Longitudinal Psychometric Analysis of the Hypertrophic Cardiomyopathy Symptom Questionnaire (HCMSQ) Using Outcomes from the Phase III EXPLORER-HCM Trial

Matthew Reaney1 · Prithvi Addepalli1 · Veleka Allen2 · John A. Spertus3 · Chantal Dolan4 · Amy J. Sehnert5 · Jennifer T. Fine5

Accepted: 16 May 2022 / Published online: 20 June 2022 © The Author(s) 2022

Abstract
Background Hypertrophic cardiomyopathy (HCM) symptoms include shortness of breath (SOB), fatigue, chest pain, palpitations, dizziness, and fainting. The HCM Symptom Questionnaire (HCMSQ), the only patient-reported outcome instrument designed to specifically measure HCM symptoms, yields four domain scores (SOB, tiredness, cardiovascular symptoms, syncope) and a total score. We evaluated the longitudinal psychometric properties of the HCMSQ using baseline to week 30 data from the phase III EXPLORER-HCM trial (NCT03470545).

Methods Test–retest reliability was assessed via intraclass correlation of patients with stable Patient Global Impression of Change (PGIC) and Patient Global Impression of Severity (PGIS) responses. Sensitivity to change was assessed via Spearman correlations with the Kansas City Cardiomyopathy Questionnaire (KCCQ-23) and the EuroQoL visual analogue scale (EQ VAS), and via one-way ANOVA comparing change groups defined on clinical (New York Heart Association [NYHA] class, left ventricular outflow tract [LVOT] gradient, peak oxygen consumption [pVO2]) and patient-reported (PGIS, PGIC) variables. Meaningful change thresholds were established via PGIC/PGIS.

Results All HCMSQ scores showed strong evidence of test–retest reliability (intraclass correlation coefficient > 0.70). Sensitivity to change was demonstrated with mostly strong/moderate correlations with KCCQ-23 and EQ VAS, and significant differences (p ≤ 0.05) in PGIS, PGIC, pVO2, and NYHA (except tiredness domain) change categories, but not LVOT gradient. Clinically meaningful score reductions were ≥1 point for tiredness and cardiovascular symptoms domains, ≥2.5 points for SOB domain, and ≥2 points for total score.

Conclusions Results suggest that HCMSQ is fit for purpose in capturing HCM symptoms and may provide evidence of treatment benefit from the patients’ perspectives.

Key Points for Decision Makers
- A proportion of patients with hypertrophic cardiomyopathy (HCM) report limiting symptoms, including shortness of breath, tiredness, chest pain, palpitations, dizziness, and fainting.
- This study assessed the validity of the HCM Symptom Questionnaire (HCMSQ), a new patient-reported outcome instrument designed to measure symptoms and treatment benefit from the perspective of patients with HCM.
- The HCMSQ demonstrated acceptable longitudinal measurement properties, supporting it to be a reliable and valid tool to capture change in symptoms associated with treatments for HCM, and supporting its use in capturing patients’ symptoms of HCM.
1 Introduction

Hypertrophic cardiomyopathy (HCM) is a primary myocardial disorder defined by left ventricular hypertrophy that cannot be explained by another cardiac or systemic disease [1]. Although fewer than one in 3000 people receive a diagnosis of HCM globally, it is thought that between one in 200 and one in 500 people in the general population have the disorder [1–4]. HCM is classified as obstructive or nonobstructive based on the presence or absence of left ventricular outflow tract (LVOT) obstruction [3, 5, 6]. Approximately two-thirds of patients with HCM are affected by obstructive HCM [6], and it is estimated that 50% of all patients diagnosed with obstructive HCM have one or more symptoms related to the condition [7].

Recent qualitative and quantitative research resulted in the development of a conceptual model to gain an in-depth understanding of the cardinal signs, symptoms, and impacts of HCM [8]. Core HCM symptoms are: shortness of breath, especially with physical exertion; fatigue; chest pain; palpitations; dizziness; and fainting (syncope) [2, 3, 6, 8–10]. These symptoms have physical, emotional, and social impacts on patients, and contribute to significant impairment in quality of life [9, 11].

The primary treatment goal for HCM is to improve patients’ symptoms and health status [3, 12]. Given the nature of HCM symptoms and their impact on patients’ function and quality of life, it is critical to understand the effect of interventions on these symptoms. Data generated from a fit-for-purpose patient-reported outcome (PRO) instrument can provide robust and direct evidence of treatment benefit from the patient’s perspective [13]. Such data can be used by regulators to understand the benefit–risk profile of the intervention; by payer and health technology assessment agencies to understand differences between treatments; by healthcare professionals to identify interventions from which patients are likely to gain benefit; and by patients to engage in shared treatment decision-making [1, 14, 15]. For a PRO instrument to be considered ‘fit for purpose’ it must be well defined and generate reliable and interpretable data. Specifically, the PRO instrument must be content valid, and it must demonstrate adequate psychometric measurement properties in the context in which it will be used, including being a valid measure of the domains it seeks to quantify, being reproducible in stable patients, and being sensitive to clinical changes.

