Heart Failure Reviews

Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures

--Manuscript Draft--

Manuscript Number:	
Full Title:	Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures
Article Type:	Unsolicited Articles
Keywords:	heart failure, quality of life, systematic review, standardization
Corresponding Author:	Jordi Alonso, PhD IMIM-Institut Hospital del Mar d'Investigacions Mèdiques Barcelona, SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	IMIM-Institut Hospital del Mar d'Investigacions Mèdiques
Corresponding Author's Secondary Institution:	
First Author:	Olatz Garin, MPH
First Author Secondary Information:	
Order of Authors:	Olatz Garin, MPH
	Michael Herdman, MSc
	Gemma Vilagut, BSc
	Montse Ferrer, MD, PhD
	Aida Ribera, PhD
	Luís Rajmil, MD, PhD
	Jose María Valderas, MD, PhD
	Francis Guillemin, MD, PhD
	Denis Revicki, PhD
	Jordi Alonso, PhD
	EMPRO -HF Group
Order of Authors Secondary Information:	
Abstract:	Aim To systematically evaluate available health-related quality of life (HRQL) instruments for use in patients with heart failure (HF). Methods and Results Seven HF-specific HRQL questionnaires and associated studies of their metric properties were identified by systematic review: the Chronic Heart Failure Assessment Tool, the Cardiac Health Profile-Congestive Heart Failure, the Chronic Heart Failure Questionnaire (CHFQ), the Kansas City Cardiomyopathy Questionnaire (KCCQ), the Left Ventricular Disease Questionnaire (LVDQ), the Minnesota Living with Heart Failure Questionnaire (MLHFQ), and the Quality of Life in Severe Heart Failure Questionnaire. Each instrument was assessed by four experts using a standardized tool for evaluating patient reported outcomes (EMPRO; scores from 0 to 100). Four questionnaires were given adequate scores (median >50) for the attribute "conceptual model". The LVDQ had the highest rated-median for "reliability" (72.8). The CHFQ, the KCCQ, and the MLHFQ all got reasonable scores for "validity" (from 54.4 to 76.4). The reviewers rated the KCCQ the highest in terms of "sensitivity to change" (median 94.4). Only the CHFQ (50.0) and the KCCQ (72.2) received adequate scores for the "interpretability" attribute. The most highly rated instruments based on the overall EMPRO score were the KCCQ (64.4) and the MLHFQ (60.7),

	followed by the CHFQ (59.2). Conclusions Based on the first systematic and reliable expert-based evaluation of available HF-specific HRQL questionnaires; the evidence seems to support the choice of the MLHFQ, the KCCQ and the CHFQ over the others, which require further research on metric properties. Keywords: heart failure, quality of life, systematic review, standardization

Author Disclosure

O.Garin, M.Herdman, G.Vilagut, M.Ferrer, A.Ribera, L.Rajmil, JM.Valderas, F.Guillemin, D.Revicki, J.Alonso, and all other members from the EMPRO-HF Group have no conflicts of interest or financial ties to disclose.

Dear Sir or Madam,

We are pleased to submit the manuscript entitled "**Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures**", which we would like to be considered for publication in the *Heart Failure Reviews.*

The paper describes the assessment of existing health-related quality of life measures for use in patients with heart failure through the application of a standardized, expert-based method. The results presented may help to minimize professionals' concerns about these measures. Therefore, we think this manuscript would be of interest to the readers of the European Journal of Heart Failure.

None of the manuscript contents have been previously published and the manuscript is not under consideration at any other journal.

All authors have actively participated in the study, and have read and approved the submitted manuscript. Olatz Garin, Michael Herdman and Jordi Alonso were involved in the conception and design of the study, analysis and interpretation of data, and were responsible for drafting the manuscript; Olatz Garin and Gemma Vilagut were involved in the analysis and interpretation of data and critically reviewed the manuscript; Montse Ferrer was involved in the conception and design of the study, the analysis and interpretation of data, and critically reviewed the manuscript; Aida Ribera, Luis Rajmil, José Maria Valderas, Francis Guillemin and Dennis Revicki were involved in the conception and design of the study and critically reviewed the manuscript. All authors gave their final approval of the manuscript submitted.

We hope that you will consider evaluating our manuscript for publication in the *Heart Failure Reviews* and we look forward to hearing from you soon.

Sincerely,

Jordi Alonso

Address for correspondence: Jordi Alonso, Health Services Research Unit, IMIM (Institut Hospital del Mar d'Investigacions Mèdiques), Doctor Aiguader 80, 08003 Barcelona, Spain; e-mail: *jalonso@imim.es*

Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures

Olatz Garin MPH [1,2,3], Michael Herdman MSc,[1,2] Gemma Vilagut BSc, [1,2] Montse Ferrer MD, PhD,[1,2,4] Aida Ribera MD, PhD, [1,5] Luis Rajmil MD, PhD, [1,6] José M Valderas MD, PhD, [1,7] Francis Guillemin MD, PhD,[8] Dennis Revicki PhD, [9] Jordi Alonso MD, PhD[1,2,3] on behalf of the EMPRO-HF Group*

1. CIBER en Epidemiologia y Salud Pública (CIBERESP), Spain
2. Health Services Research Unit, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
3. Universitat Pompeu Fabra (UPF), Barcelona, Spain
4. Universitat Autònoma de Barcelona (UAB), Barcelona, Spain
5. Unidad de Epidemiología, Servicio de Cardiología, Hospital Vall d'Hebron, Barcelona
6. Agència d'Informació, Avaluació i Qualitat en Salut (AIAQS), Barcelona, Spain
7. Health Services and Policy Research Group, Department of Primary Health Care Sciences, University of Oxford, UK
8. Service d'epidémiologie et évaluation cliniques, Centre d'Epidémiologie Clinique, INSERM CIE6, CHU de Nancy, France
9. Center for Health Outcomes Research, United BioSource Corporation, Bethesda, Maryland, USA

