Development of a Preference-based Heart Disease-specific Health State Classification System using MacNew Heart Disease Related Quality of Life Instrument

Sanjeewa Kularatna PhD[1], Donna Rowen PhD[2], Clara Mukuria PhD [2],Steven McPhail PhD[1,3], Gang Chen PhD[4], Brendan Mulhern PhD[5], Jennifer A. Whitty PhD[6,7], Joshua Byrnes PhD[8], Paul Scuffham PhD[8], John Atherton[9], Stefan Höfer PhD[10], William Parsonage BM BS, DM, MRCP [1,9]

[1]Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, School of Public Health and Social Work, Queensland University of Technology

[2]School of Health and Related Research, University of Sheffield

[3]Clinical Informatics Directorate, Metro South Health

[4]Centre for Health Economics, Monash University

[5]University of Technology Sydney

[6] Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, Norfolk, United Kingdom

[7] National Institute for Health Research (NIHR) Applied Research Collaboration (ARC), East of England, United Kingdom

[8]Centre for Applied Health Economics, Griffith University

[9]Royal Brisbane and Women's Hospital, Metro North Hospital and Health Services

[10]Medical University Innsbruck, Innsbruck, Austria

Corresponding Author : Sanjeewa Kularatna, Australian Centre for Health Services Innovation, 60 Musk Avenue, Kelvin Grove Queensland 4156, Australia.

Email: sanjeewa.kularatna@qut.edu.au

Word Count: 4200

Abstract

Purpose

The MacNew Heart Disease Health Related Quality of Life Instrument (MacNew) is a validated, clinically sensitive, 27-item disease-specific questionnaire. This study aimed to develop a new heart-disease-specific classification system for the MacNew amenable for use in health-state valuation.

Methods

Patients with heart-disease attending outpatient clinics and inpatient wards in Brisbane, Australia, completed MacNew.The development of the new disease-specific classification system included three stages. First, a principal component analysis (PCA) established dimensionality. Second, Rasch analysis was used to select items for each dimension. Third, Rasch analysis was used to explore response-level reduction. In addition, clinician and patient judgement informed item selection.

Results

Participants included 685 patients (acute coronary 6%, stable coronary 41%, chronic heart failure 20%). The PCA identified 4 dimensions (restriction, emotion, perception of others, and symptoms). The restriction dimension was divided into physical and social dimensions. One item was selected from each to be included in the classification system. Three items from the emotional dimension and two symptom items were also selected. The final classification system had seven dimensions with four severity levels in each: physical restriction; excluded from doing things with other people; worn out or low in energy; frustrated, impatient or angry; unsure and lacking in self-confidence; shortness of breath; and chest pain.

Conclusion

This study generated a brief heart disease-specific classification system, consisting of seven dimensions with four severity levels in each. The classification system is amenable to valuation to enable the generation of utility value sets to be developed for use in economic evaluation.

Introduction

Heart disease is the most costly of all chronic disease. The burden of heart disease is highlighted by Centre for Disease Control and Prevention in the US indicating costs of $219 billion (1) including healthcare, medicine and loss of productivity between 2014 and 2015 (2). With ageing populations in many developed countries, these costs are expected to continue to increase significantly in the future (3). New research developments and a continuous influx of new technology also generate increasing demands and expectations for new interventions for the prevention and treatment of heart disease, which is challenging in the context of healthcare resource scarcity (4). This is posing a growing challenge for decision makers in allocating resources and subsidising important care among competing demands. Health economic evaluation assesses the incremental benefits and costs of new interventions which can provide crucial information on value for money to support decision-making (5).

One well established outcome measure in economic evaluation is the quality-adjusted life year (QALY). The "quality" represents the health state utility value attached to different health states which lie on a 0-1 death-full health QALY scale. These utility values are normally derived from national-level valuation surveys of the general public which assess the preference for each health state (6). Typically, QALYs are generated using multi-attribute utility instruments (MAUIs) (also referred to as preference-based measures) which consist of a health state classification system with a corresponding utility value set (7). By far the most widely used generic MAUI is the EQ-5D (8). However, there are compelling arguments that, for some diseases, generic MAUIs are not sensitive enough to measure the change in utility (9, 10). A recent review of methods to derive QALYs in trial-based cost-utility analysises in heart failure found that EQ-5D was the most widely used measure, but that overall there were similar proportions of studies finding significant, non-significant and unreported significance in differences between interventions in QALYs (10). Whilst this may be because the interventions

were ineffective, there is a concern that the use of generic preference-based measures such as EQ-5D in economic evalution in heart failure may mean that the full impact of the intervention has not been captured due to the psychometric performance of EQ-5D in this patient population (11). Whilst condition-specific measures are used in trials, to generate utility values these are often mapped to the generic preference-based measure to generate utility values. However, in this mapping process the sensitivity of the condition-specific measure will not be reflected unless this is also captured in the generic preference-based measure(12, 13). Use of a condition-specific measure that is sensitive to generate utility values for QALYs directly is therefore an advantage.

