Measuring Health Related Quality of Life in Heart Failure

Olatz Garin Boronat

TESI DOCTORAL UPF / 2012

DIRECTORA DE LA TESI

Dra. Montserrat Ferrer Forés, (Departament de Pediatria, d'Obstetrícia i Ginecologia, i de Medicina Preventiva, UAB)

TUTOR DE LA TESI

Dr. Josep M. Antó Boqué

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

UNIVERSITAT POMPEU FABRA

A los "aitas"

Acknowledgements

La tesis fue una excusa genial para VIVIR en Barcelona (¡gracias tío!). Desde entonces la URSS se ha convertido en mi familia y os quiero agradecer que hayamos aprendido, reído e incluso sufrido, siempre juntos. En particular, gracias a Puri, por ser una canguro de lujo; a Sali, por ejercer de madre putativa; a Àngels, por su incondicional ayuda; a Gemmy, por la seguridad que me dan sus respuestas; a Montse, por encontrar siempre algo positivo y a Jordi, por lo contrario (de corazón); y a Antonia, por sus cuidados.

A lo largo de este trayecto también he contado con Maika, "la hermandad", "las chicas", "la peña…", los de "la borda y la manta" y , sobre todo, con los Garin-Boronat. A todos, gracias por aguantarme y animarme, por haberos reído de mí y conmigo. ¡Por haber estado ahí!

Algunos ya sabéis que en terrenos irregulares disfruto más si voy la última: gracias Joan, Chus y Manuel por tomar la delantera.
Y finalmente, gracias, Juan Ignacio Mestre Pintó, por decir: "PROU!".

Dicen que hay vida después de la tesis así que no os vayáis muy lejos. Os voy a seguir necesitando a todos.

MOLTES GRÀCIES

Esta tesis doctoral ha sido posible gracias a la participación de miles de pacientes, al trabajo de los coautores de los manuscritos, a la ayuda recibida por parte de la Fundació Institut Mar d'Investigacions Mèdiques-IMIM, y a la paciencia de Natàlia Ras.

Abstract

The global aim of this dissertation was to identify the strengths and limitations of the Health Related Quality of Life measures specifically developed for patients suffering from heart failure. A systematic review found 7 heart failure specific Health Related Quality of Life questionnaires: Chronic Heart Failure Assessment Tool (CHAT), Cardiac Health Profile congestive heart failure (CHPchf), Kansas City Cardiomyopathy questionnaire (KCCQ), Left Ventricular Disease Questionnaire (LVDQ), Minnesota Living with Heart Failure questionnaire (MLHFQ), and Quality of Life in Severe Heart Failure Questionnaire (QLSHFQ). Evidence published until 2006 pointed out the MLHFQ as the most extended and recommendable questionnaire. The Spanish version of the MLHFQ showed similar characteristics to its original version; as did the pooled evaluation of 20 country versions. Furthermore, the measurement model assessment confirmed the unidimensionality of the MLHFQ total score, and its physical, emotional and social domains. The 2011-update of the systematic review together with an experts' standardized evaluation showed the KCCQ to have better metric properties than the MLHFQ.

Resum

L'objectiu global d'aquesta tesis doctoral va ser identificar les fortaleses i limitacions de les mesures de Qualitat de Vida Relacionada amb la Salut desenvolupades específicament per pacients amb insuficiència cardíaca. Una revisió sistemàtica va trobar 7 qüestionaris de Qualitat de Vida Relacionada amb la Salut específics per insuficiència cardíaca: *Chronic Heart Failure Assessment Tool* (CHAT), *Cardiac Health Profile congestive heart failure (CHPchf), Kansas City Cardiomyopathy questionnaire* (KCCQ), *Left Ventricular Disease Questionnaire* (LVDQ), *Minnesota Living with Heart Failure questionnaire* (MLHFQ), i *Quality of Life in Severe Heart Failure Questionnaire* (QLSHFQ). L'evidència publicada fins 2006 senyalava el MLHFQ com el qüestionari més estès i recomanable. La versió espanyola del MLHFQ va mostrar característiques similars a l'original, així com també l'avaluació conjunta de versions de 20 països. A més, l'avaluació del model de mesura va confirmar la unidimensionalitat de la puntuació total del MLHFQ, i les seves dimensions física, emocional i social. L'actualització al 2011 de la revisió sistemàtica juntament amb l'avaluació estandarditzada per experts, varen destacar les propietats mètriques del KCCQ per sobre de les del MLHFQ.

Preface

Heart failure (HF) is a serious, costly and increasingly common condition that has a significant impact on health-related quality of life (HRQL), comparable to or greater than conditions such as diabetes and arthritis.

Traditional end-points in HF have been: death, recurrent hospitalization, haemodynamic parameters, exercise and functional capacities. Nowadays they are still in force in clinical and research settings, despite their well-known limitations; and are usually combined with Patient Reported Outcomes Measures (PROMs) which incorporate a comprehensive account of the real limitations experienced by patients. Among these measures, HRQL instruments are the most used for HF patients' management. The WHO and several professional societies and associations support the fact of going beyond traditional outcomes for treatment or intervention evaluation among HF patients. There exist several HF-specific HRQL and they are worldwide administered as an outcome measure to gather information on the burden of the condition, both in research studies and clinical practice; especially now that HF prevalence has increased due to interventions' effectiveness and ageing population.

The adequate instrument should be identified based on the application aim, the resources availability, and the properties that showed. In the last decades, the exponential development and validation of HRQL measures has made the selection more complicated, and many professionals remain sceptical to their use because of the lack of clear, rigorous and standardised data. The project presented here contributes to meet that need by summarizing the strengths and limitations of the existing specific HRQL instruments developed for patients suffering from HF.

This dissertation is presented, following the instructions of the Department of Experimental and Health Sciences, as a "compendium of original publications". A narrative review about HRQL and HF has been place in first term to introduce the issue despite the fact that its corresponding publication is an encyclopaedia entry which is now in press:

Garin O, Ferrer M. *Health-Related Quality of Life and Heart Failure.* "Encyclopaedia of Quality of Life Research" Ed. Springer (in press)

For the achievement of the principal aims focused on HRQL instruments' assessment, this project summarizes the available evidence (until 2006) on the existing disease-specific HRQL instruments and their metric properties trough a systematic review with meta-analyses:

Garin O, Ferrer M, Pont A, Rué M, Kotzeva A, Wiklund I, Van Ganse E, Alonso J. *Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses.* Qual Life Res, 2009, 18: 71-85

In a further step, this dissertation goes deeper in the evaluation of the most widely used HF-HRQL questionnaire, the MLHFQ. On one side with the validation of its Spanish version:

Garin O, Soriano N, Ribera A, Ferrer M, Pont A, Alonso J, Permanyer G, on behalf of the HF-QoL Group*. *Validación de la versión española del Minnesota Living with Heart Failure Questionnaire [Validation of the Spanish version of Minnesota Living with Heart Failure Questionnaire].* Rev Esp Cardiol, 2008, 61: 251-259

and, on the other, with the confirmation of the measurement model of its original and adapted versions for the first time:

Garin O, Ferrer M, Pont A, Wiklund I, Van Ganse E, Vilagut G, Almansa J, Ribera A, Alonso J. *Evidence on the global measurement model of the Minnesota Living with Heart Failure Questionnaire.* (under review)

Finally, the project presents the evaluation of the existing HF-specific HRQL instruments. These appraisals were conducted by experts, based on the available published evidence, and under standard criteria:

Garin O, Herdman M, Vilagut G, Ferrer M, Ribera A, Rajmil L, Valderas JM, Guillemin F, Revicki D, Alonso J, on behalf of the EMPRO-HF Group* *Heart Failure Measures for Assessing Health Related Quality of Life: a systematic standardized comparison.* (submitted)

Table of Contents

1. BACKGROUND

1.1. Heart Failure

In Western Europe and North America, more than 5 million people suffer from Heart Failure (HF), with almost 400.000 new cases per year(1). The prevalence of HF was calculated to increase steeply with age(2): from one per cent in those aged 50-59 years, to 10 per cent in those aged 80-89 years; and it doubles with each decade of life.

Data from NHANES (*National Health and Nutrition Examination Survey*) 2007–2008 reveals that 2.0% of Americans self-reported having congestive heart failure(3). It is the third leading cause of cardiovascular mortality, after coronary disease and stroke (Figure 1).

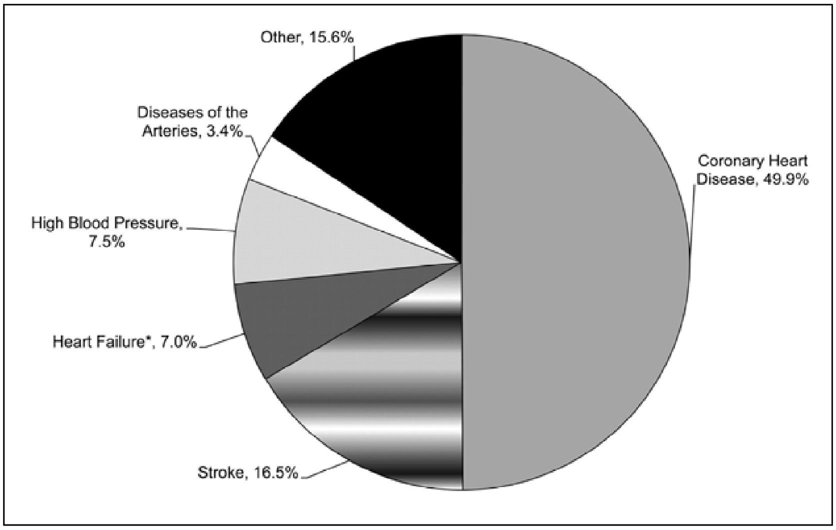


Figure 1(3) - Percentage breakdown of deaths attributable to cardiovascular disease (United States 2008)

In the same report it is stated how after pneumonia, the most common heart-related conditions for hospitalization (in descending order) were HF, chest pain, hardening of the arteries, and heart attack. Heart Failure hospitalization was higher among blacks, Hispanics, and American Indian/Alaska Natives than among whites; and also greater among women (according to analysis of several data sets by the CDC (Centers for Disease Control and prevention)in adults ≥ 18 years of age)(3).

In a recent epidemiological study undertaken in Spain(4), the prevalence of heart failure was found to be at about 7%-8% (similar in males and females), and appeared to increase with age.

The outlook for an individual patient depends on the patient's age, severity of heart failure, overall health, and a number of other factors(3). Taking into account that life expectancy for both men and women is growing; it is easy to guess how the magnitude of the problem is going to get worse.

Heart failure (HF) is a chronic and progressive disorder in which damage to the heart causes abnormalities in its contractile and relaxing functions, thus causing the inability of pumping enough blood to fulfil the metabolic requirements of the body(5). It is clinically manifested by fluid congestion or symptoms derived from the inadequate delivery of blood flow to tissues. Heart failure usually develops slowly, often over years, as the heart gradually loses its pumping ability and works less efficiently. Some people may not become aware of their condition until symptoms appear years after their heart began its decline. How severe the condition is depends on how much pumping capacity the heart has lost. The severity of the condition determines the impact it has on a person's life. At one end of the spectrum, the mild form of HF may have little effect on a person's life; at the other end, severe heart failure can interfere with even simple activities and prove fatal. Between these extremes, treatment and special management of patients often helps people lead full lives(6).

A number of symptoms are associated with HF, but none is specific for the condition. Perhaps the best known symptom is shortness of breath or “dyspnea”, which in HF may result from excess of fluid in the lungs. The breathing difficulties may occur at rest or during exercise. In some cases, congestion may be severe enough to prevent or interrupt sleep. Fatigue or easy tiring is another common symptom: as the heart's pumping capacity decreases, muscles and other tissues receive less oxygen and nutrition, which translates into fatigue(7). Fluid accumulation, or oedema, may cause swelling of the feet, ankles, legs, and occasionally, the abdomen. Excess fluid retained by the body may result in weight gain, which sometimes occurs fairly quickly. As HF progresses, the effects can become quite severe, and patients often lose the ability to perform even modest physical activity. Eventually, the heart's reduced pumping capacity may interfere with routine functions, and patients may become unable to care for themselves. Therefore, the loss in functional ability can occur quickly if the heart is further weakened by heart attacks or the worsening of other conditions that affect HF, such as diabetes and coronary heart disease. Thus, the impact HF can have on a patient is also important psychologically and socially (8).

Morbidity in HF patients mainly differs depending on how much pumping capacity the patient has lost. However, in general, HF not only comprises physical symptoms (shortness of breath and fatigue), impaired exercise capacity, and recurrent hospitalization, but also psychological problems, iatrogenic adverse effects and curtailment in social activities(9).

Heart failure caused by anatomical problems, such as a heart valve defect, is curable as it can be surgically corrected. However, for the common forms of HF --those due to damaged heart muscle-- no known cure exists, but treatment may be quite successful. These patients can minimize the effects of HF by controlling the risk factors for heart disease. Obvious steps include quitting smoking, losing weight if necessary, abstaining from alcohol, and making dietary changes to reduce the amount of salt and fat consumed. Regular, modest exercise is also helpful for many patients, though the amount and intensity should be carefully monitored by a physician. Despite lifestyle changes, most HF patients must take medication.

Many patients receive two or more drugs. Several types of drugs have been proven useful in the treatment of heart failure(10). Diuretics help reduce the amount of fluid in the body and are useful for patients with fluid retention and hypertension, similar to what aldosterone antagonists do reducing oedema and the cardiac workload. Digitalis increases the force of the heart's contractions, helping to improve circulation. Angiotensin Converting Enzyme Inhibitors (ACEI) are primarily used for the treatment of high blood pressure; and betablockers are nowadays known for their reductive effect on heart rate and heart oxygen demand. Several large studies with HF patients have indicated that these drugs reduce their mortality and morbidity. In some patients with heart failure, new devices such as implantable cardioverter defibrillators (ICD) and cardiac resinchronization devices may entail a significant improvement in their outcomes. When drug therapy and lifestyle changes fail to control symptoms, a heart transplant may be the only treatment option. However, due to the shortage of donors, candidates for transplantation often have to wait for months or even years. Some of these patients can have a circulatory assisting device implanted as a bridge to transplantation, or even as a destination therapy. However, most patients are not eligible for these advanced therapies due to concomitant comorbidities or higher age. Therefore, most of the current interventions seek to improve patients' quality of life and length of survival through lifestyle change and drug therapy.

1.2. Patient Reported Outcomes Measures

Patient Reported Outcomes Measures (PROMs) is an umbrella term that covers any outcome based on data provided by the patient or patient proxy(11). According to the Food And Drug Administration's definition, "a PROM is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else"(12).

PROMs are designed to measure a specific concept (that is a construct) in a standardised way. The principal aim of this research area is to avoid the application and administration of measures/questionnaires developed as a matter of common sense, with little scientific consideration, which may lead to poor quality results of limited value(13). Different types of outcomes are covered by the term PROMs (Figure 2)(14), from symptoms to quality of life or health related quality of life; and are usually measured either by different instruments or instruments that combine several of those concepts. PROMs that only assess symptoms or functional limitations are of primary interest to the clinician as indicative of disease severity.

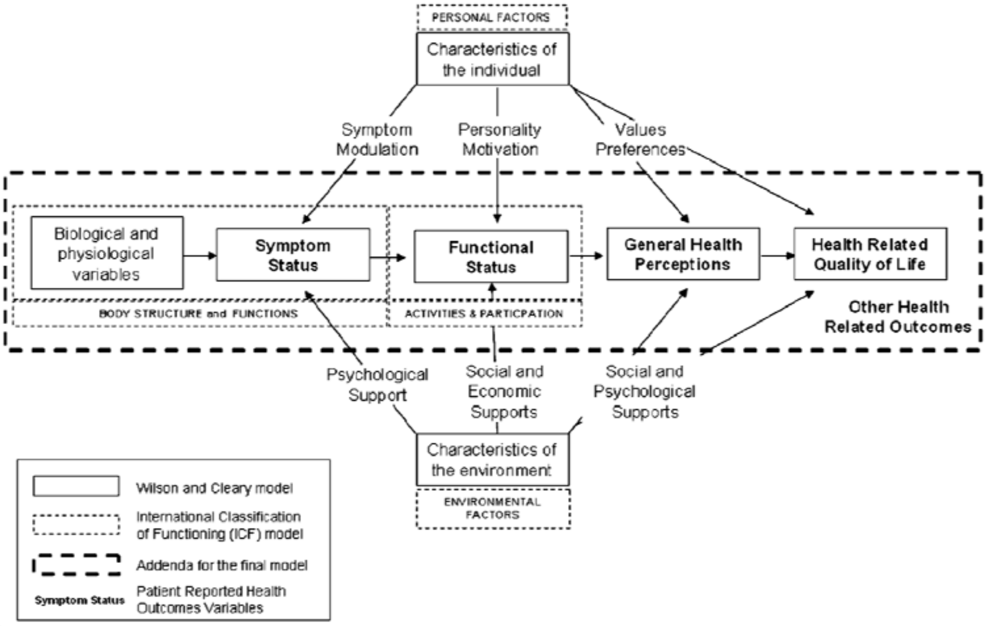


Figure 2(14) - An integrate model for health outcomes.

In general, PROMs can be used to help in decision making at the physical level, but they also are widely used in clinical trials to evaluate an intervention's effects(13). Most PROMs administered nowadays assess Health Related Quality of Life (Figure 3).

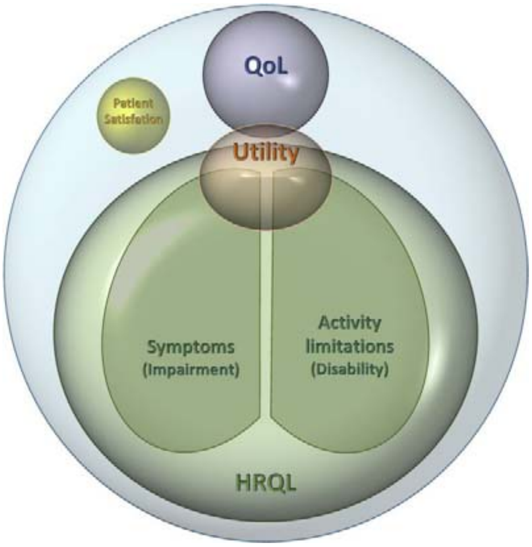


Figure 3(13) – Types of PROMs currently used in medical research

a) Health Related Quality of Life

In 1948, the World Health Organization defined health as being not only the absence of disease and infirmity, but also the presence of physical, mental and social well-being(15). Since then, the definition of HRQL and related concepts such as quality of life (for many outcomes researchers, limited to what is of primary concern to the patient(16)), health status and perceived health, has been disputed, without reaching a consensus(11;17). In 1993 Patrick and Erickson(18) defined HRQL as "the value assigned by individuals, groups, or society to the duration of survival as modified by impairments, functional states, perceptions, and social opportunities influenced by disease, injury, treatment, or policy". However, as shown in a systematic review(19), the definitions for HRQL in published articles differ in their content, or are usually not clearly provided.

In general, it is accepted that HRQL measures refer to the physical, psychological and social domains of health seen as distinct areas that are influenced by a person's experience, beliefs, expectations, and perceptions. Since the early 70s, HRQL instruments have been developed to assess a person's interpretation of his or her own health status in comparison to how he or she might hope to be(20;21).

HRQL measures have traditionally been differentiated as generic or specific, each group presenting its own characteristics (Table 1). Generic measures can be used for patients with any type of disorder or for general population(22). Their broad applicability is in general derived from their coverage of the complete spectrum of function, disability and distress that is relevant to health related quality of life (symptoms, emotional function, or social relations). According to Guyatt et al(16) generic instruments include single indicators, health profiles and utility measures. A similar classification was recently published by Khanna et al (Figure 4)(23).

Health profiles are instruments that attempt to measure all the important aspects of HRQL. The Sickness Impact Profile is an example of a health profile and includes a physical dimension (with categories of ambulation, mobility, body care and movement); a psychosocial dimension (with categories including social interaction, alertness behaviour, communication, and emotional behaviour); and five independent categories including eating, work, home management, sleep and rest, and recreation. One of the most popular generic HRQL profiles is a collection of instruments developed as part of the Medical Outcomes Study (SF-36)(24). The SF-36 includes 36 questions which measure physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health. Major advantages of health profiles include dealing with a variety of areas and being useable in any population, regardless of the underlying condition.

Table 1 – Characteristics of measures of HRQL(16)

Approach	Strengths	Weaknesses
Generic instruments		
Health Profile	- Sigel instrument - Detects differential effects on different aspects of health status - Comparisons across interventions, conditions possible	- May not focus adequately on area of interest - May not be responsive
Utility Measure	- Single number representing net impact on quantity and quality of life - Cost-utility analysis possible - Incorporates death	- Difficulty determining utility values - Does not allow examination of effect on different aspects of quality of life - May not be responsive
Specific instruments		
Disease specific Population specific Function specific Condition/problem specific	- Clinically sensible - May be more responsive	- Does not allow cross condition comparisons - May be limited in terms of population and interventions.

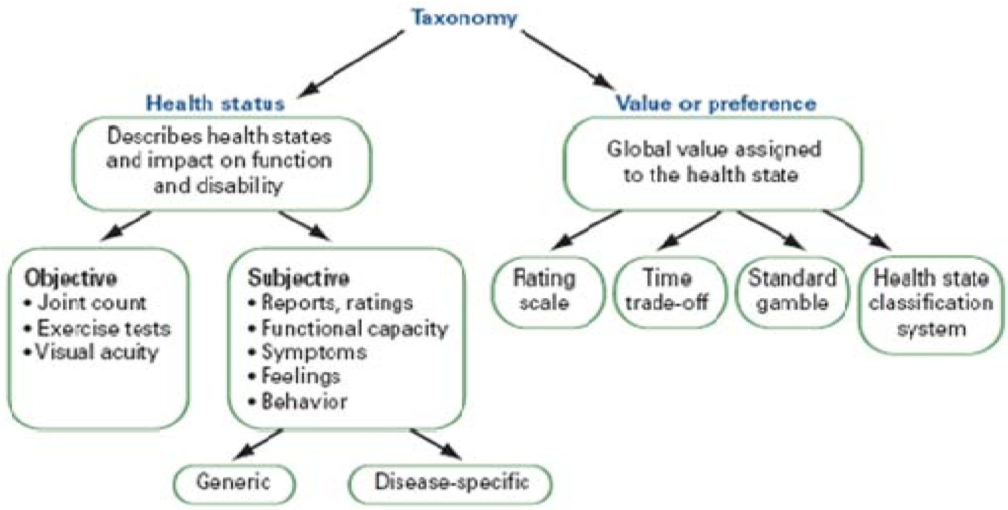


Figure 4 (23)– Variety of measures created to measure HRQL

Utility measures, the other type of generic instrument, are derived from economic and decision theory; they reflect the preferences of patients for treatment process and outcome(16). The key elements of utility measures are that they incorporate preference measurements and relate health states to death. Hence, they can be used in cost-utility analyses that combine duration and quality of life. In utility measures, HRQL is summarized as a single number along a continuum that usually extends from death (0.0) to full health (1.0) (although scores lower than zero, representing states worse than death, are possible(25)). The preferences in utility measurements may come directly from individual patients who are asked to rate the value of their health state. Alternatively, patients can rate their health status using a multiattribute, health-status classification system. A previously estimated scoring function (derived from results of preference measurements from groups of other patients or from the community) is then used to convert health status into a utility score(16;26). The most commonly used utility measure now is the EQ-5D, which consists of 2 parts: the EQ-5D descriptive system (with the EQ-5D index) and the EQ visual analogue scale (EQ VAS)(27). The EQ-5D descriptive system comprises 5 dimensions: Mobility, Self-care, Usual activities, Pain/Discomfort and Anxiety/Depression. Each dimension has one specific question and 3 levels of response: 1 “no problems”, 2 “some problems” and 3 “severe problems”. The instrument therefore defines 243 distinct health states from all the possible combinations of dimensions and levels of severity. Considering the responses to the descriptive system, each health state is converted into a utility index applying the general population preference values. Utility indexes can be used to calculate Quality Adjusted Life Years (QALYs).

Generic HRQL instruments allow to determinate the effects of the intervention on different aspects of HRQL without the use of multiple instruments; and using them one can compare the effects on HRQL of similar interventions in different diseases. However, they may not focus on the HRQL aspects of specific interest to the investigator. Inadequate focus on the quality of life issues of a specific trial is likely to result in an unresponsive instrument that may miss small but still clinically important changes in HRQL.

On the other hand, specific measures focus on aspects of HRQL that are specific to the area of primary interest. The instruments may be specific to the disease (such as heart failure), specific to a population of patients (such as children), specific to a certain function (such as sexual function), or specific to a given condition or problem (such as pain). The disadvantages of specific measures are that they are not comprehensive and cannot be used to compare across sub-populations or conditions. Nevertheless, as disease-specific instruments are designed to focus on elements of a specific condition, they may be more responsive to the effects of health care, and relate more closely to clinical symptoms(28).

b) Metric Properties

Even if one is choosing an already existing instrument or if his/her study requires a new PROM to be developed, the scaling and metric properties of the measure should be evaluated (i.e. dimensionality, reliability, validity and responsiveness). Besides, in the second case, the development process may include the identification of the measurement model and the content generation (13). To conduct that validation under a common standard, the Medical Outcomes Trust in 1994 created an independently functioning Scientific Advisory Committee. In the mid-1990s, this Committee defined a set of attributes and rigorous criteria to carry out instrument assessments(29). Five years later, those experts updated and revised the materials to take account of the expanding theories and technologies upon which PROMs should be validated (Table 2)(24).