The HCM Symptom Questionnaire (HCMSQ) has recently been developed as a novel PRO instrument for measuring the symptoms of HCM, in line with US Food and Drug Administration (FDA) guidance for developing PROs [13, 16]. In brief, cognitive debriefing interviews and a card-sorting task were conducted in 33 and 22 patients with HCM, respectively, and baseline blinded data from two trials (EXPLORER-HCM and MAVERICK-HCM) were pooled (N = 299) to develop the scoring algorithm [16]. 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 [16]. During development of the tool, robust evidence of content validity was shown in populations with nonobstructive HCM as well as obstructive HCM and empirical evidence from both populations was used to develop the scoring algorithm [16]. Preliminary evidence of psychometric measurement properties in HCM has been shown, from cross-sectional data of patients with obstructive HCM and short-term longitudinal data of a small population with nonobstructive HCM; this included item-to-item correlation, item–scale analyses, test–retest reliability, convergent validity, known-groups validity, and responsiveness [16]. The known-groups analysis, performed to determine the degree to which the HCMSQ can distinguish between defined groups of patients, showed significant differences for many domains by Patient Global Impression of Severity (PGIS) category and by New York Heart Association (NYHA) classification [16]. To support the appropriateness of the HCMSQ instrument for defining disease severity and for use in clinical studies in obstructive HCM, longitudinal confirmation of psychometric measurement properties in an obstructive HCM population are required. Here, we report the results of these longitudinal analyses, based on patient data from the phase III EXPLORER-HCM trial (ClinicalTrials.gov identifier NCT03470545).

2 Methods

2.1 Study Design and Patient Population

The EXPLORER-HCM trial was a phase III, double-blind, 1:1 randomized study to evaluate the safety, tolerability, and efficacy of mavacamten (starting at 5 mg) versus placebo in patients with symptomatic obstructive HCM [17]. The study included a 5-week screening period, a 30-week treatment period, and an 8-week post-treatment follow-up period (±7 days). A total of 251 patients with obstructive HCM were enrolled in EXPLORER-HCM (Table 1 shows demographic characteristics of the study sample). Additional details of the EXPLORER-HCM methodology, patient characteristics, and key clinical and safety results have been described elsewhere [17–20].
The HCMSQ was designed based on an HCM conceptual model of symptoms and impacts [8], and measures the core symptoms of HCM (covering tiredness/fatigue, heart palpitations, chest pain, dizziness, and shortness of breath) over a 24-hour recall period. The HCMSQ is intended to be administered on 7 consecutive days with item scores calculated as the mean of the 7-day (weekly) scores divided by the number of scores available for that week. The HCMSQ produces four domain scores (shortness of breath [4 items], tiredness [1 item], cardiovascular symptoms [3 items], syncope [1 item]), which are combined to create a total score (sum of the first three domains [8 items]) ranging from 0 to 12.5 [16]. The complete HCMSQ instrument is available in Supplementary Table S1 in the electronic supplementary material (ESM).

In EXPLORER-HCM, the HCMSQ was administered electronically on provisioned handheld devices daily during screening and through the first 6 weeks of the treatment period, and then for 7 consecutive days prior to the week 10, 14, 18, 22, 26, 30 (end of treatment), and 38 (end of study) study visits. Patients were only able to submit a daily entry if all items were completed. For HCMSQ data to be included in the evaluation, patients were required to have completed a daily entry on at least 4 out of the 7 consecutive days prior to the visit. If fewer than 4 days were available in the 7-day period, then a weekly score was not calculated and the entries for the week were treated as missing. HCMSQ scoring was performed as outlined in Supplementary Table S2 in the ESM. For example, for shortness of breath, the score is the sum of four shortness of breath items, with a potential score range of 0–18; lower scores indicate less shortness of breath. To guide the handling of missing data, a missing-data simulation study for HCMSQ domain and total scores has been performed previously [16].

In addition to the HCMSQ, several other PRO instruments were administered in the EXPLORER-HCM study, including the Kansas City Cardiomyopathy Questionnaire-23 (KCCQ-23) [17, 18, 21], the EuroQol visual analogue scale (EQ VAS) [22], the PGIS, and the Patient Global Impression of Change (PGIC). The KCCQ-23 asks respondents to describe the frequency and severity of symptoms, physical and social limitations, and how they perceive that their heart failure symptoms impact their quality of life. It includes 23 items, has a recall period of 2 weeks, and produces individual scores for each of the seven domains (symptom burden, symptom frequency, physical limitation, quality of life, social limitation, self-efficacy, and symptom stability). In addition, it generates a total symptom score, which combines symptom burden and symptom frequency scales, and two summary scores—the clinical summary score is the sum of the total symptom score and physical limitation scales; the overall summary score is the sum of the total symptom score, clinical summary score, social limitation, and quality of life scales [21]. All scores are transformed into a score ranging from 0 to 100, with higher scores reflecting better health status. The EQ VAS records respondents’ self-rated health on a vertical, graduated (0–100) VAS, whereby 0
represents worst health imaginable and 100 represents the best health imaginable. Higher scores indicate better quality of life [22]. The PGIS is a single-item questionnaire that asks respondents to rate their overall symptom severity in the past week on a 5-point rating scale (from 1 [no symptoms] to 5 [very severe]). Higher scores indicate greater severity (see also Supplementary Methods in the ESM). The PGIC is a single-item questionnaire that asks respondents to rate their overall change in symptom severity compared with their status at randomization over time on a 7-point rating scale (from 1 [very much improved] to 7 [very much worse]). Higher scores indicate a change for the worse (see also Supplementary Methods in the ESM).

The PGIS was administered electronically at screening, and both the PGIS and the PGIC were administered electronically at the week 6, 10, 14, 18, 22, 26, 30, and 38 study visits. The other PROs (KCCQ-23 and EQ VAS) were administered electronically at the baseline, week 6, 12, 18, 30, and 38 study visits on the same devices. The scoring for the KCCQ-23 instrument and the EQ VAS was carried out according to the procedures described by their respective scoring manuals; patients with missing forms at a visit were excluded from the analysis at that time point.

