

Reproducibility and responsiveness of quality of life assessment and six minute walk test in elderly heart failure patients

S T O'Keeffe, M Lye, C Donnellan, D N Carmichael

Abstract

Objective—To examine the reproducibility and responsiveness to change of a six minute walk test and a quality of life measure in elderly patients with heart failure.

Design—Longitudinal within patient study.

Subjects—60 patients with heart failure (mean age 82 years) attending a geriatric outpatient clinic, 45 of whom underwent a repeat assessment three to eight weeks later.

Main outcome measures—Subjects underwent a standardised six minute walk test and completed the chronic heart failure questionnaire (CHQ), a heart failure specific quality of life questionnaire. Intra-class correlation coefficients (ICC) were calculated using a random effects one way analysis of variance as a measure of reproducibility. Guyatt's responsiveness coefficient and effect sizes were calculated as measures of responsiveness to change.

Results—24 patients reported no major change in cardiac status, while seven had deteriorated and 14 had improved between the two clinic visits. Reproducibility was satisfactory (ICC > 0.75) for the six minute walk test, for the total CHQ score, and for the dyspnoea, fatigue, and emotion domains of the CHQ. Effect sizes for all measures were large (> 0.8), and responsiveness coefficients were very satisfactory (> 0.7). Effect sizes for detecting deterioration were greater than those for detecting improvement.

Conclusions—Quality of life assessment and a six minute walk test are reproducible and responsive measures of cardiac status in frail, very elderly patients with heart failure.

(Heart 1998;80:377-382)

Keywords: six minute walk; elderly people; heart failure; quality of life

The main aim in treating congestive heart failure has been to prolong the life of patients. However, as in other chronic conditions, symptom control and effect on functional capacity and quality of life are also important goals of treatment. The development in recent years of standardised measures of exercise capacity and quality of life in heart failure reflects the growing perception of the importance of these outcomes in patients.

Such tests have been used as outcome measures in drug trials in relatively young heart failure patients.¹⁻⁴

Responsiveness (the sensitivity of a measure to a clinically relevant change in health) and reproducibility (the stability of a test when no important change in health has occurred) are essential properties of outcome measures for intervention studies.⁵ It is important that reproducibility and responsiveness should be demonstrated in all populations in which the measures will be used. Heart failure is particularly common in elderly people⁶; however, validation studies showing that quality of life questionnaires and exercise tests are reproducible and responsive have mainly been conducted in middle aged and young elderly patients (60 to 75 years).⁷⁻⁹ Coexistent pathology, particularly cognitive impairment and chronic physical disability, might be expected to reduce the value of these tests in very elderly patients. The aim of the present study was to examine the reproducibility and responsiveness to change of a six minute walk test and a heart failure specific quality of life measure in elderly patients with heart failure attending a geriatric outpatient clinic.

Methods

PATIENTS

Consecutive patients attending a geriatric outpatient clinic with a clinical diagnosis of chronic heart failure, as defined by the Framingham criteria,¹⁰ were eligible for the study. Patients were excluded if they declined to participate, if they were unable to provide informed consent owing to severe communication disorder or cognitive impairment, if they were unable to walk without physical assistance (not including mobility aids), or if they were considered unlikely to be able to complete a six minute walking test for any reason other than heart failure.

ASSESSMENT PROTOCOL

At the baseline visit, the examiner calculated a clinical heart failure score based on findings from the history, physical examination, and current chest x ray. The quality of life questionnaire was administered to the patient by the same clinician. Finally, the patient performed a six minute walk test. Other tests or changes in drug treatment were ordered as judged clinically necessary.

We followed the same procedure at a repeat clinic visit within the following three to eight weeks, except that the patient was asked at the outset to select a response from a five point

Department of Geriatric Medicine, University of Liverpool, Liverpool, UK

S T O'Keeffe
M Lye
C Donnellan
D N Carmichael

Correspondence to:
Dr S T O'Keeffe,
Department of Geriatric
Medicine, St Michael's
Hospital, Dun Laoghaire, Co
Dublin, Republic of Ireland.

