ORIGINAL RESEARCH ARTICLE

Development of the Hypertrophic Cardiomyopathy Symptom Questionnaire (HCMSQ): A New Patient-Reported Outcome (PRO) Instrument

Matthew Reaney[1] · Veleka Allen[2] · Amy J. Sehnert[3] · Liang Fang[3] · Albert A. Hagège[4] · Srihari S. Naidu[5] ·
Iacopo Olivotto[6]

Abstract

Background Currently, there is no patient-reported outcome (PRO) instrument specifically designed to evaluate hypertrophic cardiomyopathy (HCM).

Objective We present the development and psychometric validation of a novel PRO measure, the HCM Symptom Questionnaire version 1.0 (HCMSQv1.0).

Methods Cognitive debriefing interviews and a card-sorting task were conducted in 33 patients with HCM to support development of the HCMSQv1.0, showing the scale to be interpretable and relevant to patients' experiences. Baseline blinded data from two trials (EXPLORER-HCM and MAVERICK-HCM) were pooled ($N = 299$) to develop the scoring algorithm of HCMSQv1.0. Measurement properties were examined, followed by a meaningful-change analysis to interpret scores. Rasch modeling, mixed-model repeated measures, exploratory factor analysis, confirmatory factor analysis, and missing-data simulation analysis informed the number of domains and the items in each domain.

Results The scoring algorithm for HCMSQv1.0 consists of four domains: shortness of breath, tiredness, cardiovascular symptoms, and syncope; plus a total score, with higher scores indicating more severe symptoms. Item characteristics, internal consistency, test–retest reliability, construct validity, and responsiveness were acceptable. A clinically meaningful responder definition of 1–2 points on the HCMSQv1.0 score for shortness of breath and total score, and approximately 1 point on the tiredness and cardiovascular symptom scores, was calculated based on distribution- and anchor-based methods.

Conclusion Our findings support the HCMSQv1.0 as a fit-for-purpose PRO instrument for assessing treatment benefit in patients with HCM. Studies in larger patient populations are ongoing to confirm responder definition and scoring approaches encompassing key HCM symptoms.

✉ Matthew Reaney
matthew.reaney@iqvia.com

1 IQVIA, 3 Forbury Place, 23 Forbury Road, Reading RG1 3JH, UK

2 IQVIA, New York, NY, USA

3 MyoKardia, Inc. (a wholly owned subsidiary of Bristol Myers Squibb), Brisbane, CA, USA

4 Cardiology Department, Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges-Pompidou, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

5 Westchester Medical Center, WMC Health Network, New York Medical College, Valhalla, NY, USA

6 Cardiomyopathy Unit, Careggi University Hospital, University of Florence, Florence, Italy

1 Introduction

Hypertrophic cardiomyopathy (HCM) is a chronic, progressive myocardial disorder characterized by left ventricular hypertrophy [1, 2]. It is caused by dysfunction in the basic functional unit of the heart muscle, called the sarcomere, resulting in excess myosin–actin cross-bridging that leads to hypercontractility and impaired diastolic function, which are hallmarks of the disease [3–5]. Approximately two-thirds of patients with HCM present with a rest or provoked (particularly with exercise) dynamic obstruction to left ventricular outflow (obstructive HCM) [6]. Based on patients' experiences with obstructive and nonobstructive HCM, core symptoms include exercise dyspnea, chest pain, dizziness or syncope, and fatigue [7]. These symptoms are not specific

Key Points for Decision Makers

Treatment goals in hypertrophic cardiomyopathy (HCM) focus on the relief of symptoms; however, there is no HCM-specific measure to evaluate patient symptoms for consideration during drug development and in clinical practice.

This study describes the development and evaluation of the HCM Symptom Questionnaire version 1.0 (HCMSQv1.0), a new patient-reported outcome (PRO) measure, using qualitative and quantitative methods in line with guidance for industry on PRO measures from the US Food and Drug Administration.

The results from this study demonstrate that the HCMSQ is a fit-for-purpose PRO instrument for the assessment of treatment benefit in patients with HCM following the removal of items 4 and 5, making a 9-item scale (the HCMSQv2.0). This is of key relevance for future studies in such patient populations.

to HCM and many patients remain without a diagnosis throughout life, resulting in a much higher disease burden [8]. The primary treatment goal in HCM is directed toward symptom relief [9]; however, there is no HCM-specific measure to evaluate the symptoms for drug development or clinical practice. Here, we describe the development and validation of a novel patient-reported outcome (PRO) measure, using qualitative and quantitative methods in line with guidance for industry on PRO measures from the US Food and Drug Administration (FDA) [10]. This novel PRO measure, the HCM Symptom Questionnaire version 1.0 [HCM-SQv1.0], was included in clinical trials of mavacamten, a first-in-class small molecule myosin–actin inhibitor developed to address the underlying pathological mechanism of symptomatic obstructive and nonobstructive HCM [11, 12].

2 Methods

The HCMSQ instrument was drafted based on a conceptual model previously published by Zaiser and colleagues [7] and was refined through cognitive interviews with 33 patients with HCM (16 patients with nonobstructive HCM, 17 with obstructive HCM) recruited in the UK, Italy, France, and the US. Participants were recruited through referrals from established physician and nursing panels, patient advocacy organizations, and one clinical site in the US. Cognitive interviews were conducted in five waves between 2017 and 2019 to iteratively develop and cognitively debrief the HCMSQv1.0, specifically its relevance and coverage of common HCM

symptoms, the interpretation of item language and response options, and the understandability of the instrument overall (Fig. S1 in the electronic supplementary material [ESM]). Following the cognitive interviews, the HCMSQv1.0 was finalized for use in EXPLORER-HCM and MAVERICK-HCM as an 11-item PRO instrument measuring the core symptoms of HCM from the conceptual model (Table 1).

The HCMSQv1.0 was administered as a daily electronic diary in two clinical trials: the phase III EXPLORER-HCM trial (NCT03470545) [11, 12] among 251 patients with obstructive HCM, and the phase II MAVERICK-HCM trial (NCT03442764) [13] among 48 patients with nonobstructive HCM (although 59 patients were enrolled in this trial, 48 were eligible based on protocol amendment). In both trials, the HCMSQv1.0 was completed daily during screening (for a minimum of 7 days prior to enrollment), at baseline (the last 7 days of entries during screening prior to day 1) through week 6, then for 7 consecutive days prior to the week 10 and 14 study visits. Further completion occurred in the MAVERICK-HCM study in the 7 days prior to the week 16 and 24 study visits, and in the EXPLORER-HCM study in the 7 days prior to the week 18, 22, 26, 30, and 38 study visits. The daily electronic diary did not allow patients to skip items.

