

HFSA Working Group

Clinical and Analytical Considerations in the Study of Health Status in Device Trials for Heart Failure

SHARON-LISE T. NORMAND, PhD,¹ THOMAS S. RECTOR, PhD,² JAMES D. NEATON, PhD,³ ILEANA L. PIÑA, MD,⁴ RONALD M. LAZAR, PhD,⁵ SCOTT E. PROESTEL, MD,⁶ DINA J. FLEISCHER, MBA,⁷ JAY N. COHN, MD,⁷ AND JOHN A. SPERTUS, MD MPH⁸

Boston, Massachusetts; Minneapolis, Minnesota; Cleveland, Ohio; New York, New York; Rockville, Maryland; Kansas City, Missouri

ABSTRACT

Background: Measures of health status (including symptoms, functional status, or quality of life) assess patients' experiences of their disease, and may therefore be used to quantify the benefits and risks of treatment. The aim of this article is to provide recommendations to regulatory agencies and research sponsors regarding the use of health status measures in medical device trials.

Methods and Results: A workshop jointly planned by the Heart Failure Society of America and the US Food and Drug Administration was convened in October 2003 in Washington, DC. A Working Group to address health status measures initiated its collaboration at the workshop and continued its efforts throughout the next year. The Working Group recommended assessment of health status in all studies of heart failure therapy. Standardized instruments known to be valid, reliable, responsive to changes, and available in the languages of target populations should be used. Minimizing bias may be accomplished by using blinded, independent evaluators; collecting multiple health status measures; using valid statistical methods; and creating a health status resource bank.

Conclusion: Assessment of health status should be part of any device trial and should occur regardless of whether the device is intended as destination or bridging therapy. Health status endpoints should be chosen, collected, and analyzed with the same level of scientific rigor as traditional clinical endpoints. Regulatory agencies should require use of analytic methods that handle the complexity of health status data in addition to usual protocol protections.

Key Words: Design, endpoints, quality of life, regulatory.

The objective of this article is to make several recommendations to the US Food and Drug Administration and to research sponsors regarding the use of health status measures in medical device trials. These recommendations arose from

discussions at a Heart Failure Society of America meeting during October 2003 in Washington, DC. A Health Status Working Group was formed and focused on 5 specific regulatory issues: justification for measuring health status, identification of effective measures of health status, determination of the number of health status measures, design considerations in health status device trials, and analytic considerations in using health status information. Although the information presented in this article is not meant to be comprehensive, it provides the fundamentals to support the recommendations for inclusion of health status in device trials.

What Is Health Status?

Heart failure is a disorder that includes left or right ventricular dysfunction and neurohormonal imbalances that produce physical symptoms of fatigue, edema, and dyspnea. Psychologic symptoms, such as depression and anxiety, occur frequently as well. These symptoms and other aspects

The opinions expressed in this paper are those of the authors and do not necessarily represent those of the editor or the Heart Failure Society of America (HFSA). All writing group members were required to complete and submit a Faculty Disclosure Questionnaire before the workshop was held, October 8–9, 2003. The workshop was sponsored by the HFSA.

From the ¹Harvard Medical School and School of Public Health, Boston, Massachusetts; ²Minneapolis VA Medical Center, Minneapolis, Minnesota; ³University of Minnesota School of Public Health, Minneapolis, Minnesota; ⁴Case Western Reserve University, Cleveland, Ohio; ⁵Columbia University College of Physicians and Surgeons, New York, New York; ⁶US Food and Drug Administration, Rockville, Maryland; ⁷University of Minnesota Medical School, Minneapolis, Minnesota; ⁸Mid America Heart Institute and University of Missouri Kansas City, Kansas City, Missouri.

Reprint requests: Sharon-Lise T. Normand, 180 Longwood Avenue, Boston, MA 02115.

1071-9164/\$ - see front matter

© 2005 Elsevier Inc. All rights reserved.

doi:10.1016/j.cardfail.2005.04.002

of the syndrome of heart failure can affect an individual’s physical, mental, and social function and affect the ability to perform defined tasks. Functional limitations and symptoms, in turn, can affect an individual’s quality of life or physical, psychological, and social well-being. Because the primary goal of health care is to improve and to maintain health, health status is a key outcome for the evaluation of all medical therapies, including devices. Assessments of a patient’s health status thus focus on 1 or more of these components—symptoms, functional status, or quality of life. Measures of health status quantify patients’ experiences of their disorder and the potential benefits and risks of treatment. We note that, although health status and quality of life have been used interchangeably, they can refer to different concepts.^{1–3} In our conceptualization, health status includes the range of ways in which a disease affects a patient’s life, including symptoms, function, and overall quality of life.