### 2.3 HCMSQ Psychometric Analyses

Longitudinal psychometric analyses of the HCMSQ and derivation of a responder definition specific to the HCM population were conducted in line with FDA guidance [13]. This psychometric analysis outlines the longitudinal measurement properties focusing on change from baseline, using baseline and week 30 data from the EXPLORER-HCM trial; supplemental analyses also included data from weeks 6 and 18. The following scores were evaluated per the scoring algorithm (see also Supplementary Tables S1 and S2 in the ESM): shortness of breath (items 1–3 and 6), tiredness (item 7), cardiovascular symptoms (items 8–10), and total score. Item 11 (syncope) is not included in the total score. The decision to exclude syncope from the total score was due to its low occurrence in the patient population; from a psychometric evaluation perspective there was not enough variation in this item response to be included in the total score. However, because it provides relevant information about patient experiences with HCM, it remains part of the scale. Following results from Rasch modeling, items 4 and 5 are not included in the scoring of any domain or total scores, and will be removed in future administrations of the HCMSQ [16].

#### 2.3.1 Test–Retest Reliability

Test–retest reliability is the extent to which an instrument yields consistent, reproducible estimates of a patient’s health status in a stable population. Test–retest reliability of the HCMSQ was evaluated by calculating agreement-based intraclass correlation coefficients (ICCs) using a two-way mixed model among a subgroup of stable patients. Stable patients were defined separately in two analyses: first using the PGIC (i.e., patients who reported “no change” on the PGIC at baseline and week 6, week 6 and week 18, and week 18 and week 30), and second using the PGIS (i.e., patients who endorsed the same response at both baseline and week 6, week 6 and week 18, and week 18 and week 30). ICC values > 0.7 were considered evidence of test–retest reliability in both analyses [23, 24].

#### 2.3.2 Sensitivity to Change/Responsiveness

Sensitivity to change refers to the extent to which the instrument can detect true change in patients known to have changed in clinical status. It was examined via correlations between change scores and via tests of differences between ordinal change categories.

#### 2.3.2.1 Correlations Between Change Scores

Spearman correlation coefficients were calculated between changes from baseline to week 30 in the HCMSQ domain/total scores, and changes from baseline to week 30 in the domain scores for the KCCQ-23 overall summary, clinical summary, and total symptom scores, and EQ VAS score. Moderate correlations (r ≥ 0.3) were considered evidence for sensitivity to change.

#### 2.3.2.2 Differences Between Ordinal Change Categories

Mean change in HCMSQ domain/total scores from baseline to week 30 across ordinal change groups was assessed via the following variables and ordinal change categories:

1. Change in NYHA classification from baseline to week 30 (i.e., improvement, no change, worsening).
2. Change in PGIS score from baseline to week 30 (i.e., worsened, no change, improved by one category, improved by two or more categories).
3. Change in PGIC score from baseline to week 30 (i.e., worsened, no change, minimally improved, much improved, very much improved).
4. Change in peak oxygen consumption (pVO₂) from baseline to week 30 quartile group (i.e., decreased, no change, increased).
5. Change in post-exercise LVOT gradient scores from baseline to week 30 (i.e., decrease by two percentile groups [0–25th, 26–50th, 51–75th, and 76–100th], decrease by one percentile group, no change, increase by one percentile group, increase by two percentile groups).
Change group differences were tested via one-way analysis of variance (ANOVA). To assist in interpretation, two effect size measures are reported for each change category level: Cohen’s $d$, whereby a category’s mean change score is divided by the pooled baseline standard deviation (SD); and standardized response mean (SRM), whereby a category’s mean change score SD. The magnitude of these effect size measures is most meaningfully and intuitively interpreted in terms of non-overlap ($u$) between distributions: when the effect size is 0, populations’ distributions are perfectly superimposed and $u = 0$; when the effect size is 0.2 (i.e., a small effect), normally distributed populations of equal size and variability have only 14.7% of their combined area which does not overlap; when the effect size is 0.5 (i.e., a medium effect), 33% of the combined areas covered by the two distributions do not overlap; and when the effect size is 0.8 (i.e., a large effect), the two populations are so separated that almost half ($u = 47.4\%$) of their areas do not overlap [25].

### 2.3.3 Meaningful Change Analysis (Responder Definition)

Meaningful change (i.e., the amount of individual-level change over a predefined period that could be interpreted as a meaningful clinical benefit) was evaluated by applying anchor-based methods to establish thresholds for meaningful change, using baseline and week 30 data and PGIC and PGIS scores as anchors [26]. To explore the adequateness of the anchor, the correlation (Spearman’s rho) of these anchors with the change in HCMSQ weekly scores was assessed. The anchor was implemented only if the correlation coefficient was > 0.30. Groups were formed from the “improved” PGIC responses, and mean change values were calculated for the HCMSQ weekly domain and total scores. These values represented the responder thresholds. The PGIS baseline to week 30 change scores were used to identify a sample of patients whose PGIS scores improved by at least one level. Mean change values for this sample represented a second estimate of responder thresholds.
3 Results

A total of 251 patients were enrolled in the EXPLORER-HCM study. Baseline demographic characteristics as well as baseline severity scores are shown in Table 1. Patients had a mean age of 58.5 (SD: 11.9) years, 41% (102/251) were women, and most (183/251; 73%) were in NYHA class II. More than 80% of patients had evaluable HCMSQ data at baseline; the compliance rates stayed above 70% across all further time points (weeks 3, 6, 14, 18, 22, 26, 30), with the exception of week 10 when rates were > 50% (Supplementary Table S3, see ESM).