The conceptual model of a PROM is the description of the concepts and the populations that it is intended to assess, and the relationship between those concepts. A measurement model operationalises the conceptual model and reflects an instrument's scale and subscale structure and the procedures followed to create the scale and subscale scores. The adequacy of the measurement model can be evaluated by examining evidence that: (1) a scale measures a single conceptual domain or construct; (2) multiple scales measure distinct domains; (3) the scale adequately represents variability in the domain; and (4) the intended level of measurement of the scale (i.e. ordinal, interval, or ratio) and its scoring procedures are justified. Classical test theory approaches may employ, for example, principal components analysis, factor analyses, and related techniques for evaluating the empirical measurement model underlying an instrument and for examining dimensionality(30-32). Methods based on modern test theory may use approaches including confirmatory factor analysis, structural equation modelling, and methods based on Item Response Theory (IRT)(33).

The principal definition of PROM reliability is the degree to which an instrument is free from random error. Classical approaches for examining test re-liability include (a) internal consistency reliability, typically using Cronbach's coefficient(34), and (b) reproducibility (e.g., test–retest or inter-observer -interviewer- reliability). The first approach requires one administration of the instrument; the latter requires at least two administrations. Regarding internal consistency realiability, Cronbach's coefficient provides an estimator of reliability based on all possible split-half correlations for a multi-item scale. For instruments employing dichotomous response choices, an alternative formula, the Kuder–Richardson formula (KR-20)(35), is available. A second approach to reliability can be obtained by judging the reproducibility or stability of an instrument over time (test–retest). Test–retest reproducibility is the degree to which an instrument yields stable scores over time among respondents who are assumed not to have changed on the domains being assessed. As with internal consistency reliability, minimal standards for reproducibility coefficients are also typically considered to be 0.70 for group comparisons and 0.90–0.95 for individual measurements over time(24). The influence of test administration on the second administration may overestimate reliability. Conversely, variations in health, learning, reaction, or regression to the mean may yield test–retest data that underestimate reproducibility.

The validity of an instrument is defined as the degree to which the instrument measures what it purports to measure. There are different ways to classify the evidence for the validity of a PROM, an extended one defines three types of this attribute(24):

1) Content-related: Evidence that the content domain of an instrument is appropriate regarding its intended use. Methods commonly used to obtain evidence about content-related validity include the use of lay and expert (clinician) panel judgments of the clarity, comprehensiveness, and redundancy of items and scales of an instrument. Often, the content of newly developed self-report instruments is best elicited from the population being assessed or experiencing the health condition.

2) Construct-related: Evidence that supports proposed interpretation of scores based on theoretical implications associated with the constructs being measured. Common methods to obtain construct-related validity data include examining the logical relations that should exist with other measures and/or patterns of scores for groups known to differ on relevant variables. Ideally, developers should generate and test hypotheses about specific logical relationships among relevant concepts or constructs.

Table 2- Attributes and criteria for evaluating PROMs, a modified version from the Scientific Advisory Committee of the Medical Outcome Trust(24).

1. **Conceptual and measurement model:** The rationale for, and description of, the concept and populations that a measure is intended to assess and the relationship between these concepts.
2. **Reliability:** The degree in which an instrument is free from random error.
    - **a.** *Internal consistency*: The precision of a scale, based on the homogeneity (inter-correlations) of the scale's items at one point in time.
    - **b.** *Reproducibility:* Stability of an instrument over time (test–retest) and inter-rater agreement at one point in time.
3. **Validity:** The degree to which the instrument measures what it purports to measure.
    - **a.** *Content-related*: Evidence that the domain of an instrument is appropriate regarding its intended use.
    - **b.** *Construct-related*: Evidence that supports a proposed interpretation of scores based on theoretical implications associated with the constructs being measured.
    - **c.** *Criterion-related*: Evidence that shows the extent to which scores of the instrument are related to a criterion measure.
4. **Responsiveness:** An instrument's ability to detect change over time.
5. **Interpretability:** The degree to which one can assign easily understood meaning to an instrument's quantitative scores.
6. **Respondent and administrative burden:** The time, effort, and other demands placed on those to whom the instrument is administered (respondent burden) or on those who administer the instrument (administrative burden).
7. **Alternative forms:** These include self-report, interviewer-administered, trained observer rating, computer-assisted interviewer-administered, evidence on reliability, validity, responsiveness, interpretability, and burden for each mode of administration performance-based measures.
8. **Cultural and language adaptations**: This refers to the assessment of conceptual and linguistic equivalence, as well as to the evaluation of measurement properties.

3) Criterion-related: Evidence that shows the extent to which scores of the instrument are related to a criterion measure. Criterion measures are measures of the target construct that are widely accepted as scaled, valid measures of that construct. In the area of self-reported health status assessment, criterion-related validity is rarely tested because of the absence of widely accepted criterion measures, although exceptions occur such as testing shorter versions of measures against longer versions.

An instrument's ability to detect change is known as responsiveness or sensitivity to change (highly valued at longitudinal construct validation process). The criterion of responsiveness requires asking whether the measure can detect differences in outcomes, even if those differences are small. Responsiveness can be conceptualized also as the ratio of a signal (the real change over time that has occurred) to the noise (the variability in scores seen over time that is not associated with true change in status)(36). Assessment of responsiveness involves statistical estimation of a statistic effect size (37)– that is, an estimate of a measure of the magnitude of change in health status (sometimes denoted as the 'distance' or difference between "before" and "after" scores). These statistics are the amount of observed change over the amount of observed variance. In these statistics, the numerator is always a change score, but the denominator differs depending on the coefficient used (e.g., standardized response mean, responsiveness statistic, SE of the mean).

The latest review of the PROMs attributes by the Scientific Advisory Committee of the Medical Outcomes Trust, considered “interpretability” and “administration burden” as characteristics that should also be taken into account when evaluating an instrument. Interpretability is defined as the degree to which one can assign easily understood meaning to an instrument’s quantitative scores(24). Interpretability calls for explanation of the rationale for the external measure, the change scores, and the ways that those scores are to be interpreted in relation to the external measure. Several types of information can aid in the interpretation of scores: comparative data from a sample of the general population; results from a large pool of studies, the relationship of scores to clinically recognized conditions, or how well scores predict known relevant events.

Finally, respondent burden is defined as the time, effort, and other demands placed on those to whom the instrument is administered. Demands placed on those who administer the instrument are recognized as the administrative burden of that instrument. This attribute can be evaluated by the average and range of time needed to complete the instrument on a self-administered basis or as an interviewer-administered instrument, the reading and comprehension level needed, special requirements (such as the need to consult health care records), or by the acceptability of the instrument (e.g. by indicating the level of missing items).

1.3. Health Related Quality of Life and Heart Failure

The New York Heart Association (NYHA) classification(38) and the walking distance (by asking patients how far they can walk, or performing the Six Minute Walking Test, 6MWT(39)) have historically been the most widely used ways of measuring HF functional status. Briefly, the NYHA classification helps doctors to place patients into one of four severity categories, based on their impression of limitations related to a patient's ordinary activities, in regards to normal breathing and varying degrees in shortness of breath and/or angina pain during physical activity. Classes range from I (no limitation) to IV (very limited). The walking tests only focus on functional capacity as a reliable procedures to measure exercise tolerance(40;41).

In the late eighties, investigators and clinicians mainly relied on these relatively unsophisticated instruments and ad hoc measures in order to assess functional status of HF patients. However, some authors remarked the need of incorporating new instruments in HF evaluation(42;43), and suggested them to: a) incorporate not only a widespread physical but emotional function so as to have comprehensive account of the real limitations experienced by patients; and b) to fulfil a minimum of psychometric criteria.

Among these measures, patient reported outcomes (PRO), and specifically HRQL instruments, are the most used nowadays for HF patient management(6;44). This holistic goal is considered very relevant under theWHO's current definition of health(15), and also by several professional societies and associations(45;46) which support the fact of going beyond hard-clinical outcomes for treatment or intervention evaluation among HF patients. Several decades ago, Wenger suggested HRQL to be adequate for HF evaluation in certain circumstances(47):

1) When there is little likelihood of one treatment showing a major improvement in survival over another in a clinical trial. In such a trial HRQL measurement might point towards the choice of the therapy showing the greatest benefit for patients.
2) When a treatment is effective in reducing mortality, but has toxic or unacceptable side effects. Quality of life measurement in this case may help physicians and their patients weigh the benefits and risks of such a treatment.
3) When patients are asymptomatic or have mild symptoms, the morbidity and mortality rates are low, and the therapy is long term.

On one hand, cardiologists and researchers started to incorporate generic HRQL instruments to the evaluation of HF patients. al Kaade and Hauptman conducted a search of the measures used in HF clinical trials before 2000 and founded 13 generic instruments(48): the Sickness Impact Profile, the General Well-Being Index, and the Nottingham Health Profile among other. Nowadays generic HRQL instruments (such as the EQ-5D(49) and the SF-36(50)) are used(51) in patients with HF mainly to be able to compare their burden with other populations (disorders) or to calculate loss of QALYs(52). On ther other hand, several HF-specific instruments were developed at the late eighties(42;43;53), and their use has been consolidated in the last decade. Generic and HF-specific HRQL instruments are administered together for evaluating treatment or intervention efficacy under targeted constructs and at the same time have the opportunity to compare the HRQL of patients suffering from HF with that of other chronic patients or general population.

Compared to general population, HF patients present significantly reduced HRQL(54), and they do also when compared to other chronic diseases. Evidence has shown HF as one of the chronic conditions that provokes the highest impairment in physical and mental health, comparable to or greater than arthritis and chronic lung disease(9;44). In addition to this already dramatically reduced physical health, especially patients with advanced HF have presented similar scores on mental health to patients with major depression (54).

Lower HRQL in patients with HF has been found to be independently associated with sociodemographic characteristics of the individuals, such as advanced age or lower socioeconomic status(55-57). The associations between HRQL scores and clinical variables or functional capacity measures have been evaluated in several studies, most of which achieved inconclusive results(54).On one hand, it seams that HRQL decreases as functional level worsens, as happens with the peak oxygen uptake and walking capacity(57;58). In contrast, there is some controversy about reliance on the results of physiologic measures, such as the ejection fraction(58) which was found to be non-associated with patient reported HRQL(54;56).

HRQL of HF patients has also been independently associated with less common end-points in clinical practice such as mental health or social support. Some of the results showed how neuroticism(58), emotional distress(59), and depression(60;61) were significant risk factors for worse HRQL. Moreover, as Coelho et al(62) mentioned at their review, some of these studies showed higher incidence of depression in the young, which may be suggesting that depression is due to a larger disparity between perception of functional status and expectation. That depression state may lead physicians to classify these patients as more severely compromised and rate their NYHA functional class worse(62).Taking the HRQL as an exposure, Havraneck et al(63) showed how lower reported scores were predictors of developing depressive symptoms. Moreover, some studies have reported that lack of social support (informal carer) and social isolation (maintenance of contacts and activities) are associated with increased risk of mortality among HF patients(57;64).

The assessment of HRQL is being increasingly included nowadays in medical research in HF, and also by the pharmaceutical industry for the evaluation of new treatments(65). HRQL is a major goal in the context of preventive and therapeutic cardiology, from comprehensive management to exercise programs(66;67). There are many examples of studies (clinical trials or observational) were HRQL has been the principal or at least a secondary end point(65;68) (e.g. on beta-blockers(69-72), or exercise (73) (74)). Some special management programmes have demonstrated to significantly improve HRQL--home monitoring(75), outpatient HF clinic(76), or brief cognitive therapy interventions(77). In the same way, resynchronization therapies and pacing devices have also been evaluated in terms of HRQL improvement(52;78;79). There are also evidences of HRQL as the main outcome in evaluations on sub-populations of HF patients, like iron-deficient patients treated with intravenous ferric carboxymaltose(80) or cognitive therapy(77).

In addition, not only the necessity of enhancing patients' daily functioning and well-being has made HRQL a recognized and relevant outcome, but it has also recently confirmed its ability for predicting HF mortality and hospitalisation(57;81-83), and for working as a significant indicator of prognosis in advanced heart failure patients(84). Because of these multiple factors, many professionals argue how the evaluation of HRQL in HF, as a global comprehensive outcome, deserves special attention(85). The driving force behind this trend may not only have been the recognition that traditional (biological or physiological) endpoints don't reflect the impact of the disease and interventions on patients' lives, but also the increasing role of patients in decision making, especially in chronic conditions(48).

Nevertheless, even though there has been an exponential use of both HF-specific and generic HRQL measures in the last decades(8), many professionals remain sceptical to the use of HRQL instruments, and have doubts because there is a lack of rigorous and standardised data regarding HF-specific HRQL instruments' characteristics(13).

2. THESIS RATIONALE

Morbidity in heart failure (HF) patients mainly differs depending on how much pumping capacity has the patient lost. However, in general, HF not only comprises physical symptoms (short of breath and fatigue), impaired exercise capacity, and recurrent hospitalization, but psychological problems, iatrogenic adverse effects and curtailment in social activities(9). Many professionals argue how the evaluation of Health Related Quality of Life (HRQL) in HF patients, as a global comprehensive outcome, deserves special attention(85) especially now that HF prevalence has increased due to interventions' effectiveness and ageing population(86).

The driving force behind this trend may not only had been the recognition that traditional endpoints (biological or physiologic) don't reflect the impact of disease and interventions on patients' life, but the increasing role of patients in decision making, especially in chronic conditions(48). In fact, HRQL is considered one of the 3 measureable elements which matter in elderly patients, who probably have multiple co-morbidities(87) (88). They are also gaining in importance as a predictor of mortality and hospitalization in HF settings(57;81-84).

Cardiologist and researchers started to incorporate generic HRQL instruments to the evaluation of HF patients. Generic measures can be used for patients with any type of disorder or for general population(22). However, they may not focus on the HRQL aspects of specific interest, and findings are likely to miss small but still clinically important changes in HRQL. In contrast, disease-specific instruments are designed to focus on elements of a specific condition, and may be more responsive to the effects of health care, and relate more closely to clinical symptoms(28).

In the last decades there has been an exponential development and validation of both HF-specific and generic HRQL measures (from 5 articles in 1973 to 4000 articles in early 2012 with QOL as a key word in MEDLINE)(8), which difficulties the knowledge and generalized understanding for making a selection. Moreover, many professionals remain sceptical to the use of HRQL instruments for other reasons. On one side, the development and application of HRQL instruments is often regarded as a matter of "common sense" requiring little scientific consideration. On the other, rigorous and standardised data regarding HF-specific HRQL instruments' metrics deserve a clarification(13).

Therefore, a contribution to meet that need by summarizing the strengths and limitations of the existing instruments developed for patients suffering from HF may help professionals to identify the most suitable HF-specific HRQL questionnaire.

3. OBJECTIVES OF THE DOCTORAL THESIS

The global aim of this dissertation was to identify the strengths and limitations of the Health Related Quality of Life measures specifically developed for patients suffering from heart failure.

Specific objectives:

1. To review the evidence of Health Related Quality of Life measures use in patients suffering from heart failure.

2. To identify all the available disease-specific questionnaires to measure the Health Related Quality of Life of heart failure patients, and to evaluate the available data on their underlying conceptual model and metric properties (reliability, validity, and sensitivity to change).

3. To assess the metric properties of the Spanish version of the Minnesota Living with Heart Failure Questionnaire.

4. To assess the measurement model of the Minnesota Living with Heart Failure Questionnaire; and to evaluate the reliability and validity of its original and 20 country-versions.

5. To systematically compare all existing heart failure specific Health Related Quality of Life instruments using the EMPRO tool (Evaluating the Measurement of Patient Reported Outcomes).

4. PUBLICATIONS

4.1 Encyclopaedia entry

Garin O, Ferrer M. *Health-Related Quality of Life and Heart Failure*.
"Encyclopaedia of Quality of Life Research" Ed. Springer (in press)

Health Related Quality of Life and Heart Failure

Synonyms

None

Definition:

The Health Related Quality of Life reported by patients suffering from Heart Failure.

Description:

Morbidity in Heart Failure (HF) patients mainly differs depending on how much pumping capacity the patient has lost. However, in general, HF not only comprises physical symptoms (shortness of breath and fatigue), impaired exercise capacity, and recurrent hospitalization, but also psychological problems, iatrogenic adverse effects and curtailment in social activities[1]. Because of these multiple factors, many professionals argue how the evaluation of HRQL in HF, as a global comprehensive outcome, deserves special attention[2] now that HF's prevalence has increased due to the interventions' effectiveness and an ageing population[3;4]. In fact, HRQL is considered one of the 3 measureable elements which matter in elderly patients who probably have multiple co-morbidities; being disability and cognitive impairment the other two[5].

The driving force behind this trend may not only have been the recognition that traditional (biological or physiological) endpoints don't reflect the impact of disease and interventions on patients' lives, but the increasing role of patients in decision making, especially in chronic conditions[6].

1. Impact of HF on patients' HRQL

Compared to general population, HF patients present significantly reduced HRQL[7], and they do also when compared to other chronic diseases. Evidence has shown HF as one of the chronic conditions that provokes the highest impairment in physical and mental health, comparable to or greater than arthritis and chronic lung disease[1;8]. That impairment was especially reflected on the physical functioning, role physical and general health scores of the SF-36 when compared to those individuals with no chronic conditions[8]. Another study showed how the majority of patients with HF report problems in mobility (65.9%), usual activities (75.9%) and pain/discomfort (67.8%), based on the EQ-5D[9]. In addition to this already dramatically reduced physical health, patients with advanced HF have presented similar scores in "mental health" to patients with major depression [7].

2. HRQL determinants of patients with HF

As it happens in general population or in some other disorders, lower HRQL in patients with HF has been found to be independently associated with certain sociodemographic characteristics of the individuals, such advanced age or lower socioeconomic status[10-12]. Furthermore, for HF patients some of the HRQL dimensions are also related to other physical or mental clinical indicators.

The associations between HRQL scores and clinical variables or functional capacity measures have been evaluated in several studies, most of which achieved inconclusive results[7]. On one hand, it seems that HRQL decreases as functional level worsens, as happens with the peak oxygen uptake and walking capacity[12;13]. On the other hand, there is some controversy about reliance on the results of physiologic outcomes, such as the ejection fraction[13] which was found to be non-associated with patient reported HRQL[7;11].

HRQL of HF patients has also been independently associated with less common end-points in clinical management such as mental health or social support. Some studies showed how neuroticism[13], emotional distress[14], and depression[15;16] were significant risk factors for worse HRQL. Moreover, as Coelho et al[17] mentioned at their review, some of these studies showed higher incidence of depression in the young, which may be suggesting that depression is due to a larger disparity between perception and expectation of functional status. That depression state may lead physicians to classify these patients as more severely compromised and rate their functional class worse[17].Taking the HRQL as an exposure, Havraneck et al[18] showed how lower reported scores were predictors of developing depressive symptoms.

Finally, some studies reported that lack of social support (informal carer) and social isolation (loss of contacts and activities) are associated with increased risk of mortality among HF patients[12;19].

3. HRQL as an outcome in HF:

Traditional end-points in HF have been: death, recurrent hospitalization, haemodynamic parameters, and exercise and functional capacities. However, patient management nowadays is also carried out through patient reported outcomes (PRO)[8;20]. This holistic perspective is considered very relevant under WHO's current definition of health[21], and also by several professional societies and associations[22;23], which support the fact of going beyond hard-clinical outcomes for treatment or intervention evaluation among HF patients. Several decades ago, Wenger suggested HRQL to be adequate for HF evaluation in certain circumstances[24]:

1) When there is little likelihood of one treatment showing a major improvement in survival over another in a clinical trial. In such a trial HRQL measurement might point towards the choice of therapy showing the greatest benefit for patients.

2) When a treatment is effective in reducing mortality, but has toxic or unacceptable side effects. Quality of life measurement in this case may help physicians and their patients weigh the benefits and risks of such a treatment.

3) When patients are asymptomatic or have mild symptoms, the morbidity and mortality rates are low, and the therapy is long term.

Following these instances, the assessment of HRQL is being increasingly included nowadays in clinical research in HF, and also by the pharmaceutical industry for the evaluation of new treatments[25]. In the same way, HRQL is a major goal in the context of preventive and therapeutic cardiology, from comprehensive management to exercise programs[26;27].

There are many examples of studies (clinical trials or observational) were HRQL has been the principal or a least a secondary end point[25;28] (e.g. studies on beta-blockers[29-32], or exercise -- [33;34]. Some special management programmes have demonstrated to significantly improve HRQL-- home monitoring[4], outpatient HF clinic[35], or brief cognitive therapy interventions[36]. In the same way, resynchronization therapies and pacing devices have also been evaluated in terms of HRQL improvement[37-39]. There are also studies of HRQL as the main outcome in evaluations on sub-populations of HF patients, like iron-deficient patients treated with intravenous ferric carboxymaltose[40] or cognitive therapy[36].

In addition, not only the necessity of enhancing patients' daily functioning and well-being has made HRQL a recognized and relevant outcome, but its prognostic ability has recently been confirmed for predicting HF mortality and hospitalisation[12;41-43], and for working as < significant indicator in advanced heart failure patients[44].

4. HRQL measurement instruments in HF:

The New York Heart Association (NYHA) classification[45] and the walking distance (by asking patients how far they can walk, or performing the Six Minute Walking Test, 6MWT[46]) have historically been the most widely used ways of measuring HF functional status. Nowadays they are still in use in clinical and research settings, despite their well-known limitations[47;48]. Briefly, the NYHA classification (Table 1) helps doctors to place patients into one of four severity categories, based on their impression of limitations related to a patient's ordinary activities, in regards to normal breathing and varying degrees in shortness of breath and/or angina pain during physical activity. Classes range from I (no limitation) to IV (very limited). The walking tests only focus on functional capacity as a reliable procedure to measure exercise tolerance[49;50].

In the late eighties, investigators and clinicians mainly relied on these relatively unsophisticated instruments and ad hoc measures in order to assess functional status of HF patients. However, some authors remarked the need of incorporating new instruments in HF evaluation[47;51], and suggested: a) to incorporate not only a widespread physical but emotional function so as to have comprehensive account of the real limitations experienced by patients; and b) to fulfil a minimum of psychometric criteria.

Since then, a wide range of instruments have been increasingly used to assess the burden of the condition, both in clinical trials and clinical practice[25]. On one hand, cardiologists and researchers started to incorporate generic HRQL instruments to the evaluation of HF patients. A search of the measures used in HF clinical trials before 2000 identified 13 generic instruments[6]: the Sickness Impact Profile, the General Well-Being Index, and the Nottingham Health Profile among other. Nowadays generic HRQL instruments (such as the EQ-5D[52] and the SF-36[53]) are used in patients with HF mainly to be able to compare their burden with other populations (disorders) or to calculate loss of QALYs[37]. They are frequently used together with a disease-specific instrument for evaluating treatment or intervention efficacy under more targeted constructs.

HF-specific instruments were developed at the late eighties and nineties, but their use has been consolidated in the last decade. A recent systematic review[28] identified 5 standardised and structured questionnaires specifically developed for HF patients (Table 2):

- The Minnesota Living with Heart Failure Questionnaire (MLHFQ)[47] was developed in 1987 in the US by Thomas Rector to assess the impact of HF on HRQL. The questionnaire is self-administered and it consists of 21 items: physical domain (8 items), emotional domain (5 items), and 8 items which were not originally factored into physical or emotional domains but are part of the MLHFQ's total score. Response options are presented as a 6-point Likert scale (0-5), from "no impairment" to "very much impairment". The questionnaire is summarised in 3 scores: total (range 0-105, from better to worse HRQL), physical (range 0-40), and emotional (range 0-25).
- The Quality of Life Questionnaire for Severe Heart Failure (QLQ-SHF)[54] was developed in 1987 focusing on patients suffering from sever HF, in Sweden. The questionnaire has a total of 26 items, and 5 specific scores (psychological, physical activity, life-dissatisfaction, and somatic symptoms), plus a total one. Some of the items are answered on a visual analogue scale (VAS), and other using 6-point Likertscales. The QLQ-SHF is self-administered, and higher scores indicate worse HRQL.
- The Chronic Heart Failure Questionnaire (CHFQ)[51] was developed in Canada by Gordon Guyatt and his group, who aimed to measure subjective aspects of health status in clinical trials in heart failure, and specifically longitudinal change over time within patients. The CHFQ is interviewer-administered and has 16 items divided into 3 different domains: dyspnoea (5 items, each patient eliciting 5 specific activities), fatigue (4 items) and emotional (7 items). Each of the

questions uses a 7-point scale for the responses which are simply added together (without weighting) to obtain the score (higher score=better HRQL).

- The Kansas City Cardiomyopathy Questionnaire (KCCQ)[55] aimed to cover the limitations of the already existing specific instruments for HF. It was developed by John Spertus and his group in the US in early 2000 as a self-administered questionnaire. The KCCQ, with a total of 23 items, covers 5 different domains (physical limitation, symptoms, self-efficacy, social limitation, and quality of life); and provides two summaries: clinical and functional status. All the response options are based on Likert scales, of 5, 6, or 7 points; and lower scores indicate worse HRQL.
- The Left Ventricular Dysfunction (LVD-36) questionnaire[56] was developed in UK in 2000, also with the intention of overcoming some of the problems with the existing questionnaires. The LVD-36 is a self-administered 36 item questionnaire, with one single global score. Responses are dichotomous (true or false). True responses are summed and the sum is expressed as a percentage, so that 100 is the worst possible score and 0 the best possible score.

The development process of these instruments included literature reviews and generic HRQL instruments, together with patient and expert panels in the most recent cases. Item reduction was based on factor analysis (MLHFQ, QLQ-SHF and LVD-36), or on patient-rated clinical impact (CHFQ and KCCQ). All questionnaires are multidimensional, except the LVD-36, and include at least a physical domain and a domain assessing emotional/psychological aspects. Only the CHFQ presents an individualized patient profile, addressing some questions under specific activities chosen by each patient.