Many techniques and measures are used to quantify patients’ health status.^{4,5} Potential metrics include generic health status measures that address patients’ overall health and disease-specific health status measures that quantify the specific manifestations of a given disease. The selection of the appropriate measures requires an understanding of the proposed benefits of the device, an understanding of the study population, consideration of trial design, and prespecification of the data to be collected, the potential analyses, and the desired trial interpretations. **Table 1** briefly describes 2 instruments developed to specifically quantify the health status of patients with heart failure.

Why Measure Health Status in a Device Trial?

Ultimately, the primary goals of medical care are to make patients live longer and to optimize their health status. As medical options continue to expand, and now include resource-intensive device therapies, patient preference may be driven by the impact of any 1 therapy on patient’s health status in all its dimensions. The first step in determining how best to quantify device effectiveness involves considering how the device might alter the health status of heart failure

patients. Hence a clear conceptualization of why a particular health status instrument was chosen and the outcome it is designed to measure should be explicitly presented in the study protocol.⁶ If regulatory approval is sought because the device can improve patients’ quality of life, then a valid measure of quality of life should be a primary outcome measure in the trial. Likewise, sponsors might seek an indication for the device that reduces the symptoms that are important to patients with heart failure, in which case a valid measure of such symptoms should be used.

Although trials of biventricular pacing devices that are intended to improve the health and function of heart failure patients need to explicitly evaluate the health status of study patients, other devices, such as left ventricular assist devices (LVAD), may not need to include health status assessments. For example, in some trials, the time of device treatment may be short, such as use of an LVAD implanted as a bridge to transplantation. In this situation, quality of life may be a less relevant outcome to measure as compared with “survival to transplantation.” On the other hand, for a device inserted as a destination therapy, quality of life may be the most important clinical outcome. For example, a patient with advanced heart failure who is not a candidate for transplantation may select an LVAD for the health status benefits it confers with respect to enhancing quality of life and reducing symptoms. It is important to note, however, that as waiting times to transplant continue to increase, the lines that delineate the intent of an LVAD insertion will “blur” and the health status effects of device therapy will become increasingly important in bridging therapy trials as well. Without health status assessments, there are no opportunities for clinicians to understand, from patients’ perspectives, the impact of treatment or to apply the results of trials to better match therapies to the needs and values of individual patients. For these reasons, the Working Group recommended that measurement of health status occur in all trials, regardless of the intent to insert the device as a bridging or destination therapy.

What Is an Effective Measure of Health Status?

Five key attributes of health status measures need to be present to enhance confidence in trial results.⁷ Relevant psychometric properties of an instrument (**Table 2**) include its validity, reliability, responsiveness to change, interpretability, and availability of translations in other languages. The Working Group recommended that, in addition to a clear explanation of the outcome being measured, the study protocol explicitly address these attributes for each measure of health status for the population under study. To help establish validity, the protocol should (1) characterize relevant manifestations of the device on heart failure and (2) demonstrate that scores on the chosen measure(s) of health status discriminate patients with varying manifestations of the disorder affected by the device.^{8–11} The content validity of the instrument should capture the important ways a device could affect

Table 1. Examples of 2 Health Status Instruments for Heart Failure Patients

Instrument	Self-Administered	No. of Items	Domains
Kansas City Cardiomyopathy Questionnaire ⁹	Yes	23	1. Physical limitation 2. Symptoms 3. Quality of life 4. Social interference 5. Self-efficacy 6. Overall summary
Minnesota Living with Heart Failure Questionnaire ²¹	Yes	21	1. Quality of life

Table 2. Key Psychometric Properties of a Health Status Instrument

Measurement Property	Description
Validity	The measure quantifies what it is intended to ([face validity], represents all important content of the construct [content validity], and empirically predicts the construct [predictive or criterion validity])
Reliability	Reproducible results are obtained when the measure is repeatedly given to stable patients
Responsiveness	The measure is sensitive to clinical change
Interpretable	A clinical framework is available to interpret cross-sectional and change scores
Translations exist	Linguistically and culturally appropriate translations are available

a patient's health or quality of life, and thus may include measures of device safety. Because heart failure treatment is often directed at improving the symptoms, function and quality-of-life limitations imposed by heart failure, disease-specific measures are often preferred over general health status measures for clinical trials of heart failure therapy.