To examine the adequacy of the instrument in the clinical trial setting, exit interviews were planned. The exit interviews—conducted following administration of the HCM-SQv1.0 in the EXPLORER-HCM and MAVERICK-HCM clinical trials in five patients with nonobstructive HCM and 17 with obstructive HCM—included a card-sorting task to ensure patients could consistently order the HCMSQv1.0 response options as intended (i.e., from least to most severe), and to examine whether patients could distinguish between 'very severely' and 'too short of breath to do the activity' for the shortness-of-breath items. For the card-sorting task, items 1, 2, 3, 5, and 6 were selected because they represented each unique response option type included in the HCMSQv1.0 instrument (Table 1).

The HCMSQv1.0 measures severity of tiredness/fatigue, dizziness/light-headedness, heart palpitations, and chest pain symptoms on single 5-point verbal rating scales from 'not at all' to 'very severe(ly)'. Shortness of breath is measured by six items, four of which use the same response scale and ask about severity of shortness of breath in general, during light physical activity, during moderate physical activity, and during heavy physical activity. The fifth question asks respondents to describe when their shortness of breath occurred (in relation to activity levels) at its worst, and the sixth question measures frequency of shortness of breath ('never' to 'almost always'). Syncope is measured as a binary (yes/no) response. A 24-h recall period was chosen due to the anticipated symptom variability and to limit recall bias. This instrument was translated and linguistically validated in 16

Table 1 The HCMSQ instrument

Core symptom	Item	Question	Response scale
Shortness of breath (shortness-of-breath domain)	1	Were you short of breath during the past 24 hours?	5-point categorical rating scale from 0 (not at all) to 4 (very severely)
	2	Were you short of breath during light physical activity such as walking slowly or cooking during the past 24 hours?	6-point categorical rating scale from 0 (not at all) to 5 (too short of breath to do the activity) 'I did not attempt to do the activity' was also offered as a response option
	3	Were you short of breath during moderate physical activity such as cleaning house or lifting heavy objects?	6-point categorical rating scale from 0 (not at all) to 5 (too short of breath to do the activity) 'I did not attempt to do the activity' was also offered as a response option
	4 (removed)	Were you short of breath during heavy physical activity such as jogging or playing sports during the past 24 hours?	6-point categorical rating scale from 0 (not at all) to 5 (too short of breath to do the activity) 'I did not attempt to do the activity' was also offered as a response option
	5 (removed)	Describe your shortness of breath at its worst during the past 24 hours	5-point categorical rating scale from 0 (no shortness of breath) to 4 (short of breath when resting)
	6	How often did you have shortness of breath during the past 24 hours?	5-point categorical rating scale from 0 (never) to 4 (almost always)
Tiredness/fatigue (tiredness domain)	7	Were you tired during the past 24 hours?	5-point categorical rating scale from 0 (not at all) to 4 (very severely)
Heart palpitations (cardiovascular symptoms domain)	8	Did your heart beat rapidly or flutter (palpitations) during the past 24 hours?	5-point categorical rating scale from 0 (not at all) to 4 (very severely)
Chest pain (cardiovascular symptoms domain)	9	Did you have chest pain during the past 24 hours?	5-point categorical rating scale from 0 (not at all) to 4 (very severely)
Dizziness (cardiovascular symptoms domain)	10	Were you dizzy or light-headed during the past 24 hours?	5-point categorical rating scale from 0 (not at all) to 4 (very severely)
Syncope (syncope domain)	11 (not contributing to the total score)	Did you faint or lose consciousness during the past 24 hours?	2-point categorical rating scale: no (0) or yes (1)

HCMSQ Hypertrophic Cardiomyopathy Symptom Questionnaire

The HCMSQ is not presented in full so as to ensure no improper use. MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb wishes to maintain version control of the scale and, as future iterations are made, to ensure that established and tested versions are used in clinical research and practice. ©MyoKardia 2020. For questions relating to the use of this questionnaire, please contact Amy Sehnert at MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb: Amy.Sehnert@bms.com

languages. The translation was conducted by TransPerfect as per the principles outlined by the 2005 Professional Society for Health Economics and Outcomes Research (ISPOR) Task Force [14].

The psychometric measurement properties of the HCMSQv1.0 were examined by utilizing the population datasets available from the MAVERICK-HCM and EXPLORER-HCM trials. Screening and baseline data from both trials were used and full analysis sets (FAS) were analyzed (MAVERICK-HCM FAS [FAS-M] and EXPLORER-HCM FAS [FAS-E]). Post-baseline data were only available from the MAVERICK-HCM study at the time of this analysis; data were blinded because treatment status was unknown. Blinded baseline data from MAVERICK-HCM

and EXPLORER-HCM were pooled to provide a sufficiently large sample size ($N = 299$) [15, 16] on which to conduct a series of analyses to inform the development of the scoring algorithm for the HCMSQv1.0. Analyses included Rasch modeling, mixed-model repeated measures, exploratory factor analysis, confirmatory factor analysis, and missing-data simulation analysis. The analyses guided the number of domains used in the scoring framework, the items that were included in each domain, and the handling of missing data.

The relationship and degree of association among items and between items and domains was evaluated at baseline using item-to-item correlation and item–scale analyses in the FAS-M and FAS-E. Items were considered to be performing adequately if correlations between items in the same

domain were at least moderate (correlation coefficient [r] > 0.3) [17], and items had higher correlations with their respective domains compared with other domains.

To examine the internal consistency for each multi-item scale and total score, Cronbach's alpha was calculated using the baseline data from the FAS-M and FAS-E.

The test–retest reliability for HCMSQv1.0 was examined at two time points.

(1) Using 2 weeks of screening data, with week 1 being the 'test' and week 2 the 'retest', all patients were included in the analysis on the assumption that they had no change in clinical status during screening. This analysis was done in both studies using the FAS-M and FAS-E.