3.1 HCMSQ Longitudinal Psychometric Analyses

3.1.1 Test–Retest Reliability

In PGIC-based analyses, all HCMSQ domain scores had acceptable test–retest reliability using data from baseline and week 6 (n = 51), week 18 (n = 87), and week 30 (n = 106) (shortness of breath, ICC = 0.78–0.79; tiredness, ICC = 0.76–0.82; cardiovascular symptoms, ICC = 0.72–0.97; total score, ICC = 0.82–0.90; see Table 2). Similarly, all HCMSQ domain scores had acceptable test–retest reliability (ICC > 0.70) for PGIS-based analyses (Table 2).

3.1.2 Sensitivity to Change/Responsiveness

3.1.2.1 Correlations Between Change Scores

HCMSQ domain scores and the total score had moderate or strong Spearman correlations with all KCCQ-23 and EQ VAS change scores (r ≥ 0.3) (Table 3).

3.1.2.2 Differences Between Ordinal Change Categories

HCMSQ domain and total change scores significantly differed across ordinal change levels for several measures of interest (Tables 4–8). There were statistically significant differences for all HCMSQ change scores by change in NYHA category (all p < 0.05), except for the tiredness score (Table 4); by change in PGIS score (all p < 0.0001; Table 5); by change in PGIC score (all p < 0.001; Table 6); and by change in pVO2 percentile group (all p ≤ 0.05; Table 7). There were no significant differences by change in LVOT gradient percentile group (Table 8). Effective size measures were consistently of greater magnitude for the improvement groups than for the no-change groups.

3.1.3 Meaningful Change Analysis (Responder Definition)

Thresholds with improvement categories derived from the PGIC (n = 136) and PGIS (n = 90) analyses were 1.00 and 1.37, respectively, for cardiovascular symptoms; 2.57 and

Table 4  Assessment of sensitivity to change: mean (SD) change from baseline to week 30 on HCMSQ domains, by NYHA category change

| Domain                  | n  | Mean  | SD   | Omnibus test | Pooled SD | Effect sizea |
|-------------------------|----|-------|------|--------------|-----------|--------------|
|                         |    |       |      | F  | p     | Baseline | Change score | d  | SRM |
| Cardiovascular symptoms | 3.18| 0.0445| 1.69 | 1.40 | | | | | |
| Worsened                | 7  | −0.58 | 0.78 | −0.34 | −0.41 |
| No change               | 73 | −0.38 | 1.68 | −0.23 | −0.28 |
| Improved                | 73 | −0.98 | 1.16 | −0.58 | −0.70 |
| Tiredness               | 2.58| 0.0793| 0.73 | 0.69 | | | | | |
| Worsened                | 7  | −0.33 | 0.47 | −0.46 | −0.48 |
| No change               | 73 | −0.29 | 0.73 | −0.39 | −0.41 |
| Improved                | 73 | −0.55 | 0.68 | −0.75 | −0.79 |
| Shortness of breath     | 8.41| 0.0003| 2.89 | 2.73 | | | | | |
| Worsened                | 7  | −0.36 | 1.77 | −0.13 | −0.13 |
| No change               | 73 | −1.14 | 2.79 | −0.40 | −0.42 |
| Improved                | 73 | −2.80 | 2.54 | −0.97 | −1.03 |
| Total score             | 6.04| 0.0030| 1.74 | 1.58 | | | | | |
| Worsened                | 7  | −0.62 | 1.03 | −0.35 | −0.39 |
| No change               | 73 | −0.70 | 1.76 | −0.40 | −0.44 |
| Improved                | 73 | −1.57 | 1.38 | −0.90 | −0.99 |

Bold values indicate significant differences (p ≤ 0.05) between at least smallest and largest mean value for a given domain

d Cohen’s d. HCMSQ Hypertrophic Cardiomyopathy Symptom Questionnaire, NYHA New York Heart Association, SD standard deviation, SRM standardized response mean

a Effect size: d = change score mean divided by the pooled baseline SD; SRM = change score divided by the pooled change score SD
Psychometric Analysis of the HCMSQ from the Phase III EXPLORER-HCM Trial

3.40, respectively, for shortness of breath; and 1.47 and 2.04, respectively, for total score (see Table 9). For tiredness, a PGIC threshold was not calculated because correlation between change scores and anchor was < 0.3; with PGIS as anchor, a threshold of 0.73 was derived (Table 9).

4 Discussion

The HCMSQ, which was designed to measure the patient experience in the context of HCM, had already demonstrated good cross-sectional measurement properties (internal consistency reliability, convergent validity, known-groups validity) in obstructive HCM using baseline data from the EXPLORER-HCM trial, as well as adequate cross-sectional and longitudinal measurement properties (internal consistency reliability, convergent validity, known-groups validity, test–retest reliability, and sensitivity to change) in a relatively small population with nonobstructive HCM in the MAVERICK-HCM trial [16]. This present analysis was performed to supplement these analyses with longitudinal data from a larger data set, so that the HCMSQ could be considered reliable, valid, and interpretable for use in defining changes in obstructive HCM outcomes. Responder definitions aid the ease with which the data can be interpreted.