Accepted for publication
16 February 1998

scale to the question: "Overall, how have you been from the point of view of your heart disease since I last saw you?" ("much better," "a bit better," "about the same," "a bit worse," "much worse"). Answers to this question were used as a global measure of change in cardiac status.

No specific intervention was tested in this study. Test-retest reliability of a measure can only be tested in patients who do not have a clinically significant change in status between two assessments. Conversely, data from patients who have experienced a significant change in status are needed to assess responsiveness of a measure. We presumed that most of our study population of unselected heart failure patients would have a stable cardiac status between two clinic visits three to eight weeks apart. However, some patients would experience a decline in cardiac status—for example, because of the natural history of chronic heart failure or development of intercurrent illnesses. Similarly, other patients would probably have an improvement in cardiac status at the second visit—for example, because of changes in drug treatment at the first visit.

Six minute walk test

We conducted the six minute walk tests using a standardised approach.^{9–11} A 25 metre course was marked in a level enclosed corridor and a chair was placed at each end. Patients were transported to the start of the course by wheelchair. Patients were instructed to walk from end to end at their own pace while attempting to cover as much ground as possible in six minutes. They were allowed to use their usual mobility aids. A study doctor timed the walk test, calling out the time every two minutes. This doctor encouraged patients every 30 seconds in a standardised manner, facing the patient and using one of two phrases: "You're doing well" or "Keep up the good work." Patients were allowed to slow or stop and rest during the walk, but were asked to resume walking as soon as they felt they are able to. After six minutes, the distance walked was measured to the nearest metre.

Quality of life questionnaire

A heart failure specific quality of life questionnaire, the chronic heart failure questionnaire (CHQ), developed by workers in McMaster University, was used in this study.⁸ This questionnaire examines three major areas of impairment caused by heart failure: dyspnoea (five items), fatigue (four items), and emotional function (seven items). Items are measured on a seven point Likert scale, and the scores in each dimension are added together. Thus the minimum (worst function)/maximum (best function) scores in the three domains are: dyspnoea 5/35; fatigue 4/28; and emotional function 7/49. In this study, subjects were asked to consider the last four weeks when completing the CHQ at the baseline visit. At the follow up visit, subjects were asked to consider the period between the two clinic visits.

Table 1 Baseline characteristics of study population

Number	60
Female/male	38/22
Median age (years) (range)	81 (74 to 92)
NYHA I	14%
NYHA II	33%
NYHA III	43%
NYHA IV	10%
Ischaemic heart disease	82%
Hypertension	43%
Diabetes mellitus	30%
ACE inhibitors	65%
Mean (SD) daily frusemide equivalent (mg)	68 (12)
Mean (SD) CHQ score	
Dyspnoea	21.1 (6.3)
Fatigue	18.5 (4.8)
Emotion	34.1 (8.3)
Mean (SD) six minute walk distance (m)	238.8 (51.5)
Mean (SD) heart failure score	5.4 (2.0)

ACE, angiotensin converting enzyme; CHQ, chronic heart failure questionnaire; NYHA, New York Heart Association functional class.

Clinical heart failure score

We used the clinical heart failure score developed and validated by Lee *et al*, which simulates the clinical judgment of the severity of heart failure.¹ The score combines findings from the history and physical examination with the chest *x* ray appearance. The maximum (worst) score is 13.

STATISTICS

Validity

We used Pearson's correlation coefficient to examine the interrelations between different test scores. We predicted that, as evidence of validity, there would be reasonably close relations ($r \geq 0.5$) at the baseline assessment between the six minute walk distance and both the total quality of life score and the dyspnoea quality of life domain. We also predicted that the global change rating would be closely related ($r \geq 0.5$) to changes in the three quality of life domains, the total quality of life score, and the six minute walk distance.

Reproducibility

Data from patients who reported no major overall change in cardiac status between the first and second visits were used to assess the reproducibility of the health measures. Intra-class correlation coefficients (ICC) were calculated using a random effects one way analysis of variance.^{12–14} We specified before the study that ICC values of 0.75 or more would represent satisfactory reproducibility.