(2) Using baseline ('test') and week 6 ('retest') data in MAVERICK-HCM, for the FAS-M only, patients were included in the analysis if they reported 'no change' on the Patient Global Impression of Change (PGIC) scale (a single-item questionnaire asking respondents to rate their overall change in symptom severity over time) at week 6. A second analysis was conducted among patients who reported the same severity level on the Patient Global Impression of Severity (PGIS) scale (a single-item questionnaire asking respondents to rate their overall symptom severity in the past week) at both baseline and week 6.

For the HCMSQv1.0 in the FAS-M and FAS-E baseline data, two types of construct validity were examined: convergent validity and known-groups validity.

Convergent validity was examined by evaluating the degree of association between the HCMSQv1.0 domain and total scores, the Kansas City Cardiomyopathy Questionnaire-23 (KCCQ-23; a well-established PRO instrument [18]) domain and summary scores, and the 5-dimension 5-level EuroQol questionnaire (EQ-5D-5L) health-related quality-of-life instrument [19]. Correlations were used to test for associations.

Known-groups validity was examined to determine the degree to which the HCMSQv1.0 can distinguish between defined groups of patients. Groups were formulated using PGIS scores and New York Heart Association (NYHA) class II or III, and were compared using analysis of variance (ANOVA).

Differential item functioning was conducted using a pooled data set to evaluate the extent to which each item may differ systematically between the two study populations (MAVERICK-HCM and EXPLORER-HCM).

The responsiveness of the HCMSQv1.0 was examined in the FAS-M by analyzing change from baseline to week 16.

To determine the threshold of within-patient treatment benefit, responder definitions for the HCMSQv1.0 scores were derived using distribution-based methods (in which the variability of scores is used to determine meaningful change) for MAVERICK-HCM and EXPLORER-HCM using FAS-M and FAS-E baseline data. In addition, anchor-based analyses (in which an external indicator ['anchor'] that reflects the patient's view of change is used to determine meaningful change; PGIC and PGIS served as anchors) were conducted in MAVERICK-HCM using longitudinal (baseline and week 16) data.

All aspects of the EXPLORER-HCM (NCT03470545) [11, 12] and MAVERICK-HCM (NCT03442764) [13] studies were conducted in compliance with the Health Insurance Portability Accountability Act of 1996 (USA). Institutional review board and relevant ethics committee approvals were obtained prior to participant enrollment. All participants provided informed (written and verbal) consent prior to their participation. The study protocols were approved by a US institutional review board: The New England Independent Review Board (197 First Avenue, Suite 250, Needham, MA 02494). Both studies fully adhered to the ethical principles of the Declaration of Helsinki and the specifications of the International Council on Harmonisation and Good Clinical Practice.

3 Results

3.1 Content Validity

Content validity of the HCMSQv1.0 was established through the cognitive interviews and the exit interviews. Demographic characteristics for the cognitive interviews are presented in Table 2, while demographic characteristics for the exit interviews are presented in Table 3.

In the cognitive interviews, the patients endorsed the ease and clarity of the instrument; they found the instructions to be appropriate and clear, and reported that the 24-h recall period was suitable for the HCMSQv1.0 items. They also found the phrasing of response options for all 11 items to be clear, were able to select a response option to fit their experiences, and distinguished the distinct response options. These interviews with patients with HCM established the preliminary content validity of the HCMSQv1.0.

The card-sorting exercise in the exit interviews showed that most patients were able to sort and interpret the sets of response options as intended, although some patients did have difficulty in placing certain response options. Patients understood the ordering hierarchy as intended, which supports the integrity of the content validity of the HCM-SQv1.0; an exception was item 5.

Item 5 asks about shortness of breath at its worst during the past 24 h, and the response options correspond to different levels of activity, which is a different response scale

Table 2 Sociodemographic characteristics of patients in the cognitive interviews

Demographic characteristic		Patients with cognitive interviews
Number of patients	N	33 (16 with nHCM, 17 with oHCM)
Age, years	Mean (range)	46.0 (22–74)
Sex	Male	13 (39)
	Female	20 (61)
Marital status	Single, never married	8 (24)
	Living with partner	6 (18)
	Married	15 (45)
	Separated	–
	Divorced	3 (9)
	Widowed	1 (3)
Education level	Less than university	9 (28)
	College/university degree, or higher	18 (56)
	Other (i.e., associate degree, vocational degree and other)	5 (16)
Current employment status	Employed, full time	16 (48)
	Employed, part time	5 (15)
	Homemaker	1 (3)
	Student	1 (3)
	Retired	5 (15)
	Disabled	3 (9)
	Other	2 (6)

Data are expressed as n (%) unless otherwise specified

nHCM nonobstructive hypertrophic cardiomyopathy, oHCM obstructive hypertrophic cardiomyopathy

Table 3 Demographic characteristics of patients in the exit interviews

Demographic characteristic		Patients with exit interviews
Number of patients	N	22 (5 with nHCM, 17 with oHCM)
Age, years	Mean (range)	63 (47–80)[a]
Sex	Male	12 (55)[a]
	Female	9 (41)[a]
Race	White	21 (95)[a]
Ethnicity	Hispanic	2 (9)[a]
	Non-Hispanic	19 (86)[a]
Time since end of treatment, days	Mean (range)	15 (0–29)[a]
Education level	Less than high school	–[a,b]
	High school diploma	2 (9)[a,b]
	Some college	4 (18)[a,b]
	Bachelor's degree	6 (27)[a,b]
	Graduate degree	8 (36)[a,b]

Data are expressed as n (%) unless otherwise specified

nHCM nonobstructive hypertrophic cardiomyopathy, oHCM obstructive hypertrophic cardiomyopathy

[a]Demographics for one of the EXPLORER patients were not obtained due to restriction from the recruiting site

[b]Education level was not provided for a further EXPLORER patient

from the other items. Many patients did not order this item as intended, mainly because they considered the level of difficulty of the activity rather than the expected severity of the shortness of breath experienced while doing the activity.