Regarding reliability, overall scores (total or global) and scores on the physical dimension showed the highest Cronbach's alpha in most of the questionnaires (ranging from 0.81 to 0.95). The intraclass correlation coefficients that authors found in the different publications were high and similar across questionnaires and domains (ranging from 0.78 to 0.95). For the QLQ-SHF, only test-retest correlation coefficients were reported.

The SF-36, the Six Minute Walking Test (6MWT), and the NYHA class were the common measures to be comparable across all questionnaires for their construct validity assessment. 'Physical Functioning' dimension of the SF-36 was found to be strongly associated with the physical domains of the MLHFQ and the KCCQ, and with the total score of the LVD-36 (r=0.65-0.84). The CHFQ showed the highest correlations with the Six Minute Walking Test (6MWT) (0.6-0.7); and so did the KCCQ physical domain with NYHA class (0.65), while the CHFQ domains of fatigue and dyspnoea presented the lowest correlations with this functional capacity classification (0.19 and 0.22, respectively). The total scores of the MLHFQ and LVD-36 presented the highest correlations with the 'Social Functioning' dimension of the SF-36 (r=0.70). Validity criteria used in the QLQ-SHF evaluation had limited comparability with those of the other questionnaires.

Responsiveness of the MLHFQ was evaluated at the systematic review conducting meta-analyses that also incorporate information from effectiveness studies and clinical trials. Findings showed Effect Size (ES) coefficients close to 0.5 (moderate effect) for exercise programs and beta-blockers. Special management programs, angiotensin-converting-enzyme inhibitors (ACEI), and pacing devices produced ES of approximately 0.8 (a large effect size). The coefficients shown by the physical and emotional domains in the special management groups were quite lower (0.3-0.2). The highest ES found for the CHFQ were observed in outpatients who rated their overall health as “very much better” after the intervention (ES 1.2 - 1.9). The KCCQ has shown to detect changes after discharge (ES 0.6 - 3.2), while the QLQ-SHF produced low and similar ES for beta-blockers and ACEIs (estimates between –0.1 and 0.3).

The patient’s perception of HF’s impact on his or her HRQL is a unique construct and must be measured separately from other domains, such as functional status or clinical signs and symptoms. Despite all the information available regarding the HRQL instruments for patients suffering from HF, there is still the need of further research involving the adequacy and performance of some of the HF-specific HRQL instruments based on newer, more rigorous methodology, following the current standards (e.g. appropriate sample sizes, designs…). Moreover, if a methodological goal for HF clinical research is to construct a toolkit of outcome measures, it would be most efficient to have each outcome measurement make a unique contribution to the whole and not duplicate what other measures accomplish.

Table 1. Description of the New York Heart Association functional classification[45].

NYHA classes	Patients' functional limitations (to be judged by clinicians)
I	Ordinary physical activity does not cause undue fatigue, palpitation, dyspnoea or anginal pain.
II	Comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnoea or anginal pain.
IIIa-IIIb	Comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, dyspnoea or anginal pain.
IV	Symptoms of cardiac insufficiency or of the angina syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.

Table 2. Disease-specific HRQL questionnaires for heart failure patients[28].

Questionnaire, Acronym Authors (year publ.)	**Mode of administration (time frame)**	**Response options**	**Domains**	**(No. items)**	**Range of scores**
Minnesota Living with Heart Failure Questionnaire, MLHFQ Rector et al (1987)	Self-administered (last month)	6p Likert scale (0-5)	Physical Emotional Total	(8) (5) (21)	0-105 Best to worst
Quality of Life Questionnaire for Severe Heart Failure, QLQ-SHF Wiklund et al (1987)	Self-administered (- -)	VAS 6p Likert scale	Psychological Physical activity Life-dissatisfaction Somatic Symptoms Total	(7) (7) (5) (7) (26)	0-130 Best to worst
Chronic Heart Failure Questionnaire, CHFQ Guyatt et al (1989)	Interviewer-administered (last two weeks)	7p Likert scale (1-7)	Dyspnoea Fatigue Emotional Total	(5) (4) (7) (16)	16-112 Worst to best
Kansas City Cardiomyopathy Questionnaire, KCCQ Green et al (1999)	Self-administered (last two weeks)	5-6-7p Likert scale (1-5/6/7)	Physical limitation Symptoms Self-efficacy Social limitation Quality of Life Total	(6) (8) (2) (4) (3) (23)	0-100 Worst to best
Left Ventricular Dysfunction Questionnaire-36, LVD O'Learly et al (1998)	Self-administered (*these days*)	Dichotomous (true, false)	Total	(36)	0-100 Worst to best

Cross-references

EMPRO, Generic, Disease-specific, QALY, VAS, Individualized PRO, conceptual model, reliability, Cronbach’s alpha, reproducibility, ICC, construct validity, responsiveness, effect size coefficient

References:

1. Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry SD, McGlynn EA, Ware JE, Jr. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. *JAMA*. 1989;262:907-913.

2. Jaarsma T, Lesman-Leegte I. The value of the quality of life. *Eur J Heart Fail*. 2010;12:901-902.

3. Konstam MA. Heart failure training: a call for an integrative, patient-focused approach to an emerging cardiology subspecialty. *J Am Coll Cardiol*. 2004;44:1361-1362.

4. Konstam MA. Home monitoring should be the central element in an effective program of heart failure disease management. *Circulation*. 2012;125:820-827.

5. Taylor J, Stott DJ. Chronic heart failure and cognitive impairment: co-existence of conditions or true association? *Eur J Heart Fail*. 2002;4:7-9.

6. al-Kaade S, Hauptman PJ. Health-related quality of life measurement in heart failure: challenges for the new millennium. *J Card Fail*. 2001;7:194-201.

7. Juenger J, Schellberg D, Kraemer S, Haunstetter A, Zugck C, Herzog W, Haass M. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. *Heart*. 2002;87:235-241.

8. Alonso J, Ferrer M, Gandek B, Ware JE, Aaronson NK, Mosconi P, Rasmussen NK, Bullinger M, Fukuhara S, Kaasa S, Leplège A, and the IQOLA Project Group. Health-related quality of life associated with chronic conditions in eight

countries. Results from the International Quality of Life Assessment (IQOLA) project. *Qual Life Res*. 2004;13:283-298.

9. Calvert MJ, Freemantle N, Cleland JG. The impact of chronic heart failure on health-related quality of life data acquired in the baseline phase of the CARE-HF study. *Eur J Heart Fail*. 2005;7:243-251.
10. Philbin EF, Dec GW, Jenkins PL, DiSalvo TG. Socioeconomic status as an independent risk factor for hospital readmission for heart failure. *Am J Cardiol*. 2001;87:1367-1371.
11. Rosen RC, Contrada RJ, Gorkin L, Kostis JB. Determinants of perceived health in patients with left ventricular dysfunction: a structural modeling analysis. *Psychosom Med*. 1997;59:193-200.
12. Iqbal J, Francis L, Reid J, Murray S, Denvir M. Quality of life in patients with chronic heart failure and their carers: a 3-year follow-up study assessing hospitalization and mortality. *Eur J Heart Fail*. 2010;12:1002-1008.
13. Westlake C, Dracup K, Creaser J, Livingston N, Heywood JT, Huiskes BL, Fonarow G, Hamilton M. Correlates of health-related quality of life in patients with heart failure. *Heart Lung*. 2002;31:85-93.
14. Lainscak M, Keber I. Patient's view of heart failure: from the understanding to the quality of life. *Eur J Cardiovasc Nurs*. 2003;2:275-281.
15. Rumsfeld JS, Havranek E, Masoudi FA, Peterson ED, Jones P, Tooley JF, Krumholz HM, Spertus JA. Depressive symptoms are the strongest predictors of short-term declines in health status in patients with heart failure. *J Am Coll Cardiol*. 2003;42:1811-1817.
16. Gottlieb SS, Khatta M, Friedmann E, Einbinder L, Katzen S, Baker B, Marshall J, Minshall S, Robinson S, Fisher ML, Potenza M, Sigler B, Baldwin C, Thomas SA. The influence of age, gender, and race on the prevalence of depression in heart failure patients. *J Am Coll Cardiol*. 2004;43:1542-1549.
17. Coelho R, Ramos S, Prata J, Bettencourt P, Ferreira A, Cerqueira-Gomes M. Heart failure and health related quality of life. *Clin Pract Epidemiol Ment Health*. 2005;1:19.

18. Havranek EP, Spertus JA, Masoudi FA, Jones PG, Rumsfeld JS. Predictors of the onset of depressive symptoms in patients with heart failure. *J Am Coll Cardiol*. 2004;44:2333-2338.

19. Murberg TA, Bru E. Social relationships and mortality in patients with congestive heart failure. *J Psychosom Res*. 2001;51:521-527.

20. Adamson PB, Abraham WT, Love C, Reynolds D. The evolving challenge of chronic heart failure management: a call for a new curriculum for training heart failure specialists. *J Am Coll Cardiol*. 2004;44:1354-1357.

21. Testa MA, Simonson DC. Assesment of quality-of-life outcomes. *N Engl J Med*. 1996;334:835-840.

22. Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJ, Ponikowski P, Poole-Wilson PA, Stromberg A, van Veldhuisen DJ, Atar D, Hoes AW, Keren A, Mebazaa A, Nieminen M, Priori SG, Swedberg K. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the diagnosis and treatment of acute and chronic heart failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). *Eur J Heart Fail*. 2008;10:933-989.

23. Cohn J, Cleland JG, Lubsen J, Borer JS, Steg PG, Perelman M, Zannad F. Unconventional end points in cardiovascular clinical trials: should we be moving away from morbidity and mortality? *J Card Fail*. 2009;15:199-205.

24. Wenger NK. Quality of life: can it and should it be assessed in patients with heart failure? *Cardiology*. 1989;76:391-398.

25. Anand IS, Florea VG, Fisher L. Surrogate end points in heart failure. *J Am Coll Cardiol*. 2002;39:1414-1421.

26. Sullivan M, Levy WC, Russo JE, Spertus JA. Depression and health status in patients with advanced heart failure: a prospective study in tertiary care. *J Card Fail*. 2004;10:390-396.

27. Wielenga RP, Erdman RA, Huisveld IA, Bol E, Dunselman PH, Baselier MR, Mosterd WL. Effect of exercise training on quality of life in patients with chronic heart failure. *J Psychosom Res*. 1998;45:459-464.

28. Garin O, Ferrer M, Pont A, Rue M, Kotzeva A, Wiklund I, Van GE, Alonso J. Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. *Qual Life Res*. 2009;18:71-85.

29. Fung JW, Chan SK, Yeung LY, Sanderson JE. Is beta-blockade useful in heart failure patients with atrial fibrillation? An analysis of data from two previously completed prospective trials. *Eur J Heart Fail*. 2002;4:489-494.

30. Edes I, Gasior Z, Wita K. Effects of nebivolol on left ventricular function in elderly patients with chronic heart failure: results of the ENECA study. *Eur J Heart Fail*. 2005;7:631-639.

31. Gonzalez-Juanatey JR, Alegria EE, Garcia S, V, Perez OG, Ruiz Ros JA, Espinosa Caliani JS, Anguita SM. [Use of bisoprolol in heart failure. The BISOCOR observational study]. *Rev Esp Cardiol*. 2003;56:873-879.

32. Little WC, Wesley-Farrington DJ, Hoyle J, Brucks S, Robertson S, Kitzman DW, Cheng CP. Effect of candesartan and verapamil on exercise tolerance in diastolic dysfunction. *J Cardiovasc Pharmacol*. 2004;43:288-293.

33. Harris S, LeMaitre JP, Mackenzie G, Fox KA, Denvir MA. A randomised study of home-based electrical stimulation of the legs and conventional bicycle exercise training for patients with chronic heart failure. *Eur Heart J*. 2003;24:871-878.

34. Dall'Ago P, Chiappa GR, Guths H, Stein R, Ribeiro JP. Inspiratory muscle training in patients with heart failure and inspiratory muscle weakness: a randomized trial. *J Am Coll Cardiol*. 2006;47:757-763.

35. Hole T, Grundtvig M, Gullestad L, Flonaes B, Westheim A. Improved quality of life in Norwegian heart failure patients after follow-up in outpatient heart failure clinics: results from the Norwegian Heart Failure Registry. *Eur J Heart Fail*. 2010;12:1247-1252.

36. Dekker RL, Moser DK, Peden AR, Lennie TA. Cognitive therapy improves three-month outcomes in hospitalized patients with heart failure. *J Card Fail*. 2012;18:10-20.

37. Cleland JG, Calvert MJ, Verboven Y, Freemantle N. Effects of cardiac resynchronization therapy on long-term quality of life: an analysis from the CArdiac Resynchronisation-Heart Failure (CARE-HF) study. *Am Heart J*. 2009;157:457-466.

38. Molhoek SG, Bax JJ, van Erven L, Bootsma M, Boersma E, Steendijk P, van der Wall EE, Schalij MJ. Effectiveness of resynchronization therapy in patients with end-stage heart failure. *Am J Cardiol*. 2002;90:379-383.

39. Auricchio A, Stellbrink C, Butter C, Sack S, Vogt J, Misier AR, Bocker D, Block M, Kirkels JH, Kramer A, Huvelle E. Clinical efficacy of cardiac resynchronization therapy using left ventricular pacing in heart failure patients stratified by severity of ventricular conduction delay. *J Am Coll Cardiol*. 2003;42:2109-2116.

40. Comin-Colet J, Lainscak M, Dickstein K, Filippatos GS, Johnson P, Luscher TF, Mori C, Willenheimer R, Ponikowski P, Anker SD. The effect of intravenous ferric carboxymaltose on health-related quality of life in patients with chronic heart failure and iron deficiency: a subanalysis of the FAIR-HF study. *Eur Heart J*. 2012.

41. Zuluaga MC, Guallar-Castillon P, Lopez-Garcia E, Banegas JR, Conde-Herrera M, Olcoz-Chiva M, Rodriguez-Pascual C, Rodriguez-Artalejo F. Generic and disease-specific quality of life as a predictor of long-term mortality in heart failure. *Eur J Heart Fail*. 2010;12:1372-1378.

42. Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS, Peterson ED, Masoudi FA, Krumholz HM, Havranek EP, Conard MW, Williams RE. Health status identifies heart failure outpatients at risk for hospitalization or death. *J Am Coll Cardiol*. 2006;47:752-756.

43. Rodriguez-Artalejo F, Guallar-Castillon P, Pascual CR, Otero CM, Montes AO, Garcia AN, Conthe P, Chiva MO, Banegas JR, Herrera MC. Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. *Arch Intern Med*. 2005;165:1274-1279.

44. Sullivan MD, Levy WC, Russo JE, Crane B, Spertus JA. Summary health status measures in advanced heart failure: relationship to clinical variables and outcome. *J Card Fail*. 2007;13:560-568.

45. Rossi G. [Nomenclature and diagnostic criteria in cardiology set by the New York Heart Association. Considerations on the 6th edition]. *Cuore Circ*. 1967;51:287-293.

46. Demers C, McKelvie RS, Negassa A, Yusuf S. Reliability, validity, and responsiveness of the six-minute walk test in patients with heart failure. *Am Heart J*. 2001;142:698-703.

47. Rector TS, Kubo SH, Cohn JN. Patients' self-assessment of their congestive heart failure. Part 2: content, reliability and validity of a new measure, The Minnesota Living with Heart Failure Questionnaire. *Heart Failure*. 1987;198-209.

48. Raphael C, Briscoe C, Davies J, Ian WZ, Manisty C, Sutton R, Mayet J, Francis DP. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. *Heart*. 2007;93:476-482.

49. Strijbos JH, Postma DS, van AR, Gimeno F, Koeter GH. A comparison between an outpatient hospital-based pulmonary rehabilitation program and a home-care pulmonary rehabilitation program in patients with COPD. A follow-up of 18 months. *Chest*. 1996;109:366-372.

50. Du H, Newton PJ, Salamonson Y, Carrieri-Kohlman VL, Davidson PM. A review of the six-minute walk test: its implication as a self-administered assessment tool. *Eur J Cardiovasc Nurs*. 2009;8:2-8.

51. Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. *J Gen Intern Med*. 1989;4:101-107.

52. Flynn KE, Lin L, Moe GW, Howlett JG, Fine LJ, Spertus JA, McConnell TR, Pina IL, Weinfurt KP. Relationships between changes in patient-reported health status and functional capacity in outpatients with heart failure. *Am Heart J*. 2012;163:88-94.

53. Naveiro-Rilo JC, Diez-Juarez MA, Flores-Zurutuza L, Rodriguez-Garcia MA, Rebollo-Gutierrez F, Romero BA. [Quality of life in patients with heart failure: the perspective of primary health care]. *Gac Sanit*. 2012;26:436-443.

54. Wiklund I, Lindvall K, Swedberg K, Zupkis RV. Self-assessment of quality of life in severe heart failure. An instrument for clinical use. *Scand J Psychol*. 1987;28:220-225.

55. 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;35:1245-1255.

56. O'Leary CJ, Jones PW. The left ventricular dysfunction questionnaire (LVD-36): reliability, validity, and responsiveness. *Heart*. 2000;83:634-640.

4.1 Peer-reviewed papers

a) Paper 1:

Garin O, Ferrer M, Pont A, Rué M, Kotzeva A, Wiklund I, Van Ganse E, Alonso J. *Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses.* Qual Life Res. 2009; 18(1): 71-85

Garin O, Ferrer M, Pont A, Rué M, Kotzeva A, Wiklund I, Van Ganse E, Alonso J. Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. Supplementary material. Qual Life Res. 2009; 18(1): 71-85.

b) Paper 2:

Garin O, Soriano N, Ribera A, Ferrer M, Pont A, Alonso J, Permanyer G, on behalf of the HF-QoL Group*. *Validación de la versión española del Minnesota Living with Heart Failure Questionnaire [Validation of the Spanish version of Minnesota Living with Heart Failure Questionnaire]*. Rev Esp Cardiol. 2008, 61: 251-259

c) Paper 3:

Garin O, Ferrer M, Pont A, Wiklund I, Van Ganse E, Vilagut G, Almansa J, Ribera A, Alonso J. *Evidence on the global measurement model of the Minnesota Living with Heart Failure Questionnaire.* (under review)

Evidence on the global measurement model of the *Minnesota Living with Heart Failure Questionnaire*.

Olatz Garin MPH[1,2,3]; Montse Ferrer MD, PhD[1,2,4]; Àngels Pont Dip Stat[1]; Ingela Wiklund MD, PhD[5]; Eric Van Ganse MD, PhD[6]; Gemma Vilagut BSc[1,2]; Josué Almansa BSc, PhD[1,2]; Aida Ribera PhD[7,2] ; Jordi Alonso MD, PhD[1,2,3].

[1] Health Services Research Group, IMIM-Research Institute Hospital del Mar, Spain

[2] CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.

[3] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[4] Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

[5] GlaxoSmithKline Research & Development Limited, Greenford, UK

[6] Unité de Pharmaco-épidémiologie, CHU-Lyon, France

[7] Unitat d'Epidemiologia, Servei de Cardiologia, Hospital Universitari Vall d'Hebron, Barcelona, Spain

Corresponding author: Montse Ferrer ; IMIM-Research Institute Hospital del Mar, Doctor Aiguader 88; 08003 Barcelona, Spain. Phone. +34 93 3160 740; Fax: +34 93 3160 797; e-mail address: mferrer@imim.es

Words: 3,173

Abstract

Words: 246

Purpose**:** Both in clinical and research settings, the Minnesota Living with Heart Failure Questionnaire (MLHFQ) is the most widely used measure. Nevertheless, its measurement model has never been confirmed. This study aimed to cover that gap within a large international sample. Methods**:** Data from 8 studies was merged (3,847 patients with heart failure) from 21 countries. Common variables included: MLHFQ scores, functional capacity, cardio-vascular risk factors and patient's socio-demographic characteristics. MLHFQ's measurement model was assessed by exploratory and confirmatory factor analyses (EFA-CFA). Reliability of the MLHFQ scores was evaluated through Cronbach's alpha coefficient and MLHFQ's ability to differentiate among known-groups was assessed through severity levels. Results: Findings from the EFA and CFA suggested the MLHFQ total and domain specific scores to stand under a bi-factor model. The physical and emotional scores were supported within the sample, as also was the original total score. Furthermore, a third factor had risen regarding social indicators. Reliability coefficient reached 0.9 for almost all physical and total scores. All the MLHFQ mean scores showed ability to differentiate among functional capacity groups, with most of the Effect Size coefficients reaching 0.8. Conclusions**:** Beyond the adequate reliability and validity shown by the MLHFQ scores in the different country versions, results confirmed for the first time the unidimensionality of the most used score in HF patients, the MLHFQ total score. Moreover, the identified Social domain may be considered hereafter when assessing these patients' HRQL, especially as a relevant outcome with regard to their management.

Keywords: Patient Reported Outcome, Heart Failure, Measurement Model; Factor analysis; Metric Properties

INTRODUCTION

The assessment of Health Related Quality of Life (HRQL) in Heart Failure (HF) patients deserves special attention since both, interventions' effectiveness and population ageing have increased[1, 2]. Morbidity in HF patients not only comprises recurrent hospitalization, impaired exercise capacity and physical symptoms (short of breath and fatigue), but psychological problems, iatrogenic adverse effects and curtailment in social activities. The improvement of all these dimensions of patients' life is becoming a priority for cardiologists[3], and has increased the role of patient reported outcomes (PRO) in clinical research and practice[4,5]. In addition, HRQL has recently been shown to be a good predictor of mortality and hospitalisation in patients with HF[6-8].

A systematic review, published in 2009[9], confirmed the availability of at least five standardised and structured disease-specific instruments for measuring HRQL in HF patients: the Minnesota Living with Heart Failure Questionnaire(MLHFQ)[10], the Chronic Heart Failure Questionnaire (CHFQ)[11], the Quality of Life Questionnaire for Severe Heart Failure (QLQ-SHF)[12], the Kansas City Cardiomyopathy Questionnaire (KCCQ)[13] and the Left Ventricular Dysfunction (LVD-36) questionnaire[14]. On the whole, the review suggested that most of the questionnaires studied met minimum psychometric criteria for assessing HRQL. Nonetheless, the MLHFQ was shown to be the most commonly used disease-specific measure of HRQL in patients with HF (estimated more than 100 publications the last 20 years). In fact, there are at least 34 linguistic versions of the MLHFQ[15]. The original USA-English version was developed by Thomas Rector in 1987 to assess the impact of HF on HRQL[10]. The questionnaire consists of 21 items which were proposed, from the beginning, to conform a total score. Moreover, a physical and an emotional domain have typically been also calculated, with 8 and 5 of the 21 items respectively. The other 8 items (until 21) are summed up only for the total score calculation.

The MLHFQ total score was conceptually designed to be a summary of all the important issues for HF patients' HRQL. Its world wide professional use[16], and the fact that more actual instruments' development has been inspired on the MLHFQ indicators [13], supports the adequacy of its content or underlying latent construct. However, the unidimensionality of this total score has never been methodologically confirmed. In addition, even the structure of the MLHFQ in terms of specific domains has been explore in some studies[17-20], there is no evidence about the whole measurement model by factor analysis.

Consequently, the aim of the present study was to verify the unidimensionality of the MLHFQ total score by exploring and confirming the questionnaire's global measurement model; and to evaluate the reliability and validity of the MLHFQ in 21 country versions.

METHODS

The Minnesota Living with Heart Failure Questionnaire (MLHFQ)

The MLHFQ is self-administered, and the response options of its 21 items are presented as a 6-point scale (0-5), from “no impairment” to “very much impairment”. As mentioned before, the questionnaire is summarised in 3 scores: total (range 0-105, from better to worse HRQL), physical (range 0-40), and emotional (range 0-25). As proposed by the original version’s authors, these scores are computed summing up the corresponding items’ responses; an alternatively mean imputation of missing values was conducted if missing items were less than half of those used to compute the scale [21].

The majority of the MLHFQ’s linguistic adaptations had been conducted following the standard forward and backward translation process, information on metric properties in only available for some them .

Study design

The evaluation of the MLHFQ properties was one of the aims of the International Quality of Life Outcomes Database” (IQOD) project[22], founded by the European Commission. To reach it, cross-sectional data from three observational studies and five clinical trials were merged in a common database, including 3,847 patients with HF from 21 countries (Table 1). Common variables to all studies included: MLHFQ responses, functional capacity measured by the New York Heart Association class (NYHA)[23], and cardio-vascular risk factors such as the body mass index (BMI) and smoking status; together with their socio-demographic characteristics. These variables were compared across countries using ANOVA and Bonferroni post hoc comparisons for continuous variables and Chi-square test for categorical ones.

Measurement model:

The factorial structure of the MLHFQ was assessed on the international sample which was randomly divided into 2 sub-sample to this purpose. One for conducting the exploratory factor analysis (EFA) which results were afterwards tested via confirmatory factor analysis (CFA) on the other random sub-sample[24]; both conducting category factor analysis. In the EFA, the most appropriate model (description of number of factors and item location) was selected following 2 main criteria: a) non-negative residual variances, and b) factor loadings near or above 0.4. To test the factors identified at the EFA (specific domains within the MLHFQ), and also the existence of a general factor (the MLHFQ total score), a bi-factor model structure was imposed in the CFA. This model allows all items to load in a general factor, independently of whether they are part of one of the specific domains or not. This premise was fundamental to confirm the MLHFQ global measurement model, as the questionnaire contains several items that only account for the total score. The CFA was performed with Weighted Least Squares Method, and its goodness of fit was assessed by the following indices:

a) Confirmatory Fit Index (CFI) and Tucker-Lewis Index (TLI) which should be above 0.95; and b) Root Mean Square Error of Approximation (RMSEA) which indicates adequate fit if it is below 0.08. Both the EFA and the CFA were conducted with MPlus 4.2[25].

