made by the individual clinician will not be based only on life expectancy calculations and risk of cerebral bleeding. Other clinical information (such as previous history of angioplasty or bypass surgery, previous administration of streptokinase, recent surgery, anticipated problems with vascular access etc.), patient's preference and many practical considerations (such as, for example, ongoing procedures in the catheterization laboratory) will also influence the decision. As always, guidelines or practical recommendations have to be interpreted within the proper clinical context. If not, they may become dangerous, as put forward by Wood in 1950[4]. 'Yet there is plenty of evidence to show that we are in danger of losing our clinical heritage and pinning too much faith in figures thrown up by a machine. Medicine must suffer if this tendency is not checked.'

Thus, Boersma et al. are to be congratulated for synthesizing available survival data from clinical trials into useful tables. These tables will have to be updated when the results of newer and better reperfusion therapies become available, and their practical use for selecting the most appropriate therapy will always have to be adjusted to the local hospital situation and to the complex clinical presentation of the individual patient who needs reperfusion. They can then provide a sensible framework with which to make treatment decisions in a consistent manner.

F. VAN DE WERF
University of Leuven, Leuven, Belgium

References

European Heart Journal (1996) 17, 4-6

Using exercise endpoints in heart failure trials: design considerations

See page 120 for the article to which this Editorial refers

Over the last decade exercise testing has been used as a tool to evaluate the efficacy of various treatments for heart failure. This approach has been based upon the expectation that changes in exercise time are a more objective way of assessing functional status, compared to a subjective assessment of symptoms, and that there is a relationship between changes in exercise time and changes in morbidity and mortality. These general considerations, as well as specific aspects of study design of exercise trials which may yield a more reliable result, can be evaluated by reviewing trials of various interventions in heart failure.

The assumption that an exercise test is more ‘objective’ than reports from a patient of improved symptoms or increased activity has not been subjected to rigorous, blinded evaluation. Furthermore, improvements in symptoms themselves may have a mood elevating effect which could affect exercise tolerance independent of physical changes in cardio-pulmonary function. The analysis by Narang et al. based on the angiotensin converting enzyme (ACE) inhibitor trials indicates a high degree of concordance between exercise tolerance and improvements in New York Heart Association Functional Class (NYHA-FC), a global and simple measure based upon symptoms at various activity levels[11]. Similar results have been seen in trials of other interventions such as digoxin. It could also be reasonably argued that there is no need to perform exercise tests and that careful and unbiased documentation of symptomatic improvement may be all that is needed. It is, however, likely that both exercise tests and symptomatic functional status provide complementary information, but reflect different domains of patient status, since both appear to be independent predictors of morbidity and mortality[11]. Given the importance of this issue, a formal evaluation of this question in a careful prospective study is needed.
Having accepted that formal exercise testing is 'worth' using, what is the best way to design a trial? In the analysis by Narang et al. they assumed (quite reasonably) that in reality ACE inhibitors improve exercise capacity. Therefore, any differences in results observed were likely to be due to differences in study design. Study size, duration of follow-up, method of exercise testing and perhaps initial functional disability appeared to affect trial results. Another methodological issue, that was not considered, however, is whether patients are allowed to use handrails during treadmill testing. Handrail holding has been shown to change \( \Delta^2 \) uptake by up to 31%\(^3\). This information is not available from most trials. The observation that larger studies were more likely to demonstrate differences is hardly surprising because of greater statistical power, but larger studies may also have been designed better in other ways. Longer durations of treatments may reflect both the fact that increases in the dose of ACE inhibitor were gradual and that it usually takes several weeks for ACE inhibitors to manifest their full 'pharmacological' effects. It appeared that treadmill testing was more sensitive than bicycle testing, but this is based on an indirect comparison across different trials and there are few studies where several methods of exercise tolerance are performed simultaneously in the same patients. Therefore, the comparative value of these two methods of exercise testing is still unclear. In the Narang et al. review, duration of exercise on a treadmill was found to be the most useful; measures of physiological gas exchange did not appear to add more information. This would argue for simplification of exercise protocols.