The instrument should also provide reproducible scores when stable patients are measured repeatedly over time and should detect or respond to changes as they occur.¹²⁻¹⁴ A quantitative summary of test-retest reliability, such as the intraclass correlation coefficient, should be reported, as should the internal reliability of the instrument,¹⁵ using Cronbach's alpha or suitable statistics. Because both reliability and validity can vary between study populations, the Working Group noted that examination of these psychometric properties in the pivotal study population is warranted. More important, the Working Group felt that a common resource bank of the performance characteristics of commonly used instruments in different patient populations, such as New York Heart Association class I/II compared with class III/IV for advanced heart failure patients and for post-myocardial infarction heart failure patients, would serve an important role in advancing the field and assisting in the planning of future studies. Furthermore, whenever possible, the reliability of the health status estimates should be reported separately for each treatment arm because such estimates are specific to each population and to the method of measurement.¹⁶

Given the increasing numbers of international trials and the diversity of such populations, health status measures should be translated using appropriate methods to ensure proper linguistic and cultural versions are used in the multiple countries.¹⁷ Last, although there is some controversy about how to define "clinical" significance, the differences in health status scores should be meaningful and interpretable. It is important to have references with which to interpret, both cross-sectional and longitudinal changes in scores over time, so that the magnitude of benefits observed in clinical trials may be more readily appreciated. Efforts to more directly correlate the magnitude of changes in scores

with clinically significant changes have been undertaken.^{18,19} The Working Group recommended the establishment of a health status resource bank that stores information regarding the attributes of health status measures. The resource bank should include the psychometric characteristics of the instrument within clinically relevant disease states, such as advanced heart failure, inpatients with decompensated heart failure, or outpatients with heart failure. The **Appendix** provides examples of the type of information for such a resource bank for heart failure patients.

How Many Measures of Health Status?

Although a complete discussion of what has been learned about measuring health status in clinical trials is beyond the scope of this article, it is clear that a single measure of health status should not serve as the sole component of total efficacy assessment in a device trial. Much has already been learned from pharmacologic clinical trials in heart failure.²⁰ In several studies, health status, with its domains of physical function, social interaction and emotional well-being, has not consistently correlated with other measures of health status, such as exercise capacity.²¹⁻²⁶ Such correlations, however, are not necessarily expected depending on the domain of health status assessed and the anatomic, physiologic, or hemodynamic test to which it is being compared.

Some trials have shown improvements in health status and exercise capacity with various drugs, and yet an increase in mortality. Others have shown marked improvements in mortality and no changes in health status.²⁷⁻³⁰ These findings highlight the importance of directly measuring health status as well as outcomes, such as mortality, so that insights into the tradeoffs between quantity versus quality of life may be assessed and described.

The Working Group recommended that a comprehensive assessment of health status be part of a device trial and the assessment quantify the full spectrum of domains that might be affected by the device, such as symptoms, functional assessment, and quality of life. One component of health status could serve as the primary health status endpoint or the sponsor may choose to use a combination of components as the primary endpoint. Regardless, the sponsor should prespecify the primary health status outcomes to be used and the study should be powered to detect meaningful differences in that endpoint. If a single component serves as the primary health status endpoint, the other measures of health status should be reported and be supportive, and product labeling should reflect precisely these measurements. In addition to assessing different domains of health status, in trials in which blinding is not possible, it may be important to assess a specific domain by both patient self-report and by a clinician assessment. The Working Group recognized that "questionnaire fatigue" is an important consideration, especially in an elderly heart failure population. Thus the measures chosen should be selected with sensitivity to the response burden on patients.

Of note, simple assessments of global change may not adequately capture the range of health status with sufficient detail to be useful as endpoints in clinical trials. Such measures generally ask patients to compare how they are doing now with how they were doing at the start of the study. Problems with these assessments include the difficulty in patients remembering how they were in the past (recall bias), lack of clarity about which symptoms or disease manifestations are accounting for the observed changes, and lack of correlation between patient and clinician assessments of recent change.³¹ Serial “cross-sectional” assessments with a focus on all relevant domains are a better strategy for quantifying the benefits of therapy.

What Are the Design Issues in Health Status Device Trials?

The ability of measures of health status to detect effects of devices in trials depends on the reliability and completeness of the measurements made by investigators, the number of patients studied, the participants’ baseline health status, the study duration, the effects of deaths and withdrawals from the study, the effects observed in the control group, and the consistency and magnitude of the device’s true effect on patients’ health, including any adverse effects. There are several noteworthy design considerations in device trials that use health status measures as primary or secondary endpoints, relating primarily to the lack of blinding and the nature of the health status measures. In many device trials, blinding of patients and physicians is not possible. To help mitigate this bias, the Working Group made 3 specific recommendations.