*EMPRO-HF Group: Allá F, Arnould B, Becerra V, Barblaglia G, Castellví P, Clayson D, Erhart M, Fairclough D, Fleck M, Garcia-Forero C, Gonçalves DC, Gotay C, las Hayas C, Lloyd A, Mathias S, Permanyer-Miralda G, Pont A, Prieto L, Ravens-Sieberer U, Schmidt S, Villalonga-Olives E

Address for correspondence:

Dr. Jordi Alonso

Health Services Research Unit, IMIM (Hospital del Mar Medical Research Institute), Doctor Aiguader 80, 08003 Barcelona, Spain

Tel.: +34 93 316 0740; Fax: +34 93 221 40 02; e-mail: *jalonso@imim.es*

This work was supported by a grant from the Instituto de Salud Carlos III FEDER, (PI06/90483) and Recercaixa (2010ACUP 00158).

Abstract

Aim To systematically evaluate available health-related quality of life (HRQL) instruments for use in patients with heart failure (HF).

Methods and Results Seven HF-specific HRQL questionnaires and associated studies of their metric properties were identified by systematic review: the Chronic Heart Failure Assessment Tool, the Cardiac Health Profile-Congestive Heart Failure, the Chronic Heart Failure Questionnaire (CHFQ), the Kansas City Cardiomyopathy Questionnaire (KCCQ), the Left Ventricular Disease Questionnaire (LVDQ), the Minnesota Living with Heart Failure Questionnaire (MLHFQ), and the Quality of Life in Severe Heart Failure Questionnaire. Each instrument was assessed by four experts using a standardized tool for evaluating patient reported outcomes (EMPRO; scores from 0 to 100). Four questionnaires were given adequate scores (median >50) for the attribute "conceptual model". The LVDQ had the highest rated-median for "reliability" (72.8). The CHFQ, the KCCQ, and the MLHFQ all got reasonable scores for "validity" (from 54.4 to 76.4). The reviewers rated the KCCQ the highest in terms of "sensitivity to change" (median 94.4). Only the CHFQ (50.0) and the KCCQ (72.2) received adequate scores for the "interpretability" attribute. The most highly rated instruments based on the overall EMPRO score were the KCCQ (64.4) and the MLHFQ (60.7), followed by the CHFQ (59.2).

Conclusions Based on the first systematic and reliable expert-based evaluation of available HF-specific HRQL questionnaires; the evidence seems to support the choice of the MLHFQ, the KCCQ and the CHFQ over the others, which require further research on metric properties.

Keywords: heart failure, quality of life, systematic review, standardization

Introduction

Medical care is increasingly focused on the management of chronic diseases, including heart failure (HF). Management programs are developed not only to prolong life, but also to relieve symptoms and improve overall well-being.[1] In conjunction with traditional clinical endpoints, patient reported outcomes (PROs), such as health-related quality of life (HRQL) measures, have proven useful in evaluating the achievement of these goals from the patient's perspective, and are also gaining importance as predictors of mortality and hospitalization in patients with HF.[2-4]

Both generic instruments (i.e., those covering a wide range of constructs and applicable to patients with different conditions and diseases and the general population) and disease-specific measures have been used to evaluate HRQL in relation to HF.[5;6] Disease-specific measures have, however, generally been shown to be more sensitive to changes, as they better capture issues which are relevant to these patients,[7] and a number of HF-specific HRQL questionnaires are now available. Nonetheless, as E.F. Lewis argued in a recent editorial, "many clinicians and non-outcome researchers remain skeptical and often shy away from targeting QOL as the primary endpoint in clinical studies." One of the issues that limits the use of HRQL measures is the lack of information available with which to choose the most suitable instrument, and the unconfirmed reliability of the published data.[1]

In 2009, a systematic review compiled all available published information about the measurement properties of the different HF-HRQL questionnaires.[8] However, the comparison was mainly limited due to the lack of a standardized method for evaluating the different measurement criteria of the measures assessed. A standardized tool to assess HRQL instruments has to meet three fundamental requirements: a) well-described and solidly established attributes; b) expert reviewers to conduct the assessment, and c) a scoring system which allows for direct comparisons among instruments. The EMPRO tool (evaluating measures of patient reported outcomes) was recently developed to facilitate this process; is based on an exhaustive series of

recommendations regarding the ideal attributes of PRO instruments,[9] and has been shown to be valid and reliable.[10]

The objective of this study was to systematically compare all existing HF-HRQL instruments using the EMPRO tool, to reveal the strengths and weaknesses of the constructs covered, rigor of the development process, psychometric properties and associated administrative burden. This information can act as a first step in the consolidation of healthcare professionals' acceptance of HF-HRQL measures and will allow them to base their instrument selection decisions on rigorous findings.[1]

Methods

Identification of HF questionnaires and evidence to be evaluated

Both the measures to be evaluated and the materials on which the evaluation was to be based were identified using several different methods: a) an update of a previous systematic review performed until May 2011,[8] b) a specific search in an on-line library of existing PRO instruments (www.proqolid.org), c) citation tracking of selected documents, and d) a review of instrument-specific websites. Briefly, the systematic review consisted of a broad search in NLM Gateway PubMed, using "heart failure" and "quality of life" as MeSH terms and free text. Each step was conducted separately by two reviewers (from titles to full text review). Disagreements were resolved by consensus or through the involvement of a third researcher. Only articles published in English, Spanish, French, German or Russian were included (detailed information provided elsewhere).[8]

