Editorial Comment

Pitfalls in Crossover Studies

As Lin et al correctly point out in an article published in this issue of *Current Therapeutic Research*®, one of several problems in a crossover study is a carryover effect occurring when patients are switched to comparator treatments. To establish that their patients with nonischemic heart failure had achieved roughly the same status at the start of the second treatment in the sequence (the crossover design called for felodipine followed by fosinopril or vice versa), these investigators inserted a sufficiently long washout period. They also measured a number of physiologic and clinical end points. Fortunately, the patients succeeded in returning to near-baseline levels, as assessed by results in the specified parameters. However, as a friend of mine once said to me, "Once having received a medication, the patient is never really the same."

A second and very important problem often encountered while carrying out a crossover study concerns the number of dropouts. A large percentage of the patients in the study by Lin et al dropped out or were removed for noncompliance. Dropouts complicate the statistical analysis of study data because investigators have only partial information with which to compare treatments. In fact, many investigators either incorrectly ignore the missing data or insert values such as the mean of the available data or the worst possible score. Thus, the benefit of the crossover design to decrease the sources of statistical variation is lost. In some cases, the problem is so great that the trial cannot be analyzed as a crossover, but instead must be assessed as a parallel-group study based only on the data from the first treatment period. This approach invariably deals with diminished treatment-group numbers (the purpose of crossover designs) so that it will be difficult to draw meaningful conclusions.

Extrapolating from animal studies to human clinical trials remains a difficult and complex task. However, assuming that genetic differences exist between patients in Asia and patients in Western countries may create an even greater problem in the interpretation of study results. Many authors have stated that the differences between individuals within these patient groups are greater than the between-group differences. Is the prescription of lower doses of cardiac medications to Chinese patients than those given to Westerners (as noted by Lin et al) due more to genetics or to differences in the etiology of nonischemic heart failure, diet, relations between patients and their physicians, lower body weight, compliance, or the philosophy on disease treatment? Only more research to establish a better understanding of these factors, performed in conjunction with genetic studies (and resultant phenotypes), will lead us closer to the truth.

To maintain a true double-blind study design, researchers usually strive for comparability between the dosage forms in terms of size, color, odor, and

presentation form. However, when creating tablets or capsules that are as identical as possible, important mistakes may be made. Sometimes, repackaging the medications can lead to bioavailability problems. The investigators must be aware of these difficulties and either show that bioavailability has been maintained or, perhaps more simply, should carry out a double-dummy design because it would be more efficient.

Why would *Current Therapeutic Research*[®] decide to publish a study with a problematic design, study groups that may have been too small, or, most of all, treatments that had been shown to be ineffective for a particular disease when there are already several standard therapeutic interventions available? To determine the answer, one need only consider the use of beta-blockers for heart failure. Once, beta-blockers were considered less than useful. They were even thought to contribute to patient deterioration. Now, however, with new research available, beta-blockers have been shown to be effective for many patients with heart failure. Additionally, even if this study demonstrated no differences between the 2 treatments under evaluation, it still may have been well worth undertaking. The investigators' work has added to our knowledge of the 7-point clinical scale, 2-dimensional echocardiography, exercise tests, and neurohumoral measures in nonischemic heart failure.

Michael Weintraub, MD
Editor-in-Chief