Responsiveness

There is no consensus regarding how best to assess the responsiveness to change of measures; hence various different approaches are reported in this study:

- (1) Observed change = mean ($test_1 - test_2$). Responsiveness was assessed by examining whether the mean change scores followed the expected pattern in patients with global ratings of change in cardiac status from "much worse" to "much better."
- (2) Effect size (ES) = observed change/standard deviation of test. Effect size is calculated by dividing the mean change scores by the standard deviation of the baseline score in the same subjects. Sepa-

Table 2 Baseline and change in scores ($time_2 - time_1$) with different global ratings of change

	Much worse (n = 2)	A bit worse (n = 5)	About the same (n = 24)	A bit better (n = 10)	Much better (n = 4)
Dyspnoea QOL (max 35)					
Baseline	22.5 (2.5)	19.4 (1.1)	21.5 (1.0)	21.4 (1.5)	21.0 (1.6)
Change	-5.0 (1.0)	-3.2 (0.7)	0.9 (0.7)	3.0 (0.2)	5.5 (1.1)
Fatigue QOL (max 28)					
Baseline	19.0 (1.0)	17.6 (1.5)	18.3 (0.9)	18.3 (1.2)	19.0 (1.8)
Change	-3.0 (1.0)	-3.2 (0.4)	0.3 (0.5)	3.3 (0.7)	4.5 (0.9)
Emotion QOL (max 49)					
Baseline	36.0 (1.0)	34.4 (1.2)	34.3 (1.6)	31.8 (1.4)	29.3 (1.5)
Change	-4.5 (0.5)	-3.0 (2.2)	1.25 (1.0)	2.9 (0.6)	4.0 (0.7)
Total QOL (max 92)					
Baseline	77.5 (4.5)	71.4 (2.6)	74.0 (2.6)	71.5 (2.5)	69.3 (3.6)
Change	-12.5 (0.5)	-9.4 (2.3)	2.4 (1.5)	9.2 (0.8)	14.0 (1.4)
Walk distance					
Baseline	246.0 (18.0)	251.6 (13.6)	236.5 (10.0)	239.7 (13.4)	227.8 (10.3)
Change	-48.6 (9.1)	-43.0 (8.7)	12.2 (3.5)	24.0 (4.0)	47.0 (6.1)
CHF score (max 13)					
Baseline	5.5 (0.5)	5.8 (0.9)	5.4 (0.3)	5.5 (0.4)	6.0 (0.4)
Change	3.0 (0)	1.6 (1.5)	-0.3 (0.2)	-1.4 (0.6)	-1.8 (0.7)

Data are mean (SEM). Higher chronic heart failure (CHF) scores represents more severe impairment. Higher quality of life (QOL) scores denote better function.

rate effect sizes for each measure were calculated in patients who deteriorated and in patients who improved, as scores in the two directions are not necessarily the same.¹⁴ Cohen suggests that an effect size > 0.8 is large, 0.5 to 0.8 is moderate, and 0.2 to 0.5 is small.¹⁵

(3) Responsiveness coefficient (RC) = minimum important difference/standard deviation of ($test_1 - test_2$). The responsiveness coefficient, developed by Guyatt and colleagues, relates the minimally important difference on a measure to the within subject variability in score in stable subjects (that is, those patients who report “about the same” in the global rating).¹⁶ The minimum important difference is the smallest difference in a measure that signifies a clinically significant change rather than a trivial change in patient symptoms. Previous reports suggest that minimum important difference is 30 metres for the walk test² and 0.5 for individual items of the CHQ (measured on seven point Likert scales).^{17, 18} Within subject variability is represented by the standard deviation of the change scores in stable subjects.

The higher the responsiveness coefficient, the smaller the sample size required in clinical trials to detect a minimum clinically significant change in test score with an intervention.¹⁶ A responsiveness coefficient of 0.6 or more for a test suggests that a parallel group study would require about 50 patients in each group to show a minimum important difference in test scores following an intervention.