This study, conducted using longitudinal data from the EXPLORER-HCM data set, showed strong evidence for the longitudinal psychometric properties of all HCMSQ scores. The majority of HCMSQ domain scores had acceptable (ICC > 0.70) test–retest reliability, as demonstrated by analyses of PGIC- and PCIS-based stable patients across multiple time frames. The HCMSQ was sensitive to change: there were strong or moderate correlations for change from baseline to week 30, with change scores for measures of health-related quality of life (EQ VAS) and health-related quality of life for heart failure (KCCQ-23); and significant differences in HCMSQ change scores by NYHA class (except tiredness), PGIS, PGIC, and pVO2 categories, but not by post-exercise LVOT gradient percentiles. It is possible that non-significant differences by LVOT gradient were observed because a post-exercise gradient (not a resting or Valsalva LVOT gradient) was used, and patients’ symptoms that were experienced and then reported on the HCMSQ may not have been due to high levels of exertion. Most of the means for differences in HCMSQ change scores are in a logical order despite small sample sizes for some of the categories. A small number

Table 5  Assessment of sensitivity to change: mean (SD) change from baseline to week 30 on HCMSQ domains, by PGIS category change

| Domain                  | n  | Mean  | SD   | Omnibus test | Pooled SD | Effect sizea |
|-------------------------|----|-------|------|--------------|-----------|--------------|
| Cardiovascular symptoms |    |       |      |              |           |              |
| Worsened                | 12 | 0.75  | 1.71 | 19.11        | <0.0001   | 0.44         |
| No change               | 70 | −0.22 | 0.86 |              |           | −0.13        |
| Improved 1 cat          | 58 | −1.14 | 1.19 |              |           | −0.68        |
| Improved 2+ cats        | 13 | −2.35 | 2.32 |              |           | −1.40        |
| Tiredness               |    |       |      |              |           |              |
| Worsened                | 12 | 0.16  | 0.66 | 17.44        | <0.0001   | 0.22         |
| No change               | 70 | −0.18 | 0.59 |              |           | −0.25        |
| Improved 1 cat          | 58 | −0.62 | 0.63 |              |           | −0.84        |
| Improved 2+ cats        | 13 | −1.28 | 0.62 |              |           | −1.74        |
| Shortness of breath     |    |       |      |              |           |              |
| Worsened                | 12 | 1.21  | 2.61 | 27.24        | <0.0001   | 0.42         |
| No change               | 70 | −0.93 | 1.95 |              |           | −0.32        |
| Improved 1 cat          | 58 | −2.88 | 2.36 |              |           | −1.00        |
| Improved 2+ cats        | 13 | −5.56 | 2.77 |              |           | −1.93        |
| Total score             |    |       |      |              |           |              |
| Worsened                | 12 | 0.71  | 1.61 | 33.41        | <0.0001   | 0.41         |
| No change               | 70 | −0.49 | 1.04 |              |           | −0.28        |
| Improved 1 cat          | 58 | −1.72 | 1.26 |              |           | −0.98        |
| Improved 2+ cats        | 13 | −3.45 | 1.88 |              |           | −1.98        |

Bold values indicate significant differences (p ≤ 0.05) between at least smallest and largest mean value for a given domain

cat category, d Cohen’s d, HCMSQ Hypertrophic Cardiomyopathy Symptom Questionnaire, PGIS Patient Global Impression of Severity, SD standard deviation, SRM standardized response mean

aEffect size: d = change score mean divided by the pooled baseline SD; SRM = change score divided by the pooled change score SD
of rows in the assessment of the HCMSQ versus NYHA contradict the expected order; however, this can be attributed to the small group of patients with worsened symptoms \((n = 7)\) in this assessment. Although some of the correlations between change scores of HCMSQ with KCCQ-23 and EQ VAS scores were not as high as desirable, they did cross the threshold of 0.3 and are therefore acceptable [25].

The psychometric analyses presented here validate the HCMSQ tool for use in patient populations with obstructive HCM. Taken together with the analyses performed during the development of this instrument [16], the HCMSQ demonstrated content validity and psychometric support in nonobstructive HCM and obstructive HCM as well as well-defined meaningful change estimates in both populations, and the instrument is therefore fit for purpose to support an understanding of treatment benefit in HCM clinical trials. This makes it the first HCM-specific symptom PRO instrument to measure the patient experience in HCM. As such, it has potential to be used more widely by pharmaceutical companies, payer and health technology assessment agencies, and healthcare professionals to evaluate and track salient and meaningful symptoms of HCM. The data from the HCMSQ can provide reliable, valid, and interpretable data to support decision-making.

The similar thresholds derived from the PGIC and PGIS and the narrow ranges across time points allow us to propose the following a priori responder definitions, derived from preliminary minimal meaningful change thresholds: a reduction in score of at least 1 point for tiredness and cardiovascular symptoms domains; a reduction in score of at least 2.5 points for the shortness of breath domain; and a reduction in score of at least 2 points for the total score are considered clinically relevant for patients. These estimates