First, investigators should use independent evaluators to collect health status information. Patients may respond differently to a health care member of the investigating team than to an evaluator who is not directly involved in the trial. Evaluators should be carefully trained, supervised, and monitored throughout data collection. They should be blinded to treatment group whenever possible. In other areas this has proved feasible with appropriate instructions given to trial participants about not revealing their treatment status.

Second, the health status instrument should include questions that are as specific as possible to ensure the instrument is measuring what is intended, thereby maximizing content validity. The questions should represent cross-sectional assessments of specific domains and, if necessary, employ only a short recall period. For example, asking if a patient feels better is much less specific than inquiring if the patient gets short of breath when walking up stairs. Non-specific overall health questions will not provide appropriate endpoints in a device trial. Avoidance of such questions will help ensure that the conclusions are directly related to the study device.

Third, the mode and timing of health status assessment must be consistent throughout the data collection period.

Thus, if the instrument is self-administered, then that same methodology should be used at all protocol-specified time points. If possible, multiple measurements should be made at baseline (prerandomization) to document reliability and patient stability and at multiple prespecified intervals after randomization. The frequency of evaluations should depend on the number of evaluations needed to obtain an adequate estimate of device effectiveness given the expected mortality and dropout rates as well as the amount of patient inconvenience. The Working Group also noted that cognitive impairment may exist before device implantation. Determining the validity of the measurement in an impaired population may be necessary to appropriately interpret device effectiveness.

What Are the Analytical Considerations in Using Health Status Information?

The characteristics of health status measurements demand special attention in the analysis stage. By its nature, most health status outcomes consist of a collection of responses to multiple items that comprise a single scale or several scales. Data can be missing for particular items within a patient, referred to as *item nonresponse*, or all items can be completely missing for patients, referred to as *unit nonresponse*. Clear documentation for unit nonresponse is necessary to understand the reasons for the missing data, as for any trial endpoint. For item nonresponse, complete-case analyses are valid only in the rare circumstance when data are missing completely at random. Multiple imputation methods will ultimately be more useful and statistically valid.^{32,33}

Because of the need to collect multiple health status measurements, the Working Group recommended adoption of appropriate models for longitudinal data.^{34,35} However, the choice of a longitudinal model will depend on the choice of a summary statistic for device effectiveness, such as the average change or a slope, as well as the nature and extent of missing data. When the fraction of missing data is small, multivariate analyses, such as mixed-modeling or growth-curve modeling, should be used if the data are missing at random. If data are not missing at random, then pattern-mixture models³⁶ that stratify the incomplete data by the pattern of missing values and formulate distinct models within each stratum or selection models should be considered.

In heart failure trials involving patients with advanced disorder, missing data during follow-up may arise because of death. Some patients may have treatment intensified because of the progression of their disorder. Control patients may have a device implanted or have a heart transplant. These changes complicate longitudinal comparisons of the outcome. Several approaches to analyzing such data have been described, including a longitudinal summary of the data ignoring outcomes after the event of interest. This approach may be satisfactory assuming that the missing data can be explained by the observed data.³⁷ Schlucter and colleagues³⁸ and Touloumi³⁹ describe statistical models for summarizing

Table 3. Example of Design and Analytical Considerations for Device Evaluation Based on Health Status Outcomes

Design Step	Example
Define primary health status outcome	Change in quality of life after 6 months
Specify global statistic for primary and any secondary health status outcomes using O'Brien's rank statistic ⁴³	Change in: overall quality-of-life change, symptoms, and physical limitations
Specify composite safety outcome	Death and serious adverse events

individual patient rates of change in the presence of informative events.

Another method for accommodating deaths is to analyze health status data and survival in a joint, bivariate, model⁴⁰ using an integrated quality-adjusted survival curve in which continuous quality-of-life adjustment is made based on repeated measurement of health status. In this case the summary statistic for measuring device effectiveness may be a comparison of the area under the health status curve. Similar methods that combine item response models with Cox's proportional hazard models to estimate the area under the quality-of-life curve⁴¹ are also available.

Assignment of a "worst rank" score, such as zero, to deaths or interfering events is yet another strategy. Lachin⁴² demonstrated that this is a valid approach if the event, such as death, is truly informative. However, sudden deaths that occur when patients might be experiencing relatively good quality of life make this assumption less tenable in studies of heart failure. If deaths occur at random, then adopting this procedure could result in a loss of statistical power. Table 3 provides an example of a design to assess a device on the basis of health status.