3.2 Scoring of the HCMSQv1.0

Results from Rasch modeling, mixed-model repeated measures, exploratory factor analysis, confirmatory factor analysis, and missing-data simulation analysis (data shown in the ESM) informed the number of separate domains, the items that should be included in each domain, and the handling of missing data. The scoring algorithm was also refined through the testing of various imputations and data structures, taking into account the underlying pathophysiology. These results, in conjunction with the qualitative insights from the card-sorting task, informed the final scoring algorithm for HCMSQv1.0.

The shortness-of-breath domain score is calculated as the sum of the responses to items 1 (overall severity of shortness of breath), 2 (severity of shortness of breath during light activity), 3 (severity of shortness of breath during moderate activity), and 6 (frequency of shortness of breath) over a 7-day period. Items 4 and 5 were removed as described below. If the 'I did not attempt' response choice is selected for item 2, then the shortness-of-breath score is treated as missing for that day. If the 'I did not attempt' response choice is selected for item 3, then item 3 is imputed as the mean of items 1, 2, and 6 for that day; this approach was appropriate for item 3 because imputing was not associated with a bias that minimized the severity of shortness of breath (ESM). Item 4 (shortness-of-breath severity during heavy activity) was removed from the v1.0 scoring algorithm because of a misfit between the item response and the severity of patients' shortness of breath as displayed during the Rasch modeling analysis. Item 5 (description of worst shortness of breath) was also removed from the scoring, partly because the card-sorting data questioned content validity and partly because the Rasch modeling results indicated that this item may not be functioning as expected. Thus, the potential range of scores for the shortness-of-breath domain is from 0 to 18, with lower scores indicating a lower presence of shortness of breath (Table 1).

The domain score for tiredness is the score for item 7 (severity of tiredness) over a 7-day period. The potential range is from 0 to 4, with lower scores indicating less tiredness.

The cardiovascular symptoms domain score is calculated as the sum of responses to items 8 (heart palpitation severity), 9 (chest pain severity), and 10 (dizziness severity) over a 7-day period. The potential range of scores is from 0 to 12, with lower scores indicating fewer cardiovascular symptoms.

The domain score for syncope is the score (yes/no) for item 11 (syncope) over a 7-day period. During the exploratory factor analysis and confirmatory factor analysis, this item was rarely chosen by patients (with 0% of patients fainting in MAVERICK-HCM and 4.5% in EXPLORER-HCM); it was therefore kept as additional information but excluded from the total score.

The total score is calculated as the equally weighted sum of the three domains, i.e., the sum of the shortness-of-breath domain divided by four (mean of items 1, 2, 3, and 6), the tiredness domain (item 7), and the cardiovascular symptoms domain divided by three (mean of items 8, 9, and 10) over a 7-day period. The potential range of scores is 0–12.5, with lower scores indicating a lower presence of symptoms overall.

Exploratory and confirmatory factor analyses showed that it is appropriate to calculate a total score from the three domains given that they each represent distinct and important aspects of the patient symptom experience of HCM and are statistically supported as a single score. The three domains thus provide a comprehensive assessment of HCM symptoms. The HCMSQv1.0 conceptual framework is shown in Fig. 1.

To determine the weighting of the items and domains that contribute to the total score, multiple sources, including patient interviews and patient surveys, were examined to understand the relative prevalence and importance of these symptoms in HCM. Findings suggested an equal 1:1:1 weighting ratio of the domains (data shown in the ESM). The final total score for HCMSQv1.0 therefore uses the mean of the three domain scores as a single score to measure the overall symptoms of HCM.

Missing-data simulation analysis confirmed the prespecified rule that weekly scores can be calculated if at least 4 of 7 days are completed (Table S1 in the ESM).
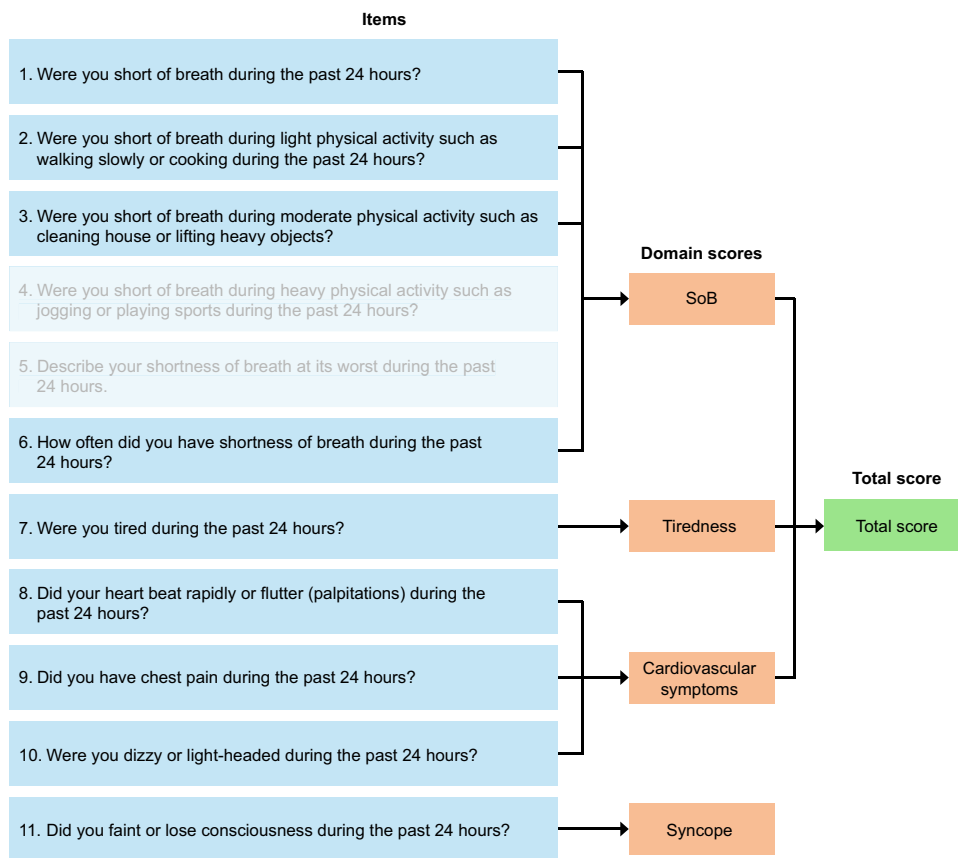
3.3 Psychometric Performance of the HCMSQv1.0 Scores

3.3.1 Item Performance

For the MAVERICK-HCM study, based on day 7 of baseline item scores, results showed correlations among the shortness-of-breath items (items 1–6) to be at least moderate ($r > 0.3$). Moderate correlations were also seen among shortness-of-breath items with tiredness and heart palpitations. Lower correlations were seen among the shortness-of-breath items with chest pain and dizziness. All items had the highest correlations with their respective domains compared with other domains.