Reliability and validity of the MLHFQ scores

The MLHFQ constructs confirmed at the previous step were evaluated in terms of reliability and validity following basic recommended methodology[26]. These assessments were conducted for the *overall sample* (the 3,847 individuals), and for the *particular case* of each country. Distribution of scores was evaluated in terms floor and ceiling effects (percent of patients with worst and best possible score, respectively). Cronbach's alpha coefficient[27] was calculated for the different constructs to assess internal consistency.

Know groups were defined by the NYHA classification in order to assess the MLHFQ discriminant validity. Different classes were collapsed in NYHA classes I-II and III-IV (a restriction already present in several of the pooled studies). As a first step, it was explored if all the countries presented similar MLHFQ's scores within the same NYHA group (ANOVA with post hoc pair wise test using Bonferroni's method to adjust for multiple comparisons). After testing this homogeneity, MLHFQ' scores between NYHA groups were compared by t-test; and their magnitude by effect size coefficient (ES=score mean difference / pooled SD). ES from 0.2 to < 0.5 was considered small, while ES from 0.5 to 0.8 and above 0.8 were respectively considered moderate, and large[28].

RESULTS

- TABLE 1, ABOUT HERE-

Mean age of the patients included in the *overall sample* was 63.5 (12.1), from 52.8 (12.6) in Brazil to 69.9 (7.3) in Sweden (Table 1). Most of the patients were male (74.2% *overall sample*), with the exception of Switzerland (47.6%). Mean BMI ranged from 29.8(6.5) in the US to 25.9(4.7) in Brazil, being 27.3(4.7) kg/m^2 at the *overall sample*. On average, 23.9% of patients were smokers, but the percentage range from 9.3% in Israel to 95.2% in Switzerland. The two NYHA groups were similarly represented at the overall sample 49.1% for I-II and 50.9% for III-IV; but had heterogeneous presence in across countries (i.e. 86.6% *vs* 13.4% in Brazil and 0% vs 100 in Switzerland).

At the EFA, the 3-factor solution with quartimin rotation (Table 2) showed better results than the 2 and 4 factor outputs. This structure was fixed at the CFA, where the model presented excellent goodness of fit coefficients: CFI=0.949, TLI=0.988 and RMSEA=0.065. This measurement model had 3 specific factors and a general factor (Figure 1). Factor 1 included 8 items, like had the original physical score, but included item 1 (*swelling in your ankles, legs*) instead of item 7 (*relating to or doing things with your friends or family difficult*), based on the statistically significant loading from the CFA (arrows on Figure 1). The emotional domain (Factor 2) was constituted by 5 items, the same ones that have been considered to be strictly emotional when scoring the MLHFQ. Factor 3 included 4 items, which could be considered as social indicators: *working to earn a living*; *recreational pastimes, sports or hobbies*; *sexual activities*; and *money for medical care*. Therefore still 5 items did not clearly load on any factor (*relating to or doing things with your friends or family difficult, eat less of foods you like, stay in a hospital*, and *side effects from medications*). Finally the most important result from the CFA was the confirmation of a single one-dimension latent construct (could be called HRQL-in-HF) for the total score, that included the 21 items. In fact, after controlling for the general factor, the specific domains provided little additional measurement precision.

-TABLE 2 AND FIGURE 1, ABOUT HERE-

Regarding the performance of this scores at the global international sample and stratifying by country, floor and ceiling effects were negligible, only in 3 *particular cases*, the best possible score was reached for more than 20% of the patients (Table 3-marked bold). At the *overall sample*, Cronbach's alpha coefficient was 0.9, 0.84, 0.72, and 0.92 for the physical, emotional, social and total scores, respectively. This estimator was in all countries *particular cases* above 0.8 for the physical score and above 0.7 for the emotional score. Cronbach's alpha for the social score was lower (0.40-0.82. The internal consistency coefficient for the MLHFQ total score almost reached 0.9 in all *particular cases*.

MLHFQ's scores means by NYHA group were similar across countries (data not shown), with the exception of the social scores of Polish patients (NYHA I-II vs III-IV, $p<0.05$). Besides this homogeneity by functional group, the MLHFQ scores were lower in patients at NYHA classes I-II than in those at classes III-IV ($p<0.001$) (Figure 2), and presented ES above 0.5, being especially high for the physical score (ES=0.95). Consequently, country stratification was only made for this score, and the magnitude of the differences between NYHA groups remained (Figure 3), with the exception of those from Brazil, Finland, and Hungary (ES from 0.39 to 0.47). Switzerland was not included in these analyses due to the lack of patients in NYHA classes I-II.

DISCUSSION

Results have confirmed for the first time the original measurement model of the MLHFQ, giving more lights to the possibility of assessing a third specific domain concerning patients' social life. Furthermore, the reliability and validity of the MLHFQ scores have been shown to be adequate among different country versions. These findings, together with the already proved simplicity, clarity, and good performance of the MLHFQ, definitely support the use of the questionnaire as an outcome for HF patients, as well as improve the knowledge about it, hence its interpretation.

Nevertheless, the study presents some limitations which deserve consideration. The first, and most relevant from authors perspective, is the lack of structural invariance assessment across country versions[29]. This evaluation was limited by the size of some countries' sample. The inclusion of as many individuals as possible, to assure enough sample for exploratory and confirmatory factor analysis and heterogeneity on HF patients characteristics, was prioritized. Second was the restricted validity assessment of the MLHFQ scores, due to the nature of the sample. The pooled studies had different primary aims, and presented only few common co-variables (not including a common walking test or generic HRQL measure). Moreover, the NYHA classification is a controversial measure, of which reliability may vary across countries. However, the collapsing strategy applied in this study, may had counteracted those problems, as the main discordances are typically around classes IIIa, IIIb, and IV[30]. Also regarding MLHFQ's psychometric properties, follow-up data was not available for reproducibility or for responsiveness evaluation. Finally, to take advantage of the international sample, authors would have liked to conduct the assessment of differential item functioning (DIF) across countries. However, the small size of some samples and the absence of common patients' characteristics among the different countries studies limited these analyses. Further country-specific works may be the most advisable designs to evaluate possible DIF between the original and each adapted version.

The MLHFQ's measurement model has been explored for the first time using EFA in a large international and heterogeneous sample. This methodology has come up with a proposal on specific domains covered by the questionnaire almost equal to the originally suggested by Rector. Only two differences rose: a minor modification regarding the content of the physical score, and the possibility of computing a social score. The original physical score included the item "*relating to or doing things with your friends or family*". However, the model showed at this study excludes this item from the physical score and adds one that originally did not belong to any specific domain: *swelling in the ankles or legs*. This slightly change may add face validity to the physical score at the same time that allows the equilibrium of the coefficients that depend on the number of items (thus, both the original domain and the suggested here contain 8 items).

Another issue is the understandable concerns with regard to the third factor beyond which suggesting the use of this new specific MLHFQ social score. This decision was mainly made up based on the better fit of the model compared to the 2 factor structure, in first place, followed by the conceptual point of view, the confirmation at the CFA, and the possibility of working and exploring this domain of HF patients in further studies (without damaging or changing any of the other well established scores). Nowadays, individuals, especially elder ones, even being supposed to live better than those in previous decades, report worse HRQL[31]. Consequently, chronic patients give more importance to social issues and to lack of side effects from being under treatment than used to, apart from the relevance of physical and emotional health. Also for that reason, other disease specific instruments for HF have included social and other construct's summaries or scores from their initial development stages[13].

The two suggestions emerging from this study regarding the specific domains (one item replacement and the calculation of a third specific score) do not represent a trouble for allowing the comparison with already published data, nor prevent comparisons on follow-up studies. Hence, even there should be no reason for avoiding their implementation hereafter, these recommendations could be dismissed as already happened to other studies' results[17, 20]. The modification of other instruments has been proposed without success, arguing poor methodology of the studies from where the changes are suggested[32], or limited dissemination of results, as well as a common general resistance to change established approaches. In the concrete case of the MLHFQ, another reason for the scarce dissemination of recommendations made on the specific domains might be the wide use of its total score, exclusive in many clinical publications that present HRQL as one more endpoint among many. Here, however, comes the importance of the MLHFQ total score. In this study two main characteristics have made possible to address the crucial issue of the total score unidimensionality (i.e. the real existence of one-dimensional latent construct underlying the MLHFQ total score) On one side, the construct, that may be called "HRQL-in-HF", has been confirmed through a bi-factor model[33], which has allowed single items to directly load, both, in the general factor (or total score) as well as in one of the specific domains. In a second order factor (a common model used in CFA of PRO measures) the general factor is constructed through its correlations with the specific domains (e.g. physical, emotional, social,…) which are considered first order factors, and all the items must load in one or other of those specific factors[34]. In the MLHFQ measurement model (the original or the one presented in this study) there are some items that made up for the total score without belonging to any specific domain. Consequently, there was a need of applying an updated factorial analysis model to support the idea of the MLHFQ total score being an overall measure of quality of life in HF patients.

On the other side, the large sample studied has also contributed to the possibility of confirming the total score. In fact, by random split, the sample allowed, in first term, to perform an exploratory factor analysis with MPlus (which was never published before), and to confirm the whole measurement model afterwards. Moreover, this sample contained a wide spectrum regarding the MLHFQ country versions, and did not shown an heterogeneous patter of missing values (data not shown).

The fact that the structure has been validated within an international sample may constitute the first step to a cross-cultural validity of the MLHFQ. Among all the MLHFQ existing country versions, there is published evidence of their validation only for a few [35-41]. The majority of the available information regarding the MLHFQ performance is on the original version and under classical test theory. The lack of a minimum validation across MLHFQ country versions is in contrast with the wide acceptance and use of this instrument in international clinical trials, effectiveness studies, and more recently, in clinical practice. As mentioned previously, this study did not addressed the issue of cross-cultural validation as it should be understood, but it shows results about what for most MLHFQ country versions may be its first psychometric evaluation.

In general it could be stated that all country versions had showed capacity to capture the entire range of HRQL impairment suffered by HF patients (low percentages of floor and ceiling effects for the 4 MLHFQ scores in all versions). Regarding both, internal consistency and construct validity, results confirmed what was already found at other studies mainly for the original version[10]. The total and the physical scores seem to be able to compare individuals, presenting the highest Cronbach's alpha coefficients (near the standard of 0.9). Moreover, these two scores also showed the best ability to differentiate patients' functional capacity, being moderately associated with NYHA classes (mean r= 0.6).

Conclusions

Findings from this study support the MLHFQ for assessing HRQL of HF patients, and specially the robustness of its total score. In contrast with its worldwide common use, this is the first time that MLHFQ model has been confirmed It has been done in a large international sample and through an up-to-date methodology, 20 years after its initial development. The MLHFQ scores have shown adequate reliability and validity among different country versions; hence their use has been reinforced compared to what was known from most of them.

ACKNOWLEDGEMENTS

The authors are very grateful to the members of the IQOD Project, managed by the MAPI Research Institute; and to all the institutions and patients that have made it possible. Furthermore, we would like to thank JM Valderas for his special support on the initial preparation of this article; and to Carlos Garcia-Forero for his contribution to the interpretation of results.

This study was supported by: European Commission contract QLRI-CT-2000-00551 (IQOD), CIRIT (2001SGR 00405), SAF2002-11009-E, and , the CIBER de Epidemiología y Salud Pública, CIBERESP.

References

1) Jaarsma T, Lesman-Leegte I. (2010) The value of the Quality of Life. Eur J Heart Fail, 12:901-2
2) Konstam MA. (2004) Heart failure training: a call for an integrative, patient-focused approach to an emerging cardiology subspecialty. J Am Coll Cardiol, 44:1361-1362.
3) Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJ, Ponikowski P, Poole-Wilson PA, Strömberg A, van Veldhuisen DJ, Atar D, Hoes AW, Keren A, Mebazaa A, Nieminen M, Priori SG, Swedberg K. (2008) ESC Committee for Practice Guidelines (CPG). Eur J Heart Fail, 10:933-89.
4) Adamson PB, Abraham WT, Love C, Reynolds D. (2004) The evolving challenge of chronic heart failure management: a call for a new curriculum for training heart failure specialists. J Am Coll Cardiol, 44:1354-1357.
5) Alonso J, Ferrer M, Gandek B, Ware JE Jr, Aaronson NK, Mosconi P, Rasmussen NK, Bullinger M, Fukuhara S, Kaasa S, Leplège A. (2004)IQOLA Project Group. Health-related quality of life associated with chronic conditions in eight countries. Results from the International Quality of Life Assessment (IQOLA) project. Qual Life Res, 13:283-298.
6) Zuluaga MC, Guallar-Castillón P, López-García E, Banegas JR, Conde-Herrera M, Olcoz-Chiva M, Rodríguez-Pascual C, Rodriguez-Artalejo F. (2010) Generic and disease-specific quality of life as a predictor of long-term mortality in heart failure. Eur J Heart Fail, 12:1372-8.
7) Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS, Peterson ED, Masoudi FA, Krumholz HM, Havranek EP, Conard MW, Williams RE. (2006) Cardiovascular Outcomes Research Consortium Health status identifies heart failure outpatients at risk for hospitalization or death. J Am Coll Cardiol, 47:752-6
8) Rodríguez-Artalejo F, Guallar-Castillón P, Pascual CR, Otero CM, Montes AO, García AN, Conthe P, Chiva MO, Banegas JR, Herrera MC. (2005) Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. Arch Intern Med, 165:1274-9.
9) Garin O, Ferrer M, Pont A, Rué M, Kotzeva A, Wiklund I, Van Ganse E, Alonso J. (2008) Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. Qual Life Res, 18:71-85.
10) Rector TS, Kubo SH, Cohn JN. (1987) Patients' self-assessment of their congestive heart failure. Part 2: content, reliability and validity of a new measure, The Minnesota Living with Heart Failure Questionnaire. Heart Failure, 198-209.
11) Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. (1989) Development and testing of a new measure of health status for clinical trials in heart failure. J Gen Intern Med, 4:101-107.

12) Wiklund I, Lindvall K, Swedberg K, Zupkis RV. (1987) Self-assessment of quality of life in severe heart failure. An instrument for clinical use. Scand J Psychol, 28:220-225.

13) Green CP, Porter CB, Bresnahan DR, Spertus JA. (2000) Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol, 35:1245-1255.

14) O'Leary CJ, Jones PW. (2000) The left ventricular dysfunction questionnaire (LVD-36): reliability, validity, and responsiveness. Heart, 83:634-640.

15) http://www.proqolid.org/. Accessed 3 August 2012.

16) Rector TS, Carson PE, Anand IS, McMurray JJ, Zile MR, McKelvie RS, Komajda M, Kuskowski M, Massie BM; I-PRESERVE Trial Investigators. Assessment of long-term effects of irbesartan on heart failure with preserved ejection fraction as measured by the minnesota living with heart failure questionnaire in the irbesartan in heart failure with preserved systolic function (I-PRESERVE) trial. Circ Heart Fail. 2012 Mar 1;5(2):217-25

17) Heo S, Moser DK, Riegel B, Hall LA, Christman N. (2005) Testing the psychometric properties of the Minnesota Living with Heart Failure questionnaire. Nurs Res, 54:265-272

18) Cruz FD, Issa VS, Ayub-Ferreira SM, Chizzola PR, Souza GE, Moreira LF, Lanz-Luces JR, Bocchi EA. (2010) Effect of a sequential education and monitoring programme on quality-of-life components in heart failure. Eur J Heart Fail, 12:1009-15

19) Franzen K, Blomqvist K, Saveman BI. (2006) Impact of chronic heart failure on elderly persons' daily life: a validation study. Eur J Cardiovasc Nurs, 5:137-145.

20) Middel B, Bouma J, de Jongste M, van Sonderen E, Niemeijer MG, Crijns H, van den Heuvel W. (2001) Psychometric properties of the Minnesota Living with Heart Failure Questionnaire (MLHF-Q). Clin Rehabil, 15:489-500.

21) Pietri G, van Ganse E, Ferrer M, Garin O, Wiklund I, Rector T. (2004) Minnesota Living with Heart Failure Questionnaire. User Manual. European Commission Report. Copyright© MAPI Research Institute.

22) Mouly M, Lobo-Luppi L for the IQOD Group. (2002) International Health-related Quality of Life Outcomes Database (IQOD) Programme. Phase I – Progress report. Cross-cultural reference values for PRO: WHQ achieved, PGWBI and MLHF in progress. Quality of Life Newsletter, 29.

23) Rossi G. (1967) [Nomenclature and diagnostic criteria in cardiology set by the New York Heart Association. Considerations on the 6th edition]. Cuore Circ, 51:287-293.

24) Flora DB, Curran PJ. (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods, 9:466-491.

25) Muthen BO. (1998-2004). Mplus Technical Appendices. Los Angeles,CA: Muthén & Muthén.

26) Scientific Advisory Committee of the Medical Outcomes Trust. (2002) Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res, 11:193-205.

27) Cronbach, LJ. (1951) Coefficient alpha and the internal structure of test. Psychometrika, 16: 297-334.
28) Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
29) Meredith W, Teresi JA. (2006) An essay on measurement and factorial invariance. Med Care, 44:S69-S77.
30) Raphael C, Briscoe C, Davies J, Ian Whinnett Z, Manisty C, Sutton R, Mayet J, Francis DP. (2007) Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. Heart; 93(4):476-82.
31) Herdman M, Ferrer M, Rajmil L, Brugulat P, Vilagut G, Medina A et al. (2008) Monitoring Health-Related Quality of Life in the Catalan Health Interview Survey using the EQ-5D. 15th Annual Conference of the International Society for Quality of Life Research (ISOQOL), 22-25 October 2008. Montevideo, Uruguay. Qual Life Res suppl, A-26.
32) Vet HC, Ader HJ, Terwee CB, Pouwer F. (2005) Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. Qual Life Res, 14:1203-1218.
33) Reise SP, Morizot J, Hays RD. (2007) The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Qual Life Res;16 Suppl 1:19-31.
34) Chen FF, West SG, Sousa KH. (2006) A Comparison of Bifactor and Second-Order Models of Quality of Life. Multivariate Behav Res, 41:189-225.
35) Quittan M, Wiesinger GF, Crevenna R, Nuhr MJ, Posch M, Hülsmann M, Müller D, Pacher R, Fialka-Moser V. (2001) Cross-cultural adaptation of the Minnesota Living with Heart Failure Questionnaire for German-speaking patients. J Rehabil Med, 33:182-186.
36) Briancon S, Alla F, Méjat E, Guillemin F, Villemot JP, Mertes PM, Zannad F. (1997) [Measurement of functional inability and quality of life in cardiac failure. Transcultural adaptation and validation of the Goldman, Minnesota and Duke questionnaires]. Arch Mal Coeur Vaiss, 90:1577-1585.
37) Ho CC, Clochesy JM, Madigan E, Liu CC. (2007) Psychometric evaluation of the Chinese version of the Minnesota Living with Heart Failure Questionnaire. Nurs Res, 56:441-448.
38) Garin O, Soriano N, Ribera A, Ferrer M, Pont A, Alonso J, Permanyer G; Grupo IC-QoL. (2008) Validación de la versión española del Minnesota Living with Heart Failure Questionnaire[Validation of the Spanish version of Minnesota Living with Heart Failure Questionnaire]. Rev Esp Cardiol, 61:251-259.
39) Miani D, Rozbowsky P, Gregori D, Pilotto L, Albanese MC, Fresco C, Fioretti PM. (2003) The Kansas City Cardiomyopathy Questionnaire: Italian translation and validation. Ital Heart J, 4:620-6.

40) Saccomann IC, Cintra FA, Gallani MC. (2007) Psychometric properties of the Minnesota Living with Heart Failure--Brazilian version--in the elderly. Qual Life Res, 16:997-1005.

41) Carvalho VO, Guimarães GV, Carrara D, Bacal F, Bocchi EA. (2009) Validation of the Portuguese version of the Minnesota Living with Heart Failure Questionnaire. Arq Bras Cardiol, 93:39-44.

Figure 1. The Bi-factor model tested through Confirmatory Factor Analysis. General and specific factors (and the given names), with the observed items' loadings.

Figure 2. MLHFQ's score means calculated at the *overall sample* for patients on NYHA I-II (grey) and on NYHA III-IV (white). Together with 95% confidence interval and the corresponding effect size coefficient of the differences across functional capacity group.

Figure 3. MLHFQ's Physical domain's score means by country; for patients on NYHA I-II (grey) and on NYHA III-IV (white). Together with 95% confidence interval and the corresponding effect size coefficient of the differences across functional capacity group.

Foot note: Aus: Australia, Brz: Brazil, Can: Canada, CzR: Czech Republic, Den: Denmark, Fin: Finland, Fra: France, Ger: Germany, GB: Great Britain, Hun: Hungary, Isr: Israel, Ita: Italy, Neth: Netherlands, Nwy: Norway, Pol: Poland, Slvk: Slovakia, Spn: Spain, Swd: Sweden, Swtz: Switzerland, US: United States, Yug: Yugoslavia

Table 1. Socio-demographic and clinical variables by country.

Country	n	Age, mean (SD)	Gender, % (male)	BMI, mean (SD)	Smokers, %	NYHA class, %	
						I-II	III-IV
Australia	73	65.7 (9.8)	87.7	28.1 (5.4)	28.8	43.8	56.2
Brazil	88	52.8 (12.6)*	76.1	25.9 (4.7)*	13.3	86.6	13.4*
Canada	533	65.0 (9.8)	84.8	27.7 (5.1)	15.9	66.4	33.6*
Czech Rep.	26	64.3 (11.5)	73.1	27.5 (5.2)	38.5	34.6	65.4
Denmark	76	65.4 (9.3)	84.2	27.5 (4.3)	47.4*	22.4	77.6*
Finland	29	60.1 (9.5)	79.3	29.6 (5.5)	27.6	20.7	79.3*
France	420	60.4 (12.9)	72.9	26.5 (4.5) *	57.6*	18.2	81.8*
Germany	872	61.5 (12.4)	75.7	27.3 (4.2) *	80.2*	36.3	63.7*
Great Britain	103	60.7 (14.3)	68.9*	26.8 (4.2) *	40.8*	17.5	82.5*
Hungary	77	61.8 (10.5)	83.1	29.3 (5.7)	20.8	31.2	68.8*
Israel	75	65.6 (11.3)	64.0*	27.7 (5.0)	9.3*	17.3	82.7*
Italy	187	61.4 (9.2)	77.5	26.0 (3.8) *	30.1	53.0	47.0
Netherlands	31	65.4 (9.4)	64.5*	26.7 (5.1)	25.8	6.5	93.5*
Norway	47	66.5 (10.4)	68.1	26.0 (4.2) *	40.4*	14.9	85.1*
Poland	214	61.7 (11.0)	80.4	27.3 (4.5) *	16.8	34.1	65.9*
Slovakia	43	62.0 (8.8)	76.7	29.2 (5.0)	16.3	7.0	93.0*
Spain	721	69.1 (12.1)	60.6*	27.2 (4.6)	15.4	70.5	29.5*
Sweden	95	69.9 (7.3)	74.7	27.3 (4.5)	18.9	12.6	87.4*
Switzerland	21	54.6 (8.7)	47.6*	26.7 (5.8)	95.2*	0.0	100.0*
US	60	64.4 (12.2)	83.3	29.8 (6.5)	21.7	51.9	48.1
Yugoslavia	56	54.3 (11.0)*	85.7	---	---	67.9	32.1
All	3847	63.5 (12.1)	74.2	27.3 (4.7)	23.9	49.1	50.9

* p<0.05 using ANOVA and Bonferroni post hoc comparisons for continuous variables and Chi-square test for qualitative ones

SD: standard deviation; NYHA : New York Heart Association

Table 2. Original MLHFQ principal component structure (1992) and Exploratory Factor Analysis (n=1936) results: 3 factors structure with factor loadings near or above 0.4 marked bold.

