

Lies, Damned Lies, and Statistics
Anthony N. DeMaria
J. Am. Coll. Cardiol. 2008;52:1430-1431
doi:10.1016/j.jacc.2008.09.013

This information is current as of July 12, 2010

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://content.onlinejacc.org/cgi/content/full/52/17/1430>

JACC
JOURNAL of the AMERICAN COLLEGE of CARDIOLOGY



EDITOR'S PAGE

Lies, Damned Lies, and Statistics*

“Facts are stubborn things, but statistics are more pliable.”

—Mark Twain (1)

It has often been said in the halls of academia that physicians would have difficulty critically reading the medical literature if they had never done research themselves. I believe this refers to the need to be able to evaluate the validity of the methods and analysis of a publication. Specifically, readers must be able to determine that a paper has included the appropriate study group, employed valid and accurate technology, eliminated confounding variables, insured adequate power, selected proper end points, and distinguished association from cause. However, I am impressed that statistical methods are playing an increasingly important role in medical (especially clinical) research, are becoming more complex, and are making it challenging for even experienced investigators to interpret the literature accurately.

I must confess to always having viewed studying statistics as similar to a screening colonoscopy; I knew that it was important and good for me, but there was little that was pleasant or fun about it. It seems that this impression was shared by many others. Derisive comments about statistics were common, as evidenced by the opening quotation from Mark Twain and the famous statement of Benjamin Disraeli (1) that there are 3 kinds of lies: lies, damned lies, and statistics. One of my mentors quipped that if he said to a desk “arise” 100 times, and it floated in the air only once, this would be very significant, but not statistically significant. Despite their critical role in science, statistics sometimes seemed to be used to prove a hypothesis rather than to test one, and this impression has persisted as the techniques have become more complicated.

When I first began to do research, the commonly employed statistical methods were relatively simple. Most data could be analyzed by the Student *t* test, chi-square, linear regression, or multivariable analysis. As time has gone on, however, both the number and complexity of statistical approaches have become greater. A description of statistical analysis now typically occupies a substantial part of the Methods section of each manuscript. We currently have 2 Associate Editors at *JACC* who are statisticians and who review every manuscript before publication. Interestingly, it has been our experience that agreement among statistical experts is often no greater than it is in other areas of cardiovascular medicine. It is not difficult to imagine the dilemma created when 2 statisticians disagree and debate in terminology understandable largely only by those in the field.

Of the statistical approaches that have been applied more recently, 2 stand out as being particularly frequent and contentious. Propensity scores are utilized in many reports in an attempt to adjust for variables inherent in the lack of randomization. Noninferiority testing is often employed to document that a new diagnostic modality is equivalent to an existing technique, or that a new therapy is equal to one that has been shown to



**Anthony N. DeMaria,
MD, MACC**

*Editor-in-Chief,
Journal of the American
College of Cardiology*

*... statistical
methods are playing
an increasingly
important role in
medical (especially
clinical) research,
are becoming more
complex, and are
making it challeng-
ing for even expe-
rienced investiga-
tors to interpret the
literature . . .*

*Brainy Quote. Benjamin Disraeli Quotes. Available at: <http://www.brainyquote.com/quotes/quotes/b/q107958.html>. Accessed September 29, 2008.

be more effective than placebo. Both have significant limitations and often engender spirited discussions at our Editorial Board meetings.

The results of a clinical trial may be due either to the intervention tested or to pre-existing confounding variables that differ between treated and control groups. Prospective trials that randomize study patients are the best way to eliminate such variables. In observational studies, however, variables that can affect prognosis usually exist and may influence the selection of management. Propensity analysis is a statistical method to correct for such confounders post-hoc by identifying factors with the potential to affect outcome and creating a score for these variables for each individual subject. The subjects can then be matched for their score so that these variables can be eliminated or minimized.