Based on baseline weekly scores (average over 7 days with at least four non-missing entries), item-to-item correlation analysis showed correlations among the

Fig. 1 HCMSQv1.0 conceptual framework. *HCMSQv1.0* Hypertrophic Cardiomyopathy Symptom Questionnaire version 1.0, *SoB* shortness of breath

Items

1. Were you short of breath during the past 24 hours?

2. Were you short of breath during light physical activity such as walking slowly or cooking during the past 24 hours?

3. Were you short of breath during moderate physical activity such as cleaning house or lifting heavy objects?

4. Were you short of breath during heavy physical activity such as jogging or playing sports during the past 24 hours?

5. Describe your shortness of breath at its worst during the past 24 hours.

6. How often did you have shortness of breath during the past 24 hours?

7. Were you tired during the past 24 hours?

8. Did your heart beat rapidly or flutter (palpitations) during the past 24 hours?

9. Did you have chest pain during the past 24 hours?

10. Were you dizzy or light-headed during the past 24 hours?

11. Did you faint or lose consciousness during the past 24 hours?

Domain scores

SoB

Tiredness

Cardiovascular symptoms

Syncope

Total score

Total score

shortness-of-breath items (items 1–6) and tiredness to be at least moderate ($r > 0.3$). Results from the item–scale analysis indicated that all items had the highest correlations with their respective domains compared with other domains.

Similar findings were observed in the EXPLORER-HCM study. Based on day 7 of baseline item scores, results showed correlations among the shortness-of-breath items (items 1–6) to all be at least moderate ($r > 0.3$). All items had the highest correlations with their respective domains compared with other domains.

Based on baseline weekly scores (average over 7 days with at least four non-missing entries), item-to-item correlation analysis showed correlations among all weekly items to be at least moderate ($r > 0.3$). However, syncope (item 11) had negligible to very low correlations with all the other weekly item scores. This was probably due to the low frequency of fainting (0.3% of entries reported a syncope episode). Results from the item–scale analysis showed that all items had the highest correlations with their respective domains compared with other domains.

3.3.2 Reliability

For the MAVERICK-HCM trial, the shortness-of-breath domain and total scores had good internal consistency

(Cronbach's alpha [standardized] 0.924 and 0.811, respectively), but the cardiovascular symptoms domain did not show good internal consistency (0.586) [Table S2 in the ESM]. However, in the EXPLORER-HCM trial (which represents a larger data set), all domain and total scores had good internal consistency (0.795, 0.962, and 0.924 for cardiovascular symptoms domain, shortness-of-breath domain, and total score, respectively).

For samples from both trials (longitudinal analysis only in the MAVERICK-HCM trial), all domain and total scores had acceptable reliability (Table 4), with an intraclass correlation between HCMSQv1.0 scores at baseline and week 6 in patients reporting the same global impression of symptoms on the PGIC and PGIS scales ranging from 0.829 to 0.865 and from 0.746 to 0.815, respectively. These test–retest assessments suggest that the domains and total score of the HCMSQv1.0 produce stable, reliable scores when administered under similar conditions.

3.3.3 Construct Validity

For convergent validity, moderate correlations (e.g., HCMSQ shortness-of-breath and KCCQ-23 clinical summary score ranged from −0.72 to −0.71) were observed between the HCMSQv1.0 scores for cardiovascular

Table 4 Test–retest reliability for HCMSQ

| Domain | Correlation between HCMSQ scores at baseline (test) and W6 (retest) in patients reporting the same global impression of symptoms on the PGIC and PGIS | | | | Correlation between scores at W1 (test) and W2 (retest) in patients with no change in clinical status during screening | |
| | ICC-Baseline_W6 (PGIC) | | ICC-Baseline_W6 (PGIS) | | ICC-Screening_W1_W2 | |
	N	Correlation	N	Correlation	N	Correlation
MAVERICK-HCM trial data set						
Cardiovascular symptoms	11	0.829	15	0.746	33	0.754
Tiredness	11	0.862	15	0.815	33	0.807
Shortness of breath	11	0.841	15	0.753	33	0.814
Total score	11	0.865	15	0.796	33	0.762
EXPLORER-HCM trial data set						
Cardiovascular symptoms	–	–	–	–	194	0.888
Tiredness	–	–	–	–	194	0.827
Shortness of breath	–	–	–	–	194	0.844
Total score	–	–	–	–	194	0.886

HCM hypertrophic cardiomyopathy, *HCMSQ* Hypertrophic Cardiomyopathy Symptom Questionnaire, *ICC* intraclass correlation coefficient, *PGIC* Patient Global Impression of Change, *PGIS* Patient Global Impression of Severity, *W* week

symptoms, tiredness, shortness of breath, and total score, and the conceptually similar or related KCCQ-23 domains and summary scores (physical limitation, symptom frequency, symptom burden, total summary score, clinical summary score, quality of life, social limitation, and overall summary score) and those in the EQ-5D-5L visual analog scale (Table S3 in the ESM).

For known-groups validity, in both the MAVERICK-HCM and EXPLORER-HCM studies, significant differences in the expected direction were observed by PGIS category for the HCMSQv1.0 scores for tiredness, shortness of breath, and total symptoms, but the score for cardiovascular symptoms was only significant in the EXPLORER-HCM study. The lack of significance for cardiovascular symptoms was most likely due to the sample size of the MAVERICK-HCM study. Significant differences were not observed among HCMSQv1.0 domain/total scores by NYHA class in MAVERICK-HCM (Fig. 2). This was also most likely due to small sample size. For the EXPLORER-HCM study, significant differences by NYHA classification were observed in the HCMSQ scores for tiredness, shortness-of-breath, and total symptoms, but not in the score for cardiovascular symptoms.

The presence of differential item functioning was not observed for any of the HCMSQ items because there were no significant differences in slopes or intercepts between the populations.

