributions), making it a multivariate linear or multivariate logistic regression model.

I believe their article invites two additional teaching points for reinforcement, which I underline here. I surveyed 22 empirical articles published in the same January 2013 issue. Of these, three articles (13.6%) used the term “multivariate” incorrectly, including one article that used the term “bivariate.” Five articles (22.7%) used “multivariate,” “multiple” (i.e., multiple regression), and “multivariable” interchangeably, including one article that used the term “bivariate.” Three articles (13.6%) used the term “multivariate” correctly in the context of repeated-measures or nested data, while eleven (50%) contained no violations.

First, the term “univariate” is most appropriate (and perhaps is unnecessarily described explicitly as such) when there is only one response variable per observation. Depending on whether there is one explanatory variable or multiple explanatory variables, the terms “univariable” and “multivariable” (i.e., multiple) would help to additionally clarify the kind of univariate analysis being conducted. A *t*-test comparing mean levels of a response variable between two subgroups is a univariable analysis; and so is a regression model of the same response variable, with the subgroup specified as the single binary explanatory variable. Use of the term “bivariate” to describe such a *t*-test, while common (and observed twice in the cursory survey described above), introduces unnecessary confusion and should be discouraged.

Second, the term “multivariate” should be understood to apply to a diverse set of methods that allow for more than one response per observation.² Hidalgo and Goodman noted certain applications of repeated measures regression, or—to retain consistency with the terminology I elaborated upon—multivariate multivariable regression. This presents a compelling rationale for why the terms “multivariate” and “multivariable” should not be used interchangeably. Other types of statistical analyses are also classified as “multivariate,” including discriminant analysis, canonical correlation, and principal components analysis.

In this letter, I extend the comments of Hidalgo and Goodman. These nuances in the use of statistical terminology, however, have not gained formal traction at most peer-reviewed

journals.³ This is likely because equally reasonable perspectives are also taught. For example, in one leading textbook for clinical practitioners, the author says that “multivariate analysis refers to simultaneously predicting multiple outcomes.”^{4(p1)} But the author also writes (contrary to the recommendation above): “I think it is more informative to restrict the term ‘univariate’ to analyses of a single variable, while restricting the term ‘bivariate’ to refer to the association between two variables.”^{4(p5)} Ultimately, achieving consensus on these issues will help to avoid further confusion and facilitate substantive progress on communicating the results of public health research in the published literature.

Acknowledgments

A.|C. Tsai acknowledges salary support from the National Institutes of Health (K23 MH-096620).

References

1. Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health*. 2013; 103(1):39–40. [PubMed: 23153131]
2. Rencher, AC. *Methods of multivariate analysis*. 2nd ed.. Wiley-Interscience; New York: 2002.
3. Peters TJ. Multifarious terminology: multivariable or multivariate? univariable or univariate? *Paediatr Perinat Epidemiol*. 2008; 22(6):506. [PubMed: 19000286]
4. Katz, MH. *Multivariable analysis: a practice guide for clinicians*. Cambridge University Press; New York: 1999.