Heart disease is a collection of complex clinical syndromes presented as heart failure, atrial fibrillation and angina pectoris, among others. Measuring quality of life in patients with heart disease with complex symptoms is challenging for any generic instrument such as the EQ-5D (11). In contrast, disease specific quality of life instruments in heart disease collect specific information such as shortness of breath/ chest pain and of frequency and impact on sleep(14). Additionally, emotional, and social impact of heart disease are difficult to quantify with the limited questions posed by the EQ-5D (10). The inability to participate in social activities, feeling excluded, worn out due to constant effect of chronic heart conditions are some of the health related quality of life domains, heart disease specific instruments are capable of capturing (14). As such, there are calls for inclusion of disease specific measures when capturing effectiveness of interventions in heart conditions(15). While available disease specific measures are sensitive to changes of quality of life heart disease, none are preference based. Generic measures may lack sensitivity to changes as they do not capture important symptoms of heart disease such as breathlessness, loss of self-control and tiredness (10).Moreover, a recent review suggested sensitivity of the instrument measuring quality of life as an important factor in determining cost effectiveness of an intervention (10) .

There are number of heart disease specific instruments covering various facets of such as heart failure (16) and angina(17). The MacNew Heart Disease Health Related Quality of Life Instrument [MacNew] is one of the most widely used heart disease-specific instruments (18) and has been found to be clinically sensitive and responsive in detecting change in heart disease (19). Reliability, validity, responsiveness and interpretability of this instrument has been tested (18-20). However, the MacNew is not a preference-based instrument and cannot be used directly to calculate QALYs which is required in economic evaluation. Additionally, in its existing form the Macnew has too many questions to be able to be reliably valued. Although there have been mapping functions reported to predict health state utility values elicited from generic preference-based instruments, since the targeting instruments are not heart disease-specific instruments, these mapping functions from MacNew may not be ideal (13). To the best of our knowledge, no single preference-based heart disease-specific instrument is available. As economic evaluations underpin billions of dollars of public and private healthcare investment and expenditure, it is imperative that accurate measurement and valuation of changes in health for heart disease patients exists. There is evidence disease specific preference based measures can accurately measure milder health states and differences have been observed in mean chnage and standard deviation of utility values between generic and disease specific (21). It is important to ascertain if heart disease specific utility values would have impact on the resource allocation decisions. This is the first step in the journey to determine if more sensitive and nuanced heart disease specific utility weights would translate to saving money for health system in the future. This study fills in the gap in the literature by developing a heart disease-specific health state classification system from the MacNew that is amenable to valuation. This will facilitate the development of a new preference-based heart disease-specific instrument. Preference based measures (PBM) have two components;

classification system and utility value set. Aim of this was to develop a heart disease specific classification system.

Methods

The study collected data from patients with cardiovascular conditions using four heart disease-specific quality of life instruments: the MacNew Heart Disease Health Related Quality of Life Instrument (18), the Minnesota Living with Heart Failure Questionnaire (16), Seattle Angina Questionnaire (17), and The Kansas City Cardiomyopathy Questionnaire (22), as well as two generic instruments the EQ-5D-5L and SF-36 with some sociodemographic information. The data were collected between June 2017 and June 2018 at the Royal Brisbane and Women's Hospital (RBWH), Brisbane, Australia. Participants were approached in the RBWH cardiac outpatient clinics and hospital wards and invited to take part in the study. Adults above 18 years age, attending cardiology clinic or warded for treatment of any chronic heart condition were eligible for the study. The self reported heart disease category was confirmed by treating physician. They were asked to complete the questionnaire at the clinic or were given the option to complete the questionnaire at home and post it to the study team. Ethical approval was obtained from the RBWH Human Research Ethics Committee (HREC/17/QRBW/418).

MacNew

The MacNew instrument has shown strong pshycometric properties(19) and unlike the other three instruments, it does not ask patients to consider a specific heart disease. The MacNew was evolved from the Quality of Life after Myocardial Infarction (QLMI) (23) which captured quality of life changes in patients who underwent cardiac rehabilitation after an acute myocardial infarction (MI). MacNew consists of 27 items within three dimensions: physical,

emotional and social. MacNew uses a Likert type item scoring system with 7 as the best and 1 as the worst quality of life.

Analysis

Development of the new classification system involved a three-stage process - Stage I: establish dimensionality; Stage II: eliminate and select items per dimension; Stage III: explore item-level reduction. This process has been used previously to derive the SF-6D from the SF-36 (24), and since used to generate several condition-specific preference based measures (25-28).

Frequency distributions of the MacNew items to measure floor and ceiling effects and domain scores (global, physical, social and emotional) were firstly investigated. Spearman's correlation coefficients were used to examine whether individual items had a substantial correlation with each other, as the reliability of a scale is based on the strength of the average inter-item correlation (29). Mean inter-item correlations and corrected item total correlations were also checked. It has been suggested that inter-item correlations above 0.3 and below 0.7 are desirable (29). Corrected item total correlations indicate how well each item correlates with the total of the other items in the scale.

Stage 1: Establish dimensionality

A principal component analysis (PCA) was used as the extraction tecnique. Items 7 ("Happy, satisfied, pleased"), 11 ("felt dependent on "), 13 ("others confidence ") and 27 ("sexual intercourse"") were removed prior to the PCA as these items were deemed to not contribute to the quality of life of patients but rather reflect on patient confidence and satisfaction. In addition, majority of respondents rated not applicable to items 27 rendering it non relevant for the analysis. Initially, a three-component solution was followed according to the original theoretical domain structure of the instrument (18). However, a more recent

empirical study resulted in five components: emotion, restriction, symptoms, perception of others, and social (19). We generated PCAs for 3-6 components in different iterations to seek the optimal solution. We followed the criteria recommended by Tabachnick and Fidell (30) in experimenting with different extraction techniques, considering various number of components, and applying both oblique and orthogonal types of rotation techniques, until the optimal and meaningfulness result dimension structure was obtained. This was following their instruction to obtain "solution with greatest utility, consistency and meaning" (30).In deciding the number of components we used Kaiser's criterion, Catell's screeplot, percentage of variance and parallel analysis (using Monte Carlo PCA) as analytical criteria. We used previous published PCA on MacNew as well as clinical and patient advice to guide the meaning of components we extracted (19). The PCA was conducted using SPSS software.