Results

Sixty of 68 consecutive heart failure patients were able to complete the baseline assessment. Eight patients were excluded because they had one or more of the following problems: severe confusion or communication disorder (n = 6), hemiparesis (n = 2), and severe arthritis (n = 2). Eighteen (30%) of the 60 patients included in the study had a history of cerebrovascular disease; seven (12%) were

receiving regular analgesia for arthritis; and five (8%) used a walking stick or frame in performing daily activities. Nineteen of the 56 patients tested (32%) had mild to moderate cognitive impairment as defined by an abbreviated mental test score > 3/10 and < 8/10.¹⁹ Other characteristics of these patients are shown in table 1.

Forty five patients underwent repeat assessment a median of four (range three to eight) weeks later. At the follow up visit, two patients felt much worse, five a bit worse, 24 about the same, 10 a bit better, and four much better. Of the 15 patients who did not have a repeat clinic assessment, there were logistical difficulties in arranging follow up within the specified period for eight patients (in three cases owing to unavailability of a study doctor), three patients had been admitted to hospital, one had died, and three refused. There were no significant differences between the baseline assessments of the 45 patients who did and the 15 patients who did not have a repeat assessment.

Reproducibility, assessed by calculating an intraclass correlation coefficient (R) from the results of the 24 patients who reported no major overall change in cardiac status, was satisfactory for the six minute walk distance (R = 0.91), for total CHQ (R = 0.83), and for the dyspnoea (R = 0.83), fatigue (R = 0.79), and emotion (R = 0.78) domains of the CHQ. Reproducibility was mediocre for the clinical heart failure score (R = 0.55). Baseline and change data corresponding to different global ratings of change are shown in table 2. Changes in the CHQ domain scores and the walk

Table 3 Effect sizes and responsiveness coefficients of health measures

	Responsiveness coefficients	Effect sizes	
		Improvement	Deterioration
Dyspnoea QOL	0.76	0.87	1.37
Fatigue QOL	0.77	0.99	1.14
Emotion QOL	0.71	0.81	1.59
Total QOL	1.13	1.38	1.70
Walk distance	1.73	0.85	2.13
Heart failure score	N/A	1.49	1.29

N/A, not applicable; QOL, quality of life.

Table 4 Correlations between change in quality of life (QOL), walk distance, and heart failure scores

	Global rating of change	Dyspnoea QOL	Fatigue QOL	Emotion QOL	Total QOL	Walk distance
Dyspnoea QOL	0.70					
Fatigue QOL	0.69	0.41				
Emotion QOL	0.46	0.38	0.42			
Total QOL	0.77	0.75	0.74	0.83		
Walk distance	0.78	0.60	0.58	0.47	0.70	
Heart failure score	-0.56	-0.33	-0.46	-0.20	-0.40	-0.42

distance at the follow up visit were in the directions and of the magnitude expected according to the global rating of change ($p < 0.01$ on a test for linear trend).

Effect sizes and responsiveness coefficients are shown in table 3. Effect sizes for all measures were large, and responsiveness coefficients were also very satisfactory. Effect sizes for detecting deterioration were greater than those for detecting improvement.

As predicted, there was a good correlation at the baseline assessment between walk distance and the total CHQ score ($r = -0.79$) and the dyspnoea CHQ dimension ($r = -0.58$). The correlations between changes in scores of different variables are shown in table 4. There were strong correlations between the global rating of change in cardiac status and changes in the dyspnoea and fatigue domains of CHQ, the total CHQ score and the walk distance. Change in the clinical heart failure score and in the emotion CHQ domain were more poorly related to the global change rating.