The EMPRO

The EMPRO is a tool developed to facilitate the standardized evaluation of PRO measures.[10] Expert reviewers use the EMPRO (after examining all the materials) to evaluate instruments on 39 items covering eight relevant attributes: conceptual and measurement model, reliability, validity, sensitivity to change, interpretability, administrative burden, alternative modes of administration, and cross-cultural and linguistic adaptations (Table 1). The last of these was not included, as only original language versions were evaluated. Reviewers express their degree of agreement with items with options from 1 (strongly disagree) to 4 (strongly agree); they can also check "no information". The response option "not applicable" is included for a small number of items. Reviewers are required to provide detailed comments justifying or explaining their ratings.

-Table 1 about here-

As a final task for completing the EMPRO, experts are asked to provide a general recommendation for the instrument being evaluated and provide a rationale for their recommendation (from "strongly recommend" to "would not recommend").

Evaluation process

Each HF-specific questionnaire was evaluated by four different expert reviewers who completed the EMPRO assessment based on the evidence available. Reviewers were selected based on their expertise in developing, validating, and/or using PRO measures, had not been involved in the development of any of the HF-HRQL measures being evaluated, and were randomly assigned.

In addition to access to their individual EMPRO tool, reviewers were provided with a copy of the original measure to be assessed, the user manual (if available), and all the identified full-text articles describing the instrument's development or psychometric property assessments, along with any relevant conference information.

After an initial independent evaluation (blind to other reviewers), experts could modify their original ratings in consensus review rounds designed to enhance agreement as much as possible. The appraisal process was carried out via a secure on-line system, and ended after a maximum of two rounds of consensus. Consensus was not necessarily reached, so the process allowed each reviewer to maintain his or her own ratings, which were the basis for the analyses.

Comments registered by appraisers during the evaluation process were reviewed to support the most relevant strengths and weaknesses of the questionnaires.

Analysis

At the end of the assessment period, there were four EMPRO appraisals for each evaluated HF-specific measure. Seven attribute scores (Table 1) and an overall score were obtained from

every EMPRO appraisal. The attribute scoring process consisted of several steps: a) items answered as "no information" were assigned a score of 1 (lowest possible score) if at least 50% of all items for one attribute were rated; b) items rated as "not applicable" were imputed by the mean of the remaining attribute items; c) attribute scores were obtained by calculating the response mean of the corresponding items; and d) attribute scores were linearly transformed to a range of 0 (worst possible score) to 100 (best possible score). In addition to the seven attribute scores, an overall score was calculated for each EMPRO appraisal. This score combined the information from the five most psychometric-related attributes (conceptual and measurement model, reliability, validity, sensitivity to change and interpretability). The overall score was calculated only when at least three out of the five attributes had a score, by imputing 0 to missing attribute scores. EMPRO scores were considered reasonably adequate or acceptable if they reached at least 50 (half of the maximum score).

To summarize the information gathered from each HF-specific measure (four EMPRO appraisals per measure) the data from the four reviewers was presented by median and range.

The reliability of the "multiple-reviewer method" (agreement among appraisers) was tested with the one-way random effects intraclass correlation coefficient (ICC).[11]

Results

Seven questionnaires for evaluating the HRQL of HF patients were identified and a wide range of materials regarding the characteristics of the original versions of those tools was collected (Table 2). The questionnaires were the Chronic Heart Failure Assessment Tool (CHAT)[12], the Cardiac Health Profile-Congestive Heart Failure (CHPchf)[13], the Chronic Heart Failure Questionnaire (CHFQ)[14], the Kansas City Cardiomyopathy Questionnaire (KCCQ)[15], the Left Ventricular Disease Questionnaire (LVDQ)[16], the Minnesota Living with Heart Failure Questionnaire (MLHFQ)[17], and the Quality of Life Questionnaire in Severe Heart Failure (QLQ-SHF)[18]. The number of documents made available to the EMPRO appraisers for the assessment of each questionnaire ranged from one (CHAT) to 21 (MLHFQ), most of which were published papers (Table 2), including full articles (range 1−11) and abstracts (range 1−6) (see Annex). User manuals and/or scoring instructions were available only for the CHFQ, KCCQ, and MLHFQ.

The questionnaires were developed between 1987 and 2007, mainly in English speaking countries (except the CHPchf and the QLQ-SHF, both developed in Sweden). In general, they all share the same objective, but feature different structures: from 36 items in a single domain/score (LVDQ) to 16 items conforming to three domains/scores (CHFQ). All the measures are self-administered, with the exception of the CHFQ.

