Exercise Testing as Outcome in Congestive Heart Failure Trials
Design Considerations When Interpreting Results

J. Lubsen
Société pour la Recherche Cardiologique SOCAR SA, Givrins, Switzerland and Erasmus University, Rotterdam, The Netherlands

Summary
In addition to standard features of clinical trial design such as randomisation and double-blinding, sensitivity to drug effects is an important consideration when conducting exercise capacity trials in patients with heart failure. Two issues need to be addressed in this context. Firstly, it is important to enrol patients who are potential responders. Patients who have, for their age and sex, normal exercise capacity are unlikely to improve, even when given a drug that has a positive effect on exercise capacity. In addition, those patients who remain clinically stable following withdrawal of their previous drug therapy are unlikely to respond subsequently to an experimental drug with a similar mechanism of action. Secondly, failure to complete scheduled exercise tests during follow-up, prompting a 'per-protocol' analysis of results, may mask the drug's actual effect. To avoid this, an 'intention-to-treat' approach to data collection and analysis, with appropriate allowance made for missing test data, should be adopted.

Elsewhere in this issue, Swedberg argues that 'while the beneficial effects of angiotension converting enzyme (ACE) inhibitors on survival are evident, their impact on exercise tests has been inconsistent' (Swedberg 1994). The author concludes that the reason why trials with ACE inhibitors have shown variable results in this regard may have been (i) that sample sizes have been too small and (ii) that ACE inhibitors lack a pronounced effect on exercise capacity.

There are several possible explanations as to why a trial assessing the effects of a drug on exercise capacity in patients with congestive heart failure may fail to show an effect:
1. The reproducibility of exercise testing may be insufficient to detect changes in exercise capacity of the magnitude that result from treatment.
2. Exercise capacity as assessed by ergometry may be insensitive to changes in the severity of heart failure.
3. The trial may have enrolled a category of patients who are insensitive to the effect of the drug.
4. The drug's effect on exercise capacity may be masked because of the way in which the follow-up exercise tests are performed, or the results are analysed.

The lack of reproducibility of exercise testing is always a potential explanation for failure to detect a therapeutic effect. If there is large within-patient variation in exercise testing parameters and/or if
they are subject to random error, then it will not be possible to reproduce reliably the measured response (i.e. it will show large variation when repeated in the same patient). In a clinical trial, this will result in a large standard error of the effect estimate, wide confidence intervals and large p values (Gardner & Altman 1986). The response variability will increase further if there is large between-patient variation and if no pretreatment exercise testing has been performed to serve as a baseline reference point.

Variability of response, whether due to random error or biological variation, is the only ill in this context for which there is a cure: increasing the sample size. As I will argue below, the possible alternative explanations are unrelated to sample size. Suffice it to say at this stage that at least one study of the effect of the ACE inhibitor enalapril on exercise capacity had a fairly large sample size (Swedberg 1994).

The second explanation, lack of sensitivity of exercise testing to changes in clinical condition, is difficult to assess on the basis of trial data alone. Ergometry is a well established diagnostic technique, and there is ample evidence that the degree of impairment of exercise capacity as assessed by this method is related to the severity of heart failure. In theory, it is possible that the relationship between impairment of exercise capacity and severity of heart failure is influenced by treatment: as pointed out elsewhere (Swedberg 1994), vasodilators have a greater beneficial effect on exercise capacity than ACE inhibitors. However, this seems unlikely. Besides, the possibility that exercise capacity might be determined by factors other than cardiac function per se is, in the present context, irrelevant. While additional factors (such as the effect of training on the efficiency of the leg musculature) may contribute to the individual variability in exercise capacity, these will ‘randomise out’ in a clinical trial. For this reason, I will assume that exercise capacity is a valid surrogate end-point in the sense that it reflects the severity of heart failure, and focus exclusively on the third and fourth possible explanations suggested above for the lack of drug effect on exercise capacity.

1. Sensitivity for Effects: An Additional Design Consideration in Clinical Trials

Clinical trials can be viewed as a means of determining the magnitude of a hitherto unknown therapeutic effect of a drug in a specific clinical context. Standard features of clinical trials are randomisation to achieve comparable patient groups and, whenever necessary, double-blinding to eliminate potential bias in data collection and decisions of the treating physicians on additional measures. These features ensure that, in case the drug tested has in fact no effect on the parameter in question, the trial will on average indicate ‘no difference’ between the treatment groups.

This property of randomised placebo-controlled trials is crucial, but it is not sufficient in itself. The trial must also be sensitive enough to detect effects of the drug if in fact they do exist. That trials need to be randomised and double-blind to ensure that a ‘no difference’ result is generated when the drug is ineffective is well known. What is perhaps less well known, and certainly less often explicitly discussed, is that sensitivity to drug effects is also an important consideration in clinical trial design.

2. The Standard Exercise Capacity Trial

In discussing sensitivity to drug effects in the context of exercise capacity trials in patients with stable congestive heart failure, I will first consider the standard parallel group trial design (crossover designs have special problems and are outside the scope of this discussion). This consists of the following features:
1. Initial selection, including a baseline exercise test to establish patient eligibility (based on clinical history and status, and exercise capacity within the range specified). This consists of the following features:
2. An interim clinical history and a second baseline exercise test, performed approximately 4 weeks later.
3. Final selection of those patients who remain clinically stable and who demonstrate an exercise capacity at the second test that is within a certain