3.3.4 Sensitivity to Change

When analyzing change from baseline to week 16 by treatment status in the MAVERICK-HCM study, HCMSQv1.0 scores for tiredness, shortness-of-breath, and total symptoms showed a statistically significant change, whereas that for cardiovascular symptoms did not (Table 5). Despite the small sample sizes, the responsiveness of the HCMSQv1.0 scores was demonstrated for shortness-of-breath when using both the PGIC and PGIS, and tiredness and total symptoms when using the PGIC (Table S4 in the ESM).

3.4 Within-Patient Meaningful Change (Responder Definition)

In the MAVERICK-HCM study, anchor-based analyses (conducted using longitudinal data) showed that the PGIC at week 16 was correlated ($r > 0.3$) with change in each HCMSQv1.0 domain score from baseline to week 16 (shortness of breath, tiredness, cardiovascular symptoms, and total symptom) and was therefore considered an appropriate anchor. The change in PGIS from baseline to week 16 was correlated above 0.3 with the change in each HCMSQv1.0 score, except for cardiovascular symptoms ($r = -0.17$). The PGIS was therefore used as an anchor to define a responder definition only in shortness of breath, tiredness, and total symptom scores.
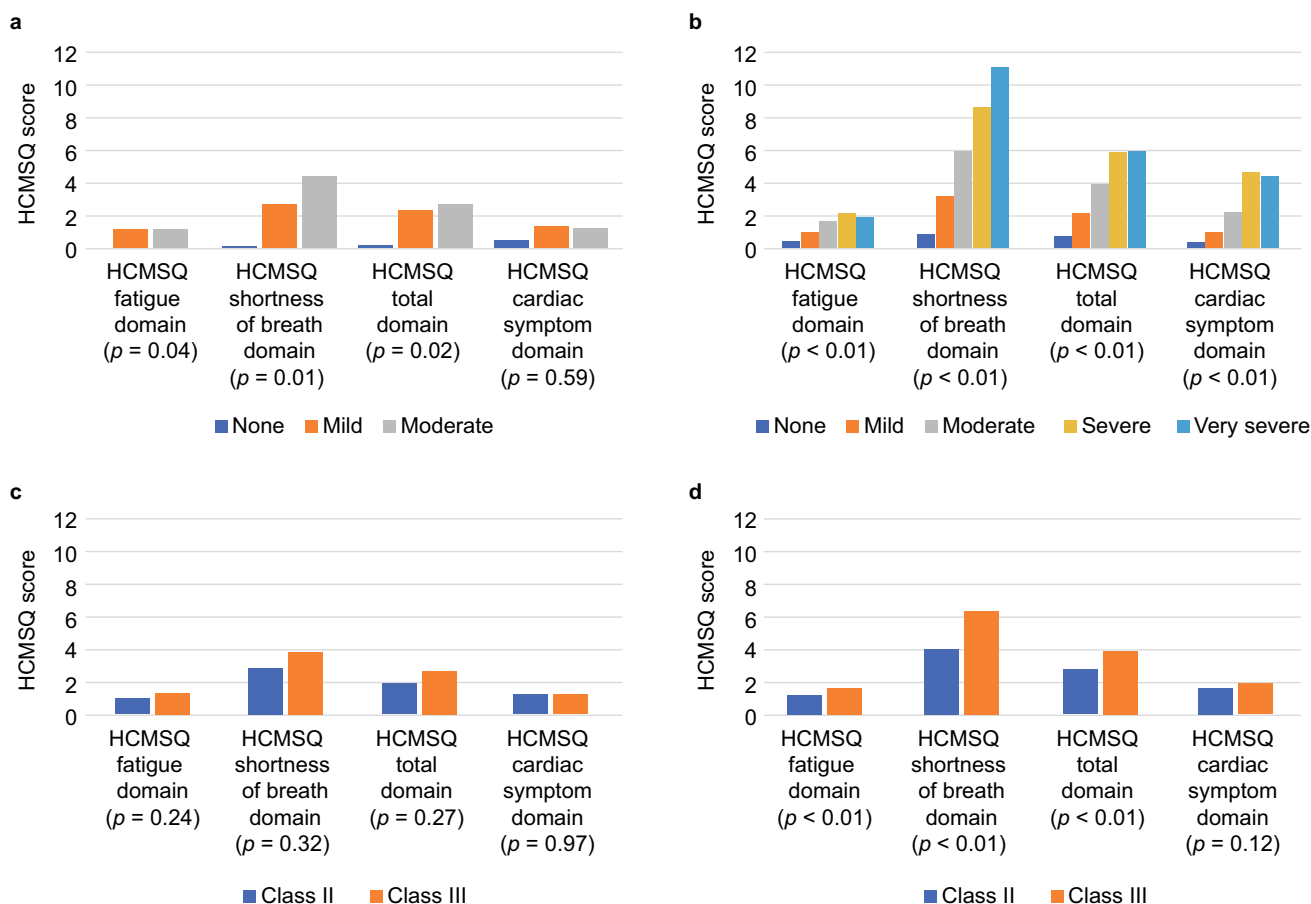
Fig. 2 HCMSQv1.0 domain and total scores by PGIS and NYHA classification for MAVERICK-HCM and EXPLORER-HCM. HCMSQ scores by **a** PGIS for MAVERICK-HCM; **b** PGIS for EXPLORER-HCM; **c** NYHA for MAVERICK-HCM; and **d** NYHA for EXPLORER. *HCM* hypertrophic cardiomyopathy, *HCMSQ* Hypertrophic Cardiomyopathy Symptom Questionnaire, *HCMSQv1.0* HCMSQ version 1.0, *NYHA* New York Heart Association, *PGIS* Patient Global Impression of Severity

Table 5 Sensitivity to change: mean change in HCMSQ scores from baseline to week 16 (test against mean change = 0) for MAVERICK-HCM trial data set

Domain	N	Mean (SD)	T-value	Prob T
Cardiovascular symptoms	21	−0.16 (1.25)	−0.60	0.557
Tiredness	21	−0.35 (0.70)	−2.3	0.035
Shortness of breath	21	−1.10 (1.61)	−3.1	0.005
Total score	21	−0.68 (1.26)	−2.5	0.023

HCM hypertrophic cardiomyopathy, *HCMSQ* Hypertrophic Cardiomyopathy Symptom Questionnaire, *SD* standard deviation

Responder definition estimates from anchor- and distribution-based approaches using MAVERICK-HCM and EXPLORER-HCM trial data sets are shown in Table 6. Based on these data, we have selected a responder-definition range of 1–2 points for the HCMSQv1.0 shortness of breath and total symptom scores, and approximately 1 point for the tiredness and cardiovascular symptom scores to identify individuals who experienced a significant change in symptoms from baseline in the EXPLORER-HCM study.