The Working Group noted that all of the analytical approaches for handling missing data involve assumptions that are not easily verified. The sponsor should prespecify methods for handling missing data, the planned sensitivity analyses, and the rationale for the selection of methods in the study protocol. Additionally, the Working Group recommended the use of multiple analytic methods when data are missing because of deaths, withdrawals, or other events, and to assess the robustness of the conclusions to different assumptions regarding the missing data or censoring mechanisms.

Conclusion

The Working Group recommended that a comprehensive assessment of health status be part of any device trial for heart failure and that such an assessment occur regardless of whether the device is intended as destination or bridge to transplantation therapy. Health status endpoints should be chosen, collected, and analyzed with the same level of scientific rigor as traditional clinical endpoints. However, because of the investigator's limited ability to blind patients

and researchers participating in device trials and of the nature of health status data, additional design and analytic steps are needed. Regulatory agencies should require use of analytical methods that handle the complexity of health status data in addition to the usual protocol protections. Although the Working Group was charged with developing recommendations specifically for device trials, the recommendations in this article apply to all trials of all heart failure interventions.

Acknowledgments

The authors are indebted to Marvin A. Konstam (president, Heart Failure Society of America [HFSA]), Cheryl Yano (executive director, HFSA), participants at an HFSA workshop, and 2 anonymous reviewers. The opinions expressed in this article are those of the authors and do not necessarily reflect the position of the HFSA.

References

- Covinsky KE, Wu AW, Landefeld CS, Connors AF, Phillips RS, Tsevat J, et al. Health status versus quality of life in older patients: does the difference matter? *Am J Med* 1999;106:435-40.
- Lepelge A, Hunt S. The problem of quality of life in medicine. *JAMA* 1997;278:47-50.
- Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. *JAMA* 1995; 273:59-65.
- Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622-9.
- Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *N Engl J Med* 1996;334:835-40.
- Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA* 1994;272:619-26.
- Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992; 45:1341-5.
- Gorkin L, Norvell NK, Rosen RC, Charles E, Shumaker SA, McIntyre KM, et al. Assessment of quality of life as observed from the baseline data of the Studies of Left Ventricular Dysfunction (SOLVD) trial quality-of-life substudy. *Am J Cardiol* 1993;71:1069-73.
- Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol* 2000;35: 1245-55.
- Bennett SJ, Oldridge NB, Eckert GJ, Embree JL, Browning S, Hou N, et al. Discriminant properties of commonly used quality of life measures in heart failure. *Qual Life Res* 2002;11:349-59.
- Rector TS, Kubo SH, Cohn JN. Validity of the Minnesota Living with Heart Failure questionnaire as a measure of therapeutic response to enalapril or placebo. *Am J Cardiol* 1993;71:1106-10.
- Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-8.
- Guyatt GH, Veldhuyzen Van Zanten SJ, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: a taxonomy and review. *Can Med Assoc J* 1989;140:1441-8.
- Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989;42:403-8.