Stage II: Eliminate and select items per dimension

Within each dimension resulting from stage I, we conducted Rasch analysis (31) to reduce the items. Rasch is the unidimensional variant of the Item Response Theory models and convert categorical responses to continuous latent scale (32). All Rasch analyses were conducted using RUMM2030 software (31). We considered goodness of fit of each item within the fitted model for each dimension with range, spread, how the response was ordered, as well as the differential item functioning (DIF) for gender (26). Goodness of fit was assessed using the individual item fit (33). Fit residual values above 2.5 and significant probability values below the Bonferroni adjustment (probability base 0.05) were considered to determine non-fitting items (34). A respondent with a high level of the latent trait within each dimension being measured would endorse high scoring responses on each item, and vice versa. This was presented in the threshold map of the analysis. DIF occurs when different groups within the sample (e.g. males versus females) respond in a different manner to an individual item despite no difference in underlying health. Generally, once an item was non-fitting to a model it was

removed, and the model was re-estimated; however, we also investigated the potential reason for the non-fitting. It could have been that the item should be better grouped into a different dimension, or perhaps it represented a standalone dimension (for example, for symptoms). Ceiling and floor effects of each item were considered during item selection for each dimension (35). We also examined item wording and its suitability for valuation. This resulted in a draft classification system.

Stakeholder engagement for item selection

Patient opinion and the collective opinion of the research team were also considered for item selection. A proposed classification system was discussed with the clinical team to ascertain the clinical importance of the selected items and to identify any omitted items that should be included due to their clinical importance. The MacNew developers were also consulted at this stage. The proposed health state classification was then presented to cardiologists, nursing and allied health practitioners (n=20) who interact with patients with heart disease daily. They were presented with unfinalized set of items in each dimension and were asked to select items based on their experience with patients. A revised classification system was next presented to a health consumer advisory group (n=12) for heart disease, consisting of senior members of the society with numerous chronic heart diseases. This part of the study was conducted as a focus group discussion with a researcher experienced in the methods facilitating. The purpose of selecting itmes was explained to them and one by one the selected items were presented for their opinion. We also asked if any important aspects of QoL in heart disease had been missed.

Stage III: Explore response-level reduction

In valuation, respondents only see one level for each item in the classification system e.g. ‘none of the time’. The MacNew has different wordings for their items, ranging from “all

of the time" to "none of the time", "extremely limited" to "not limited at all" and "extreme shortness of breath" to "no shortness of breath", with seven response levels each. The objectiveof stage III was to assess the possibility of reducing the original seven response levels to four levels in each item as seven levels may not all be distinguished in valuation. Rasch analysis and distribution of levels in selected items were used to guide the collapsing of adjacent levels and examine the ability of the respondents to identify the order of the levels. Threshold probability curves that give information on the distribution of item levels across the latent space were examined (33). Domain-specific Rasch models were conducted with all the items per dimension including the excluded items. The levels were merged, taking into consideration the threshold probability curves, frequency distribution of the levels, threshold maps and item goodness of fit statistics of the proposed merged levels (33).

Results

A total of 685 respondents participated this study (Table 1). Patients were recruited mainly from RBWH cardiology outpatient clinic with chronic heart disease conditions. The others were from wards and electrocardiography appointments. The majority (67%) were male. The age range was from 18 to 97 with a mean age of 62 (SD13.5) years. Stable coronary disease and chronic heart failure were the most common diseases reported. The mean MacNew Global and scale scores ranged from 5.3/7 (emotion) to a 5.6/7 (social) with global scale reporting 5.4/7 with higher value indicating better quality of life.

Stage I: Establish dimensionality

The best suitable method of rotation for this data set was obliquel rotational, with direct oblimin technique. Direct Oblimin is suitable here as correlation between components are assumed and this method use a parameter to control the degree of correlation between the

components. The PCA indicated a four-component solution (Table 2). In our analysis, components loaded into emotion, restriction, symptoms and perception of others (Table 2). All symptom items loaded together into one dimension (Table 2). We divided the restriction dimension into physical and social dimensions (as they are important aspects of quality of life) creating final five dimensions; physical restriction, social restriction, emotion, perception of others and symptoms.

Stage II: Eliminate and select items per dimension

Table 3 reports the Rasch and psychometric analyses. The psychometric analysis displayed that all selected items (n=23, after exclusions) performed well based on missing data (n=0), floor effects and internal consistency. However, we observed substantial ceiling effects on several items (> 50%) (Table 3), in particular items 10 (tearful), 14 (chest pain), 18 (frightened), 22 (family overprotective), 23 (burden on others), 24 (felt excluded) and 25 (unable to socialise). All items in the restriction dimension, including the social and physical components, performed well. All symptoms except chest pain also showed good psychometric properties. Items with high levels of floor and ceiling effects were further investigated if they were appropriate to include.**Physical dimension**

Items 17, 20 and 26 were selected for this physical dimension (Table 3) from the restriction component. Although item 21 (unsure of exercise) was initially also considered, it was excluded as it was about uncertainty rather than physical activity. The three remaining items were modelled using Rasch. All three items fitted well, except that item 17 (limited in doing exercise) showed a discrepancy in the threshold map showing the ordering across the seven responses. Cross-tabulation of item 20 (restricted) and item 26 (physically restricted) showed that they are suffering from local dependency, and are similar in capturing physical

restriction concerning quality of life. As the adjective “physical” is available in item 26 there was consensus among clinicians and academics that item 26 best describes the physical domain of quality of life, hence item 26 was selected.