Original	**Items** **Did your heart failure prevent you from living as you wanted during the last month by…**	**F1** **Physical**	**F2** **Emotional**	**F3** **Social**
	1. …swelling in your ankles, legs, etc?	**0.505**	-0.091	0.002
Physical	2. …sit or lie down to rest during the day?	**0.762**	-0.025	0.022
Physical	3. …walking about or climbing stairs difficult?	**0.853**	0.044	0.040
Physical	4. …working around the house or yard difficult?	**0.762**	0.076	0.208
Physical	5. …going places away from home difficult?	**0.623**	-0.129	0.103
Physical	6. …sleeping well at night difficult?	**0.487**	-0.216	-0.006
Physical	7. …relating to or doing things with your friends or family difficult?	**0.379**	-0.225	0.310
	8. …working to earn a living difficult?	-0.012	-0.095	**0.752**
	9. …recreational pastimes, sports or hobbies difficult?	0.297	0.026	**0.658**
	10. …sexual activities difficult?	0.074	-0.117	**0.523**
	11. …eat less of the foods you like?	0.130	-0.312	0.201
Physical	12. …short of breath?	**0.840**	-0.007	-0.078
Physical	13. …tired, fatigued, or low energy?	**0.833**	-0.100	-0.067
	14. …stay in a hospital?	0.111	-0.270	0.146
	15. …money for medical care?	-0.076	-0.269	**0.377**
	16. …side effects from medications?	0.042	-0.337	0.159
Emotional	17. …feel you are a burden to your family or friends?	-0.032	**-0.713**	0.173
Emotional	18. …feel a loss of self-control in your life?	-0.012	**-0.737**	0.152
Emotional	19. …worry?	0.011	**-0.841**	-0.015
Emotional	20. …difficult for you to concentrate or remember things?	0.259	**-0.475**	-0.019
Emotional	21 …feel depressed?	0.084	**-0.847**	-0.123

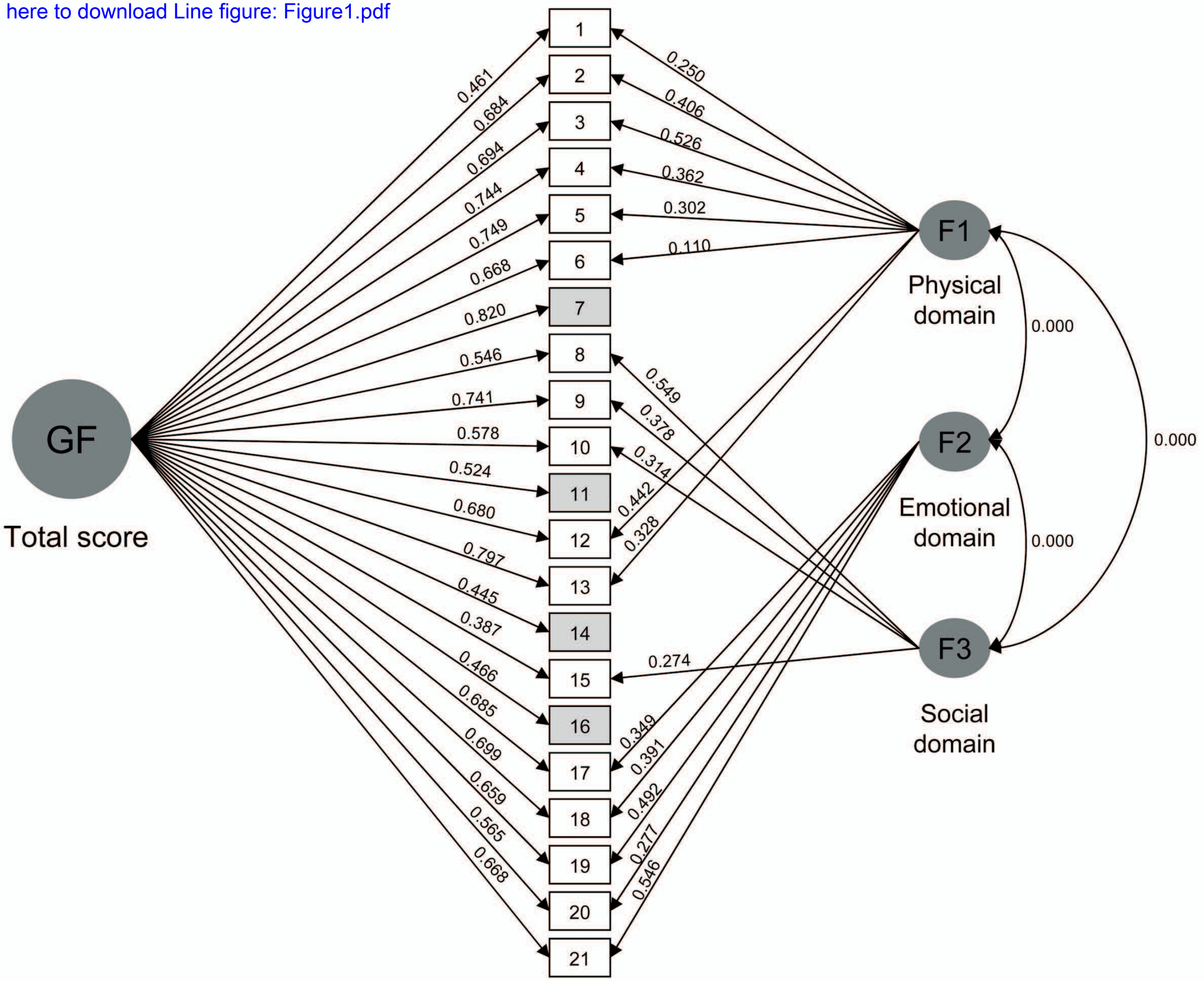
MLHFQ: Minnesota Living with Heart Failure Questionnaire

Table 3. Distribution of MLHFQ scores and internal consistency coefficients by country.

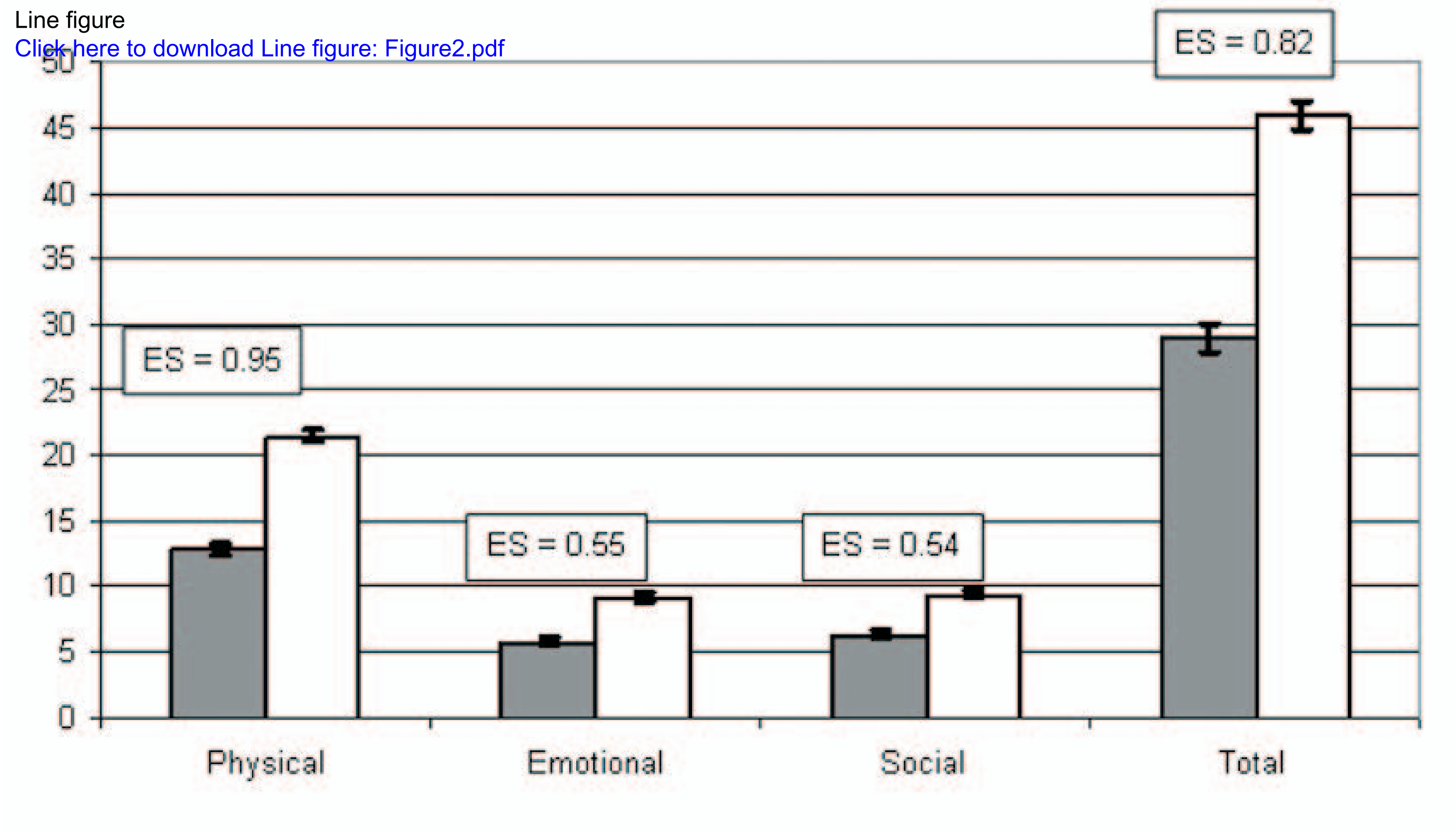
	Physical				**Emotional**				**Social**				**Total**			
	Mean (SD)	**% Floor effect**	**% Ceiling effect**	**Cronbach's alpha**	**Mean (SD)**	**% Floor effect**	**% Ceiling effect**	**Cronbach's alpha**	**Mean (SD)**	**% Floor effect**	**% Ceiling effect**	**Cronbach's alpha**	**Mean (SD)**	**% Floor effect**	**% Ceiling effect**	**Cronbach's alpha**
Australia	19.6 (10.3)	0.0	0.0	0.91	9.5 (7.5)	2.7	11.0	0.91	10.6 (6.4)	8.2	9.6	0.74	45.2 (24.6)	0.0	0.0	0.94
Brazil	15.9 (9.5)	1.1	2.3	0.88	7.8 (7.0)	0.0	17.0	0.86	9.3 (5.6)	3.4	5.7	0.69	37.4 (21.7)	0.0	0.0	0.92
Canada	14.6 (9.7)	0.6	3.2	0.88	6.0 (6.4)	1.1	**23.8**	0.86	7.0 (5.5)	1.9	15.6	0.66	31.7 (21.7)	0.2	1.7	0.92
Czech Rep.	16.7 (8.4)	0.0	0.0	0.90	5.4 (4.1)	0.0	7.7	0.73	5.8 (4.3)	0.0	15.4	0.56	30.9 (15.4)	0.0	0.0	0.88
Denmark	19.0 (9.5)	0.0	2.6	0.89	8.7 (6.2)	0.0	6.6	0.85	9.9 (5.9)	3.9	2.6	0.68	42.8 (20.9)	0.0	0.0	0.91
Finland	25.7 (9.2) *	0.0	0.0	0.88	11.7 (6.1)	0.0	3.4	0.81	12.5 (6.5) *	13.8	3.4	0.79	56.7 (23.9)*	0.0	0.0	0.94
France	14.4 (10.7)	1.2	7.4	0.91	6.6 (6.1)	1.7	12.6	0.85	5.8 (5.3)	1.0	18.3	0.74	31.5 (23.2)	0.2	3.6	0.94
Germany	16.1 (9.5)	0.5	2.8	0.90	6.2 (5.7)	0.0	15.9	0.86	7.7 (5.6)	1.9	11.4	0.70	34.0 (21.0)	0.0	1.0	0.93
Great Britain	22.5 (9.7) *	4.9	0.0	0.90	10.5 (7.4)	1.0	11.7	0.89	9.1 (5.9)	1.9	11.7	0.74	48.2 (23.9)	0.0	0.0	0.94
Hungary	20.4 (9.0)	0.0	1.3	0.91	7.2 (5.4)	0.0	3.9	0.80	10.0 (6.3)	6.5	0.0	0.82	43.0 (20.9)	0.0	0.0	0.92
Israel	19.4 (10.4)	2.7	2.7	0.93	8.2 (6.6)	2.7	5.3	0.83	9.9 (6.1)	8.0	6.7	0.73	44.2 (24.8)	0.0	0.0	0.94
Italy	11.4 (9.0)	0.0	5.3	0.90	6.1 (6.2)	1.1	16.6	0.87	4.8 (5.0)	0.0	**25.7**	0.73	25.7 (20.2)	0.0	2.1	0.93
Netherlands	21.1 (9.0)	0.0	0.0	0.85	8.6 (6.8)	0.0	16.1	0.86	9.8 (5.0)	0.0	9.7	0.67	45.0 (22.1)	0.0	0.0	0.92
Norway	18.9 (8.1)	0.0	0.0	0.82	6.9 (5.4)	0.0	12.8	0.76	6.0 (4.1)	0.0	10.6	0.40	36.5 (15.9)	0.0	0.0	0.84
Poland	18.9 (8.1)	0.0	0.0	0.86	6.9 (5.4)	0.5	8.9	0.80	12.4 (5.9) *	5.1	3.7	0.78	43.1 (18.5)	0.0	0.0	0.90
Slovakia	23.8 (6.4) *	0.0	0.0	0.83	9.3 (4.9)	0.0	2.3	0.76	8.1 (5.5)	2.3	7.0	0.69	46.8 (15.7)	0.0	0.0	0.87
Spain	17.0 (10.2)	0.6	3.7	0.90	8.5 (6.1)	0.7	8.7	0.82	6.9 (5.2)	1.3	14.2	0.66	37.8 (21.1)	0.1	0.4	0.91
Sweden	17.8 (8.2)	0.0	2.1	0.87	8.5 (5.4)	0.0	4.2	0.80	9.6 (5.1)	0.0	5.3	0.58	40.6 (18.1)	0.0	1.1	0.89
Switzerland	25.7 (7.5) *	0.0	0.0	0.81	10.0 (6.2)	0.0	4.8	0.78	9.5 (4.6)	4.8	0.0	0.50	50.5 (18.5)	0.0	0.0	0.89
US	15.9 (10.3)	0.0	6.7	0.92	7.6 (7.5)	0.0	**23.3**	0.89	7.3 (5.7)	1.7	20.0	0.67	35.6 (25.1)	0.0	5.0	0.95
Yugoslavia	21.0 (12.1)	1.8	3.6	0.91	7.2 (6.2)	0.0	16.1	0.74	10.8 (6.1)	5.4	7.1	0.71	45.1 (23.7)	0.0	0.0	0.91
All	16.6 (10.0)	0.7	3.2	0.90	7.2 (6.2)	0.7	13.4	0.84	7.8 (5.8)	2.2	12.4	0.72	36.3 (22.0)	0.1	1.2	0.92

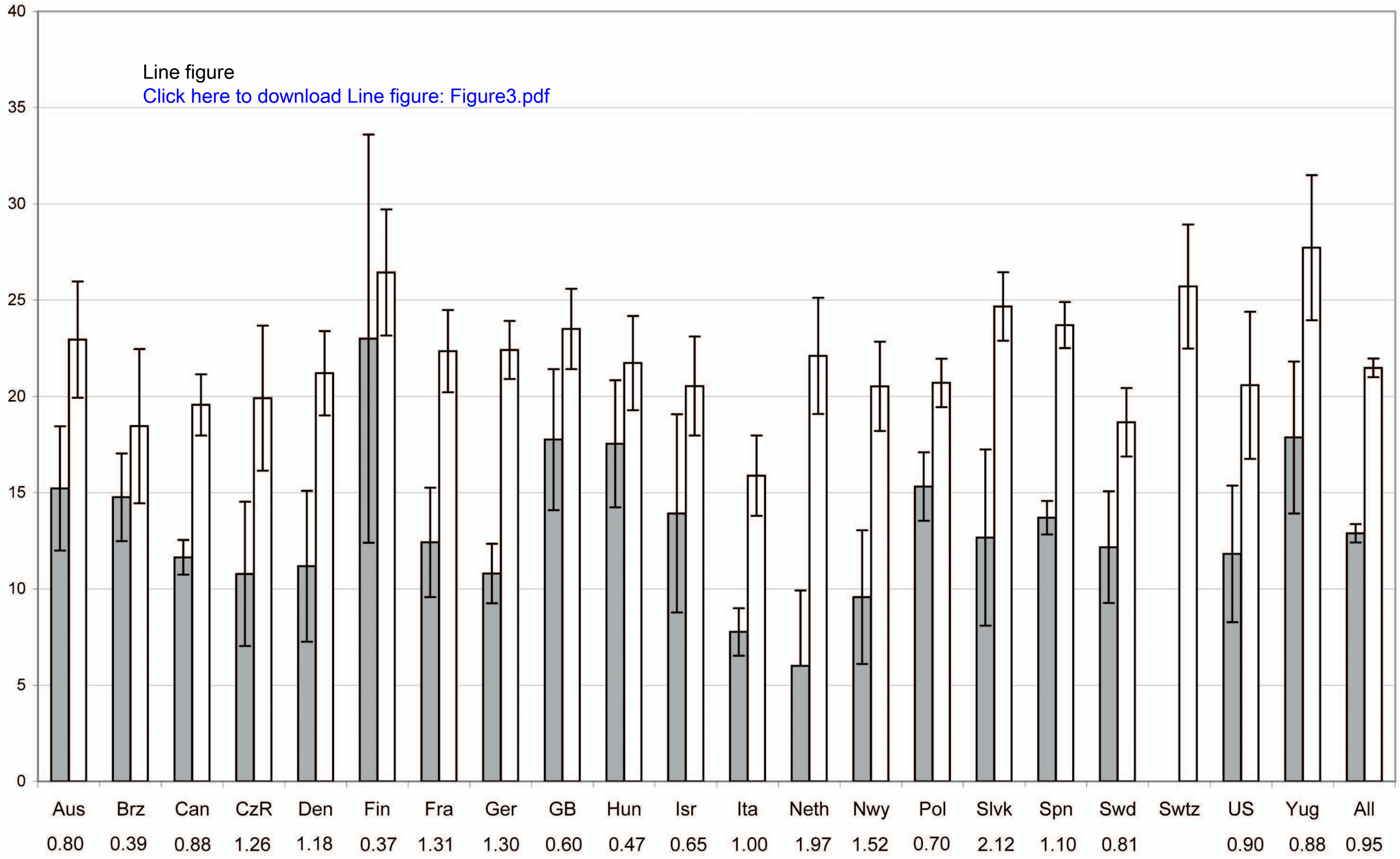
MLHFQ: Minnesota Living with Heart Failure Questionnaire

*p<0.05 using ANOVA and Bonferroni post hoc comparisons, taking US as reference.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
GF
Total score
F1
Physical domain
F2
Emotional domain
F3
Social domain
0.461
0.684
0.694
0.744
0.749
0.668
0.820
0.546
0.741
0.578
0.524
0.680
0.797
0.445
0.387
0.466
0.685
0.699
0.659
0.565
0.668
0.250
0.406
0.526
0.362
0.302
0.110
0.549
0.378
0.314
0.442
0.328
0.274
0.349
0.391
0.492
0.277
0.546
0.000
0.000
0.000

Line figure


50
45
40
35
30
25
20
15
10
5
0
ES = 0.95
ES = 0.55
ES = 0.54
ES = 0.82
Physical
Emotional
Social
Total


40
35
30
25
20
15
10
5
0
Aus 0.80
Brz 0.39
Can 0.88
CzR 1.26
Den 1.18
Fin 0.37
Fra 1.31
Ger 1.30
GB 0.60
Hun 0.47
Isr 0.65
Ita 1.00
Neth 1.97
Nwy 1.52
Pol 0.70
Slvk 2.12
Spn 1.10
Swd 0.81
Swtz
US 0.90
Yug 0.88
All 0.95

d) Paper 4:

Garin O, Herdman M, Vilagut G, Ferrer M, Ribera A, Rajmil L, Valderas JM, Guillemin F, Revicki D, Alonso J, on behalf of the EMPRO-HF Group* *Assessing Health Related Quality of Life in Heart Failure: a systematic standardized comparison of available measures.* (submitted)

Garin O, Herdman M, Vilagut G, Ferrer M, Ribera A, Rajmil L, Valderas JM, Guillemin F, Revicki D, Alonso J. Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. Heart Fail Rev. 2013 May 17.

5. GENERAL DISCUSSION

This dissertation is composed by our knowledge regarding how HRQL is measured in patients with HF focused on a narrative review of HRQL and HF (Encyclopaedia entry, 2012), and 4 manuscripts. On one hand, for the achievement of the principal aims, this project summarizes the available evidence on the existing disease-specific HRQL instruments and their metric properties (Qual Life Res, 2009)(68), and subjects their original versions to experts' evaluation under standard criteria (*EMPRO*-submitted, 2012). On the other hand, the dissertation goes deeper into the evaluation of the most widely used HF- specific HRQL questionnaire, the MLHFQ; i.e. the validation of the MLHFQ Spanish version (Rev Esp Cardiol, 2008)(89), and the confirmation of the measurement model of both the original and adapted versions of the MLHFQ for the first time (MLHFQ measurement model-submitted, 2012).

Heart Failure has been shown to be one of the chronic conditions that provokes highest impairment in physical and mental health, comparable to or greater than arthritis and chronic lung disease(9;44). HRQL in these patients is related to severity or characteristics of the disease, and usually also to age or lower socioeconomic status(55-57). Among the attempts made to measure HF functional status, the NYHA classification and the walking distance have historically been the most widely used; and they are still in force in clinical and research settings, despite their well-known limitations(43;90). In the last decades not only generic HRQL measures have been used in HF patients(48-50), but also al least 5 disease-specific questionnaires have been developed(68). Most of them presented adequate metric properties, but there were considerable differences between questionnaires in terms of the amount of evidence available on their metric properties, being the MLHFQ the one with more published data. The Spanish version of the MLHFQ presented reasonably adequate metric properties; as did the combined evaluation of 21 country versions. The measurement model assessment of those 21 versions, in a pooled sample, confirmed the unidimensionality of the MLHFQ total score, and its specific domains. Five years after the first systematic review, an update found some more bibliography to be reviewed, regarding both already known and new HF- specific HRQL questionnaires. A standardized evaluation of each questionnaires' characteristics was conducted with all the evidence, and showed the KCCQ to have similar or even better metric properties than the MLHFQ.

Apart form the specific limitations that each of the presented studies has itself, there are some limitations that are common and have to be taken into account when reporting or generalizing the results from this dissertation as a whole.

First, a list with the excluded publications and their corresponding criterion for that exclusion was not provided at the initial systematic review (until 2006), neither at its update (until 2011); and none of them included either the search in all the existing data bases. Those two steps may have added reliability to the review process. Second, in order to have a complete assessment on the HF-specific HRQL questionnaires, some additional issues may have to be addressed: a) an evaluation of the quality of the studies where data is reported(91); b) the inclusion and differentiation of all the available versions of each instrument (linguistic, country, internet, ….); c) the analyses of published data (e.g. meta-analyses) of all the studies that had used HF-specific HRQL questionnaires to compare their performance on specific groups (of severity, treatment, ….). Third, a face to face well-designed study may be the best way to really compare the HF-specific HRQL questionnaires (under the same conditions and populations), although it may not be easy to conduct. Fourth, EMPRO scoring procedure should be reviewed in order to find out a better way to reach a consensus among appraisals. Finally, regarding the specific work done on the MLHFQ, the good results on its Spanish version should be complemented with a rigorous evaluation of its sensitivity to change. The equivalence among the MLHFQ country versions should be assessed by testing Differential Item Functioning(92).

Relevant considerations and reflections:

An encyclopaedia entry should not only be clear (as it aim is to make issues understandable) and concise (normally limited space is supplied by the existence of complementary entries), but is should also be updated. The latter is one of the reasons why the "Encyclopaedia of Quality of Life Research" would be a live-project, so as to facilitate the incorporation of new evidence. However, a printed book format edition will be available in 2013, so the entry version presented as part of this dissertation is the one closed for that aim. Authors have tried to give relevant information on HF patients' HRQL, selecting evidence from the huge amount of published data, setting their perceptions versus that of general population or other chronic conditions and summarizing the studied associations between HRQL and other characteristics of these patients. Furthermore, the entry includes some outlines about the measurement of HRQL in HF.

At the first systematic review(68) we included studies using adapted versions of the HF- specific HRQL instruments. In fact, adapted versions were used at one third (32.4%) of the selected articles. Most of them were related to the MLHFQ and undoubtedly contributed to the robustness of the results and conclusions regarding this instrument. Despite the widespread use of the MLHFQ, for which there exist more than 34 country versions, validation studies were only identified for a few of them(93-97). On the other hand, at the EMPRO assessment only information on the original versions was considered, to avoid contamination between different language versions.

Based on the results raised from the first systematic review, authors made several hypotheses regarding the EMPRO assessment; however, some discrepancies have been found between them. As an example, the MLHFQ reliability was hypothesised to be higher in relation to the others in absolute terms, and according to EMPRO ratings it was not the best on this attribute. Similarly, the KCCQ showed better EMPRO scores than expected. These discrepancies on findings from the two studies might be mainly explained by three factors: a) EMPRO assessment was only conducted on the original version of the instruments; b) EMPRO appraisals were only based on documents specifically reporting information on the questionnaires' metric properties, while the initial systematic review considered evaluative studies as indicators of responsiveness; finally, c) for some of the questionnaires, those 5 years of "life" passed between the two studies may have made the difference.

Apart from allowing comparisons between the questionnaires' metric properties, the initial systematic review helped to support the use of disease-specific HRQL measures in effectiveness studies. Our use of clinicians' criteria as benchmarks for the magnitude of change supported the claim that the studied instruments adequately reflect change over time in health status; and behave differently regarding the type of intervention(8).

The Spanish version of the MLHFQ has demonstrated adequate measurement properties which are similar to those of the original version(89). The excellent findings in terms of reliability and validity were particularly noteworthy, and in general the study results support the use of the MLHFQ in Spain. Moreover, this study clearly illustrates some of the advantages of disease-specific instruments over generic ones(28), such as the suitable distribution of the MLHFQ scores on the full range of HF severity and the greater capacity (compared to the SF-36) to detect change.

At the study on the MLHFQ measurement model, the possibility of assessing a third specific domain concerning patients' social life(98) was confirmed, and results also support a structure almost equal to the one originally suggested by its developers in 1987 (MLHFQ measurement model-submitted, 2012). Furthermore, this confirmation was made for the first time using exploratory and confirmatory factor analyses in a large international and heterogeneous sample. The fact that the structure has been validated within an international sample may constitute the first step towards a cross-cultural validity of the MLHFQ. As mentioned before, among all the existing MLHFQ country versions there lacks evidence supporting their metric properties, which contrasts with the wide acceptance and use of this instrument in international clinical trials, effectiveness studies and, more recently, in clinical practice. Even though our study did not address the issue of cross-cultural validation as it should be understood, it shows results about what may be for most MLHFQ country versions a first psychometric evaluation (acceptable reliability and validity among the different country versions participating at the IQOD project).

In fact, the study on the MLHFQ measurement model responds to the pre-existing need of applying an updated factorial analysis model to support the unidimensionality of the MLHFQ total score to provide an overall measure of quality of life in HF patients. And even though it also suggests two minor changes on the specific domains (one item replacement in the physical score and the calculation of a third specific domain), they may not represent any trouble for allowing the comparison with already published data, nor prevent comparisons in follow-up studies. Hence, even though there should be no reason for avoiding their implementation hereafter, these recommendations could be dismissed as already happened to other studies' results(99;100). In the specific case of the MLHFQ, a reason for the scarce dissemination of recommendations on the specific domains might be the wide use of its total score, in many clinical publications that present HRQL as another secondary endpoint among many. Here, however, resides part of the importance of the MLHFQ's total score, and this study's strength in addressing the crucial issue of the total score unidimensionality (i.e. the real existence of one-dimensional latent construct underlying the MLHFQ total score). These findings, together with the already proved simplicity, clarity, and good performance of the MLHFQ, definitely support the use of the questionnaire as an outcome for HF patients, as well as improve the knowledge about it, and its interpretation.