-Table 2 about here-

Each of the seven questionnaires was evaluated by sets of four different reviewers, resulting in a total of 28 EMPRO appraisals. The results of these assessments are summarized (median and range of reviewer scores) per questionnaire and attribute in table 3. Four questionnaires (CHPchf, CHFQ, KCCQ, and MLHFQ) yielded adequate scores for “conceptual model”, with medians ranging from 52.4 to 61.9 on the EMPRO scale (ranging from 0 to 100) and with individual ratings from 33.3 (for the CHPchf) to 95.2 (for the CHFQ). The QLQ-SHF received

the lowest median scores for this attribute: 16.7 (range 4.8-19.1). For the "reliability" attribute, the LVDQ had the highest median (72.8), while the QLQ-SHF was given the lowest scores (median 23.8). Not enough information was found on reliability for the CHAT to calculate an EMPRO reliability score. The CHPchf, the CHFQ, the KCCQ, and the MLHFQ presented reasonable scores for the "validity" attribute: median values from 50.0 (13.3–77.8) for the CHPchf to 76.4 (33.3–100.0) for the KCCQ. The other three questionnaires were given lower scores on validity (with a median ranging from 12.5 to 40.0). It was impossible to calculate a score for "sensitivity to change" for the CHPchf and the CHAT, and those given to the QLQ-SHF were very low (median 5.6, range 0.0–11.1). In the rest of the instruments, sensitivity to change received the highest scores of all the attributes: from 66.7 (66.7–88-9) for the MLHFQ to 94.4 (55.6–100.0) for the KCCQ. Only the CHFQ and the KCCQ presented adequate scores for the "interpretability" attribute (medians 50.0 (range 22.2–55.6) and 72.2 (range 44.4–88.9), respectively).

As a summary of the metric properties (the five attributes described above), only three out of the seven questionnaires received an acceptable median on the overall score: 59.2 (52.1–71.29) for the CHFQ, 60.7 (55.2–65.9) for the MLHFQ and 64.4 (60.2–81.9) for the KCCQ.

The results of the assessments of the other two independent attributes are also shown in table 3. The EMPRO scores for "burden" ranged from a median of 33.3 (33.3–38.1) for the CHPchf to 54.8 (52.4–61.9) for the MLHFQ. Only one reviewer rated this attribute for the QLQ-SHF (0.0) and no information was available to calculate it for the CHAT. The CHAT, CHPchf, and QLQ-SHF did not have "alternative modes of administration". This attribute was rated as 0.0 for the LVDQ. The KCCQ presented the highest scores for this attribute (median 50.0), followed by the CHFQ and the MLHFQ, with median scores of 41.7 and 33.3, respectively.

- Table 3 about here-

As shown in table 4, none of the four experts who evaluated the CHPchf felt they could make a recommendation due to the lack of information about the instrument. There was also general consensus (among the corresponding set of reviewers) that the QLSHFQ should not be recommended. The LVDQ was either not recommended or recommended with provisos by the four experts that appraised it. Both the CHAT and the CHFQ were provisionally recommended by consensus, while the reviewers were evenly split between recommending the KCCQ and the MLHFQ with provisos and strongly recommending them.

-Table 4 about here-

Discussion

Of the seven existing disease-specific instruments to measure HRQL in patients with heart failure, the most highly rated by experts (following standard criteria) were the KCCQ and the MLHFQ, closely followed by the CHFQ. These three questionnaires are undeniably the most commonly used disease-specific measures for patients with HF.[8] Although knowledge of the instrument might have a greater impact on its use than the quality of the measure itself, in this case our findings on quality coincided with the degree of use of the instruments. But are these instruments being properly used? Are the professionals using them aware of their limitations? This study aimed to shed light on these fundamental issues. Detailed information on the different measurement attributes highlights the adequacy of these instruments compared to one another, and should help professionals when choosing the best tool to use for a specific study context.

The KCCQ is the most recently developed of the three recommended instruments, designed almost 10 years after the CHFQ and the MLHFQ. By the time the KCCQ was published, the two other questionnaires were already quite well known, as were some of their advantages and disadvantages, which may have helped inform the development of the KCCQ. The shorter lifespan of the KCCQ has had an effect on its use in absolute terms; however, it has quickly become quite popular and received the best ratings for validity, sensitivity to change, and interpretability. Although reviewers remarked on the lack of information regarding some aspects of the development process of the KCCQ and on the insufficient information available on the evaluation of its test-retest reliability, the results support a reasonable conceptual model and its other psychometric properties were positively rated.

The MLHFQ has been the most widely used instrument for evaluating HRQL in HF patients internationally. Much has been published on the measurement properties of the original version (reviewed in this study), as well as on versions in other languages, and a large number of studies have used the MLHFQ as a primary or secondary outcome.[19] Five years ago, the MLHFQ was

identified as the questionnaire with the best properties,[8] and even now it has the highest scores for reliability, and a good ability to both measure its objective (validity), and detect change over time. Furthermore, although some appraisers expressed concern about the MLHFQ's domain structure (lack of confirmatory factor analysis[20;21]), the ratings for this attribute were adequate. This was not the case for interpretability, an attribute for which further evaluation is warranted.

The CHFQ features the most well-defined conceptual model, a more than adequate degree of reliability and capacity to detect change over time, and acceptable validity and interpretability. Reviewers remarked on the instrument's particular suitability for use in longitudinal studies. This characteristic is supported by the CHFQ's "individualized profile". It is the only HF-HRQL instrument that allows patients to select the most important activities for them and base some of their responses on those activities.[22] Nevertheless, there are concerns that the need to administer the instrument by means of an interviewer might make it impractical and more time consuming than others, which may in part explain its relatively scarce use more than 20 years after its development.

Among the other four HF-HRQL questionnaires evaluated in this study, a complete assessment of psychometric properties was only possible for the LVDQ and the QLQ-SHF. The results showed quite high scores on reliability and sensitivity to change for the LVDQ, but these were based on only two studies and were accompanied by low ratings for the other measurement attributes. The general consensus was that the instrument was potentially interesting (especially considering its low administration burden) but that it has had only limited use and further study into its characteristics is required. A number of weak points were identified in the development and reporting of the QLSHFQ. Appraisers considered that the instrument's psychometric properties were not only inadequate but had only been "superficially studied". In general, the reviewers felt that more evidence was required in order to make recommendations about the administration of the CHAT or the CHPchf, as their development is still too recent.