Social dimension

There were three items selected from the restriction component which describe the social aspect of quality of life; items 12 (social activities), 24 (felt excluded) and 25 (unable to socialise). In item 12, the words “usual” and “family” might not relate to everyone, as some people might not have family and others might not do their social activities with family. Therefore, item 12 was excluded. Items 24 and 25 were found to be covering the same quality of life space when they were cross-tabulated, and found to have similar answers, making one of them redundant. We seleted item 24 after consulting with clincians and patients.

Emotion dimension

Ten items loaded into this dimension in the PCA (Table 2). As items 3 (confident), 4 (discouraged) and 6 (worn out /low in energy) were not fitting they were excluded, and Rasch model was fitted again using the remaining seven items. These seven items fitted well together. Both items 1 and 15 have low floor and ceiling effects, ordered responses, and no DIF with a wide range and spread. There was consensus across clinicians, patients and researchers that item 1(frustrated) should be in the final instrument. Item 15 (lack of self-confidence) also was selected based on the advice from the clinical team. Although not fitting, it was ultimately decided item 6 (worn out) should be included in the final instrument, due to its minimum floor and ceiling effects, good internal consistency and ordered responses. Item 6 description “felt worn out or low in energy” was considered as an important description of aspects of the emotional attributes of quality of life. We ultimately selected three items from this dimension; items 1, 6 and 15.

Perception of others

There were two items that loaded into 'perception of others' (burden on the family). Items 22 (family overprotective) and 23 (burden on others) were the components of this dimension. The wording "family" in item 22 may not relate to everyone, and high ceiling effects and DIF for gender in item 23 contributed to their exclusion. We concluded these aspects of the quality of life are already represented by the items of social and emotional dimensions, therefore we did not choose any item from this dimension.

Symptoms

We did not use Rasch analysis on this collection of items representing symptoms since they are not expected to be unidimensional. There were four items – 9 (shortness of breath), 14 (chest pain), 16 (aching legs) and 19 (dizzy). The clinical team decided the most important symptoms were chest pain and shortness of breath and this was confirmed by the patient group. We created two separate dimensions from symptoms item 9 and 14. This selection was entirely based on clinician advice.

Stakeholder engagement for item selection

More than half of the clinicians agreed with item 26 for physical dimension, item 24 for social and item 1 for emotional. All clinicians agreed that shortness of breath and chest pain are the most important two symptoms selecting items 9 and 14. There were still doubts about items 24 or 25, whether item 15 is necessary when 6 is included, and if we have missed any important characteristics.

The patient advisory group debated which of the two social components items - item 24 or 25 - best described their experience. The majority (9/12 patient participants) agreed that item 24 was better. Reasons for this included: "*Item 24 includes your socialisation and sports or anything else, so it's more general so it probably would capture more"* "*...[it's] is a better*

question, because it's very clear what that means to people." Participants felt strongly and unanimously that item 15 was a critical aspect impacting quality of life for people with heart conditions. Reasons for this included*"Because of the physical restriction we become anxious and lacking in self-confidence" "It sometimes gets to a point where you just don't go out, because there's too many of those things to consider*." *"... if you've got no confidence, people give up. And [people who've given up] are a cost to the health system."* Most patients felt that the proposed items captured the most important components associated with their quality of life. The only additional suggestion involved including an item on sexual and intimate relationships. The majority view was expressed in the following: "*A lot of these things are playing on most of our minds, every single day. And I don't think there's anything else ...that you've missed."*

Stage III: Explore item-level reduction

Examination of threshold probability curves for items suggested that levels "most of the time", "a good bit of the time" and "some of the time" were closest together (Figure 1).This means disordering of these curves and responders have displayed difficulty in distinguishing between levels (32). This pattern was similar across the other items not selected in the classification system. Although not present in all items, most other items have "a little of the time" and "hardly any of the time" close to each other. These observations suggested collapsing levels 2,3,4 and 5,6 together. Threshold curves of item 26 suggested collapsing levels 2,3,4,5 together, however, this was not confirmed by its frequency distribution. In the emotion dimension, item 15 showed the same clumping of levels, confirming the four-level collapse. However, threshold curves of item 1 suggested five levels, collapsing 2,3,4 together and keeping 1,5,6,7 as independent levels. Each suggested option was tested with the Rasch analysis and resulting in new frequency distributions. Collapsing levels 2,3,4 and 5.6 resulted in the best-ordered models and fit. The collapsed 2,3,4 was named "some of the time" and 5,6

was named “hardly any of the time” as they had the highest frequency in nearly 90% of the total items considered.

The final classification system MacNew-7D with four levels in each consisting of items 1(frustrated), 6 (wornout), 9(shortness of breath), 14(chest pain), 15(self confidence), 24(excluded), 26(physical restriction) is presented in Table 4.

Discussion

This study developed a new brief heart disease-specific health state classification system consisting of seven dimensions, each with four levels, from the widely used and validated MacNew instrument (18). This classification system contains physical, social, emotion, symptoms of chest pain, and shortness of breath, which are well suited for this broad group of patients. The new classification system, based on the validated MacNew, is ideal for health state valuation and generating heart disease-specific utility values across a number of specific cardiac diseases. Moreover, it is a brief 7-item questionnaire and will have less burdon on patients for completion.