Observational studies are less expensive and usually require less time than randomized clinical trials. In addition, they generally provide a wider spectrum of patients. Therefore, the potential of a propensity score to overcome the lack of randomization is very seductive. However, different approaches to propensity analysis exist, and errors in the model such as the failure to use interaction terms may result in bias. Moreover, propensity analysis typically results in a comparison of groups with a similar composite of variables, but not necessarily the same type or severity. The most important limitation of a propensity score, however, is that it addresses only overt identifiable factors and cannot account for hidden confounders. Regardless of how many factors are identified (I am aware of one analysis that evaluated 74), uncertainty will always exist as to whether you have captured the (often intangible) reasons that 1 patient was managed differently from another. For this reason, many of our reviewers, and some of our editors, feel that the propensity score is fatally flawed and give it little credence. At the very least, sensitivity analysis should be attempted to indicate the level of hidden bias that would be required to alter the conclusions of the study (2).

When a therapy exists that has proven to be superior to placebo, or a diagnostic technique is available that is acceptably accurate is available, an alternate modality of equal or nearly equal efficacy may be more desirable if it has beneficial ancillary characteristics. Thus, a therapy with fewer side effects or a less expensive diagnostic test would not need to be superior to existing modalities. In such cases, noninferiority testing is often carried out (3). Noninferiority testing requires that the efficacy of the existing modality be accurately quantified, which may be

complicated when multiple prior studies are available. Since the new study may involve patients with different characteristics than in prior reports, a noninferiority margin is defined that establishes the greatest decrease in efficacy one will accept in return for the benefits of the new modality. This noninferiority margin is usually based on clinical judgment and is therefore somewhat subjective; it is usually 10% to 20%.

In addition to the variability inherent in determining the margin, nuances in the statistical aspects of noninferiority trials have been the source of criticism. In fact, 1 paper reviewed 8 recently published cardiovascular clinical noninferiority trials and was able to confirm noninferiority in only 4 (3). When these statistic issues are combined with the frequent use of composite end points and the differing classification of side effects as adverse events or primary end points, it is not surprising that noninferiority trials are often viewed with caution.

The ability to critically read the medical literature is becoming increasingly difficult, even for investigators. The statistical aspects of the design and analysis of clinical trials have become more complex and of greater importance to the conclusions that are drawn from the data. At times it is difficult to determine if statistics are being used to analyze the data or substantiate it. Some statistical approaches, such as propensity scores and noninferiority testing, are clearly imperfect or seriously flawed. Nevertheless, they may contribute to treatment recommendations and/or reimbursement decisions. It appears that these statistical matters have resulted in some frustration and skepticism on the part of our readers. As an editor, and as a reader myself, I am sympathetic to these sentiments.

Address correspondence to:

Dr. Anthony N. DeMaria
Editor-in-Chief, *Journal of the American College of Cardiology*
3655 Nobel Drive, Suite 630
San Diego, California 92112
E-mail: ademaria@acc.org

REFERENCES

1. The Quotations Page. Available at: <http://www.quotationspage.com/quote/27556.html>. Accessed September 29, 2008.
2. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med* 2002;137:693-6.
3. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006;145:62-9.

Lies, Damned Lies, and Statistics

Anthony N. DeMaria

J. Am. Coll. Cardiol. 2008;52:1430-1431

doi:10.1016/j.jacc.2008.09.013

This information is current as of July 12, 2010

**Updated Information
& Services**

including high-resolution figures, can be found at:

<http://content.onlinejacc.org/cgi/content/full/52/17/1430>

Supplementary Material

Supplementary material can be found at:

<http://content.onlinejacc.org/cgi/content/full/52/17/1430/DC1>

Citations

This article has been cited by 1 HighWire-hosted articles:

<http://content.onlinejacc.org/cgi/content/full/52/17/1430#otherarticles>

Rights & Permissions

Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:

<http://content.onlinejacc.org/mis/permissions.dtl>

Reprints

Information about ordering reprints can be found online:

<http://content.onlinejacc.org/mis/reprints.dtl>