| Domain                        | Omnibus test | Pooled SD | Effect size* | SRM |
|-------------------------------|--------------|-----------|--------------|-----|
|                               | \(F\)        | \(p\)     | Baseline     | Change score | \(d\) | SRM |
| Cardiovascular symptoms       | 7.60         | \(<0.0001\) | 1.69         | 1.40 | 0.23 | 0.28 |
| Worsened                      | 14           | 0.39      | 1.43         | 0.23 | 0.28 |
| No change                     | 38           | −0.10     | 0.68         | −0.06 | −0.07 |
| Minimally improved            | 19           | −0.85     | 1.22         | −0.50 | −0.61 |
| Much improved                 | 61           | −0.77     | 1.33         | −0.46 | −0.55 |
| Very much improved            | 32           | −1.53     | 1.79         | −0.94 | −1.09 |
| Tiredness                     | 6.00         | \(0.0002\) | 0.73         | 0.69 | 0.20 | 0.21 |
| Worsened                      | 14           | 0.15      | 0.65         | 0.20 | 0.21 |
| No change                     | 38           | −0.30     | 0.69         | −0.41 | −0.43 |
| Minimally improved            | 19           | −0.33     | 0.59         | −0.45 | −0.47 |
| Much improved                 | 61           | −0.38     | 0.60         | −0.53 | −0.55 |
| Very much improved            | 32           | −0.82     | 0.77         | −1.12 | −1.19 |
| Shortness of breath           | 13.58        | \(<0.0001\) | 2.89         | 2.73 | 0.39 | 0.41 |
| Worsened                      | 14           | 1.13      | 2.18         | 0.39 | 0.41 |
| No change                     | 38           | −0.94     | 1.64         | −0.33 | −0.34 |
| Minimally improved            | 19           | −1.26     | 2.08         | −0.44 | −0.46 |
| Much improved                 | 61           | −2.22     | 2.65         | −0.77 | −0.81 |
| Very much improved            | 32           | −3.90     | 2.74         | −1.35 | −1.43 |
| Total score                   | 12.28        | \(<0.0001\) | 1.74         | 1.58 | 0.32 | 0.35 |
| Worsened                      | 14           | 0.56      | 1.32         | 0.32 | 0.35 |
| No change                     | 38           | −0.56     | 1.04         | −0.32 | −0.36 |
| Minimally improved            | 19           | −0.93     | 1.28         | −0.53 | −0.58 |
| Much improved                 | 61           | −1.20     | 1.42         | −0.69 | −0.76 |
| Very much improved            | 32           | −2.31     | 1.81         | −1.32 | −1.46 |

Bold values indicate significant differences \((p \leq 0.05)\) between at least smallest and largest mean value for a given domain

\(d\) Cohen’s \(d\), \(HCMSQ\) Hypertrophic Cardiomyopathy Symptom Questionnaire, \(PGIC\) Patient Global Impression of Change, \(SD\) standard deviation, \(SRM\) standardized response mean

\(^*\)Effect size: \(d = \) change score mean divided by the pooled baseline SD; \(SRM = \) change score divided by the pooled change score SD
of definitions of minimal meaningful change are conservative, and may change with time as more evidence is gathered through further studies [27]. Data on responders can be used to support interpretation of data derived from the HCMSQ, and to report the proportion of people who see a meaningful improvement or worsening in symptoms. It is important to note that the anchors on which meaningful change were defined were both patient-reported; the data set did not identify a clinical marker that could be considered an appropriate anchor. However, patient-reported anchors may be more suitable than clinical measures or those that are assessed by a clinician [28].

This psychometric analysis aimed to evaluate the longitudinal reliability and validity of the HCMSQ domains in the EXPLORER-HCM data set. Importantly, this analysis has addressed the next steps suggested in the publication discussing development of the instrument [16]. Our findings should be interpreted in the context of several potential limitations. First, while the HCMSQ was developed for patients with obstructive HCM and nonobstructive HCM, the EXPLORER-HCM data set included only patients with obstructive HCM. Similarly, the generalizability of data from clinical trials to real-world clinical practice cannot be assumed. The EXPLORER-HCM patient population was relatively small (n = 251) and, although it was demographically similar to a general population of patients with HCM [17], patient experiences of disease, treatment or care, and motivations for symptom improvement cannot be assumed to be the same. The HCMSQ can be considered content valid in a broad population owing to its development from a conceptual model developed in an observational sample, but it will be essential to confirm the psychometric measurement properties in a larger and more heterogeneous population of patients with HCM under routine clinical care. Further, the EXPLORER-HCM trial was designed to test the safety and efficacy of mavacamten, whereas analysis of the psychometric performance of the HCMSQ was a secondary consideration and the trial was not necessarily set up optimally for all analyses. Moreover, for HCMSQ data to be included in the analysis, patients were required to have a daily entry on at least 4 out of the 7 consecutive days prior to the visit. As such, as is often the case with PROs, there were missing data for which imputation was required; however, previous missing-data simulation analyses support this strategy for accounting for data gaps. Lastly, all questionnaires were administered electronically.

### Table 7: Assessment of sensitivity to change: mean (SD) change from baseline to week 30 on HCMSQ domains, by pVO₂ quartile change group