Discussion

The incidence and prevalence of heart failure increase dramatically with increasing age. For example, the Framingham study reported that the prevalence of heart failure was over 9% in subjects over 75 years of age.⁶ Heart failure is associated with impaired exercise tolerance and reduced quality of life and has a poor prognosis regardless of age.²⁰ The therapeutic objectives in treating heart failure in elderly people depend on the individual patient. Although some of the large trials excluded older patients, there is now good evidence that treatment with angiotensin converting enzyme (ACE) inhibitors reduces mortality in elderly heart failure patients.²¹ Nevertheless, because elderly people often have other conditions that increase the risk of death, the absolute gain in survival even with effective treatment is often small.²² In patients with major comorbid conditions or with cognitive impairment, symptom control and improvement in quality of life may be more important than prolonging survival. Also, elderly people are more prone to develop side effects with cardiovascular and other drugs. For example, reduction in dyspnoea with diuretic treatment must be balanced against the propensity of these agents to cause incontinence and postural dizziness. Thus the effects of interventions on exercise tolerance and quality of life are particularly important considerations in the treatment of elderly heart failure patients and should be considered as primary end points in clinical trials in this population.

The measures evaluated in this study are well established in the study of heart failure patients. The six minute walking test has been found to be a valid and reproducible measure of functional exercise capacity in chronic heart failure.^{4,7,23} Furthermore, walking distance predicts long term morbidity and mortality.²⁴ This test may be particularly suitable for elderly patients who often find it difficult to perform adequate treadmill tests, and improvement in this type of test in response to treatment may be more important and more relevant to activities of daily living than to changes in treadmill exercise capacity. The reproducibility, validity, and responsiveness to clinically significant change in the CHQ were shown in a trial of digoxin in heart failure patients in sinus rhythm.^{2,8} The heart failure score, developed by Lee *et al*, has been used in two major trials,^{1,2} and a strong correlation ($r = 0.85$) between this score and resting pulmonary capillary wedge pressure has been reported.¹

The high prevalence of comorbid conditions in elderly people might be expected to limit the value of walk tests and quality of life assessment instruments in this population. For example, the presence of neurological and musculoskeletal problems and general deconditioning might reduce the proportion of patients able to complete a six minute walk test, as well as increasing the variability of successive walk tests. Similarly, even mild cognitive impairment might impair the reproducibility and responsiveness of a quality of life questionnaire. However, our results suggest that a six minute walk test and a disease specific quality of life instrument continue to be useful in very elderly patients. Eighty eight per cent of 68 consecutive heart failure patients attending a geriatric clinic were able to perform all tests. Reproducibility, validity, and responsiveness of these tests were satisfactory according to standards defined before the study. In contrast, the reproducibility of a clinical heart failure score was poor and change in score of this test correlated poorly with the global rating of change.

Several investigators have suggested that an ICC of more than 0.75 is acceptable when studying groups of patients,^{14,25} and this was the standard adopted in this study, although—as Streiner and Norman have pointed out—there is no sound basis for making such a recommendation.⁵ McHorney and Tarlov have argued that ICC values of more than 0.90 are required if measures are to be used to assess individual, as opposed to group, data.²⁶ Only the walk test attained this level of reproducibility in our study. However, even for the walk test the span of the limits of agreement between successive tests in stable patients, calculated using the method of Bland and Altman,²⁷ is substantial at 60 metres, suggesting that this test would be of little value for assessing change in individual patients. The one way analysis of variance used to calculate ICC values in our study provides more conservative estimates than the use of two way analysis of variance, as advocated by Deyo.²⁸

With one way analysis of variance, a systematic shift between testing times, which may result from a learning effect, will result in lower values for the ICC; two way analysis of variance would eliminate the error caused by such a systematic shift.¹⁴

Responsiveness to clinically significant change is an important and often neglected aspect of measures used in clinical trials. However, it is not yet clear how best to assess responsiveness, and, like other researchers, we examined various different statistics.^{14–29} The different approaches yielded broadly similar and satisfactory results. Detecting improvement is the usual goal of intervention studies. For all measures in this study, effect sizes for detecting deterioration were greater than effect sizes for detecting improvement, although the latter were still within an acceptable range. Although the number of patients experiencing deterioration in our study was small, this finding has also been reported by investigators using other health measures.³⁰ This probably reflects the fact that most of our patients were already on optimal treatment for heart failure at the onset of the study, although a “ceiling effect” may have occurred in some patients with relatively mild heart failure.