4 Discussion

PRO instruments provide and quantify information relevant to patients' health status, such as disease and treatment outcomes [20]. A successful PRO instrument that completely captures relevant manifestations of disease, as well as treatment outcomes, can guide clinical research and decision making, and help to improve patient outcomes.

The HCMSQ was developed rigorously and iteratively in line with FDA guidance on PRO development [10]. Specifically, the HCMSQv1.0 was developed from a conceptual model informed by 27 interviews with patients with HCM from four countries [7], and then conceptually and cognitively tested for comprehensiveness and clarity through cognitive interviews with a total of 33 patients with HCM, and

Table 6 HCMSQ responder definition estimates from anchor- and distribution-based approaches using MAVERICK-HCM and EXPLORER-HCM trial data sets

Domain	Anchor-based approach		Distribution-based approach	
	PGIC improved week 16 (*N* = 10)	PGIS improved week 16 (*N* = 6)	½ SD	1 SEM
MAVERICK-HCM trial data set				
Cardiovascular symptoms	0.39	–	0.51	0.51
Tiredness	0.83	0.91	0.34	0.30
Shortness of breath	2.01	2.35	1.08	0.93
Total score	1.46	1.59	0.62	0.60
EXPLORER-HCM trial data set				
Cardiovascular symptoms	–	–	0.87	0.58
Tiredness	–	–	0.37	0.31
Shortness of breath	–	–	1.45	1.15
Total score	–	–	0.89	0.60

HCM hypertrophic cardiomyopathy, *HCMSQ* Hypertrophic Cardiomyopathy Symptom Questionnaire, *PGIC* Patient Global Impression of Change, *PGIS* Patient Global Impression of Severity, *SD* standard deviation, *SEM* standard error of the mean

For EXPLORER-HCM, only distribution-based approaches (i.e., ½ SD and 1 SEM) were conducted given that only baseline data were used

a card-sorting task with 22 patients with HCM. This PRO instrument measures the key symptoms of HCM and can complement the use of other heart disease-specific PROs, such as the KCCQ-23, and standard assessment tools used by physicians, such as NYHA functional class, to gain a comprehensive understanding of symptoms of HCM from patients with the disease.

The scoring algorithm for version 1.0 of the HCMSQ was developed empirically from the MAVERICK-HCM trial among patients with nonobstructive HCM and cross-sectionally in the EXPLORER-HCM trial among patients with obstructive HCM, using both classical and modern test theories to find support for four domains (shortness of breath [4 items], tiredness [1 item], cardiovascular symptoms [3 items], and syncope [1 item]) and a total symptom score. Items 4 and 5 from the shortness-of-breath domain were excluded from the scoring of the HCMSQ and will be removed from further iterations of the HCMSQ (HCMSQ version 2.0 [HCMSQv2.0]). Syncope does not contribute to the total score but provides relevant information about patient experiences with HCM, and will therefore remain as a domain in the instrument. The decision to exclude syncope from the total score was due to its low occurrence in the studied HCM patient population and lack of variation in the item response. The psychometric measurement properties of the HCMSQ were tested without items 4 and 5 and are thus supporting HCMSQv2.0.

The measurement properties of these HCMSQ scores (except for syncope) were evaluated longitudinally in the MAVERICK-HCM trial among patients with nonobstructive HCM and cross-sectionally in the EXPLORER-HCM trial among patients with obstructive HCM. Reliability,

validity, and sensitivity to change were established for all scores. However, further corroboration is required to be able to conclude that known-groups validity has been established for the score for cardiovascular symptoms.

Overall, these findings support the HCMSQ as a fit-for-purpose PRO instrument for the assessment of treatment benefit in patients with HCM following the removal of items 4 and 5, making a 9-item scale (the HCMSQv2.0). The shortness-of-breath subscale of the HCMSQ is particularly well supported, demonstrating strong content validity and psychometric performance.

Some limitations should be considered when interpreting these findings. Although the sample for the cross-sectional analysis was robust, the sample size for the longitudinal research was relatively small and was restricted to a short time period (16 weeks) in a population with nonobstructive HCM in the MAVERICK-HCM study. As such, to facilitate its use to define disease severity and support trial endpoints in obstructive HCM, longitudinal confirmation of psychometric measurement properties in an obstructive HCM population is required. Furthermore, there were some missing data, owing to incomplete questionnaires from the participants; however, missing-data simulation analyses were conducted to support the strategy for accounting for these data gaps (Table S1 in the ESM). The shortness-of-breath domain score is a secondary endpoint of the EXPLORER-HCM trial, and the whole instrument will be analyzed and further explored in patients with obstructive HCM, while further research will be needed in larger patient populations with nonobstructive HCM.

The calculation of the total score by aggregating the domain scores could also be revisited, with further clinical

considerations to inform the weighting of different domains rather than relying solely on the statistical weighting, which offered no differential weighting for symptoms known to be differentially prominent and bothersome per the conceptual disease model of HCM. Additional analyses correlating total score with clinical outcomes are also being performed and will be reported separately.

The HCMSQ was designed to evaluate symptoms alone in patients with HCM and did not assess the broader concepts of patient functioning, mood, or quality of life. Other PRO instruments are available to evaluate these attributes and could be used in conjunction with the HCMSQ if required.

5 Conclusions

By considering cross-sectional data in obstructive HCM and short-term longitudinal data in a small sample of patients with non-obstructive HCM, this analysis demonstrated the content validity of the HCMSQ and provided empirical evidence to support the scoring algorithm. Moreover, preliminary evidence indicated psychometric reliability and validity in HCM. Further research is needed to confirm its use in larger and longitudinal samples prior to use. Overall, the HCMSQ is the first and only HCM-specific PRO and may help to establish treatment benefit and symptom response in patients with HCM participating in clinical trials, to contribute to HCM diagnosis and monitoring, and to inform understanding of the experience of patients receiving care in clinical practice.