In two recent trials of digoxin in congestive heart failure, the 6 min walk test was found to be useful and paralleled the results obtained with formal treadmill exercise tests and change in symptomatic status\(^4,5\). The value of the 6 min walk test is further illustrated by the study by Bittner et al.\(^2\) indicating a clear and independent inverse relationship with both mortality and morbidity. The analysis by Narang et al. suggests that trials which included a high proportion of patients in NYHA-FC III, demonstrated clearer results. This would be consistent with the expectation that patients who are most limited (i.e. are the sickest) may be more responsive to treatment. This is analogous to the use of high risk patients to increase the power of a mortality study. One factor not considered by Narang et al. is how the measurements of exercise duration could be best used to increase the sensitivity of trials. First, it might be useful to only include patients with limited exercise tolerance (e.g. <500 m on a 6 min walk test or <=6-7 METS on a Naughton/or modified Bruce protocol). Second, after identifying patients with limited exercise tolerance, performing duplicate tests at baseline and at the end of follow-up and averaging each of the two results at each time point provides more reliable and precise values. The preliminary results from work in our laboratory has suggested that the use of duplicate 6 min walk tests dramatically reduced the standard deviation of measurements from 71 to 39 m. We have used this information in the design of a new trial, the Randomized Evaluation of Strategies for Left Ventricular Dysfunction (RESOLVD) study. In this trial, the use of duplicate baseline and follow-up measures will allow detection of an absolute difference (change from baseline in treatment group minus change from baseline in control group) of 16 m with 90% power (standard deviation=39 m), compared to a 28 m detectable difference at 90% power using a single measurement (and standard deviation=71 m). We do not advocate ‘constraining’ (i.e. excluding further patients if the results are outside a narrow range, such as ±15%) the two baseline tests, as this could cause methodological artifacts due to regression to the mean, thus resulting in a 'biased' baseline value that could create an apparent placebo response. Utilizing the above approaches, a 30% to 40% gain in statistical power may result for the same number of patients.

Apart from the technical issues, there are several other inherent limitations of relying solely upon exercise testing as an endpoint. First, with some interventions, there may be a discrepancy between changes in exercise capacity and impact on mortality and morbidity, e.g. in the Second Veteran's Heart Failure Trial (V-HeFT II), ACE inhibitors were superior to hydralazine plus isosorbide dinitrate in reducing mortality, yet the latter regimen had a greater impact on exercise tolerance\(^6\). Second, there may be a discrepancy between short or medium term effects vs long-term effects. For example, in long-term trials the impact of ACE inhibitors on functional status (and presumably exercise tolerance) appears to decrease with time\(^7\). This may be due to a combination of 'differential' patient loss (patients who deteriorate may be withdrawn or die; this may differ substantially in the two groups being compared), or actual loss of efficacy of the drug. Third, although for specific interventions improved exercise capacity may be a marker of reduced morbidity and mortality from worsening congestive heart failure events (e.g. digoxin), the intervention may increase morbidity and mortality from other causes (e.g. arrhythmia or ischaemia). Therefore, the net clinical risk–benefit ratio cannot be derived solely from trials that are designed only to examine the effects of a treatment on exercise capacity.

In summary, exercise tolerance is a useful method to assess one facet of the effects of some

Eur Heart J, Vol. 17, January 1996
interventions in heart failure. Although logically a 'test' should be more 'objective' and 'sensitive' than a patient report of improved symptoms or increased functional capacity, this assumption has not been adequately tested. If exercise tolerance is to be measured to evaluate treatments in heart failure, we recommend the use of simple, standardized methods (e.g. 6 min walk or treadmill test) with multiple baseline and endpoint measures, in reasonably large numbers of patients (200 to 300 subjects) followed for at least 4 to 6 months. Such an approach can sometimes be an initial (but not sole) step in the evaluation of new agents. However, changes in exercise tolerance are not necessarily a robust surrogate for a similar effect on mortality and morbidity. Therefore, trials of exercise capacity are complementary, rather than a replacement for large trials evaluating the effects of an intervention on morbidity and mortality.

S. YUSUF
R. TSUYUKI
McMaster University and Hamilton Civic Hospitals Research Centre, Ontario, Canada

The authors thank Drs R. S. McKelvie and J. B. Young for their helpful comments in preparing this Editorial.

References

Echocardiographic tissue diagnosis

See page 135 for the article to which this Editorial refers

Echocardiography has become one of the principal diagnostic tools in the evaluation of patients with known or suspected heart disease. The versatility of this ultrasonic examination is one of its major strengths. With the ability to image tissue, as well as record blood flow, this diagnostic tool offers many opportunities for making both anatomical and functional diagnoses.

One goal of echocardiography which has been studied for decades but is yet to be established as a clinically useful technique is the ability to identify changes in myocardial tissue. Ultrasound has shown modest ability in its capacity to differentiate tissue types in organs other than the heart. The echocardiographer is able to identify structures that are very echo reflective, such as calcium or dense fibrosis.

There have been subjective and somewhat anecdotal observations in subgroups of patients, such as those with amyloidosis and hypertrophic cardiomyopathy.

Now many investigators are trying to provide a more objective method for evaluating myocardial tissue using transthoracic ultrasound. One technique which has been used by different laboratories is cyclical variation in myocardial integrated backscatter of the myocardium. This has been utilized to try to identify ischaemic myocardium, and a variety of studies have attempted to examine the grey level texture of the myocardium for cardiac diagnoses.

These studies include attempts to identify amyloid heart disease and hypertrophic cardiomyopathy\[1\]. A similar grey scale analysis has been used to try to identify scarred myocardium, as well as infiltrative cardiomyopathy\[2\]. Most tissue identification efforts in recent years have used integrated backscatter to try to