Going back to the three HF-HRQL questionnaires that were recommended as a result of the expert evaluations, putting their EMPRO scores in context may help in interpreting those recommendations. The development of the EMPRO included a pilot test with the Spanish versions of the two most well-known generic HRQL questionnaires:[10] the EQ-5D and the SF-36. The EMPRO scores for "reliability" and "responsiveness" were similar or even lower (medians from 50 to 64) than those obtained for the CHFQ, the KCCQ, or the MLHFQ. High EMPRO scores are not easy to reach (i.e. reliability evaluation includes both classical and item response theories).

However, our results should be interpreted taking into account the study's limitations, which can be attributed to different sources. The first of these limitations are those inherent to any systematic retrieval or to the use of EMPRO assessments. The literature review may have failed to identify all the HF-HRQL instruments or related documents, as not all existing databases were searched. Nevertheless, we believe the systematic nature of the review and the sensitive search strategies used probably identified the great majority of available evidence on the characteristics and performance of the instruments evaluated. Evaluation with the EMPRO is based on published literature and is therefore constrained by the quantity and quality of published information available for each instrument. Newer or less frequently used instruments may be penalized because of the lack of evidence on some of their properties. That being said, most of the instruments assessed here were published before 2000, meaning that enough time has passed for considerable data on their psychometric characteristics to have been published. Secondly, the EMPRO ratings may be biased by the individual expertise of the evaluators, although the sets of four reviewers randomly assigned to each questionnaire may have attenuated this. The degree of inter-rater agreement among reviewers was reasonably adequate (ICC=0.8).

Finally, the study presented here had the limitation of only evaluating original versions of the HF questionnaires, so a specific assessment should be conducted if there is special interest

regarding any country-specific version. However, including the studies of different country versions could introduce noise to the evaluation process as poor EMPRO scores could be due to a non-rigorous adaptation process rather than to the original instrument. Moreover, the authors would like to point out the fact that such an assessment (with the EMPRO tool) may be complemented with an evaluation of the quality of the studies from which data is extracted. The COSMIN (consensus-based standards for the selection of health measurement instruments) was recently developed with the participation of a group of international experts, as a checklist for evaluating the methodological quality of studies that assess the properties of an instrument, not the quality of the instrument itself.[23] Therefore, even the EMPRO was chosen to accomplish our research objectives, in further broader projects those objectives might be expanded now that both tools are available.

In conclusion, this is the first study to provide a systematic and reliable expert-based evaluation of available disease-specific quality of life questionnaires for use in patients with heart failure. The evidence would currently support a preference for the use of three instruments: the MLHFQ, the KCCQ, and the CHFQ. Choosing between them will depend on particular study needs and requirements, and our results may facilitate decisions in the instrument selection process. Further testing and/or improved development is needed before the other instruments can be recommended based on the evidence available.

Acknowledgements

We would especially like to thank Dave McFarlane for his assistance in managing our initial databases and Àngels Pont and Áurea Martín-Morris for their help in editing the last draft of this manuscript. This work was supported by a grant from the Instituto de Salud Carlos III FEDER, (PI06/90483) and Recercaixa (2010ACUP 00158).

Author Disclosure

O.Garin, M.Herdman, G.Vilagut, M.Ferrer, A.Ribera, L.Rajmil, JM.Valderas, F.Guillemin, D.Revicki, J.Alonso, and all other members from the EMPRO-HF Group have no conflicts of interest or financial ties to disclose.

References

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

(2) Zuluaga MC, Guallar-Castillon P, Lopez-Garcia E, Banegas JR, Conde-Herrera M, Olcoz-Chiva M, Rodriguez-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(12):1372-1378.

(3) Iqbal J, Francis L, Reid J, Murray S, Denvir M (2010) 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 12(9):1002-1008.

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

(5) Flynn KE, Lin L, Moe GW, Howlett JG, Fine LJ, Spertus JA, McConnell TR, Pina IL, Weinfurt KP (2012) Relationships between changes in patient-reported health status and functional capacity in outpatients with heart failure. Am Heart J 163(1):88-94.

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

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

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

(9) 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(3):193-205.

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

(11) Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86(2):420-428.

(12) Dunderdale K, Thompson DR, Beer SF, Furze G, Miles JN (2008) Development and validation of a patient-centered health-related quality-of-life measure: the chronic heart failure assessment tool. J Cardiovasc Nurs 23(4):364-370.

(13) Mannheimer B, Andersson B, Carlsson L, Wahrborg P (2007) The validation of a new quality of life questionnaire for patients with congestive heart failure-an extension of the Cardiac Health Profile. Scand Cardiovasc J 41(4):235-241.

(14) 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(2):101-107.

(15) 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(5):1245-1255.

(16) O'Leary CJ, Jones PW (2000) The left ventricular dysfunction questionnaire (LVD-36): reliability, validity, and responsiveness. Heart 83(6):634-640.

(17) 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.

(18) 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(3):220-225.

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

(20) 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(4):265-272.

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

(22) Dijkers MP (2003) Individualization in quality of life measurement: instruments and approaches. Arch Phys Med Rehabil 84(4 Suppl 2):S3-14.

(23) Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) 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 19(4):539-549.

Table 1. EMPRO attributes: name, number of items, and definition; together with examples of scoring meaning.