Declarations

Conflicts of interest Matthew Reaney and Veleka Allen are employees of IQVIA and have received funding from MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb, for work performed on this study. Amy J. Sehnert is an employee of MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb. Liang Fang is no longer an employee of MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb, but is now an employee of Bolt Biotherapeutics Inc. Albert A. Hagège has received financial support from Amicus, Bristol Myers Squibb, Gilead, MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb, and Sanofi Genzyme. Srihari S. Naidu has served as a consultant for MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb, and Cytokinetics, and has served as an executive committee member for the clinical trial VALOR-HCM (sponsored by MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb.). Iacopo Olivotto has received grants and honoraria from MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb, Amicus, Shire Takeda, Sanofi Genzyme, and Menarini International and has served as chair of the steering committee for the clinical trial EXPLORER-HCM (sponsored by MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb).

Availability of data and material The Bristol Myers Squibb policy on data sharing may be found at https://www.bms.com/researchers-and-partners/independent-research/data-sharing-request-process.html.

Ethics approval All aspects of the EXPLORER-HCM (NCT03470545) and MAVERICK-HCM (NCT03442764) studies were conducted in compliance with the Health Insurance Portability Accountability Act of 1996 (USA). Institutional review board and relevant ethics committee approvals were obtained prior to participant enrollment. The study protocols were approved by a US institutional review board: The New England Independent Review Board (197 First Avenue, Suite 250, Needham, MA 02494). Both studies fully adhered to the ethical principles of the Declaration of Helsinki and the specifications of the International Council on Harmonisation and Good Clinical Practice.

Consent to participate All participants provided informed (written and verbal) consent prior to their participation.

Code availability Not applicable.

Consent for publication Not applicable.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by MR, VA, AJS, and LF. All authors contributed to the drafting and revision of the manuscript and approved the final manuscript.

References

1. Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, et al. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). Eur Heart J. 2014;35(39):2733–79. https://doi.org/10.1093/eurheartj/ehu284.

2. Semsarian C. Guidelines for the diagnosis and management of hypertrophic cardiomyopathy. Heart Lung Circ. 2011;20(11):688–90. https://doi.org/10.1016/j.hlc.2011.07.017.

3. Maron BJ. Clinical course and management of hypertrophic cardiomyopathy. N Engl J Med. 2018;379(7):655–68. https://doi.org/10.1056/NEJMra1710575.

4. Seferović PM, Polovina M, Bauersachs J, Arad M, Gal TB, et al. Heart failure in cardiomyopathies: a position paper from the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail. 2019;21(5):553–76. https://doi.org/10.1002/ejhf.1461.

5. Sequeira V, Bertero E, Maack C. Energetic drain driving hypertrophic cardiomyopathy. FEBS Lett. 2019;593(13):1616–26. https://doi.org/10.1002/1873-3468.13496.

6. Lu DY, Pozios I, Haileselassie B, Ventoulis I, Liu H, et al. Clinical outcomes in patients with nonobstructive, labile, and obstructive hypertrophic cardiomyopathy. J Am Heart Assoc. 2018;7(5): e006657. https://doi.org/10.1161/jaha.117.006657.

7. Zaiser E, Sehnert AJ, Duenas A, Saberi S, Brookes E, et al. Patient experiences with hypertrophic cardiomyopathy: a conceptual model of symptoms and impacts on quality of life. J Patient Rep Outcomes. 2020;4(1):102–12. https://doi.org/10.1186/s41687-020-00269-8.

8. Maron MS, Hellawell JL, Lucove JC, Farzaneh-Far R, Olivotto I. Occurrence of clinically diagnosed hypertrophic cardiomyopathy in the United States. Am J Cardiol. 2016;117(10):1651–4. https://doi.org/10.1016/j.amjcard.2016.02.044.

9. Ommen SR, Mital S, Burke MA, Day SM, Deswal A, et al. 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2020;142(25):e558–631. https://doi.org/10.1161/cir.0000000000000937.

10. US Department of Health and Human Services. Food and Drug Administration (2009) 74 FR 65132—guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims; availability. www.federalregister.gov/documents/2009/12/09/E9-29273/guidance-for-industry-on-patient-reported-outcome-measures-use-in-medical-product-development-to. Accessed 21 Feb 2021.

11. Ho CY, Olivotto I, Jacoby D, Lester SJ, Roe M, et al. Study design and rationale of EXPLORER-HCM: evaluation of mavacamten in adults with symptomatic obstructive hypertrophic cardiomyopathy. Circ Heart Fail. 2020;13(6): e006853. https://doi.org/10.1161/circheartfailure.120.006853.

12. Olivotto I, Oreziak A, Barriales-Villa R, Abraham TP, Masri A, et al. Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet. 2020;396(10253):759–69. https://doi.org/10.1016/S0140-6736(20)31792-X.

13. Ho CY, Mealiffe ME, Bach RG, Bhattacharya M, Choudhury L, et al. Evaluation of mavacamten in symptomatic patients with nonobstructive hypertrophic cardiomyopathy. J Am Coll Cardiol. 2020;75(21):2649–60. https://doi.org/10.1016/j.jacc.2020.03.064.

14. Wild D, Grove A, Martin M, Eremenco S, McElroy S, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health. 2005;8(2):94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x.

15. Kyriazos TA. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. Psychology. 2018. https://doi.org/10.4236/psych.2018.98126.

16. O'Neill TR, Gregg JL, Peabody MR. Effect of sample size on common item equating using the dichotomous Rasch model. Appl Measur Educ. 2020;33(1):10–23. https://doi.org/10.1080/08957347.2019.1674309.

17. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum; 1988.

18. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol. 2000;35(5):1245–55. https://doi.org/10.1016/s0735-1097(00)00531-3.

19. EuroQol Research Foundation website (2020) EQ-5D: about. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/. Accessed 21 Feb 2021.

20. Kelkar AA, Spertus J, Pang P, Pierson RF, Cody RJ, et al. Utility of patient-reported outcome instruments in heart failure. JACC Heart Fail. 2016;4(3):165–75. https://doi.org/10.1016/j.jchf.2015.10.015.