| Domain               | n  | Mean  | SD    | Omnibus test | Pooled SD | Effect size<sup>a</sup> |
|----------------------|----|-------|-------|--------------|-----------|-------------------------|
|                      |    |       |       | F  | p  | SD       | Change score | d  | SRM |
| Cardiovascular symptoms | 3.05 | 0.0502 | 1.69  | 1.40 |       |           |           |    |     |
| Decreased            | 32  | −0.49 | 1.57  |       | 0.29  | 0.35     |
| No change            | 91  | −0.55 | 1.27  |       | 0.33  | 0.39     |
| Increased            | 30  | −1.25 | 1.69  |       | 0.74  | 0.89     |
| Tiredness            | 3.21 | 0.0433 | 0.73  | 0.69 |       |           |           |    |     |
| Decreased            | 32  | −0.38 | 0.78  |       | 0.52  | 0.55     |
| No change            | 91  | −0.33 | 0.69  |       | 0.45  | 0.48     |
| Increased            | 30  | −0.70 | 0.62  |       | 0.95  | 1.01     |
| Shortness of breath  | 8.11 | 0.0005 | 2.89  | 2.73 |       |           |           |    |     |
| Decreased            | 32  | −1.36 | 3.22  |       | 0.47  | 0.50     |
| No change            | 91  | −1.51 | 2.28  |       | 0.52  | 0.55     |
| Increased            | 30  | −3.63 | 2.97  |       | 1.26  | 1.33     |
| Total score          | 6.38 | 0.0022 | 1.74  | 1.58 |       |           |           |    |     |
| Decreased            | 32  | −0.88 | 1.90  |       | 0.51  | 0.56     |
| No change            | 91  | −0.89 | 1.40  |       | 0.51  | 0.56     |
| Increased            | 30  | −2.02 | 1.63  |       | 1.16  | 1.28     |

Bold values indicate significant differences (p ≤ 0.05) between at least smallest and largest mean value for a given domain

<sup>a</sup>Cohen’s d. HCMSQ Hypertrophic Cardiomyopathy Symptom Questionnaire, pVO₂ peak oxygen consumption, SD standard deviation, SRM standardized response mean

<sup>1</sup>Effect size: d = change score mean divided by the pooled baseline SD; SRM = change score divided by the pooled change score SD
Further tests using pen-and-paper-based versions of the instrument could confirm the results. Despite these limitations, the HCMSQ demonstrated adequate measurement properties to support its use as a fit-for-purpose PRO to measure the symptoms of HCM.

Table 8: Assessment of sensitivity to change: mean (SD) change from baseline to week 30 on HCMSQ domains, by LVOT quartile change group

| Domain                  | n  | Mean | SD  | Omnibus test | Pooled SD | Effect size<sup>a</sup> |
|-------------------------|----|------|-----|--------------|-----------|-------------------------|
| Cardiovascular symptoms |    |      |     |              |           |                         |
| Worsened 2 cats         | 24 | −0.81| 1.42| 0.54         | 0.7031    | −0.48 −0.58             |
| Worsened 1 cat          | 20 | −1.03| 2.21| 0.45         | 1.69      | −0.61 −0.74             |
| No change               | 51 | −0.50| 1.19| 0.61         | 0.73      | −0.30 −0.36             |
| Improved 1 cat          | 39 | −0.64| 1.41| 0.61         | 0.69      | −0.38 −0.46             |
| Improved 2+ cats        | 19 | −0.68| 1.21| 0.48         | 0.76      | −0.40 −0.49             |
| Tiredness               |    |      |     |              |           |                         |
| Worsened 2 cats         | 24 | −0.33| 0.65| 0.58         | 0.6805    | −0.46 −0.48             |
| Worsened 1 cat          | 20 | −0.48| 0.99| 0.57         | 0.73      | −0.66 −0.69             |
| No change               | 51 | −0.33| 0.58| 0.54         | 0.71      | −0.44 −0.47             |
| Improved 1 cat          | 39 | −0.53| 0.73| 0.61         | 0.76      | −0.72 −0.76             |
| Improved 2+ cats        | 19 | −0.44| 0.71| 0.48         | 0.76      | −0.61 −0.64             |
| Shortness of breath     |    |      |     |              |           |                         |
| Worsened 2 cats         | 24 | −2.26| 2.59| 0.60         | 0.6641    | −0.78 −0.83             |
| Worsened 1 cat          | 20 | −2.56| 3.60| 0.57         | 0.74      | −0.89 −0.94             |
| No change               | 51 | −1.51| 2.41| 0.57         | 0.74      | −0.52 −0.56             |
| Improved 1 cat          | 39 | −2.02| 3.05| 0.57         | 0.74      | −0.70 −0.74             |
| Improved 2+ cats        | 19 | −1.51| 2.25| 0.57         | 0.74      | −0.52 −0.56             |
| Total score             |    |      |     |              |           |                         |
| Worsened 2 cats         | 24 | −1.17| 1.51| 0.57         | 1.74      | −0.67 −0.74             |
| Worsened 1 cat          | 20 | −1.47| 2.38| 0.57         | 1.74      | −0.84 −0.93             |
| No change               | 51 | −0.87| 1.28| 0.57         | 1.74      | −0.50 −0.55             |
| Improved 1 cat          | 39 | −1.24| 1.72| 0.57         | 1.74      | −0.71 −0.79             |
| Improved 2+ cats        | 19 | −1.05| 1.41| 0.57         | 1.74      | −0.60 −0.66             |

<sup>a</sup>Cat category, d Cohen’s d. HCMSQ Hypertrophic Cardiomyopathy Symptom Questionnaire, LVOT left ventricular outflow tract, SD standard deviation, SRM standardized response mean

<sup>b</sup>Effect size: d = change score mean divided by the pooled baseline SD; SRM = change score divided by the pooled change score SD

Further tests using pen-and-paper-based versions of the instrument could confirm the results. Despite these limitations, the HCMSQ demonstrated adequate measurement properties to support its use as a fit-for-purpose PRO to measure the symptoms of HCM.