Attribute	**Definition**	**Num. items**	**Higher scores mean… (ranging from 0 to 100)**
Conceptual and measurement model	The rationale for and description of the concept and the populations that a measure is intended to assess and the relationship between these concepts.	7	*- The concept is more clearly stated to be measured.* *- The empirical basis and methods for obtaining the item and for combining them are more appropriate.*
Reliability	The degree to which an instrument is free from random error.	8	*- More clearly described and more adequate methods to collect internal consistency data* *- Better values of Cronbach's alpha and/or KR-20 coefficients*
Validity	The degree to which the instrument measures what it purports to measure.	6	*- More evidence regarding content-related validity of the instrument for its intended use.*
Sensitivity to change	An instrument's ability to detect change over time.	3	*- More clearly described and more appropriate methods to assess responsiveness.* *- The estimated magnitude of change is more clearly described and the results are better.*
Interpretability	Possibility of assigning meaning to quantitative scores.	3	*- The strategies to facilitate interpretation are more clearly described and appropriate.*
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	*- The skills and time to complete the instrument are more clearly described and acceptable.*
Alternative modes of administration	Alternative modes of administration used for the administration of the instrument.	2	*- The metric characteristics and use of each alternative mode of administration are specifically described and adequate.*

Table 2. Characteristics of disease-specific instruments identified to assess health related quality of life in heart failure (HF).

Instrument	Aim: To measure…	Country and year of development	Specific domains (no. of items)	Overall score (no. items)	Score range	Original mode of administration (alternative)	No. documents used in appraisal‡
Chronic Heart Failure Assessment Tool **CHAT**	…health related quality of life in HF from the patient perspective	UK 2007	Symptoms** Activity levels** Psycho-social** Emotions**	Total (46)	(A variety of response scales are used) †	Self-administered	Full articles :2
Cardiac Health Profile congestive heart failure **CHPchf**	…how HF influences subjective perceptions of physical, psychological and social well-being	Sweden 2007	- -	Total (10)*	(Visual Analogue Scales)†	Self-administered	Full articles: 1
Chronic Heart Failure questionnaire **CHFQ**	…longitudinal change over time within persons with HF.	Canada 1989	Dyspnea (5) Fatigue (4) Emotional (7)	Total (16)	16 – 112 Worst to best	Interview-administered (Telephone)	Full articles: 7 Abstracts: 1
Kansas City Cardiomyopathy questionnaire **KCCQ**	…health related quality of life in HF.	USA 1999	Physical limitation (6) Symptoms (8) Self-efficacy (2) Social limitation (4) Quality of life (3)	Total (23)	0 – 100 Worst to best	Self-administered (Telephone)	Full articles: 11 Abstracts: 1
Left Ventricular Disease Questionnaire **LVDQ**	…the impact of left ventricular dysfunction on daily life and wellbeing	UK 1998	- -	Total (36)	0 – 100 Worst to best	Self-administered (Telephone)	Full articles: 1
Minnesota Living with Heart Failure questionnaire **MLHFQ**	…the extent to which HF prevents patients from living the way they would want to.	USA 1987	Physical (8) Emotional (5)	Total (21)	0 – 105 Best to worst	Self-administered (Telephone)	Full articles: 11 Abstracts: 6
Quality of Life Questionnaire in Severe Heart Failure **QLQ-SHF**	…self-assessment of health-related quality of life in severe HF	Sweden 1987	Psychological (7) Physical activity (7) Life-dissatisfaction (5) Somatic symptoms (7)	Total (26)	0 – 130 Best to worst	Self-administered	Full articles: 1 Abstracts: 2

*This scale is a disease-specific add-on for the Cardiac Health Profile. The heart failure specific scale includes 10 items; **Number of items per dimension not specified. †Response option type, when response range was not available. ‡ in addition to the questionnaire, its manual (if available) and the summary of results of the previous systematic review (REF)

Table 3. EMPRO attribute scores (ranging from 0 to 100) for the HF-HRQL instruments evaluated: median score (observed score ranges)

EMPRO attribute	CHAT	CHPchf	CHFQ	KCCQ	LVDQ	MLHFQ	QLQ-SHF
Concept and measurement model	45.2 (28.6 – 61.9) n=4	61.9 (33.3 – 66.7) n=4	64.3 (57.1 – 95.2) n=4	52.4 (42.9 – 76.2) n=4	42.9 (33.3 – 52.4) n=3	54.8 (42.9 – 66.7) n=4	16.7 (4.8 – 19.1) n=4
Reliability	--	27.8 (11.1 – 44.4) n=2	61.1 (44.4 – 80.0) n=4	52.8 (44.4 – 72.2) n=4	72.8 (61.1 – 80.0) n=4	69.4 (66.6 – 72.2) n=4	23.8 (6.6 – 26.6) n=3
Validity	40.0 (20.0 – 60.0) n=4	50.0 (13.3 – 77.8) n=4	54.4 (46.7 – 58.3) n=4	76.4 (33.3 – 100.0) n=4	38.3 (25.0 – 86.7) n=4	66.8 (60.0 – 66.8) n=4	12.5 (6.7 – 20.0) n=4
Sensitivity to change	--	--	72.2 (66.6 – 77.8) n=4	94.4 (55.6 – 100.0) n=4	72.2 (55.6 – 77.8) n=4	66.7 (66.7 – 88.9) n=4	5.6 (0.0 – 11.1) n=4
Interpretability	--	11.1 (– 11.1) n=1	50.0 (22.2 – 55.6) n=4	72.2 (44.4 – 88.9) n=4	11.1 (0.00 – 22.2) n=2	44.4 (33.3 – 44.4) n=4	0.0 (– 0.0) n=1
Overall score †	--	25.8 (13.8 – 37.8) n=2	59.2 (52.1 – 71.2) n=4	64.4 (60.2 – 81.9) n=4	45.8 (28.7 – 60.3) n=4	60.7 (55.2 – 65.9) n=4	12.1 (6.5 – 14.8) n=4
Burden	--	33.3 (33.3 – 47.6) n=3	38.1 (28.6 – 38.1) n=3	38.1 (24.0 – 52.4) n=4	52.4 (42.9 – 61.9) n=4	54.8 (52.4 – 61.9) n=4	0.0 (– 0.0) n=1
Alternative modes of administration	--	--	41.7 (16.7 – 50.0) n=4	50.0 (-- - 50.0) n=3	0.0 (– 0.0) n=2	33.3 (33.3 – 33.3) n=3	--

