The power of power

David Blandino, MD
UPMC Shadyside Hospital Family Practice Residency

Goutham Rao, MD, Feature Editor

The first trial of beta-blockers in myocardial infarction was entitled "The lack of prophylactic effect of propranolol in myocardial infarction"[1]—a conclusion inconsistent with our current understanding of beta-blocker therapy. The reason has to do with "statistical power"—a statistic that tells us the chance of finding a significant difference between treatments.[2]

TYPE 1 AND TYPE 2 ERRORS

We draw conclusions based on the results of clinical trials. No trial is perfect. Trials are designed with the knowledge that there is a probability of drawing a conclusion based on the results that does not represent the truth about 2 or more therapies.

If we conclude from the results of a trial that 2 therapies are of different effectiveness, when in reality they are the same, we have committed what is known as a type 1 error. The probability of making a type 1 error is termed the *alpha*. Trials are usually designed with an a of 0.05 (5%).

On the other hand, if we conclude that the 2 therapies are the same when they are actually different, we have committed a type 2 error. The probability of making a type 2 error is known as the *beta*.

Perhaps a bit more intuitively, we are often interested in knowing the probability of finding a difference when there really is one. This probability is called *power* and may be expressed as $1 - \beta$.

Correspondence: Goutham Rao, MD, 3518 Fifth Avenue, Pittsburgh, PA 15261. E-mail: grao@upmc.edu.

POWER IN STUDY DESIGN

In designing a study, the power of a study to detect differences between 2 groups depends upon the number of subjects in each group, whether the groups are equal in size, the variability of responses among subjects, the magnitude of difference one is trying to detect, and the probability of making a type 1 error.[3] Researchers can make some educated assumptions to determine the number of subjects to include in a study to assure that clinically relevant differences are found between 2 groups if they exist.

Practicing clinicians should use power to determine the impact of a negative study. For example, the propranolol study[1] was designed with a power of only 23%, meaning that there was only a 23% chance of detecting a difference. Drawing conclusions about the lack of effectiveness of propranolol based on this study, therefore, would be a mistake. In clinical trials of an active drug vs a placebo, 100 subjects in each group or more are often needed to detect clinically relevant results—so beware of negative results with small numbers of patients.

REFERENCES

1. Clausen J, Felski M, Jorgensen FS. The lack of prophylactic effect of propranolol in myocardial infarction. *Lancet* 1966; 2:920–924.
2. Freiman JA, Chalmers TC, Smith H. The importance of beta, the Type II error, and sample size in the design and interpretation of the randomized control trial. *N Engl J Med 1978; 299:690–694.*
3. Cohen J. A power primer. *Psychological Bulletin* 1992; 112:155–159.