15. Knapp TR. The reliability of measuring instruments. Ed 2. Edgeworth Laboratory for Quantitative Educational and Behavioral Science Series. Vancouver, BC, Canada; 2002.
16. Alwin DF. Problems in the estimation and interpretation of the reliability of survey data. *Qual Quant* 1989;23:277–331.
17. Linguistic validation manual for patient reported outcomes instruments. Lyon, France: MAPI Research Trust, 2004.
18. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469–78.
19. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Contr Clin Trials* 1989;10:407–15.
20. Santanello NC, Baker D, Cappelleri JC, Copley-Merriman K, DeMarinis R, Gagnon JP, et al. Regulatory issues for health-related quality of life—PhRMA Health Outcomes Committee Workshop, 1999. *Value Health* 2002;5:14–25.
21. Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. Pimobendan Multicenter Research Group. *Am Heart J* 1992;124:1017–25.
22. Feldman AM, Baughman KL, Lee WK, Gottlieb SH, Weiss JL, Becker LC, et al. Usefulness of OPC-8212, a quinolinone derivative, for chronic congestive heart failure in patients with ischemic heart disease or idiopathic dilated cardiomyopathy. *Am J Cardiol* 1991;68:1203–10.
23. Packer M, Colucci WS, Sackner-Bernstein JD, Kinhal V, Gottlieb SS, Gregory JJ, et al. Double-blind, placebo-controlled study of the effects of carvedilol in patients with moderate to severe heart failure. The PRECISE Trial. Prospective Randomized Evaluation of Carvedilol on Symptoms and Exercise. *Circulation* 1996;94:2793–9.
24. Massie BM, Berk MR, Brozna SC, Elkayam U, Plehn JF, Kukin ML, et al. Can further benefit be achieved by adding flosequinan to patients with congestive heart failure who remain symptomatic on diuretic, digoxin, and an angiotensin converting enzyme inhibitor? Results of the flosequinan-ACE inhibitor trial (FACET). *Circulation* 1993;88:492–501.
25. Giannuzzi P, Temporelli PL, Corra U, Tavazzi L. Antiremodeling effect of long-term exercise training in patients with stable chronic heart failure: results of the Exercise in Left Ventricular Dysfunction and Chronic Heart Failure (ELVD-CHF) Trial. *Circulation* 2003;108:554–9.
26. Deswal A, Bozkurt B, Seta Y, Pariliti-Eiswirth S, Hayes A, Blosch C, et al. Safety and efficacy of a soluble P75 tumor necrosis factor receptor (Enbrel, etanercept) in patients with advanced heart failure. *Circulation* 1999;99:3224–6.
27. Rector TS, Johnson G, Dunkman WB, Daniels G, Farrell L, Henrick A, 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. The V-HeFT VA Cooperative Studies Group. *Circulation* 1993;87:VI71–7.
28. Rector TS, Tschumperlin LK, Kubo SH, Bank AJ, Francis GS, McDonald KM, et al. Use of the Living With Heart Failure questionnaire to ascertain patients' perspectives on improvement in quality of life versus risk of drug-induced death. *J Card Failure* 1995;1:201–6.
29. Baxter AJ, Spensley A, Hildreth A, Karimova G, O'Connell JE, Gray CS. Beta blockers in older persons with heart failure: tolerability and impact on quality of life. *Heart* 2002;88:611–4.
30. Cohn JN, Goldstein SO, Greenberg BH, Lorell BL, Bourge RC, Jaski BE, et al. A dose-dependent increase in mortality with vesnarinone among patients with severe heart failure. Vesnarinone Trial Investigators. *N Engl J Med* 1998;339:1810–6.
31. Conard MW, Spertus JA, Heidenreich P, Peterson E, Poston WSC, Weintraub W. Examining the concordance between patient and physician assessments of health status in heart failure patients. *J Am Coll Cardiol* 2002;39(Suppl 2):461A.
32. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc* 1996;91:473–89.
33. Schafer JW, Graham JW. Missing data. Our view of the state of the art. *Psychol Meth* 2002;7:147–77.
34. Little RJ. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc* 1993;88:125–34.
35. Fairclough DL. Methods of analysis for longitudinal studies of health-related quality of life. In: Staquet MJ, Hayr RD, Fayer PM, editors. *Quality of life assessment in clinical trials: methods and practice*. New York: Oxford University Press; 1998.
36. Little RJ, Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* 1996;52:98–111.
37. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer, 2000.
38. Schluchter MD, Greene T, Beck GJ. Analysis of change in the presence of informative censoring: application to a longitudinal trial of progressive renal disease. *Stat Med* 2001;20:989–1007.
39. Touloumi G, Pocock SJ, Babiker AG, Darbyshire J. Estimation and comparison of rates of change in longitudinal studies with informative dropout. *Stat Med* 1999;18:1215–33.
40. Glasziou PP, Cole BF, Belger RD, Hilden J, Simes RJ. Quality adjusted survival analysis with repeated quality of life measures. *Stat Med* 1998;17:1215–29.
41. Wang C, Douglas J, Anderson S. Item response models for the joint analysis of quality of life and survival. *Stat Med* 2002;21:129–42.
42. Lachin JM. Worst-rank score analysis with informatively missing observations in clinical trials. *Control Clin Trials* 1999;20:408–22.
43. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics* 1984;40:1079–87.

Appendix

Attributes for Disease-Specific Health Status Instruments

This appendix demonstrates prototypical information that should be included in a health status resource data bank.

Kansas City Cardiomyopathy Questionnaire

The Kansas City Cardiomyopathy Questionnaire (KCCQ) is a 23-item instrument that quantifies physical function, symptoms (frequency, severity and recent change), social function, self-efficacy/knowledge, and quality of life. A summary score is derived from the physical function, symptom (frequency and severity), social function, and quality-of-life domains. Scores range from 0 to 100, where higher scores reflect better health status.