The MacNew instrument is a quality of life instrument without an associated utility algorithm. The next stage of this project will be valuing the classifaction system derived here to generate a utility value set. More importantly, the resulting algorithm will facilitate generation of utility values from the MacNew. Currently, there is a mapping (crosswalk) function that enables EQ-5D utility values to be predicted using MacNew scores (13). The mapped EQ-5D utility values will not however reflect the sensitivity of MacNew, rather they

reflect the sensitivity of EQ-5D and are dependent upon the predictive relationship between MacNew and EQ-5D which will contain error. In contrast, the results from the current program of research will provide a more robust utility value set to be used to evaluate heart disease related health interventions.

In selecting dimensions, we considered many aspects including the psychometric analysis results, the wording of the items, and opinions from research team, clinicians and patients. The five-components solution was chosen after conducting many PCAsas well as previous evidence which is similar to Dempster et al (19) but differs from what has been reported from the original developers (18). This result did not exactly match with Dempster et al. (19); social items of our analysis were within the restriction dimension. . Item 23 (burden on others) and 22 (family overprotective) loaded together in our analysis within "perception of others". In contrast, item 23 was within the emotion dimension as reported by Dempster et al (19). In our results, physical and social functions were loaded together. We decided to divide them into two separate dimensions, as it is clear that social and physical functions are separate. There were three items relating to emotional functions; worn out, frustrated and self-confidence that loaded together but were separated for the classification system based on guidance from clinicians. Inclusion of these symptoms was later confirmed by the patient advisory group. Rasch analysis and item probability curves were the main informants used to determine the number of levels, however our focus was to produce a classification system that would be amenable for a health state valuation. An instrument with seven levels would not be amendable to valuation since respondents to a valuation exercise would be unlikely be able to clearly distinguish between all of the different severity levels and the health states would become complex.

There are currently no other heart disease-specific preference-based measures. A heart disease-specific preference-based measure will be useful for resource allocation decisions

within cardiovascular budget allocations. Once the utility value set is developed, comparisons can be made with the EQ-5D-5L generic utility values on discrimination and responsiveness of the utility values in this specific group of patients and the impact on economic evaluation analysis. It will be of importance if there are any effects on the incremental cost-effectiveness ratio by using this instrument, and the impact this would have on resource allocation decisions particlaurly when cost-effectiveness estimates are close to funding thresholds.

There are several limitations to this study. The MacNew-7D was derived from the responses of patients with heart disease inpatients and patients who attended outpatient clinics. However, our sample did not consist of patients with more severe conditions, such as those recovering from myocardial infarction. Therefore, some information about the ceiling effects may be overestimated. This classification system was developed using a single data sample. Nevertheless. we used a large sample (n=685)(36) to prevent inter-sample variations. Moreover, clinicians and patients from a single site with heart disease provided face validity for the selected items, indicating the appropriateness of the classification system for heart disease.

We developed the first heart disease-specific classification system based on an existing disease-specific quality of life instrument. There are seven dimensions to the instrument with four levels in each. The classification system will be valued to enable utility values to be directly generated to estimate QALYs from all existing and future MacNew datasets.

Acknowledgements

We thank Cardiology Department, Royal Brisbane and Women's Hospital for the data collection support and our colleagues Ms Sara Baniahmadi for data collection, Dr. Sameera Senanayake for data cleaning and data management, and Ms Carly Dyer for conducting qualitative interviews. We sincerely appreciate the support of the patients who participated in

the study and staff of the Cardiology department in validating the instrument, and the Heart Failure patient advocacy group for their valuable insights. J.A. Whitty is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) East of England. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Sources of funding

This study was funded by Australian Heart Foundation Vanguard grant (2016/101407) and Heart Foundation post-doctoral fellowship for Dr. Kularatna.

Disclousres

All authours disclose no conflict of interest

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Reference

1. American Heart Association. Cardiovascular Disease: A costly burde for America. Projections through 2035.
2. Centres for Disease Control and Prevention. Heart Disease Facts U.S. Department of Health & Human Services2019 [Available from: https://www.cdc.gov/heartdisease/facts.htm.
3. Chan Y-K, Tuttle C, Ball J, Teng T-HK, Ahamed Y, Carrington MJ, et al. Current and projected burden of heart failure in the Australian adult population: a substantive but still ill-defined major health issue. BMC Health Services Research. 2016;16(1):501.
4. Weisfeldt ML, Zieman SJ. Advances in the prevention and treatment of cardiovascular disease. Health Aff (Millwood). 2007;26(1):25-37.
5. Michael F Drummond et al. Methods for the Economic Evaluation of Health Care Programmes. Fourth Edition ed. Oxford: Pxford University Press
6. Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. Valuing EQ-5D health states for Sri Lanka. Qual Life Res. 2015;24(7):1785-93.
7. Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. Review of Valuation Methods of Preference-Based Measures of Health for Economic Evaluation in Child and Adolescent Populations: Where are We Now and Where are We Going? PharmacoEconomics. 2020.
8. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Appl Health Econ Health Policy. 2017;15(2):127-37.