Jenkinson *et al* noted that the SF-36 and Dartmouth COOP, two generic measures of health related quality of life, were not responsive to self reported improvements in global health in a study of elderly heart failure patients starting treatment with an ACE inhibitor.³¹ Our results in a rather similar population using the CHQ, a heart failure specific quality of life measure, are very different. It is possible that ACE inhibitors, despite their beneficial effects on mortality, do not lead to major improvement in quality of life. However, the better responsiveness of the CHQ may simply reflect the fact that it is designed specifically for use in heart failure, with several items chosen by the patients themselves.

This study has several limitations. The numbers studied were small, particularly when examining the responsiveness to change in different directions. A quarter of the patients who had a baseline assessment did not have a repeat assessment within a relatively generous follow up period. This mainly reflects the frailty of our study group, and is an indicator of the problems to be expected when conducting research in such a population. Although our results suggest that the health measures assessed should be useful in conducting clinical trials of interventions in elderly people with heart failure, we did not examine the effects of any specific intervention in this study. Our results for the psychometric properties of the CHQ are very close to those reported by Guyatt *et al* using the same instrument in an intervention trial.⁵ Nevertheless, the discrepancy between our results and those of Jenkinson and colleagues³¹ suggests that the responsiveness of the CHQ in very elderly patients with heart failure should be confirmed in an intervention study. We used a transitional scale from “much worse” to “much better” to judge change in cardiac status between the two testing sessions,

and this scale was used as the external criterion of change for assessing responsiveness. Although this approach has been used in many similar studies,^{16–30–32} it should be noted that transitional scales are subject to bias resulting from the patient’s expectation of change—for example, after intensification of treatment at the initial visit.

In conclusion, our study suggests that the reproducibility and responsiveness of a walking test and a quality of life assessment instrument are satisfactory even in very elderly and frail patients with heart failure. Thus these measures may be useful as primary end points when conducting clinical trials in such populations.