The most recommended HF disease-specific HRQL instruments by the experts in the EMPRO assessment are the KCCQ and the MLHFQ, closely followed by the CHFQ (*EMPRO*-submitted, 2012). Probably these 3 are also the most widely used and well-known instruments among professionals, but are they selected for local and international studies because of their metric characteristics? Results shown in the last study of this dissertation have been obtained following a systematic reliable process, which at first term makes potential users of HRQL instruments aware of the range of measurement options available.

The comparisons of the metric properties within the identified HF-specific questionnaires clearly showed the reasons why the CHAT, the CHPchf, the LVDQ and the QLQ-SHF were recommend with provisos or not recommended by experts. A number of weak points were identified regarding several of their attributes, or there simply was no information on them. However, how well do the KCCQ, MLHFQ and CHFQ performed compare to other HRQL instruments? Their "conceptual model" and "interpretability" were rated lower than those of the EQ-5D, and the SF-36, the Spanish versions of which were assessed with EMPRO as part of its development process(101). However, the CHFQ, the KCCQ, or the MLHFQ have shown in our study higher or similar scores on reliability and responsiveness to these two well-known generic questionnaires.

Hence, from the 7 identified HF-specific HRQL questionnaires, at least 3 (the MLHFQ, the KCCQ, and the CHFQ) can currently be recommended. Choosing between them will depend on particular study focus, needs and requirements; and findings from this dissertation may facilitate decisions in the instrument selection process. Nevertheless, the need of further research regarding the adequacy and performance of some of the HF-specific HRQL instruments based on newer, more rigorous methodology, following the updated standard recommendations (e.g. adequate sample sizes, designs…), has also been pointed out.

6. CONCLUSIONS

1) At least seven disease-specific instruments exist designed to evaluate the health related quality of life of patients suffering from heart failure.
2) The KCCQ, the MLHFQ, and the CHFQ have good evidence based results regarding metric properties; choosing either one or the other should depend on the aim and resources availability:
    a. in certain settings, self-administration might be preferred, so the choice would be between the MLHFQ and the KCCQ.
    b. for use in longitudinal studies or daily clinical practice, where responsiveness is a priority, the KCCQ and the CHFQ would be recommended.
    c. specific dimensions addressed by particular questionnaires may be of particular interest in some studies, such as ‘self efficacy’ (KCCQ) or ‘social limitation’(KCCQ and MLHFQ).
3) The Spanish version of the MLHFQ demonstrated adequate metric properties, comparable to the original. These results support the use of the MLHFQ in Spanish heart failure patients, although it would be advisable to re-evaluate its responsiveness to change.
4) The MLHFQ scores in the different country versions have shown adequate reliability and validity.

5) The unidimensionality of the MLHFQ total score, and its physical, emotional and social domains, have been confirmed for the first time.
6) The Social domain of the MLHFQ may be considered hereafter when assessing these patients' health related quality of life, especially as a relevant outcome with regard to their management.
7) Evidence based on Item Response Theory about disease-specific instruments to evaluate the health related quality of life of patients suffering from heart failure is very scarce. Further research on this approach may improve the quality of their measurement models, interpretation strategies, and equivalence among different country-versions.

7. BIBLIOGRAPHY

(1) Kannel WB, Belanger AJ. Epidemiology of heart failure. Am Heart J 1991; 121(3 Pt 1):951-957.

(2) Eriksson H. Heart failure: a growing public health problem. J Intern Med 1995; 237(2):135-141.

(3) Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB et al. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation 2012; 125(1):e2-e220.

(4) Anguita SM, Crespo Leiro MG, de Teresa GE, Jimenez NM, Alonso-Pulpon L, Muniz GJ. Prevalence of heart failure in the Spanish general population aged over 45 years. The PRICE Study. Rev Esp Cardiol 2008; 61(10):1041-1049.

(5) Denolin H, Kuhn H, Krayenbuehl HP, Loogen F, Reale A. The definition of heart failure. Eur Heart J 1983; 4(7):445-448.

(6) Adamson PB, Abraham WT, Love C, Reynolds D. The evolving challenge of chronic heart failure management: a call for a new curriculum for training heart failure specialists. J Am Coll Cardiol 2004; 44(7):1354-1357.

(7) Jafri SM, Lakier JB, Rosman HS, Peterson E, Goldstein S. Symptoms and tests of ventricular performance in the evaluation of chronic heart failure. Am Heart J 1986; 112(1):194-196.

(8) Lewis EF. Still at the drawing board: improving quality of life in heart failure. Circ Heart Fail 2012; 5(2):137-139.

(9) Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry SD et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. JAMA 1989; 262(7):907-913.

(10) McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Bohm M, Dickstein K et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2012; 14(8):803-869.

(11) Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health 2003; 6(5):522-531.

(12) Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006; 4:79.

(13) McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. BMC Med 2011; 9:86.

(14) Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. Qual Life Res 2008; 17(9):1125-1135.

(15) Testa MA, Simonson DC. Assesment of quality-of-life outcomes. N Engl J Med 1996; 334(13):835-840.

(16) Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993; 118(8):622-629.

(17) Lenderking W. Task Force report of the Patient-Reported Outcomes Harmonization Group: too much harmony, not enough melody? Value Health 2003; 6(5):503-504.

(18) Patrick DL, Erickson P. Health Status and Health Policy. Allocating Resources to Health Care. New York: Oxford University Press, Inc., 1993.

(19) Taillefer MC, Dupuis G, Roberge MA, LeMay S. Health-related quality of life models: systematic review of literature. Soc Indic Res 2003; 64(2):293-323.

(20) Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care. I: Applications and issues in assessment. BMJ 1992; 305(6861):1074-1077.

(21) Fletcher A, Gore S, Jones D, Fitzpatrick R, Spiegelhalter D, Cox D. Quality of life measures in health care. II: Design, analysis, and interpretation. BMJ 1992; 305(6862):1145-1148.

(22) Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care 1989; 27(3 supp.):S217-S232.

(23) Khanna D, Tsevat J. Health-related quality of life--an introduction. Am J Manag Care 2007; 13 Suppl 9:S218-S223.

(24) Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002; 11(3):193-205.

(25) Boyle MH, Torrance GW, Sinclair JC, Horwood SP. Economic evaluation of neonatal intensive care of very-low-birth-weight infants. N Engl J Med 1983; 308(22):1330-1337.

(26) Feeny D, Furlong W, Barr RD, Torrance GW, Rosenbaum P, Weitzman S. A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992; 10(6):923-928.

(27) EQ-5D™ http://www.euroqol.org/home.html. [Accessed May 2012]

(28) Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 2003; 56(1):52-60.

(29) Scientific Advisory Committe Medical Outcomes Trust Bulletin. Instrument review criteria. Med Outcom Trust Bull 1995; 3.

(30) Meredith W, Teresi JA. An essay on measurement and factorial invariance. Med Care 2006; 44(11 Suppl 3):S69-S77.

(31) Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Qual Life Res 2007; 16 Suppl 1:19-31.

(32) Chen FF, West SG, Sousa KH. A Comparison of Bifactor and Second-Order Models of Quality of Life. Multivariate Behav Res 2006; 41(2):189-225.

(33) Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications, 1991.

(34) Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16(3):297-334.

(35) Tucker LR. A note on the estimation of test reliability by the Kuder-Richardson formula (20). Psychometrika 1949; 14(2):117-119.

(36) Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 1991; 12(4 Suppl):142S-158S.

(37) Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Elrbaum Associates, 1988.

(38) Rossi G. [Nomenclature and diagnostic criteria in cardiology set by the New York Heart Association. Considerations on the 6th edition]. Cuore Circ 1967; 51(6):287-293.

(39) Demers C, McKelvie RS, Negassa A, Yusuf S. Reliability, validity, and responsiveness of the six-minute walk test in patients with heart failure. Am Heart J 2001; 142(4):698-703.

(40) Strijbos JH, Postma DS, van AR, Gimeno F, Koeter GH. A comparison between an outpatient hospital-based pulmonary rehabilitation program and a home-care pulmonary rehabilitation program in patients with COPD. A follow-up of 18 months. Chest 1996; 109(2):366-372.

(41) Du H, Newton PJ, Salamonson Y, Carrieri-Kohlman VL, Davidson PM. A review of the six-minute walk test: its implication as

a self-administered assessment tool. Eur J Cardiovasc Nurs 2009; 8(1):2-8.

(42) Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. J Gen Intern Med 1989; 4(2):101-107.

(43) Rector TS, Kubo SH, Cohn JN. Patients' self-assessment of their congestive heart failure. Part 2: content, reliability and validity of a new measure, The Minnesota Living with Heart Failure Questionnaire. Heart Failure 1987;198-209.

(44) Alonso J, Ferrer M, Gandek B, Ware JE, Aaronson NK, Mosconi P et al. Health-related quality of life associated with chronic conditions in eight countries. Results from the International Quality of Life Assessment (IQOLA) project. Qual Life Res 2004; 13:283-298.

(45) Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJ, Ponikowski P, Poole-Wilson PA et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the diagnosis and treatment of acute and chronic heart failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). Eur J Heart Fail 2008; 10(10):933-989.

(46) Cohn J, Cleland JG, Lubsen J, Borer JS, Steg PG, Perelman M et al. Unconventional end points in cardiovascular clinical trials: should we be moving away from morbidity and mortality? J Card Fail 2009; 15(3):199-205.

(47) Wenger NK. Quality of life: can it and should it be assessed in patients with heart failure? Cardiology 1989; 76(5):391-398.

(48) al-Kaade S, Hauptman PJ. Health-related quality of life measurement in heart failure: challenges for the new millennium. J Card Fail 2001; 7(2):194-201.

(49) Flynn KE, Lin L, Moe GW, Howlett JG, Fine LJ, Spertus JA et al. Relationships between changes in patient-reported health status and functional capacity in outpatients with heart failure. Am Heart J 2012; 163(1):88-94.

(50) Naveiro-Rilo JC, Diez-Juarez MA, Flores-Zurutuza L, Rodriguez-Garcia MA, Rebollo-Gutierrez F, Romero BA. [Quality of life in patients with heart failure: the perspective of primary health care]. Gac Sanit 2012; 26(5):436-443.

(51) Dyer MT, Goldsmith KA, Sharples LS, Buxton MJ. A review of health utilities using the EQ-5D in studies of cardiovascular disease. Health Qual Life Outcomes 2010; 8:13.

(52) Cleland JG, Calvert MJ, Verboven Y, Freemantle N. Effects of cardiac resynchronization therapy on long-term quality of life: an analysis from the CArdiac Resynchronisation-Heart Failure (CARE-HF) study. Am Heart J 2009; 157(3):457-466.

(53) Wiklund I, Lindvall K, Swedberg K, Zupkis RV. Self-assessment of quality of life in severe heart failure. An instrument for clinical use. Scand J Psychol 1987; 28(3):220-225.

(54) Juenger J, Schellberg D, Kraemer S, Haunstetter A, Zugck C, Herzog W et al. Health related quality of life in patients with

congestive heart failure: comparison with other chronic diseases and relation to functional variables. Heart 2002; 87(3):235-241.

(55) Philbin EF, Dec GW, Jenkins PL, DiSalvo TG. Socioeconomic status as an independent risk factor for hospital readmission for heart failure. Am J Cardiol 2001; 87(12):1367-1371.

(56) Rosen RC, Contrada RJ, Gorkin L, Kostis JB. Determinants of perceived health in patients with left ventricular dysfunction: a structural modeling analysis. Psychosom Med 1997; 59(2):193-200.

(57) Iqbal J, Francis L, Reid J, Murray S, Denvir M. Quality of life in patients with chronic heart failure and their carers: a 3-year follow-up study assessing hospitalization and mortality. Eur J Heart Fail 2010; 12(9):1002-1008.

(58) Westlake C, Dracup K, Creaser J, Livingston N, Heywood JT, Huiskes BL et al. Correlates of health-related quality of life in patients with heart failure. Heart Lung 2002; 31(2):85-93.

(59) Lainscak M, Keber I. Patient's view of heart failure: from the understanding to the quality of life. Eur J Cardiovasc Nurs 2003; 2(4):275-281.

(60) Rumsfeld JS, Havranek E, Masoudi FA, Peterson ED, Jones P, Tooley JF et al. Depressive symptoms are the strongest predictors of short-term declines in health status in patients with heart failure. J Am Coll Cardiol 2003; 42(10):1811-1817.

(61) Gottlieb SS, Khatta M, Friedmann E, Einbinder L, Katzen S, Baker B et al. The influence of age, gender, and race on the prevalence of depression in heart failure patients. J Am Coll Cardiol 2004; 43(9):1542-1549.

(62) Coelho R, Ramos S, Prata J, Bettencourt P, Ferreira A, Cerqueira-Gomes M. Heart failure and health related quality of life. Clin Pract Epidemiol Ment Health 2005; 1:19.

(63) Havranek EP, Spertus JA, Masoudi FA, Jones PG, Rumsfeld JS. Predictors of the onset of depressive symptoms in patients with heart failure. J Am Coll Cardiol 2004; 44(12):2333-2338.

(64) Murberg TA, Bru E. Social relationships and mortality in patients with congestive heart failure. J Psychosom Res 2001; 51(3):521-527.

(65) Anand IS, Florea VG, Fisher L. Surrogate end points in heart failure. J Am Coll Cardiol 2002; 39(9):1414-1421.

(66) Sullivan M, Levy WC, Russo JE, Spertus JA. Depression and health status in patients with advanced heart failure: a prospective study in tertiary care. J Card Fail 2004; 10(5):390-396.

(67) Wielenga RP, Erdman RA, Huisveld IA, Bol E, Dunselman PH, Baselier MR et al. Effect of exercise training on quality of life in patients with chronic heart failure. J Psychosom Res 1998; 45(5):459-464.

(68) Garin O, Ferrer M, Pont A, Rue M, Kotzeva A, Wiklund I et al. Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. Qual Life Res 2009; 18(1):71-85.

(69) Fung JW, Chan SK, Yeung LY, Sanderson JE. Is beta-blockade useful in heart failure patients with atrial fibrillation? An analysis of data from two previously completed prospective trials. Eur J Heart Fail 2002; 4(4):489-494.

(70) Edes I, Gasior Z, Wita K. Effects of nebivolol on left ventricular function in elderly patients with chronic heart failure: results of the ENECA study. Eur J Heart Fail 2005; 7(4):631-639.

(71) Gonzalez-Juanatey JR, Alegria EE, Garcia S, V, Perez OG, Ruiz Ros JA, Espinosa Caliani JS et al. [Use of bisoprolol in heart failure. The BISOCOR observational study]. Rev Esp Cardiol 2003; 56(9):873-879.

(72) Little WC, Wesley-Farrington DJ, Hoyle J, Brucks S, Robertson S, Kitzman DW et al. Effect of candesartan and verapamil on exercise tolerance in diastolic dysfunction. J Cardiovasc Pharmacol 2004; 43(2):288-293.

(73) Harris S, LeMaitre JP, Mackenzie G, Fox KA, Denvir MA. A randomised study of home-based electrical stimulation of the legs and conventional bicycle exercise training for patients with chronic heart failure. Eur Heart J 2003; 24(9):871-878.

(74) Dall'Ago P, Chiappa GR, Guths H, Stein R, Ribeiro JP. Inspiratory muscle training in patients with heart failure and inspiratory muscle weakness: a randomized trial. J Am Coll Cardiol 2006; 47(4):757-763.

(75) Konstam MA. Home monitoring should be the central element in an effective program of heart failure disease management. Circulation 2012; 125(6):820-827.

(76) Hole T, Grundtvig M, Gullestad L, Flonaes B, Westheim A. Improved quality of life in Norwegian heart failure patients after follow-up in outpatient heart failure clinics: results from the

Norwegian Heart Failure Registry. Eur J Heart Fail 2010; 12(11):1247-1252.

(77) Dekker RL, Moser DK, Peden AR, Lennie TA. Cognitive therapy improves three-month outcomes in hospitalized patients with heart failure. J Card Fail 2012; 18(1):10-20.

(78) Molhoek SG, Bax JJ, van Erven L, Bootsma M, Boersma E, Steendijk P et al. Effectiveness of resynchronization therapy in patients with end-stage heart failure. Am J Cardiol 2002; 90(4):379-383.

(79) Auricchio A, Stellbrink C, Butter C, Sack S, Vogt J, Misier AR et al. Clinical efficacy of cardiac resynchronization therapy using left ventricular pacing in heart failure patients stratified by severity of ventricular conduction delay. J Am Coll Cardiol 2003; 42(12):2109-2116.

(80) Comin-Colet J, Lainscak M, Dickstein K, Filippatos GS, Johnson P, Luscher TF et al. The effect of intravenous ferric carboxymaltose on health-related quality of life in patients with chronic heart failure and iron deficiency: a subanalysis of the FAIR-HF study. Eur Heart J 2012.

(81) Zuluaga MC, Guallar-Castillon P, Lopez-Garcia E, Banegas JR, Conde-Herrera M, Olcoz-Chiva M et al. Generic and disease-specific quality of life as a predictor of long-term mortality in heart failure. Eur J Heart Fail 2010; 12(12):1372-1378.

(82) Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS et al. Health status identifies heart failure outpatients at

risk for hospitalization or death. J Am Coll Cardiol 2006; 47(4):752-756.

(83) Rodriguez-Artalejo F, Guallar-Castillon P, Pascual CR, Otero CM, Montes AO, Garcia AN et al. Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. Arch Intern Med 2005; 165(11):1274-1279.

(84) Sullivan MD, Levy WC, Russo JE, Crane B, Spertus JA. Summary health status measures in advanced heart failure: relationship to clinical variables and outcome. J Card Fail 2007; 13(7):560-568.

(85) Jaarsma T, Lesman-Leegte I. The value of the quality of life. Eur J Heart Fail 2010; 12(9):901-902.

(86) Konstam MA. Heart failure training: a call for an integrative, patient-focused approach to an emerging cardiology subspecialty. J Am Coll Cardiol 2004; 44(7):1361-1362.

(87) Taylor J, Stott DJ. Chronic heart failure and cognitive impairment: co-existence of conditions or true association? Eur J Heart Fail 2002; 4(1):7-9.

(88) Zanolla L, Zardini P. Selection of endpoints for heart failure clinical trials. Eur J Heart Fail 2003; 5(6):717-723.

(89) Garin O, Soriano N, Ribera A, Ferrer M, Pont A, Alonso J et al. Validación de la versión española del Minnesota Living with Heart Failure Questionnaire[Validation of the Spanish version of Minnesota Living with Heart Failure Questionnaire]. Rev Esp Cardiol 2008; 61(3):251-259.

(90) Raphael C, Briscoe C, Davies J, Ian WZ, Manisty C, Sutton R et al. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. Heart 2007; 93(4):476-482.

(91) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010; 19(4):539-549.

(92) Petersen MA, Groenvold M, Bjorner JB, Aaronson N, Conroy T, Cull A et al. Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. Qual Life Res 2003; 12(4):373-385.

(93) Briancon S, Alla F, Mejat E, Guillemin F, Villemot JP, Mertes PM et al. [Measurement of functional inability and quality of life in cardiac failure. Transcultural adaptation and validation of the Goldman, Minnesota and Duke questionnaires]. Arch Mal Coeur Vaiss 1997; 90(12):1577-1585.

(94) Ho CC, Clochesy JM, Madigan E, Liu CC. Psychometric evaluation of the Chinese version of the Minnesota Living with Heart Failure Questionnaire. Nurs Res 2007; 56(6):441-448.

(95) Quittan M, Wiesinger GF, Crevenna R, Nuhr MJ, Posch M, Hulsmann M et al. Cross-cultural adaptation of the Minnesota Living with Heart Failure Questionnaire for German-speaking patients. J Rehabil Med 2001; 33(4):182-186.

(96) Saccomann IC, Cintra FA, Gallani MC. Psychometric properties of the Minnesota Living with Heart Failure--Brazilian version--in the elderly. Qual Life Res 2007; 16(6):997-1005.

(97) Carvalho VO, Guimaraes GV, Carrara D, Bacal F, Bocchi EA. Validation of the Portuguese version of the Minnesota Living with Heart Failure Questionnaire. Arq Bras Cardiol 2009; 93(1):39-44.

(98) Suzuki M, Yamada S, Shimizu Y, Kono Y, Hirashiki A, Izawa H et al. Development of the participation scale for patients with congestive heart failure. Am J Phys Med Rehabil 2012; 91(6):501-510.

(99) Heo S, Moser DK, Riegel B, Hall LA, Christman N. Testing the psychometric properties of the Minnesota Living with Heart Failure questionnaire. Nurs Res 2005; 54(4):265-272.

(100) Middel B, Bouma J, de Jongste M, van Sonderen E, Niemeijer MG, Crijns H et al. Psychometric properties of the Minnesota Living with Heart Failure Questionnaire (MLHF-Q). Clin Rehabil 2001; 15(5):489-500.

(101) Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health 2008; 11(4):700-708.

8. ANNEXES

4.1 Annex 1: HF-HRQL questionnaires

The CHAT

Dunderdale K, Thompson DR, Beer SF, Furze G, Miles JN. Development and Validation of a Patient-Centered Health-Related Quality-of-life Measure: The Chronic Heart Failure Assessment Tool. *J Cardiovasc Nurs*. 2008; 23(4):364-70.

transition items that ask respondents to compare their overall health, breathing, symptoms, and medication with those at a specific point in time.

Nine items were developed using phrase completion scales; each scale consists of a phrase followed by an 11–point response scale.[20] Examples of items are shown in Table 1.

TABLE 1 Example of Items

Example of transition items

	Much less than other adults	Somewhat less than other adults	About the same as other adults	Somewhat more than other adults	Much more than other adults	Not applicable
I have to think through activities before doing them.	0	1	2	3	4	N/A
I am restricted in what activities I can do.	0	1	2	3	4	N/A
I am careful in what activities I do.	0	1	2	3	4	N/A

Example of Five-point response scale (during the past two weeks)

	Very often	Often	Occasionally	Almost never	Never
Concerned about the side effects of your medication.	☐	☐	☐	☐	☐
Relying on medication to keep your symptoms under control.	☐	☐	☐	☐	☐
Felt breathless/puffed out	☐	☐	☐	☐	☐
Felt breathless/puffed out when you were still or inactive and resting	☐	☐	☐	☐	☐

Example of Phrase-completion scales

	0	1	2	3	4	5	6	7	8	9	10
I have difficulty concentrating	None of the time										All of the time
My sleeping habits have changed	None of the time										All of the time
My family/partner worry about me	None of the time										All of the time

The CHFQ

Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. *J Gen Intern Med*. 1989; 4(2):101-7.

trial of captopril in refractory chronic congestive heart failure. J Am Coll Cardiol. 1983;2:755-63.

7. Massie B, Bourassa M, DiBianco R, et al. Long-term oral administration of amrinone for congestive heart failure: lack of efficacy in a multicenter controlled trial. Circulation. 1985;71:963-71.
8. DiBianco R, Shabetai R, Silverman BD, Leier CV, Benotti JR, and The Amrinone Multicenter Study Investigators. Oral amrinone for the treatment of chronic congestive heart failure: results of a multicenter randomized double-blind and placebo-controlled withdrawal study. J Am Coll Cardiol. 1984;4:855-66.
9. Sharpe DN, Murphy J, Coxon R, Hannan SF. Enalapril in patients with chronic heart failure: a placebo-controlled, randomized, double-blind study. Circulation. 1984;70:271-8.
10. Kirshner B, Guyatt GH. A methodologic framework for assessing health indices. J Chronic Dis. 1985;38:27-36.
11. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. Canad Med Assoc J. 1986; 134:889-95.
12. Brook RH, Ware JE, Davies-Avery A, et al. Overview of validity and the index of well-being. Health Serv Res. 1976;11:478-507.
13. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. Med Care 1981;19:787-805.
14. Guyatt GH, Townsend M, Berman LB, Pugsley SO. Quality of life in patients with chronic airflow limitation. Br J Dis Chest 1987; 81:45-54.
15. Guyatt GH, Sullivan MJJ, Fallen EL, et al. A controlled trial of digoxin in heart failure. Am J Cardiol. 1988;61:371-5.
16. Lee DC, 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.
17. Guyatt GH, Sullivan MJ, Fallen EL, Pugsley SO, Taylor DW, Berman LB. The six minute walk: a new measure of exercise capacity in patients with chronic heart failure. Canad Med Assoc J. 1985;132:919-23.
18. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40:171-8.

APPENDIX

Summary of the Chronic Heart Failure Questionnaire

The questionnaire begins by eliciting five activities in which the patient experiences dyspnea during day-to-day activities:

1. I would like you to think of the activities that you have done during the last two weeks that have made you feel short of breath. These should be activities which you do frequently and which are important in your day-to-day life. Please list as many activities as you can that you have done during the last two weeks that have made you feel short of breath.

[CIRCLE THE NUMBER ON THE ANSWER SHEET LIST ADJACENT TO EACH ACTIVITY MENTIONED. IF AN ACTIVITY MENTIONED IS NOT ON THE LIST, WRITE IT IN, IN THE RESPONDENT'S OWN WORDS, IN THE SPACE PROVIDED]

Can you think of any other activities you have done during the last two weeks that have made you feel short of breath?