9. Lorgelly PK, Doble B, Rowen D, Brazier J, Thomas DM, Fox SB, et al. Condition-specific or generic preference-based measures in oncology? A comparison of the EORTC-8D and the EQ-5D-3L. 2017;26(5):1163-76.
10. Rankin J, Rowen D, Howe A, Cleland JGF, Whitty JA. Valuing health-related quality of life in heart failure: a systematic review of methods to derive quality-adjusted life years (QALYs) in trial-based cost–utility analyses. Heart Failure Reviews. 2019;24(4):549-63.
11. Kularatna S, Byrnes J, Chan YK, Carrington MJ, Stewart S, Scuffham PA. Comparison of contemporaneous responses for EQ-5D-3L and Minnesota Living with Heart Failure; a case for disease specific multiattribute utility instrument in cardiovascular conditions. Int J Cardiol. 2017;227:172-6.
12. Kularatna S, Senanayake S, Chen G, Parsonage W. Mapping the Minnesota living with heart failure questionnaire (MLHFQ) to EQ-5D-5L in patients with heart failure. Health and Quality of Life Outcomes. 2020;18(1):115.
13. Chen G, McKie J, Khan MA, Richardson JR. Deriving health utilities from the MacNew Heart Disease Quality of Life Questionnaire. European journal of cardiovascular nursing : journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology. 2015;14(5):405-15.
14. 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.
15. Berry C, McMurray J. A review of quality-of-life evaluations in patients with congestive heart failure. Pharmacoeconomics. 1999;16(3):247-71.
16. Bilbao A, Escobar A, García-Perez L, Navarro G, Quirós R. The Minnesota living with heart failure questionnaire: comparison of different factor structures. Health and quality of life outcomes. 2016;14:23-.
17. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonell M, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. Journal of the American College of Cardiology. 1995;25(2):333-41.
18. Höfer S, Lim L, Guyatt G, Oldridge N. The MacNew Heart Disease health-related quality of life instrument: A summary. Health and Quality of Life Outcomes. 2004;2(1):3.
19. Dempster M, Donnelly M, O'Loughlin C. The validity of the MacNew Quality of Life in heart disease questionnaire. Health and quality of life outcomes. 2004;2:6-.
20. Alphin S, Höfer S, Perk J, Slørdahl S, Zwisler A-DO, Oldridge N. The MacNew Heart Disease Health-Related Quality of Life Questionnaire: A Scandinavian Validation Study. Soc Indic Res. 2015;122(2):519-37.
21. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment. PharmacoEconomics. 2017;35(1):33-41.
22. 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. Journal of the American College of Cardiology. 2000;35(5):1245-55.
23. Hillers TK, Guyatt GH, Oldridge N, Crowe J, Willan A, Griffith L, et al. Quality of life after myocardial infarction. Journal of clinical epidemiology. 1994;47(11):1287-96.
24. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of health economics. 2002;21(2):271-92.
25. Nguyen KH, Mulhern B, Kularatna S, Byrnes J, Moyle W, Comans T. Developing a dementia-specific health state classification system for a new preference-based instrument AD-5D. Health Qual Life Outcomes. 2017;15(1):21.
26. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. Value in Health. 2015;18(6):846-55.
27. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value Health. 2011;14(5):721-31.

28. Brazier J RD, Mavranezouli I, Tsuchiya A, Young T. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technol Assess 2012;16(32),.
29. Robert K. Gable MBW. Instrument Development in the Affective Domain: Springer 1993.
30. Tabachnick BG, Fidell LS. Using multivariate statistics, 5th ed. Using multivariate statistics, 5th ed. Boston, MA: Allyn & Bacon/Pearson Education; 2007. p. xxvii, 980-xxvii, .
31. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis and rheumatism. 2007;57(8):1358-62.
32. Young TA, Yang Y, Brazier JE, Tsuchiya A. The use of rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. Medical decision making : an international journal of the Society for Medical Decision Making. 2011;31(1):195-210.
33. Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne KJQoLR. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. 2008;18(2):253.
34. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Health. 2004;7 Suppl 1:S22-6.
35. Nguyen K-H, Mulhern B, Kularatna S, Byrnes J, Moyle W, Comans T. Developing a dementia-specific health state classification system for a new preference-based instrument AD-5D. Health and quality of life outcomes. 2017;15(1):21-.
36. Comrey AL. A first course in factor analysis. New York: Academic Press; 1973.

Table 1: Demographics of the respondents (n=685)

	N (%)
Gender	
Male	460 (67.2%)
Age	
<= 57	229 (33.4)
58-68	244 (35.6)
69+	212 (30.9)
Heart Disease	
Acute coronary syndrome	42 (6.1)
Stable coronary Heart disease	284 (41.5)
Acute decompensated Heart Failure	4 (0.6)
Chronic Heart Failure	135 (19.7)
Hypertension	73 (10.7)
Hypertropic cardiomyopathy	52(7.6)
Other	219(32.0)*
Employment	
Full time	168 (24.5)
Unemployed	39 (5.7)
Retired	179 (26.1)
MacNew scoring	**Mean (SD)**
MacNew Global	5.4 (1.2)
MacNew Physical	5.4 (1.3)
MacNew Emotional	5.3 (1.2)
MacNew Social	5.6 (1.3)