- 1 Lee DC-S, Johnson RA, Bingham JB, *et al*. Heart failure in outpatients. A randomized trial of digoxin versus placebo. *N Engl J Med* 1982;306:699–705.
- 2 Guyatt GH, Sullivan MJ, Fallen EL, *et al*. A controlled trial of digoxin in congestive heart failure. *Am J Cardiol* 1988;61:371–5.
- 3 Rector TS, Johnson G, Dunkman WB, *et al*. Evaluation by patients with heart failure of the effects of enalapril compared with hydralazine plus isosorbide dinitrate on quality of life: V-HeFT II. *Circulation* 1993;87:VI-71–7.
- 4 Packer M, Gheorghade M, Young JB, *et al*. Withdrawal of digoxin from patients with chronic heart failure treated with angiotensin converting enzyme inhibitors. *N Engl J Med* 1993;329:1–7.
- 5 Streiner DL, Norman GR. *Health measurement scales*, 2nd ed. Oxford: Oxford Medical Publications, 1995.
- 6 Kannel WB, Belanger AJ. Epidemiology of heart failure. *Am Heart J* 1991;121:951–7.
- 7 Lipkin D, Scriven AJ, Crake T, Poole-Wilson PA. Six minute walking test for assessing exercise capacity in chronic heart failure. *BMJ* 1986;292:653–5.
- 8 Guyatt GH, Nogradi S, Halcrow S, *et al*. Development and testing of a new measure of health status for clinical trials in heart failure. *J Gen Intern Med* 1989;4:101–7.
- 9 Guyatt GH, Pugsley SO, Sullivan MJ, *et al*. Effect of encouragement on walking test performance. *Thorax* 1984;39:818–22.
- 10 McKee PA, Castelli WP, McNamara M, *et al*. The natural history of congestive heart failure: the Framingham study. *N Engl J Med* 1971;285:1441–6.
- 11 Guyatt GH, Sullivan MJ, Thompson PJ, *et al*. The 6 minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J* 1985;132:919–23.
- 12 Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966;19:3–11.
- 13 Shrout PE, Fleiss JL. Intraclass correlations: uses in inter-rater reliability. *Psychol Bull* 1979;86:420–8.
- 14 Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997;50:79–93.
- 15 Cohen J. *Statistical power analysis for the behavioural sciences*, 2nd ed. New Jersey: Lawrence Erlbaum Associates, 1988.
- 16 Guyatt G, Walter S, Norman G. Measuring change over time—assessing the usefulness of evaluative instruments. *J Chron Dis* 1987;40:171–8.
- 17 Jaeschke R, Singer J, Guyatt GH. Measurement of health-status—ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
- 18 Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms—a comparison of 2 techniques. *J Clin Epidemiol* 1996;49:1215–19.
- 19 Jitapunkul S, Pillay I, Ebrahim S. The abbreviated mental test—its use and validity. *Age Ageing* 1991;20:332–6.
- 20 O’Keeffe ST, Lye M. Heart failure in the elderly: the same syndrome as the clinical trials? In: McMurray JVV, Cleland JGF, eds. *Heart failure in clinical practice*. London: Martin Dunitz, 1996:47–71.
- 21 Garg R, Yusuf S, for the Collaborative Group on ACE Inhibitor Trials. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. *JAMA* 1995;273:1450–6.
- 22 Welch HG, Albertsen PC, Nease RF, *et al*. Estimating treatment benefits for the elderly—the effect of competing risks. *Ann Intern Med* 1996;124:577–84.
- 23 Yusuf S, Tsuyuki R. Using exercise end points in heart failure trials: design considerations. *Eur Heart J* 1996;17:4–6.
- 24 Bittner V, Weiner DH, Yusuf S, *et al*. Prediction of mortality and morbidity with a 6-minute walk test in patients with left-ventricular dysfunction. *JAMA* 1993;270:1702–7.
- 25 Fleiss JL. *The design and analysis of clinical experiments*. New York: John Wiley and Sons, 1986.
- 26 McHorney CA, Tarlov AR. Individual patient monitoring in clinical practice: are available health status surveys enough? *Qual Life Res* 1995;4:293–307.

- 27 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical assessment. *Lancet* 1986;i:307-10.
- 28 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health-status measures—statistics and strategies for evaluation. *Control Clin Trials* 1991;12:S142-58.
- 29 Bombardier C, Raboud J. A comparison of health-related quality-of-life measures for rheumatoid-arthritis research. *Control Clin Trials* 1991;12:S243-56.
- 30 Siu AL, Ouslander JG, Osterweil D, et al. Change in self-reported functioning in older persons entering a residential care facility. *J Clin Epidemiol* 1993;46:1093-101.
- 31 Jenkinson C, Jenkinson D, Shepperd S, et al. Evaluation of treatment for congestive heart failure in patients aged 60 years and older using generic measures of health status (SF-36 and COOP charts). *Age Ageing* 1997;26:7-13.
- 32 Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical-change—an analogy to diagnostic-test performance. *J Chron Dis* 1986;39:897-906.

Heart—[http:// www.heartjnl.com](http://www.heartjnl.com)

Visitors to the world wide web can now access *Heart* either through the BMJ Publishing Group's home page (<http://www.bmjpg.com>) or directly by using its individual URL (<http://www.heartjnl.com>). There they will find the following:

- Current contents list for the journal
- Contents lists of previous issues
- Members of the editorial board
- Subscribers' information
- Instructions for authors
- Details of reprint services

A hotlink gives access to:

- BMJ Publishing Group home page
- British Medical Association web site
- Online books catalogue
- BMJ Publishing Group books

The web site is at a preliminary stage and there are plans to develop it into a more sophisticated site. Suggestions from visitors about features they would like to see are welcomed. They can be left via the opening page of the BMJ Publishing Group site or, alternatively, via the journal page, through "about this site".