† Overall score calculated as the mean of all the items of 5 psychometric related attributes (conceptual and measurement model, reliability, validity, sensitivity to change and interpretability). The overall score was calculated only when at least 3 out of the 5 attributes had a score; and in those cases if an attribute was missing a 0 was imputed.

Table 4. Overall recommendation and reviewers' comments for each HF-HRQL instrument

Instrument	Overall recommendation	Selected verbatim reviewers comments
CHAT	Recommend with provisos	• "poor description about the samples and methods for calculating coefficients" **Overall comments:** Additional information required regarding interpretation, longitudinal assessment, scoring, etc.
CHPchf		• "need more information for scoring" • "lack of measurement model evaluation" **Overall comments:** Not enough information to determine the degree of recommendation
CHFQ	Recommend with provisos	• "No information about methods for including items in given dimensions" • "No sample size estimation or information on sample characteristics" • "No intraclass correlation coefficient [for reliability estimates]" **Overall comments:** Particularly recommended for longitudinal studies. Take into account the mode of administration (interviewer).
KCCQ	Strongly recommend or recommend with provisos	• "not sure about item reduction, factor analysis, used to derive final 23 items [in] 5 domains" • "detailed test-retest description" • "poor description of a priori hypotheses were made [in testing construct validity]" **Overall comments:** The KCCQ seems to be more responsive regarding improvement than deterioration.
LVDQ	Recommend with provisos or do not recommend	• "Unidimensionality of the underlying concept is questionable" • "One third of patients reported the instrument was missing important aspects of their well-being" • "More work needs to be done before scores on this instrument can be reliably interpreted" **Overall comments:** Additional work required before it can be usefully applied. Replication in different sets of patients is needed.
MLHFQ	Strongly recommend or recommend with provisos.	• "Limited patient involvement" • "Insufficient documentation of how items, domain and conceptual framework were identified" • "Investigated in many papers, settings and population, with correct approach" **Overall comments:** More work necessary to determine whether 16 or 21 item is preferable. Practical and widely-used instrument.
QLQ-SHF	Not recommended	• "Target population was not involved at all" • "No information on Cronbach's alpha for the subscales" • "No information provided on magnitude of change" **Overall comments:** Although the instrument is potentially interesting, it was only superficially studied.

Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures

Olatz Garin MPH [1,2,3], Michael Herdman MSc,[1,2] Gemma Vilagut BSc, [1,2] Montse Ferrer MD, PhD,[1,2,4] Aida Ribera MD, PhD, [1,5] Luis Rajmil MD, PhD, [1,6] José M Valderas MD, PhD, [1,7] Francis Guillemin MD, PhD,[8] Dennis Revicki PhD, [9] Jordi Alonso MD, PhD[1,2,3] on behalf of the EMPRO-HF Group*

1. CIBER en Epidemiologia y Salud Pública (CIBERESP), Spain
2. Health Services Research Unit, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
3. Universitat Pompeu Fabra (UPF), Barcelona, Spain
4. Universitat Autònoma de Barcelona (UAB), Barcelona, Spain
5. Unidad de Epidemiología, Servicio de Cardiología, Hospital Vall d'Hebron, Barcelona
6. Agència d'Informació, Avaluació i Qualitat en Salut (AIAQS), Barcelona, Spain
7. Health Services and Policy Research Group, Department of Primary Health Care Sciences, University of Oxford, UK
8. Service d'epidémiologie et évaluation cliniques, Centre d'Epidémiologie Clinique, INSERM CIE6, CHU de Nancy, France
9. Center for Health Outcomes Research, United BioSource Corporation, Bethesda, Maryland, USA

*EMPRO-HF Group: Allá F, Arnould B, Becerra V, Barblaglia G, Castellví P, Clayson D, Erhart M, Fairclough D, Fleck M, Garcia-Forero C, Gonçalves DC, Gotay C, las Hayas C, Lloyd A, Mathias S, Permanyer-Miralda G, Pont A, Prieto L, Ravens-Sieberer U, Schmidt S, Villalonga-Olives E

Address for correspondence:

Dr. Jordi Alonso

Health Services Research Unit, IMIM (Hospital del Mar Medical Research Institute), Doctor Aiguader 80, 08003 Barcelona, Spain

Tel.: +34 93 316 0740; Fax: +34 93 221 40 02; e-mail: *jalonso@imim.es*

Reviewed articles of existing HRQL questionnaires for HF patients:

1. 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;23:364-370.