*Some patients had more than one diagnosis- hence does not add to 685 or 100%

Table 2: Principal component analysis

MacNew items	Restriction	Emotion	Symptoms	Perception of others
Item 1- Frustrated		0.822		
Item 2- Worthless		0.843		
Item 3 – Confident		0.494		
Item 4 - Down in the dumps		0.866		
Item 5- Relaxed		0.881		
Item 6- Worn out		0.467		
Item 8 - Restless		0.740		
Item 9- Short of breath			0.608	
Item 10- Tearful		0.611		
Item 12- Social activities	0.478			
Item 14- Chest pain			0.788	
Item 15 – Lack of self-confidence		0.656		
Item 16 – Aching legs			0.543	
Item 17- Limited in exercise	0.775			
Item 18 – Frightened		0.507		
Item 19- Dizzy			0.576	
Item 20- Restricted	0.742			
Item 21 – Unsure of exercise	0.621			
Item 22- Family overprotective				0.669
Item 23- Burden on others				0.493
Item 24- Felt excluded	0.814			
Item 25- Unable to socialise	0.746			
Item 26- Physically restricted	0.811			

Table 3: Rasch and psychometric analysis for item reduction

Item	Items selected for the classification	Floor effect %	Ceiling effect %	Response ordered	No DIF	Good fit	Range	Spread
Physical								
Item 17- Limited in exercise		9.9	33.7	×	✓	✓	-0.28 to 1.90	0.57 to 0.21
Item 20- Restricted		2.5	42.6	×	✓	×	-1.61 to 1.31	0.83 to 0.13
Item 21- Unsure of exercise		5.7	37.4	✓	×	×	-0.64 to 1.13	0.66 to 0.24
Item 26- Physically restricted	x	2.8	44.2	×	✓	×	-1.62 to 1.19	0.84 to 0.23
Social								
Item 12- Social activities		3.6	42.3	✓	✓	✓	-0.77 to 1.63	0.68 to 0.16
Item 24- Felt excluded	x	1.8	57.2	✓	✓	×	-1.1 to 0.76	0.76 to 0.32
Item 25- Unable to socialise		1.3	63.5	✓	✓	×	-1.61 to 0.41	0.83 to 0.40
Emotional								
Item 1 – Frustrated	x	1.9	18.4	✓	×	✓	-1.4 to 2.4	0.81 to 0.08
Item 2 - Worthless		2.0	45.0	✓	✓	✓	-1.1 to 0.7	0.76 to 0.33
Item 3 –Confident		4.8	43.0	✓	×	×	0.5 to 1.21	0.62 to 0.23
Item 4 - Down in the dumps		1.5	31.5	✓	✓	×	-1.7 to 1.6	0.85 to 0.16
Item 5- Relaxed		4.7	11.1	✓	×	✓	-1.4 to 3.4	0.80 to 0.03

Item 6- Worn out	x	8.0	8.2	✓	×	×	-0.5 to 2.9	0.64 to 0.05
Item 8- Restless		1.6	31.4	✓	×	✓	-1.0 to 1.5	0.79 to 0.18
Item 10- Tearful		1.3	54.3	×	×	✓	-0.98 to 0.06	0.73 to 0.48
Item 15 – Lack of self confidence	x	2.5	41.9	✓	✓	✓	-1.05 to 0.89	0.74 to 0.29
Item 18 – Frightened		0.7	57.2	✓	×	✓	-1.9 to 0.10	0.88 to 0.47
Perception of others								
MacNew 22- Family overprotective		3.6	51.2	✓	✓	×	-0.27 to 0.56	0.57 to 0.36
MacNew 23- Burden on others		3.9	53.7	✓	×	×	-0.33 to 0.28	0.58 to 0.43
Symptoms								
MacNew 9- Shortness of breath	x	3.2	29.5					
MacNew 14- Chest pain	x	1.2	59.9					
MacNew 16 – Aching legs		7.6	28.8					
MacNew 19- Dizzy		1.8	36.4					

DIF= differential item functioning

Table 4 : New classification system MacNew-7D

	Dimension	Level	Description
1	Physical restriction	1	You are not limited at all physically
		2	You are somewhat limited physically
		3	You are moderately limited physically
		4	You are extremely limited physically
2	Excluded from doing things with other people	1	You do not feel excluded from doing things with other people
		2	You feel excluded from doing things with other people hardly any of the time
		3	You feel excluded from doing things with other people some of the time
		4	You feel excluded from doing things with other people all of the time
3	Worn out or low in energy	1	You do not feel worn out or low in energy
		2	You feel worn out or low in energy hardly any of the time
		3	You feel worn out or low in energy some of the time
		4	You feel worn out or low in energy all of the time
4	Frustrated, impatient or angry	1	You do not feel frustrated, impatient, or angry
		2	You feel frustrated, impatient, or angry hardly any of the time
		3	You feel frustrated, impatient, or angry some of the time
		4	You feel frustrated, impatient or angry all of the time
5	Unsure and lacking in self confidence	1	You do not feel unsure and lacking in self-confidence
		2	You feel unsure and lacking in self-confidence hardly any of the time
		3	You feel unsure and lacking in self-confidence some of the time
		4	You feel unsure and lacking in self-confidence all of the time

6	Shortness of Breath *While doing your day to day physical activities*	1	You have no shortness of breath
		2	You have some shortness of breath
		3	You have moderate shortness of breath
		4	You have extreme shortness of breath
7	Chest Pain	1	You have no chest pain
		2	You have chest pain hardly any of the time
		3	You have chest pain some of the time
		4	You have chest pain all of the time