Validity. When compared with the physician-assessed New York Heart Association (NYHA) class, the mean KCCQ summary scores are 82 ± 16 for NYHA I, 69 ± 20 for NYHA II, 50 ± 20 for NYHA III, and 35 ± 21 for NYHA IV.

Reliability/reproducibility. The test-retest reproducibility of the KCCQ was originally established in an outpatient cohort of 39 stable patients (mean age = 64 years, 69% male, mean NYHA = 2.0 ± 0.59) where the mean 3-month change = -2.1 ($P = .36$). The intraclass correlation coefficient in 320 stable patients from 13 centers at 6 ± 2 weeks apart was 0.88.

Responsiveness/sensitivity to change. In a cohort of 39 patients (mean age = 68, 62% male, mean NYHA = 3.3 ± 0.46) admitted to the hospital for decompensated heart failure and re-evaluated 3 months later (when their condition

had significantly improved), baseline, and 3-month KCCQ overall summary scores were 31.8 and 56.1 (mean change = 24.3, $P < .001$).

Interpretability. Several mechanisms for establishing standards for interpreting scores are available. One is to examine the prognostic significance of KCCQ scores and the other is to benchmark score changes against clinical assessments of change. To facilitate the interpretation of cross-sectional KCCQ scores, 1516 patients assessed 3 months after a myocardial infarction complicated by heart failure were followed for 1-year survival and heart failure hospitalization (Fig. 1).

A cohort of 460 patients from 13 centers was followed for 6 ± 2 weeks, at which point physicians assessed their clinical change (blinded to KCCQ scores). Among these patients, the magnitude and direction of change is shown in Fig. 2. KCCQ change scores were exquisitely reflective of clinical changes both in terms of its directionality (improvement versus deterioration) and proportionality of change (magnitude). These findings suggest that a mean difference between groups ≥ 5 points on the KCCQ Overall Summary Scale reflects a clinically significant difference in heart failure status.

An alternative approach to interpreting clinical changes is to appreciate the prognostic significance of changes in scores. In a cohort 659 subjects assessed 3 and 6 months after a myocardial infarction complicated by congestive heart failure, those whose KCCQ overall summary scores declined by ≥ 10 points had a 107% increased risk of dying or being hospitalized over the next 3 months (event rate = 11.4% versus 5.5%, $P < .001$) and a 94% increased relative risk at 1 year (27.7% versus 14.3%, $P < .001$). These data suggest that a 10-point decline in KCCQ scores has important prognostic significance.

Available translations. English (US and UK), Czech, Dutch, Danish, Estonian, Flemish, Finnish, French (Canada and France), Georgian, German, Greek, Hebrew, Italian, Norwegian, Polish, Portuguese, Russian, Spanish (Argentina, Mexico [US], and Spain) Serbian, Swedish, and Turkish. See www.cvoutcomes.org for current information.

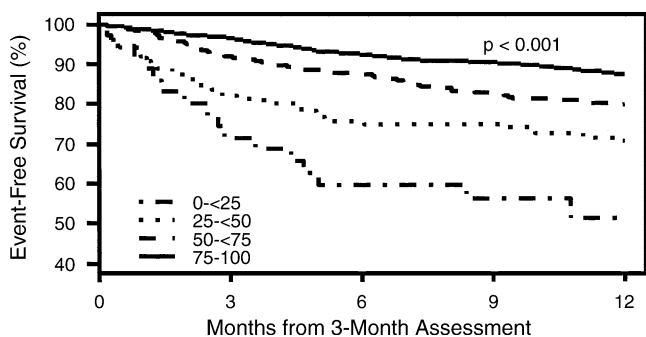


Fig. 1. Kaplan-Meier curves for study of 1516 heart failure patients assessed 3 months after a myocardial infarction using the Kansas City Cardiomyopathy Questionnaire.

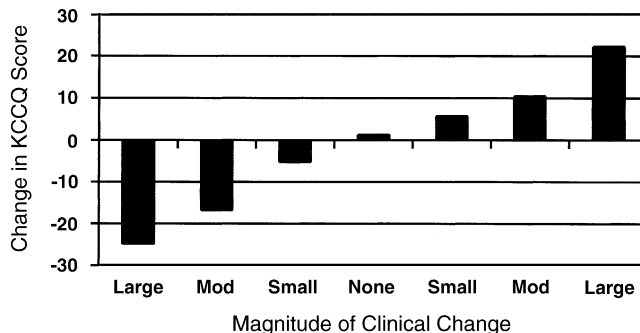


Fig. 2. Kansas City Cardiomyopathy Questionnaire change scores in a cohort of 460 patients from 13 centers.