2. Dunderdale, K., Furze, G., Thompson, D. R., Beer, S. F., and Miles, J. N. Health-related quality of life from the perspective of patients with chronic heart failure. The British Journal of Cardiology 14(4), 207-212. 2012.

3. Mannheimer B, Andersson B, Carlsson L, Wahrborg P. The validation of a new quality of life questionnaire for patients with congestive heart failure-an extension of the Cardiac Health Profile. *Scand Cardiovasc J*. 2007;41:235-241.

4. 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.

5. Wyrwich KW, Spertus JA, Kroenke K, Tierney WM, Babu AN, Wolinsky FD. Clinically important differences in health status for patients with heart disease: an expert consensus panel report. *Am Heart J*. 2004;147:615-622.

6. O'Keeffe ST, Lye M, Donnellan C, Carmichael DN. Reproducibility and responsiveness of quality of life assessment and six minute walk test in elderly heart failure patients. *Heart*. 1998;80:377-382.

7. Wolinsky FD, Wyrwich KW, Nienaber NA, Tierney WM. Generic versus disease-specific health status measures. An example using coronary artery disease and congestive heart failure patients. *Eval Health Prof*. 1998;21:216-243.

8. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care*. 1999;37:469-478.

9. Oka RK, De Marco T, Haskell WL, Botvinick E, Dae MW, Bolen K, Chatterjee K. Impact of a home-based walking and resistance training program on quality of life in patients with heart failure. *Am J Cardiol*. 2000;85:365-369.

10. Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P, McCullough PA, Pina I, Tooley J, Weintraub WS, Rumsfeld JS. Monitoring clinical changes in patients with heart failure: a comparison of methods. *Am Heart J*. 2005;150:707-715.

11. Soto GE, Jones P, Weintraub WS, Krumholz HM, Spertus JA. Prognostic value of health status in patients with heart failure after acute myocardial infarction. *Circulation*. 2004;110:546-551.

12. 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.

13. Hauptman PJ, Masoudi FA, Weintraub WS, Pina I, Jones PG, Spertus JA. Variability in the clinical status of patients with advanced heart failure. *J Card Fail*. 2004;10:397-402.

14. Spertus JA, Jones PG, Kim J, Globe D. Validity, reliability, and responsiveness of the Kansas City Cardiomyopathy Questionnaire in anemic heart failure patients. *Qual Life Res*. 2008;17:291-298.

15. 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.

16. Flynn KE, Lin L, Ellis SJ, Russell SD, Spertus JA, Whellan DJ, Pina IL, Fine LJ, Schulman KA, Weinfurt KP. Outcomes, health policy, and managed care: relationships between patient-reported outcome measures and clinical measures in outpatients with heart failure. *Am Heart J*. 2009;158:S64-S71.

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

18. Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. Pimobendan Multicenter Research Group. *Am Heart J*. 1992;124:1017-1025.

19. 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.

20. Rector TS, Kubo SH, Cohn JN. Validity of the Minnesota Living with Heart Failure questionnaire as a measure of therapeutic response to enalapril or placebo. *Am J Cardiol*. 1993;71:1106-1107.

21. Gorkin L, Norvell NK, Rosen RC, Charles E, Shumaker SA, McIntyre KM, Capone RJ, Kostis J, Niaura R, Woods P, . Assessment of quality of life as observed from the baseline data of the Studies of Left Ventricular Dysfunction (SOLVD) trial quality-of-life substudy. *Am J Cardiol*. 1993;71:1069-1073.

22. Bennett LS, Barbour C, Langford B, Stradling JR, Davies RJ. Health status in obstructive sleep apnea: relationship with sleep fragmentation and daytine sleepiness, and effects of continuous positive airway pressure treatment. *Am J Respir Crit Care Med*. 1999;159:1884-1890.

23. Bennett SJ, Oldridge NB, Eckert GJ, Embree JL, Browning S, Hou N, Chui M, Deer M, Murray MD. Comparison of quality of life measures in heart failure. *Nurs Res*. 2003;52:207-216.

24. Chua R, Keogh AM, Byth K, O'Loughlin A. Comparison and validation of three measures of quality of life in patients with pulmonary hypertension. *Intern Med J*. 2006;36:705-710.

25. 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;54:265-272.

26. Riegel B, Moser DK, Glaser D, Carlson B, Deaton C, Armola R, Sethares K, Shively M, Evangelista L, Albert N. The Minnesota Living With Heart Failure Questionnaire: sensitivity to differences and responsiveness to intervention intensity in a clinical population. *Nurs Res*. 2002;51:209-218.

27. Sneed NV, Paul S, Michel Y, Vanbakel A, Hendrix G. Evaluation of 3 quality of life measurement tools in patients with chronic heart failure. *Heart Lung*. 2001;30:332-340.

28. Ni H, Toy W, Burgess D, Wise K, Nauman DJ, Crispell K, Hershberger RE. Comparative responsiveness of Short-Form 12 and Minnesota Living With Heart Failure Questionnaire in patients with heart failure. *J Card Fail*. 2000;6:83-91.

29. 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.

30. Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health Qual Life Outcomes*. 2006;4:89.

31. Masoudi FA, Rumsfeld JS, Havranek EP, House JA, Peterson ED, Krumholz HM, Spertus JA. Age, functional capacity, and health-related quality of life in patients with heart failure. *J Card Fail*. 2004;10:368-373.

32. Myers J, Zaheer N, Quaglietti S, Madhavan R, Froelicher V, Heidenreich P. Association of functional and health status measures in heart failure. *J Card Fail*. 2006;12:439-445.

33. 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.