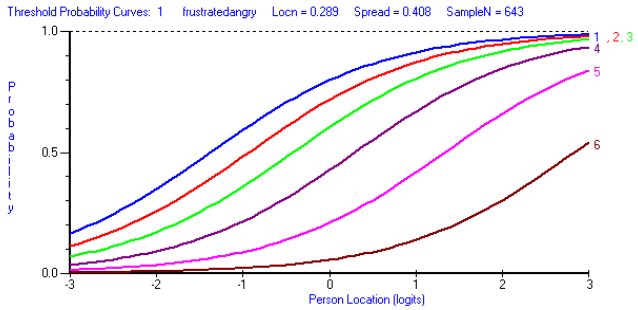

Figure 1A: Threshold probability curve for item 1before rescoring

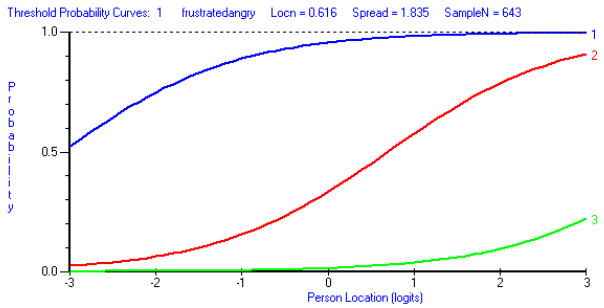

Figure 1B: Threshold probability curve for item 1after rescoring

Stage 1 analysis

1. First used factor analysis without specifying number of factors.
a. Results using Pearson correlation treating data as cardinal. No rotation.

Condition	Results
Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO)	0.961
Bartlett's Test of Sphericity	Sig=0.0000
Correlation matrix	90%>0.3 correlation*
Kaisers criterion	2 factors**
Scree plot	2 factors
Parallel analysis	2 factors

*More than 90% of correlation coefficients were above 0.3 ; **2 factors had eigenvalue above 1

Parallel analysis- using Monte Carlo for Parallel analysis

Number	PCA Eigenvalue	Parallel analysis Eigenvalue	Decision
1	11.37	1.346	Accept
2	1.958	1.286	Accept
3	0.944	1.246	Reject
4	0.899	1.2087	Reject
5	0.764	1.764	Reject

b. Results using polychoric correlation to determine number of factors

This analysis was conducted using FACTOR software. First polychoric an analysis was run indicating 7 factors to extract (23/3=7) without rotation. The analysis using parallel analysis indicated the correct number of dimensions to be extracted are 2 agreeing with previous results.

Number	Eigenvalue	95 percentile of random % of variance	Decision
1	13.1	9.6	Accept
2	2.0	8.9	Accept
3	0.85	8.3	Reject
4	0.76	7.9	Reject
5	0.65	7.5	Reject

Both analyses confirmed 2 component model.

So similar to Person correlation 2 factor extraction was conducted for polychoric.

Condition	Results
Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO)	0.96032
Bartlett's Test of Sphericity	Sig=0.000010
Correlation matrix	90%>0.3 correlation*
Parallel analysis	2 factors

*More than 90% of correlation coefficients were above 0.3 ;

2. Specifying 3 component analysis as given in Hofer et 2004 with PCA direct obilimin extraction

Pearson

	Pearson			Polychoral		
MacNew items	Restriction	Emotion	Symptom	Restriction	Emotion	Symptoms
Item 1- Frustrated		0.813			0.852	
Item 2- Worthless		0.829			0.853	
Item 3 – Confident		0.437			0.447	
Item 4 - Down in the dumps		0.870			0.901	
Item 5- Relaxed		0.852			0.893	
Item 6- Worn out		0.438			0.477	
Item 8 - Restless		0.755			0.800	
Item 9- Short of breath			0.637			0.521
Item 10- Tearful		0.723			0.727	
Item 12- Social activities	0.450			0.477		
Item 14- Chest pain			0.707			0.488
Item 15 – Lack of self-confidence		0.703			0.703	
Item 16 – Aching legs			0.595	*0.304*		
Item 17- Limited in exercise	0.697			0.786		
Item 18 – Frightened		0.588			0.616	
Item 19- Dizzy			0.498	*0.331*		
Item 20- Restricted	0.748			0.839		
Item 21 – Unsure of exercise	0.650			0.711		
Item 22- Family overprotective	0.544			0.503		
Item 23- Burden on others	0.647			0.622		
Item 24- Felt excluded	0.892			0.964		
Item 25- Unable to socialise	0.789			0.812		
Item 26- Physically restricted	0.748			0.809		

2. Specifying 4 component analysis as given in Dempster et al 2004 with PCA direct obilimin extraction

	Pearson				Polychoral			
MacNew items	Restriction	Emotion	Symptom	Perception	Restriction	Emotion	Symptoms	Perception
Item 1- Frustrated		0.822				0.838		
Item 2- Worthless		0.843				0.884		
Item 3 – Confident		0.494				0.449		
Item 4 - Down in the dumps		0.866				0.889		
Item 5- Relaxed		0.881				0.888		
Item 6- Worn out		0.467				0.432		
Item 8 - Restless		0.740				0.757		
Item 9- Short of breath			0.608				0.447	
Item 10- Tearful		0.611				0.627		
Item 12- Social activities	0.478				0.472			
Item 14- Chest pain			0.788				0.589	
Item 15 – Lack of self-confidence		0.656				0.626		
Item 16 – Aching legs			0.543		0.380			
Item 17- Limited in exercise	0.775				0.866			
Item 18 – Frightened		0.507				0.528		
Item 19- Dizzy			0.576				0.395	
Item 20- Restricted	0.742				0.823			
Item 21 – Unsure of exercise	0.621				0.601			
Item 22- Family overprotective				0.669				0.545
Item 23- Burden on others				0.493				0.575
Item 24- Felt excluded	0.814				0.752			
Item 25- Unable to socialise	0.746				0.700			
Item 26- Physically restricted	0.811				0.926			