Minnesota Living with Heart Failure (MLHF) Questionnaire

The MLHF Questionnaire was designed in 1984 to measure the effects of heart failure and treatments for heart failure on an individual’s quality of life. The content of the questionnaire was selected to be representative of the ways heart failure and treatments can affect the key physical, emotional, social, and mental dimensions of quality of life without being too long to administer during clinical trials or practice.

Validity. The concept measured is the effect of symptoms and functional limitations associated with heart failure on each patient’s quality of life. The content covers 21 ways heart failure most commonly affects patients’ lives. Each patient rates how much each item prevented them from living as he or she desired during the past month. The magnitude of correlation of the score with other measures directly corresponds with the extent of conceptual and content overlap (Table 4).

Table 4. Minnesota Living with Heart Failure (MLHF) Questionnaire

Measure	Correlation With MLHF Score
Chronic Heart Failure score	0.81
Dyspnea subscore	0.63
Fatigue subscore	0.78
Emotional subscore	0.74
Functional Status Scale—physical limitations	0.75
Emotional distress	0.64
SF-12 overall score	0.61
SF-12 physical subscore	0.57
SF-12 mental subscore	0.68
Time tradeoff utility willingness to trade current health for less time in perfect health)	0.56
Dyspnea scale	0.52
Clinician perceptions of patient’s health	0.44
Duration of submaximal bicycle exercise	0.43
Six-minute walk test	0.39, 0.26
Anerobic threshold	0.37
Peak oxygen consumption	0.30, <math>< 0.10</math>
Ejection fraction	0.03, 0.03, <math>< 0.10</math>

Reliability. To establish the internal consistency of the MLHF Questionnaire, Cronbach's alpha has been estimated to be 0.92 or better in several studies of patients with varying severity of heart failure (alpha = 1.0 indicates perfect consistency of patients' responses). The test-retest validity has been established by demonstrating that the correlation between scores from repeated assessments of clinically stable patients have been 0.87 or better.

Responsiveness. The standard error of the measure (the measurement error in individual scores) has been estimated to be approximately 6 to 7 points on the scale that ranges from 0 to 105. In randomized controlled trials, significant improvements in quality of life have been attributed to a variety of treatments including angiotensin-converting enzyme inhibitors, investigational positive inotropes, cardiac resynchronization, exercise training, and disease management programs. Significant effects have not been observed in studies of calcium channel blockers, and most studies of β -adrenergic receptor blockers or digoxin. These latter findings are consistent with the notion that most patients treated with these medications do not experience dramatic changes in their symptoms and functional limitations and may experience bothersome side effects.

Interpretability. In scoring the MLHF Questionnaire, the same response format is used for all items so one can see how a patient weighs each item and use the sum of responses as a measure of the effect of heart failure on a patient's life. It has been noted that the majority of patients in 1 study who were aware of their current score indicated that a 5-point improvement would be worthwhile if the

treatment had no side effects or out-of-pocket costs. Approximately 40% indicated they would risk a slight increase in mortality for this amount of improvement in quality of life. Others suggested that changes of 1 standard error, 6 to 7 points for this measure, might be clinically meaningful. Changes need to be greater than 2.77 times the standard error to be 95% confident the change was not the result of random variation in individual scores. To facilitate the interpretation of scores, it has been noted that average baseline scores from several studies were 21, 37, 53, and 69 for groups of patients assigned to NYHA classes I, II, III and IV, respectively. A longitudinal study suggested that a change of 1 NYHA class corresponded to a mean change of 10 points. Investigational inotropic agents and angiotensin-converting enzyme inhibitors improved quality-of-life scores by approximately 3 to 7 points on average. Improvements in quality-of-life scores attributed to cardiac resynchronization averaged 6 to 14 points. In unblinded studies of disease management and exercise programs, quality-of-life scores improved by 8 to 14 points more than control groups. A 10-point worsening in the quality-of-life measure was associated with a higher risk of hospitalization or death in another study.

Available translations. English (Australia, Canada, Ireland, and UK), Spanish (Spain, Argentina, and US); French (France, Belgium, Switzerland, and Canada), German; Polish; Italian; Dutch (Belgium); Portuguese; Swedish; Finnish; Danish; Norwegian; Latvian; Lithuanian; Estonian; Russian; Romanian; Slovakian; Croatian; Czech; Hungarian; Greek; and Hebrew. See www.mlhfq.org for a more detailed overview with references.