[RECORD ADDITIONAL ITEMS]

2. I will now read a list of activities that make some people with lung problems feel short of breath. I will pause after each item long enough for you to tell me if you have felt short of breath doing that activity during the last two weeks. If you haven't done the activity during the last two weeks, just answer "no." The activities are:

[READ ITEMS, OMITTING THOSE WHICH RESPONDENT HAS VOLUNTEERED SPONTANEOUSLY. PAUSE AFTER EACH ITEM TO GIVE RESPONDENT A CHANCE TO INDICATE WHETHER HE/SHE HAS BEEN SHORT OF BREATH WHILE PERFORMING THAT ACTIVITY DURING THE LAST TWO WEEKS. CIRCLE THE NUMBERS ADJACENT TO APPROPRIATE ITEMS ON THE ANSWER SHEET]

1. Being **angry** or upset
2. Having a **bath** or shower
3. **Bending**
4. **Carrying,** such as carrying groceries
5. **Dressing**
6. **Eating**
7. **Going** for a walk
8. Doing your **housework**
9. **Hurrying**
10. **Lying** flat
11. **Making** a bed
12. **Mopping** or scrubbing the floor
13. **Moving** furniture
14. **Playing** with children or grandchildren
15. **Playing** sports
16. **Reaching** over your head
17. **Running,** such as for a bus
18. **Shopping**
19. **Talking**
20. **Vacuuming**
21. **Walking** around your own home
22. **Walking** uphill
23. **Walking** upstairs
24. **Walking** with others on level ground
25. **Preparing** meals
26. Trying to **sleep**

If more than five items have been listed, the interviewer then helps the subject determine the five activities that are most important in the subject's day-to-day life.

3. Of the items you have listed, which is the most important to you in your day-to-day life? I will read through the items, and when I am finished, I would like you to tell me which is the most important.

[READ THROUGH ALL ITEMS SPONTANEOUSLY VOLUNTEERED AND THOSE FROM THE LIST THAT THE PATIENT MENTIONED]

Which of these items is most important to you in your day-to-day life?

[LIST ITEM ON RESPONSE SHEET]

This process is continued until the five most important activities are determined. The interviewer then proceeds to find out how much shortness of breath the subject has experienced during the preceding two weeks. Throughout the questionnaire, response options are printed on cards of different colors, which are given to the subject.

4. I would now like you to describe how much shortness of breath you have experienced during the last two weeks while doing the five most important activities you have selected. Please indicate how much

shortness of breath you have had during the last two weeks while [INTERVIEWER: INSERT ACTIVITY LIST IN 3a] by choosing one of the following options from the card in front of you: [GREEN CARD]

1. Extremely short of breath
2. Very short of breath
3. Quite a bit short of breath
4. Moderate shortness of breath
5. Some shortness of breath
6. A little shortness of breath
7. Not at all short of breath

This process continues until the subject's degrees of dyspnea on all five of his or her most important activities has been determined. The remainder of the questionnaire asks 11 standard questions that are the same for all subjects. The wording is deliberately repetitious, experience having taught us that the repetition ensures subjects' understanding. Response options are consistently presented as seven-point scales. An example of the way the questions are structured follows:

5. In general, how much of the time during the last two weeks have you felt frustrated or impatient? Please indicate how often during the last two weeks you have felt frustrated or impatient by choosing one of the following options from the card in front of you: [BLUE CARD]

   1. All of the time
   2. Most of the time
   3. A good bit of the time
   4. Some of the time
   5. A little of the time
   6. Hardly any of the time
   7. None of the time

The wording structures of the other questions are identical, and appropriate seven-point scales are offered for each question. The content of the remaining ten questions is as follows:

6. What abut fatigue? How tired have you felt over the last two weeks?
7. How often during the past two weeks have you felt inadequate, worthless, or as if you were a burden on others?
8. How much energy have you had in the last two weeks?
9. In general, how much of the time did you feel upset, worried, or depressed during the last two weeks?
10. How much of the time during the last two weeks did you feel relaxed and free of tension?
11. How often during the last two weeks have you felt low in energy?
12. In general, how often during the last two weeks have you felt discouraged or down in the dumps?
13. How often during the last two weeks have you felt worn out or sluggish?
14. How happy, satisfied, or pleased have you been with your personal life during the last two weeks?
15. In general, how often during the last two weeks have you felt restless, tense, or uptight?

Scoring of the CHQ

The questions are divided into three areas, or dimensions:

Dyspnea (questions 4a–4e)
Fatigue (question 6, 8, 11, 13)
Emotional function (question 5, 7, 9, 10, 12, 14, 15)

The scores for all questions in each dimension are simply added together. Thus, using a seven-point scale for the responses, the minimum and maximum scores for each dimension would be:

Dimension	*Minimum Score (Worst Function)*	*Maximum Score (Best Function)*
Dyspnea	5	35
Fatigue	4	28
Emotional function	7	49

The KCCQ

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(5):1245-55.

APPENDIX

THE KANSAS CITY CARDIOMYOPATHY QUESTIONNAIRE:

The following questions refer to your **heart failure** and how it may affect your life. Please read and complete the following questions. There are no right or wrong answers. Please mark the answer that best applies to you.

1. **Heart failure** affects different people in different ways. Some feel shortness of breath while others feel fatigue. Please indicate how much you are limited by **heart failure** (shortness of breath or fatigue) in your ability to do the following activities <u>over the past 2 weeks</u>.

Place an **X** in one box on each line

Activity	**Extremely Limited**	**Quite a bit Limited**	**Moderately Limited**	**Slightly Limited**	**Not at all Limited**	**Limited for other reasons or did not do the activity**
Dressing yourself	☐	☐	☐	☐	☐	☐
Showering/Bathing	☐	☐	☐	☐	☐	☐
Walking 1 block on level ground	☐	☐	☐	☐	☐	☐
Doing yardwork, housework or carrying groceries	☐	☐	☐	☐	☐	☐
Climbing a flight of stairs without stopping	☐	☐	☐	☐	☐	☐
Hurrying or jogging (as if to catch a bus)	☐	☐	☐	☐	☐	☐

2. <u>Compared with 2 weeks ago</u>, have your symptoms of **heart failure** (shortness of breath, fatigue or ankle swelling) changed?
My symptoms of **heart failure** have become . . .

Much worse	Slightly worse	Not changed	Slightly better	Much better	I've had no symptoms over the last 2 weeks
☐	☐	☐	☐	☐	☐

3. Over the <u>past 2 weeks</u>, how many times did you have **swelling** in your feet, ankles or legs when you woke up in the morning?

Every morning	3 or more times a week, but not every day	1–2 times a week	Less than once a week	Never over the past 2 weeks
☐	☐	☐	☐	☐

4. Over the <u>past 2 weeks</u>, how much has **swelling** in your feet, ankles or legs bothered you?
It has been . . .

Extremely bothersome	**Quite a bit** bothersome	**Moderately** bothersome	**Slightly** bothersome	**Not at all** bothersome	I've had **no swelling**
☐	☐	☐	☐	☐	☐

5. Over the <u>past 2 weeks</u>, on average, how many times has **fatigue** limited your ability to do what you want?

All of the time	Several times per day	At least once a day	3 or more times per week but not every day	1–2 times per week	Less than once a week	Never over the past 2 weeks
☐	☐	☐	☐	☐	☐	☐

6. Over the <u>past 2 weeks</u>, how much has your **fatigue** bothered you?
It has been . . .

Extremely bothersome	**Quite a bit** bothersome	**Moderately** bothersome	**Slightly** bothersome	**Not at all** bothersome	I've had **no fatigue**
☐	☐	☐	☐	☐	☐

7. Over the past 2 weeks, on average, how many times has **shortness of breath** limited your ability to do what you wanted?

All of the time	Several times per day	At least once a day	3 or more times per week but not every day	1–2 times per week	Less than once a week	Never over the past 2 weeks
☐	☐	☐	☐	☐	☐	☐

8. Over the past 2 weeks, how much has your **shortness of breath** bothered you?
It has been . . .

Extremely bothersome	**Quite a bit** bothersome	**Moderately** bothersome	**Slightly** bothersome	**Not at all** bothersome	I've had **no shortness of breath**
☐	☐	☐	☐	☐	☐

9. Over the past 2 weeks, on average, how many times have you been forced to sleep sitting up in a chair or with at least 3 pillows to prop you up because of **shortness of breath**?

Every night	3 or more times a week, but not every day	1–2 times a week	Less than once a week	Never over the past 2 weeks
☐	☐	☐	☐	☐

10. **Heart failure** symptoms can worsen for a number of reasons. How sure are you that you know what to do, or whom to call, if your **heart failure** gets worse?

Not at all sure	**Not very** sure	**Somewhat** sure	**Mostly** sure	**Completely** sure
☐	☐	☐	☐	☐

11. How well do you understand what things you are able to do to keep your **heart failure** symptoms from getting worse? (for example, weighing yourself, eating a low salt diet, etc.)

Do not understand at all	Do not understand very well	Somewhat understand	Mostly understand	Completely understand
☐	☐	☐	☐	☐

12. Over the past 2 weeks, how much has your **heart failure** limited your enjoyment of life?

It has **extremely** limited my enjoyment of life	It has limited my enjoyment of life **quite a bit**	It has **moderately** limited my enjoyment of life	It has **slightly** limited my enjoyment of life	It has **not limited** my enjoyment of life at all
☐	☐	☐	☐	☐

13. If you had to spend the rest of your life with your **heart failure** the way it is right now, how would you feel about this?

Not at all satisfied	Mostly dissatisfied	Somewhat satisfied	Mostly satisfied	Completely satisfied
☐	☐	☐	☐	☐

14. Over the past 2 weeks, how often have you felt discouraged or down in the dumps because of your **heart failure**?

I felt that way **all of the time**	I felt that way **most of the time**	I **occasionally** felt that way	I **rarely** felt that way	I **never** felt that way
☐	☐	☐	☐	☐

15. How much does your **heart failure** affect your lifestyle? Please indicate how your **heart failure** may have limited your participation in the following activities over the past 2 weeks.

Please place an **X** in one box on each line

Activity	**Severely** limited	Limited **quite a bit**	**Moderately** limited	**Slightly** limited	**Did not** limit at all	Does not apply or did not do for other reasons
Hobbies, recreational activities	☐	☐	☐	☐	☐	☐
Working or doing household chores	☐	☐	☐	☐	☐	☐
Visiting family or friends out of your home	☐	☐	☐	☐	☐	☐
Intimate relationships with loved ones	☐	☐	☐	☐	☐	☐

The LVDQ

O'Leary CJ, Jones PW. The left ventricular dysfunction questionnaire (LVD-36): reliability, validity, and responsiveness. *Heart*. 2000; 83(6):634-40.

used. No placebo group was used, so it is unclear whether the rate of decline was reduced for these patients. The investigators suggest that health status may not have been measured as accurately as possible and that "some improvement in physical ability might not have been addressed by the questions that were asked."

A third questionnaire was developed in Sweden specifically for patients with severe heart failure.[3] This has been translated into an English version, but the translation does not appear to be particularly good as it contains such questions as: "Did you feel well at ease during the last week?". This raises doubts about the translation process, and also concern as to whether meanings originally intended by the developers were correctly conveyed into the English version.

To overcome some of the problems with the existing questionnaires, a new instrument—the left ventricular dysfunction questionnaire (LVD-36)—has been developed specifically for patients with chronic left ventricular dysfunction. It was designed to measure the impact of left ventricular dysfunction on daily life and wellbeing. Its development process used a hierarchical item selection procedure[6–8] to create a short questionnaire that is simple to complete and score, yet retains the necessary psychometric requirements of a health status measure. This paper is concerned with tests of the validity, reliability, and responsiveness of LVD-36 and provides a comparison of its performance with that of the LIhFE.

Table 1 The left ventricular dysfunction questionnaire (LVD-36)

Please answer the following questions as you are feeling **these days**. *Tick either true or false for each question.*

Because of my heart condition:	**True**	**False**
I suffer with tired legs	☐	☐
I suffer with nausea (feeling sick)	☐	☐
I suffer with swollen legs	☐	☐
Because of my heart condition:	**True**	**False**
I am afraid that if I go out I will be short of breath	☐	☐
I am frightened to do too much in case I become short of breath	☐	☐
I get out of breath with the least physical exercise	☐	☐
I am frightened to push myself too far	☐	☐
I take a long time to get washed or dressed	☐	☐
If you do not do these activities for any reason other than your heart condition, then please tick false		
Because of my heart condition:	**True**	**False**
I have difficulty running, such as for a bus	☐	☐
I have difficulty either jogging, exercising or dancing	☐	☐
I have difficulty playing with children/grandchildren	☐	☐
I have difficulty either mowing the lawn or hoovering/vacuum cleaning	☐	☐
Because of my heart condition:	**True**	**False**
I feel exhausted	☐	☐
I feel low in energy	☐	☐
I feel sleepy or drowsy	☐	☐
I need to rest more	☐	☐
I feel that everything is an effort	☐	☐
My muscles feel weak	☐	☐
I get cold easily	☐	☐
I wake up frequently during the night	☐	☐
I have become frail or an invalid	☐	☐
Because of my heart condition:	**True**	**False**
I feel frustrated	☐	☐
I feel nervous	☐	☐
I feel irritable	☐	☐
I feel restless	☐	☐
I feel out of control of my life	☐	☐
I feel that I can not enjoy a full life	☐	☐
I've lost confidence in myself	☐	☐
Because of my heart condition:	**True**	**False**
I have difficulty having a regular social life	☐	☐
There are places I would like to go to but can't	☐	☐
I worry that going on holiday could make my heart condition worse	☐	☐
I have had to alter my lifestyle	☐	☐
I am restricted in fulfilling my family duties	☐	☐
I feel dependent on others	☐	☐
	True	**False**
I find it a real nuisance having to take tablets for my heart condition	☐	☐
My heart condition stops me doing things that I would like to do	☐	☐

PLEASE CHECK THAT YOU HAVE ANSWERED ALL THE QUESTIONS
THANK YOU FOR YOUR TIME

Methods

All patients participating in this study gave written informed consent. The study was approved by the Wandsworth District Health Authority ethics committee.

DEVELOPMENT OF THE LVD-36: A BRIEF DESCRIPTION

Full details of the development of the LVD-36 are given elsewhere.[8 9] A pool of 179 items was created using information from a review of published reports, existing questionnaires, semi-structured interviews with patients with left ventricular dysfunction, and consultation with clinicians. The items were presented to 139 patients with left ventricular dysfunction who were asked to respond to each item as true or false (as it applied to them), along with two global questions concerning health and impairment of function. Items were then removed from the pool if they were not associated with either of the global questions, if they were endorsed by the majority of the sample, if they were associated with sex, age, or disease duration, or if the endorsement rate was low and the association with global health was significant but weak.[9] A principal components analysis was carried out on the remaining items. Those items that loaded highly on the first component and weakly on the subsequent components were retained to form the questionnaire.

QUESTIONNAIRES USED

LVD-36

The LVD-36 is a 36 item questionnaire for patients with left ventricular dysfunction (table 1). Responses are dichotomous (true or false). True responses are summed and the sum is expressed as a percentage, so that 100 is the worse possible score and 0 the best possible score. The LVD-36 takes approximately five minutes to complete.

Medical outcomes study, SF-36

The short form 36 questionnaire (SF-36) was used to measure general health status. It contains 36 discrete items that produce eight component scores and two overall summary scores. Scores run from 0 (worse possible score) to 100 (best possible score). The SF-36 takes approximately 15 minutes to complete.

LIhFE

The LIhFE is a condition specific questionnaire for patients with heart failure. It consists of 21 items, each of which is scored on a six point scale (0–5). The score runs from 0 (best possible score) to 105 (worse possible score). To aid comparisons with the other questionnaires, the LIhFE scores were expressed as a

The MLHFQ

Rector TS, Kubo SH, Cohn JN. Patients' self-assessment of their congestive heart failure. Part 2: Content, reliability and validity of a new measure, the Minnesota Living with Heart Failure questionnaire. *Heart Fail*. 1987; 3:198-209.

Table II

CONTENT OF THE MINNESOTA LIVING WITH HEART FAILURE QUESTIONNAIRE

These questions concern how your heart failure (heart condition) has prevented you from living as you wanted during the last month. These items listed below describe different ways some people are affected. If you are sure an item does not apply to you or is not related to your heart failure, then circle 0 (No) and go on to the next item. If an item does apply to you, then circle the number rating of how much it prevented you from living as you wanted. Remember to think about ONLY THE LAST MONTH.

Did your heart failure prevent you from living as you wanted during the last month by

	No	Very Little				Very Much	Relationship to LHFQ Score*
1. causing swelling in your ankles, legs, etc.?	0	1	2	3	4	5	0.41
2. making your working around the house or yard difficult?	0	1	2	3	4	5	0.81
3. making your relating to or doing things with your friends or family difficult?	0	1	2	3	4	5	0.82
4. making you sit or lie down to rest during the day?	0	1	2	3	4	5	0.79
5. making you tired, fatigued, or low on energy?	0	1	2	3	4	5	0.86
6. making your working to earn a living difficult?	0	1	2	3	4	5	0.51
7. making your walking about or climbing stairs difficult?	0	1	2	3	4	5	0.77
8. making you short of breath?	0	1	2	3	4	5	0.78
9. making your sleeping well at night difficult?	0	1	2	3	4	5	0.77
10. making you eat less of the foods you like?	0	1	2	3	4	5	0.51
11. making your going places away from home difficult?	0	1	2	3	4	5	0.74
12. making your sexual activities difficult?	0	1	2	3	4	5	0.58
13. making your recreational pastimes, sports, or hobbies difficult?	0	1	2	3	4	5	0.83
14. making it difficult for you to concentrate or remember things?	0	1	2	3	4	5	0.60
15. giving you side effects from medications?	0	1	2	3	4	5	0.35
16. making you worry?	0	1	2	3	4	5	0.69
17. making you feel depressed?	0	1	2	3	4	5	0.70
18. costing you money for medical care?	0	1	2	3	4	5	0.54
19. making you feel a loss of self-control in your life?	0	1	2	3	4	5	0.70
20. making you stay in a hospital?	0	1	2	3	4	5	0.38
21. making you feel you are a burden to your family or friends?	0	1	2	3	4	5	0.52

Copyright University of Minnesota, 1986.
Median responses from 83 patients are underlined for each item.
** Spearman rank-order correlation between sum of items 1-21 and each item.*

cur. The last of these three items, swelling, is very useful to clarify the instructions (see Appendix).

The initial median (25th to 75th percentiles) LHFQ score was 34 (16 to 54) out of a maximal possible score of 105 (Table III), which represents the maximal perceived impairment. The median difference between the first and second scores was −1 (−11 to 5), indicating there was no significant bias. Baseline variability was consistent across patient subgroups and cause of left ventricular

The QLQ-SHF

Wiklund I, Lindvall K, Swedberg K, Zupkis RV. Self-assessment of quality of life in severe heart failure. An instrument for clinical use. *Scand J Psychol.* 1987; 28 (3):220-5.

e heart failure

d ROBERT V. ZUPKIS³

lity of life in
ology, 1987,

l the natural
cal interven-
cted adverse
to assess QL
onnaire was
relevant and
des somatic,
uestionnaire

Sweden.

emotional well-being.
quantitative measures
). The introduction of
to survive, but adverse
o assess the subjective

tors, such as economic
dt, 1975; Naess, 1979;
ctive well-being are of
corded signs of disease
ll-being describes the
ion and positive/nega-
ost important factor in
l health and well-being
on well-being does not
h allows them to do in
red to objective, more
ficance for QL, but the
both kinds of indicators

s suffering from severe
sis, QL was defined as
o function in basic daily
f-assessment of health-

METHOD

Patients

Typical patients with severe heart failure were recruited. An objective evaluation of the severity of breathlessness according to the New York Heart Association (NYHA) classification and of angina pectoris, utilizing a four-graded scale, was performed by a cardiologist.

There were 18 women and 33 men with a median age of 64 years, range 44–78 years. Twenty-one of the patient were 70 years or older. Thirteen patients were hospitalized and 38 patients treated on an ambulatory basis. According to the NYHA, 16% were in grade II, 61% grade III and 23% grade IV. Forty-three per cent were symptomatic due to previous myocardial infarction, 31% due to cardiomyopathy and 25% due to valvular heart disease. Nineteen per cent were affected by dyspnoea at heavy physical exertion, 61% at light physical exertion and 19% at rest. Twenty per cent had angina pectoris at exertion and 14% at exertion, at rest or when emotionally upset.

Quality of life questionnaire in severe heart failure (QLQ-SHF)

Ninety items covered somatic, functional and emotional disturbances. There were symptoms of cardiovascular failure such as fatigue, dyspnoea and palpitation, and other cardiac/somatic symptoms, e.g. chest pain and dizziness (20 items); emotional/cognitive symptoms (21 items) derived from work on infarction patients (Wiklund *et al.*, 1984); life satisfaction reflecting subjective well-being/happiness (10 items). Self-care activity and light to moderate exertion in ambulation/mobility/physical activity (24 items) was adopted from others (Katz *et al.*, 1963; Goldman *et al.*, 1981), whereas social/recreational activity (15 items) resembled the content of other scales (Bergner *et al.*, 1981; Hunt *et al.*, 1981). An example of the questions is presented in Table 1.

Table 1. *An example of questions included in the QLQ-SHF*

QLQ aspect	Question	Scale
Somatic	"Did you suffer from BREATHLESSNESS during the last week?"	VAS
Life satisfaction	"Were you SATISFIED WITH YOUR LIFE during the last week?"	VAS
Emotional	"Did you feel NERVOUS OR ANXIOUS during the last week?"	VAS
Physical activity	"To what extent have you been able to MOVE AROUND OUTDOORS during the last week?"	Likert

Visual analogue scales (VAS), measured symptoms, mood and attitudes (Aitken, 1969); and seven-graded Likert scales social interaction and physical activity. High values indicated that the patient had severe problems.

Statistics

Prior to the data analysis, the preliminary pool of items was grouped into five content areas (i.e. subscales) (somatic/physical aspects, emotional/cognitive aspects, life satisfaction, physical activity, and social/recreational activity) based on the author's judgements. In general, the content areas addressed by the items were obvious, and classification was easy. A principal components analysis was done on each of the subscales. The purpose of the principal components analysis was simply to aid in the selection of likely candidates for the shortened questionnaire. There was no need to explore the dimensionality of the preliminary pool of items, and no exploration was undertaken. Items were selected for a shortened subscale if they represented a clinically relevant problem according to the cardiologist, and/or a frequent subjective complaint by the patient, and/or loaded high (>0.70) on the first principal component from the principal components analysis of the appropriate subscale. The above criteria provided five to seven questions for a subscale. *Somatic aspects* were represented by dyspnoea at rest, chest pain limiting daily activity, fatigue, exhaustion, weakness, dullness, poor physical health, where 33–65% of the patients indicated values above 50 on the VAS; *emotional/cognitive aspects* by indecision, nervous/anxious,

4.1 Annex 2: EMPRO versions

9. APPENDICES

9.1. Complementary peer-reviewed paper

Valderas JM, Ferrer M, Mendívil J, Garin O, Rajmil L, Herdman M, Alonso J; Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network. *Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures*. Value Health, 2008, 11(4):700-8.

4.1 On process research

Manuscript draft

Cross-cultural validation of the Minnesota Living with Heart Failure Questionnaire for 21 countries”

Olatz Garin BSc, MPH[1]; Montse Ferrer MD, PhD[1]; Àngels Pont Dip Stat[1]; Ingela Wiklund MD, PhD[2]; Eric Van Ganse MD, PhD[3]; Jordi Alonso MD, PhD[1,4].

[1] Health Services Research Unit, Institut Municipal d’Investigaciò Mèdica (IMIM-IMAS), Barcelona, Spain.

[2] Behavioral Medicine, Astra Hässle Research Laboratoris, Möelndal, Sweden.

[3] Unité de Pharmacoépidémiologie, Centre Hospitalier Lyon-Sud, Lyon, France.

[4] Universitat Autónoma de Barcelona (UAB).

Study supported by: EC contract QLRI-CT-2000-00551 (IQOD), CIRIT (2001SGR 00405), and SAF2002-11009-E. The study is free from conflicts of interests.

*Address for correspondence: Montse Ferrer, Health Services Research Unit, Institut Municipal d’Investigació Mèdica, Doctor Aiguader 80; 08003 Barcelona, Spain. Tel. +34 93 225 75 53; Fax: +34 93 221 40 02; *e-mail address:* **mferrer@imim.es**

INTRODUCTION

Heart Failure (HF) is a serious, costly, and increasingly common condition, whichs {Cowie, 1997 28 /id;Eriksson, 1995 40 /id;Kannel, 1991 31 /id} prevalence increases with age and reaches about 6-7% of the elderly{Ho, 1993 10124 /id}. It is the most frequent cause of hospitalization amongst people aged 65 and over, and is responsible for 5% of all admittances. Moreover, it has been proved that HF is one of the chronic conditions that provokes greater impairment in physical and mental health perception, compared with others like diabetes, arthritis and general population{Stewart, 1989 39 /id}. Specifically, this condition has a profound impact on patients' Health Related Quality of Life (HRQL) because there are some specific factors: physical symptoms (short of breath and fatigue), psychological problems, iatrogenic adverse effects and curtailment in social activities{Bulpitt, 1991 4791 /id}.

Furthermore, since its treatment goals are not curative, the interest in assessing HRQL in HF patients has increased in recent years. To this end, several HF-specific HRQL instruments{Rector, 1987 7496 /id;Guyatt, 1989 955 /id;Wiklund, 1987 1001 /id;O'Leary, 2000 7306 /id;Green, 2000 273 /id} have been developed and are widely used to assess the burden of the condition, as an outcome variable in clinical trials and practice. A recent study compares the characteristics of these HF-specific instruments and, weighting the quality and quantity of data on psychometric properties, the Minnesota Living with Heart Failure Questionnaire (MLHFQ) seam to be the most advisable instrument{Garin O 9999999 /id}.

This questionnaire has been translated into 34 languages (with a varied linguistic validation methodology). Although the reliability, validity and responsiveness of this instrument has been sufficiently demonstrated for the original US version, and for some other languages, equivalence between different versions has not been examined. The assessment across multiple versions of different countries and languages, is important to determine whether it is appropriate to combine results from international multicentric research projects. Therefore the purpose of this study was to evaluate the equivalence of psychometric properties and item functioning of the MLHFQ across 21 countries.

METHODS

Data sources (study population):

Data from several epidemiological studies (2) and clinical trials (6) carried out by universities and hospitals of different countries, was compiled, anonymized and depurated. The final common database, included 3266 patients from 21 different countries: Australia (n=73), Brazil (n=88), Canada (n=534), Czech Republic (n=26), Denmark (n=78), Finland (n=29), France (n=449), Germany (n=908), Great Britain (n=106), Hungary (n=84), Israel (n=77), Italy (n=188), Netherlands (n=31), Norway (n=48), Poland (n=215), Slovakia (n=47), Spain (n=48), Sweden (n=99), Switzerland (n=21), USA (n=61), and Yugoslavia (n=56). A part from the HRQL baseline assessment and patients' functional capacity (New York Heart Association class, NYHA), some socio-demographic and clinical variables were also merged to study patients characteristics on some HF risk factors like: age, gender, BMI or smoking status.

The Minnesota Living with Heart Failure Questionnaire (MLHFQ):

The MLHFQ was developed by Thomas Rector to assess HF patients' HRQL{Rector, 1987 7496 /id;Rector, 1992 7331 /id}. The questionnaire comprises 21 items focused on patients perceptions concerning the effects of HF on their physical (8 items) and emotional (5 items) lives. The other 8 items reflect the effects of HF on working to earn a living, edema, sexual activities, appetite, cost of medical care, side effects from medication and hospitalization, which were not originally factored into physical or emotional dimensions but are accounted for in the total score. Response options are presented as a 6 point Likert scale (0-5), from "no impairment" to "very much", with a total maximum score 105. Lower MLHFQ scores indicate better health status. The original US version has demonstrated excelent psychometric properties{Middel, 2001 44 /id;Bennett, 2003 133 /id}, and its easy administration has facilitated its use in many clinical research projects and clinical practice{Garin O 9999999 /id}{Pollock, 1990 7333 /id}.

Psychometric analyses:

The psychometric properties of the MLHFQ versions were examined following recommended standard methods{Scientific Advisory Committee of the Medical Outcomes Trust., 2002 7648 /id} . Firstly, the evaluation started examining item-level characteristics to ensure understandability : data completeness (% of missing < 10), and range of response choices endorsed; and whether were they comparable across countries. Secondly, because the MLHFQ uses Likert's methods of summated scores, it is important to examine this method's assumptions{Likert, 1932 888888 /id}: items in each hypothesized scale should contain approximately the same proportion of information about the construct being measured; and they also should be substantially linearly related to the total score computed from all other items in the scale. Test of these assumptions determine the appropriateness of including an item in a particular scale and whether it is appropriate to simply sum item scores to estimate scale scores. Therefore, the relationship of each item to its scale was evaluated by considering a coefficient of 0.4 the minimum item-scale correlation, and the maximum difference between items contributing the same scale.

Once item-level characteristics had been confirmed, scale distribution characteristics and reliability of scale scores were examined: mean and SD across countries, floor and ceiling effects (percent of patients with lowest and highest possible score), and internal consistency reliability (by Cronbach's alpha coefficient){Cronbach, 1951 4100 /id}.

Differential item functioning in the MLHFQ across countries:

Test of differential item functioning (DIF) was used to assess the presence of item equivalence{Petersen MA, 2003 22222222 /id}; that is, to evaluate if HF patients from different countries showed differing probabilities of success on any item, after matching on the scale score. This approach can handle both situations where an item is more difficult in one country (uniform DIF), and situations where an item is less strongly related to the score in one country (non-uniform DIF). An ordinal logistic regression approach was used to assessed DIF being the item response the dependent variable in the models; and the scale and the country (defined as dummy variable) the independent variables, together with the scale-country interaction{Zumbo BD, 1999 1111111 /id}.

A test of significance and measure of magnitude (variance increase to country differences from USA) for DIF was calculated by following the natural hierarchy of entering variables into the model: 1) the scale score, 2) the country, and finally 3) the interaction term. The Chi-squared value for each hierarchy step was obtained to compute the significance test. Simultaneous DIF significance was tested subtracting the Chi-squared value of step 1 from the Chi-squared of step 3 and comparing the resultant value with its distribution{Swaminathan H, 1990 3333333 /id}. The corresponding effect size was calculated by the R^2 (explained variance) attributable to both the county and the interaction terms simultaneously: R^2 at step 3 minus the R^2 at step 1(ΔR^2). Uniform and non-uniform DIF single magnitude were tested computing the same formula for step 2- step 1, and step2-step3, respectively.

A double criterion was applied in considering an item to exhibit DIF: statistical significance of Chi-squared tests ($p<0.05$) and magnitude of DIF (ΔR^2) higher than 2%{Bjorner JB, 2005 4444444 /id}.

RESULTS

Subject characteristics:

Mean age across countries was 62.5 years old, and the mean ages per country are mostly between 60 and 70 years (Table 1). The youngest population was found in Brazil, with a mean of 52.8; whereas the oldest population was found in Sweden which mean was 70.1year old. The majority of participants were male (76.4%), but this percentage differed from 47.6% in Switzerland to 87.7% in Australia. USA had the population with the highest BMI, whereas Brazil had the population with the lowest (with means from 30.02 to 25.90 kg/m^2). On average, 27.2% of subjects were smokers, but in Germany this variable get quite high values (80.2%). There was a considering variability across countries regarding NYHA classes' percentages, with a pattern indicating more representation of classes III and IV.

Item-level descriptive statistics by country:

Focusing on differences across countries, all but 4 showed less than 10% missing data on all items; half of the countries used the entire range for all items; and the ordering of item means was generally consistent across countries. Regarding items' systematic patterns, item 8 ("working to earn a living") showed more than 10% missing data for France, Hungary, Spain and Sweden; and more than 5% for Denmark, Germany, Norway, and Slovakia. In most countries, item 13 ("fatigue") was the one with the highest mean and item 14 ("stay in hospital") the one with the lowest.

Relationship of items to scale by country:

Of the 273 item-scale correlations, more than 86% were greater or at least 0.6 (Table 2); only four correlations were lower than the recommended standard of 0.4: one at the physical scale (item 6 in Norway), and 3 at the emotional (item 17 in Yugoslavia, item 18 in Czech Republic, and item 20 also in Norway). In general, correlations between items and their hypothesized scale were roughly equivalent for each item, with the exceptions already mentioned where the correlations also varied more than 0.4 across items.

Scale-level statistics by country:

Score means were calculated alter stratifying according to functional capacity, and participants were divided into NYHA class I.II or III-IV. The MLHFQ total scores mean for both groups were similar across countries, and as expected the mean for those with less functional capacity was higher (44.8 versus 27.2 for those with less impairment) (Table 3.1). The same happened for the physical and emotional scores (Tables 3.2. and 3.3 respectively). The percent of respondents with the lowest scale score was 0 for the three scales. The percent of respondents at ceiling varied by scales, from almost 0 for the total score in both groups, to 50% for the emotional score in The Netherlands. For most countries this last scale had the higher percentages of ceiling effect. However, the physical scale had just two countries with moderate percentage of ceiling effect (Denmark 11.8% and USA 13.8). The Cronbach's alpha coefficient was 0.9 for most countries' total and physical scores.

Differential item functioning in the MLHFQ across countries:

Six countries out of 20 showed no DIF in any of the items: Australia, Germany, France, Great Britain, Switzerland, and Canada. When evaluating simultaneously uniform and non-uniform DIF, 10 countries presented between one and three items with DIF (according to the criteria defined before); and 4 (Hungary, Spain, Slovakia, and Yugoslavia) showed DIF for 4 to 6 items.

The 260 overall ordinal logistic regressions conducted also showed that, of the 13 items evaluated (corresponding to those conforming a scale) none showed homogeneous item functioning for all countries. However, when evaluating uniform and non-uniform DIF separately, some items' DIF became non significant (Table 4). Thus, from the results of these separated analyses, it can be said that items 2, 4, 13 and, 21 presented homogeneous item functioning. Regarding uniform DIF, items 5, 17, and 18 were those with highest frequency (presented in 5 countries). And the same happened to item 19, for non-uniform DIF.

DISCUSSION

When a questionnaire is translated to another language, the first step is to examine whether it is linguistically equivalent to the source version. To achieve this objective, adaptation process consists of forward-backward translations and cognitive debriefing with panels of experts and patients. Most MLHFQ versions were linguistically adapted and evaluated by MAPI Research Institute, and were found to be fairly easy to translate, conceptually equivalent and understandable by respondents. The present study assessed the psychometric properties across different versions of the instrument and evaluated their equivalence. This second step in the overall process of adaptations, has confirmed what was conclude in the previous one; the low percentage of missing data suggests the MLHFQ easily comprehension. Moreover, most country versions have shown adequate properties, meeting the scaling assumptions, and demonstrating excellent reliability and low differential item functioning.

Item-level evaluations have shown the Czech Republic as the only country in which the majority of the MLHFQ items did not ensured the entire response range; , probably because of its small sample size (n=26). The assessment also showed "fatigue" (item 13) as provoking the largest impairment in HF patients in most countries; while “stay in the hospital” (item 14), seemed to have a very slight impact on their HRQL. However, in general items contribute substantially and approximately equal to the scale scores in most versions of the MLHFQ confirming the Likert summated rating scoring method’s assumptions. However, there were exceptions: item 6 (“sleeping”) in Norway, may be due to geographical differences{Otsuka, 2001 1 /id}; and items 17, 18, and 20 from the emotional scale (Table 2), which needs to be explored to evaluated whether is it due to translation issues or cross-cultural differences.

On the other hand, at a scale level, the MLHFQ scores were found to have good reliability and validity across countries. Differences on mean scores for NYHA class I-II and III-IV, demonstrated the instrument ability to discriminate functional capacity. Furthermore, the pattern of these means across countries showed the translations’ equivalence. The low percentages of floor and ceiling effects suggested the range of the MLHFQ scores in all versions is adequate for capturing the entire range of HRQL impairment suffered by HF patients. Regarding internal consistency, the high Cronbach’s alpha coefficients for the overall sample and across countries for the physical and total scores confirmed the MLHFQ to be useful to compare individuals in most cases {Scientific Advisory Committee of the Medical Outcomes Trust., 2002 7648 /id} since they achieved the highest standard of 0.9. Moreover, half of the countries achieved also this standard for the emotional score; and anyway all exceptions showed a coefficient of al least 0.7, the minimum required to compare groups.

The most relevant part of this study were the DIF analyses. Theses evaluations permitted two different focus. On one hand, study how well did each country’s items function in the same way as the source ones. Items of Australia, Germany, France, Great Britain, Switzerland, and Canada adapted versions presented homogenous functioning with the original ones. However, at least 4 of the 13 studied items of Spain, Slovakia, and Yugoslavia versions, exhibit DIF. This result show how these countries have more than one third of their items presenting problems on equivalence or item bias; so these adapted versions should be reviewed (conceptually, psychometrically, or with different sample sizes).

On the other hand, DIF analyses allowed us to know which of the studied items functioned worst across countries. Results showed a remarkable problem in the emotional scale, where 3 out of 5 items, exhibited DIF in five or six countries. Probably, these results are indicating a special difficulty on translating or cross-cultural adapting this issues, versus that more physical or related to symptoms. Therefore, it would be appropriate to make an special effort when adapting to a different language/culture HRQL items focus on psychological or emotional impairment.

The present study has some limitations that merit comment. On one hand, data pooled from different studies, had different characteristics (study's design, sample size or patients' severity). From those disadvantages, there are two specially remarkable for the analyses conducted: 1) the lack of both, common clinical variables and follow-up data to assess validity in a better way and responsiveness of the MLHFQ; and 2) the wide range of sample sizes that may probably had biased the DIF evaluation in some particular countries. On the other hand, an other limitation is related to the MLHFQ's structure. A factor analysis should be conducted to confirm the original structure in the different country versions. In case those analyses presented a new scale (i.e. social), the equivalence of items (that at the moment are not related to any special construct but with the total score); would be evaluated. Further research is also recommended to confirm that items that exhibited differential item functioning in the present study, also do so when using other statistical methods for DIF assessment (such as IRT).

Conclusions and applicability:

In summary, adequate properties and scaling assumptions, together with an excellent reliability and low differential item functioning, showed by most country versions, suggests an adequate level of equivalence across MLHFQ versions. These results are specially relevant to support the valid use of this HRQL specific instrument in international multicentric research projects, where data of different languages and cultures could be pooled. On the other hand, it is important to remark the inclusion of DIF analyses in this study; as a new approach to detect items' un-equivalence. The statistical method applied with this aim, has helped to validate the inclusion of patients form different countries in the same study, and thus analyze their scores simultaneously.

ACKNOWLEDGEMENTS

The authors are very grateful to the members of the IQOD Project, managed by the MAPI Research Institute. Furthermore, we would like to thank JM Valderas for his special support on the preparation of this article.

Table 1. Socio-demographic and clinical variables by country.

Country	n	Age, mean (SD)	Gender, % (male)	BMI, Mean (SD)	Smokers, %	NYHA class, %	
						I-II	III-IV
Australia	73	65.7 (9.8)	87.7	28.1 (5.4)	28.8	43.8	56.2
Brazil	88	52.8 (12.6)	76.1	25.9 (4.7)	13.3	86.6	13.4
Canada	534	65.0 (9.8)	84.8	27.6 (5.1)	15.9	66.3	33.7
Czech Rep.	26	64.3 (11.5)	73.1	27.5 (5.2)	38.5	34.6	65.4
Denmark	78	65.4 (9.2)	83.3	27.4 (4.3)	48.7	21.8	78.2
Finland	29	60.1 (9.5)	79.3	29.6 (5.5)	27.6	20.7	79.3
France	449	60.8 (12.8)	71.0	26.5 (4.5)	61.8	17.1	82.9
Germany	908	61.9 (12.3)	75.0	27.3 (4.2)	80.2	35.9	64.1
Great Britain	106	61.1 (14.3)	68.9	26.7 (4.3)	39.6	17.9	82.1
Hungary	84	62.3 (10.5)	81.0	29.3 (5.6)	19.0	29.8	70.2
Israel	77	65.4 (11.3)	64.9	27.7 (5.0)	9.1	16.9	83.1
Italy	188	61.5 (9.2)	77.7	26.0 (3.8)	30.5	52.7	47.3
Netherlands	31	65.4 (9.4)	64.5	26.7 (5.1)	25.8	6.5	93.5
Norway	48	66.7 (10.4)	68.8	26.0 (4.1)	39.6	14.6	85.4
Poland	215	61.7 (11.0)	80.0	27.3 (4.5)	16.7	34.0	66.0
Slovakia	47	63.0 (9.2)	72.3	28.9 (5.0)	14.9	6.4	93.6
Spain	48	62.8 (14.1)	56.3	27.0 (4.5)	25.0	35.4	64.6
Sweden	99	70.1 (7.3)	73.7	27.3 (4.4)	19.2	12.1	87.9
Switzerland	21	54.6 (8.7)	47.6	26.7 (5.8)	95.2	0.0	100.0
USA	61	64.4 (12.1)	83.6	30.0 (6.6)	21.3	52.7	47.3
Yugoslavia	56	54.3 (11.0)	85.7	---	---	67.9	32.1
All	3266	62.5 (11.7)	76.4	27.3 (4.7)	27.2	41.2	58.8

Table 2. Item-level descriptive statistics and relationship of items to scale, by country.

Country	Range of % missing	% items using the entire range	Range of item means	Range of item-scale correlations	
				Physical	Emotional
Australia	0.0 - 2.7	100.0	0.5 - 3.3	0.6 - 0.8	0.7 - 0.8
Brazil	0.0 - 0.0	100.0	0.3 - 2.7	0.5 - 0.8	0.5 - 0.8
Canada	0.0 - 1.3	100.0	0.3 - 2.5	0.5 - 0.8	0.6 - 0.8
Czech Rep.	0.0 - 3.8	47.6	0.4 - 3.1	0.5 - 0.8	0.2 - 0.6
Denmark	0.0 - 6.4	100.0	0.5 - 3.2	0.5 - 0.8	0.6 - 0.7
Finland	0.9 - 12.7	100.0	0.8 - 2.3	0.6 - 0.8	0.5 - 0.8
France	0.1 - 6.1	100.0	0.6 - 2.6	0.6 - 0.8	0.6 - 0.7
Germany	0.0 - 3.4	85.7	1.1 - 4.0	0.5 - 0.8	0.5 - 0.7
Great Britain	0.0 - 4.7	100.0	0.6 - 3.6	0.6 - 0.8	0.6 - 0.8
Hungary	0.0 - 22.6	100.0	0.5 - 3.2	0.6 - 0.8	0.4 - 0.7
Israel	0.0 - 3.9	95.2	0.9 - 3.0	0.7 - 0.9	0.6 - 0.7
Italy	0.0 - 2.1	100.0	0.3 - 2.0	0.5 - 0.8	0.6 - 0.8
Netherlands	0.0 - 3.2	95.2	0.2 - 3.6	0.5 - 0.8	0.6 - 0.8
Norway	0.0 - 6.3	100.0	0.5 - 3.2	0.3 - 0.9	0.3 - 0.7
Poland	0.0 - 3.3	95.2	0.6 - 3.3	0.5 - 0.7	0.5 - 0.7
Slovakia	0.0 - 8.5	81.0	0.7 - 3.7	0.4 - 0.8	0.4 - 0.7
Spain	0.0 - 10.4	95.2	0.6 - 3.5	0.5 - 0.8	0.4 - 0.8
Sweden	0.0 - 10.1	95.2	0.7 - 3.4	0.5 - 0.8	0.5 - 0.7
Switzerland	0.0 - 4.8	71.4	0.7 - 4.1	0.5 - 0.8	0.4 - 0.7
USA	0.0 - 1.6	90.5	0.2 - 3.0	0.7 - 0.9	0.6 - 0.8
Yugoslavia	0.0 - 1.8	95.2	0.7 - 3.7	0.6 - 0.9	0.3 - 0.7
All	0.3 - 5.4	100.0	0.8 - 2.6	0.6 - 0.8	0.6 - 0.7

Table 3.1. Scale means (SD), percent of respondents at lowest and highest possible MLHFQ scores, and internal consistency reliability coefficients by country; for MLHFQ total score.

Country	Mean (SD)		% floor effect	% ceiling effect	Cronbach's alpha
	NYHA I-II	NYHA III-IV			
AUS	34.1 (20.8)	54.0 (24.1)	0.0	0.0	0.94
BRA	35.2 (22.6)	41.5 (17.4)	0.0	0.0	0.92
CAN	25.8 (18.5)	41.9 (24.0)	0.0	1.7	0.92
CZE	19.8 (14.0)	36.7 (14.1)	0.0	0.0	0.88
DEN	26.4 (18.0)	46.6 (19.8)	0.0	0.0	0.91
FIN	52.3 (34.8)	59.6 (20.8)	0.0	0.0	0.94
FRA	22.0 (10.0)	45.0 (18.8)	0.2	3.3	0.94
GER	20.4 (16.6)	45.4 (24.2)	0.0	1.0	0.93
GBR	40.1 (16.7)	51.4 (23.9)	0.0	0.0	0.94
HUN	38.9 (18.5)	39.4 (20.1)	0.0	0.0	0.93
ISR	33.2 (21.8)	45.4 (24.5)	0.0	0.0	0.95
ITA	16.5 (12.6)	35.1 (21.8)	0.0	2.1	0.93
NET	7.0 (1.4)	48.5 (20.2)	0.0	0.0	0.92
NOR	22.7 (11.9)	38.8 (15.8)	0.0	0.0	0.84
POL	34.9 (17.5)	45.8 (17.8)	0.0	0.0	0.90
SLO	25.0 (5.3)	48.4 (15.4)	0.0	0.0	0.87
SPA	26.1 (10.3)	45.2 (18.7)	0.0	0.0	0.88
SWE	32.7 (13.1)	40.7 (18.2)	0.0	1.0	0.90
SWI	---	51.7 (18.2)	0.0	0.0	0.89
USA	27.5 (24.6)	45.0 (24.9)	0.0	4.9	0.94
YOU	41.4 (25.6)	54.4 (16.3)	0.0	0.0	0.91
All	27.7 (19.8)	44.8 (21.7)	0.0	1.3	0.93

Table 3.2. Scale means (SD), percent of respondents at lowest and highest possible MLHFQ scores, and internal consistency reliability coefficients by country; for MLHFQ physical score.

Country	**Mean (SD)**		**% floor effect**	**% ceiling effect**	**Cronbach's alpha**
	NYHA I-II	NYHA III-IV			
Australia	15.6 (10.1)	24.5 (10.1)	0.0	1.4	0.94
Brazil	15.1 (10.3)	18.9 (7.2)	0.0	2.3	0.89
Canada	12.1 (8.8)	20.3 (10.7)	0.0	3.9	0.90
Czech Rep.	10.9 (6.2)	20.1 (8.0)	0.0	0.0	0.90
Denmark	11.0 (8.6)	21.9 (9.0)	0.0	2.6	0.92
Finland	23.2 (13.1)	26.7 (8.8)	0.0	6.9	0.92
France	12.2 (5.3)	22.7 (8.0)	0.0	2.8	0.92
Germany	10.6 (7.7)	23.1 (10.3)	0.0	0.0	0.91
Great Britain	19.4 (7.2)	25.0 (9.9)	0.0	0.0	0.92
Hungary	17.6 (8.7)	22.3 (9.4)	0.0	1.2	0.92
Israel	14.5 (10.1)	21.0 (10.8)	0.0	2.6	0.94
Italy	7.9 (6.3)	16.1 (10.1)	0.0	5.3	0.90
Netherlands	5.5 (3.5)	23.2 (9.0)	0.0	0.0	0.88
Norway	9.6 (4.6)	21.3 (8.0)	0.0	0.0	0.85
Poland	15.5 (8.1)	21.0 (8.0)	0.0	0.0	0.87
Slovakia	13.0 (4.4)	25.2 (6.3)	0.0	0.0	0.84
Spain	12.9 (6.4)	23.2 (9.2)	0.0	0.0	0.88
Sweden	12.7 (5.3)	19.5 (8.9)	0.0	2.0	0.89
Switzerland	---	26.9 (8.5)	0.0	0.0	0.88
USA	12.4 (10.2)	21.7 (11.0)	0.0	6.6	0.94
Yugoslavia	17.8 (12.0)	27.1 (8.8)	0.0	1.8	0.91
All	12.7 (9.0)	21.8 (9.7)	0.0	3.1	0.92

Table 3.3. Scale means (SD), percent of respondents at lowest and highest possible MLHFQ scores, and internal consistency reliability coefficients by country; for MLHFQ emotional score.

Country	Mean (SD)		% floor effect	% ceiling effect	Cronbach's alpha
	NYHA I-II	NYHA III-IV			
Australia	6.3 (5.7)	12.0 (7.9)	0.0	11.0	0.91
Brazil	7.5 (6.9)	7.2 (7.3)	0.0	17.0	0.86
Canada	4.5 (5.1)	8.4 (7.5)	0.0	23.8	0.86
Czech Rep.	2.8 (2.4)	6.8 (4.3)	0.0	7.7	0.73
Denmark	6.6 (5.3)	9.2 (6.4)	0.0	6.4	0.85
Finland	11.0 (9.1)	12.1 (5.3)	0.0	3.4	0.81
France	4.1 (4.0)	9.8 (5.9)	0.0	10.9	0.85
Germany	3.9 (4.5)	8.3 (6.5)	0.0	15.6	0.86
Great Britain	8.9 (6.8)	10.8 (7.6)	0.0	11.3	0.89
Hungary	7.7 (5.6)	7.4 (5.9)	0.0	3.6	0.81
Israel	5.7 (5.4)	8.5 (6.8)	0.0	5.2	0.83
Italy	3.9 (4.4)	8.8 (6.9)	0.0	16.5	0.87
Netherlands	1.0 (1.4)	9.2 (6.7)	0.0	16.1	0.86
Norway	4.4 (3.3)	7.2 (5.6)	0.0	14.6	0.77
Poland	5.1 (4.6)	7.9 (5.6)	0.0	8.8	0.80
Slovakia	3.0 (3.0)	10.0 (4.8)	0.0	2.1	0.77
Spain	5.7 (5.0)	11.4 (6.1)	0.0	6.3	0.77
Sweden	7.1 (4.0)	8.8 (5.7)	0.0	4.0	0.80
Switzerland	---	10.0 (6.2)	0.0	4.8	0.78
USA	5.5 (7.2)	10.1 (7.7)	0.0	24.6	0.89
Yugoslavia	7.3 (6.6)	6.9 (5.7)	0.0	16.1	0.74
All	5.2 (5.5)	8.9 (6.6)	0.0	14.2	0.85

Table 4. Effect Size of uniform and non-uniform DIF (shadowed) of each item by country. Ordered from lower to higher frequency.

Country	ΔR^2								
	Item 3	**Item 20**	**Item 6**	**Item 12**	**Item 7**	**Item 5**	**Item 18**	**Item 19**	**Item 17**
Brazil						0.09			
Denmark								0.053	
Italy			0.033						
Norway		0.069							
Poland							0.040		
Finland					0.045				0.043
Holland						0.034			0.068
Israel						0.033			0.029
Sweden				0.030				0.069	
Czech Rep.					0.074		0.128		
Hungary			0.038				0.077		0.131
Spain	0.053				0.036	0.078		0.017/0.068	
Yugoslavia				0.049			0.107	0.038	0.051
Slovakia					0.033	0.048	0.047	0.029	0.067/0.031

All the DIF presented in the table were also significant (Chi-squared <0.05).