Table 9: Meaningful change analysis (responder definition) for the HCMSQ, from baseline to week 30

| HCMSQ domain               | Anchor: PGIC improved at week 30 (n = 136) | Anchor: PGIS improved at week 30 (n = 90) |
|----------------------------|--------------------------------------------|------------------------------------------|
| Cardiovascular symptoms    | 1.00                                       | 1.37                                     |
| Tiredness                  | —a                                        | 0.73                                     |
| Shortness of breath        | 2.57                                       | 3.40                                     |
| Total score                | 1.47                                       | 2.04                                     |

<sup>a</sup>Threshold was not calculated because correlation between change scores and anchor was <0.3

5 Conclusions

The HCMSQ is the only instrument designed to specifically measure the symptoms of HCM. It demonstrates acceptable longitudinal measurement properties, supporting that it is a reliable and valid tool for use in HCM clinical trials to capture change in symptoms associated with treatments. These results support the use of the HCMSQ
as a fit-for-purpose instrument to capture the experience of patients with symptoms of HCM.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41669-022-00340-8.

Acknowledgments The authors wish to thank all patients who participated in the EXPLORER-HCM trial and their families. The authors acknowledge Jeffrey Walter of IQVIA and Johanna Scheinost of Oxford PharmaGenesis for assistance in the preparation of this manuscript.

Declarations

Funding Support of this analysis and manuscript was provided by MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb. Medical writing support was provided by IQVIA and PharmaGenesis Oxford Central, with funding from MyoKardia, Inc.

Conflict of interest MR, PA, and VA 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. JAS has served as a consultant and holds a research grant from MyoKardia, Inc.; he also provides consulting services for Bayer, Janssen, Merck, Novartis, Pfizer, and Terumo; JAS owns the copyright to the Kansas City Cardiomyopathy Questionnaire. CD is a consultant for Bristol Myers Squibb and the following companies: Alexion Pharmaceuticals, Elekta, Genentech, Gilead Sciences, Global Blood Therapeutics, Lyell Immunopharma, Portola Pharmaceuticals, Inc., Puma Biotechnology, and Regenxbio Inc. AJS is an employee of MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb, and holds stocks. JTF is an employee of MyoKardia, Inc., a wholly owned subsidiary of Bristol Myers Squibb, and holds stocks/options.

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

Code availability Not applicable.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Matthew Reaney, Veleka Allen, and Amy Sehnert. All authors contributed to the drafting and revision of the manuscript. All authors approved the final manuscript.

Ethics approval All aspects of the EXPLORER-HCM study and its analysis were conducted in compliance with the Health Insurance Portability Accountability Act of 1996 (USA). Institutional review board and ethics committee approvals were obtained prior to participant enrollment. Study protocols were approved by a US institutional review board: The New England Independent Review Board (197 First Avenue, Suite 250, Needham, MA 02494); appropriate approvals were also obtained for each participating country. The study and its analysis 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 and publish All participants provided informed (written and verbal) consent prior to their participation. All data is deidentified and no individual patient is presented; therefore, consent to publish was not required.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

References

1. Ommen SR, Mital S, Burke MA, Day SM, Deswal A, Elliott P, et al. 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy. Circulation. 2020;142(25):e558–631. https://doi.org/10.1161/CIR.0000000000000937.

2. Ku L, Feiger J, Taylor M, Mestroni L. Cardiology patient page. Familial dilated cardiomyopathy. Circulation. 2003;108(17):e118–21. https://doi.org/10.1161/01.cir.0000097493.70422.50.

3. Gerbs BJ, Maron BJ, Bonow RO, Dearnari JA, Fifer MA, Link MS, et al. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2011;124(24):2761–96. https://doi.org/10.1161/CIR.0b013e318223e230.

4. Sensarian 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.

5. Maron MS, Olivotto I, Betocchi S, Casey SA, Lesser JR, Losi MA, et al. Effect of left ventricular outflow tract obstruction on clinical outcome in hypertrophic cardiomyopathy. N Engl J Med. 2003;348(4):295–303. https://doi.org/10.1056/NEJMoa021323.

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

7. Jain SS, Li SS, Xie J, Sutton MB, Fine JT, Edelberg JM, et al. Clinical and economic burden of obstructive hypertrophic cardiomyopathy in the United States. J Med Econ. 2021;24(1):1115–1123. https://doi.org/10.1080/13696998.2021.1978242.

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

9. Christiaans I, van Langen IM, Birnie E, Bonsel GJ, Wilde AA, Smets EM. Quality of life and psychological distress in hypertrophic cardiomyopathy mutation carriers: a cross-sectional cohort study. Am J Med Genet A. 2009;149a(4):602–12. https://doi.org/10.1002/ajmg.a.32710.

10. Hershberger RE, Morales A, Siegfried JD. Clinical and genetic issues in dilated cardiomyopathy: a review for genetics professionals. Genet Med. 2010;12(11):655–67. https://doi.org/10.1097/GIM.0b013e3181f2481f.

11. Cox S, O’Donoghue AC, McKenna WJ, Steptoe A. Health related quality of life and psychological wellbeing in patients with hypertrophic cardiomyopathy. Heart. 1997;78(2):182–7. https://doi.org/10.1136/hrt.78.2.182.
12. Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, Charron P, 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.

13. FDA. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. 2009. https://www.fda.gov/media/77832/download. Accessed 16 June 2021.

14. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Educ. 2017;17(4):137–44. https://doi.org/10.1093/bjaed/mkw060.

15. Cruz Rivera S, McMullan C, Jones L, Kyte D, Slade A, Calvert M. The impact of patient-reported outcome data from clinical trials: perspectives from international stakeholders. J Patient-Rep Outcomes. 2020;4(1):51. https://doi.org/10.1186/s41687-020-00219-4.

16. Reaney M, Allen V, Sehnert AJ, Fang L, Hagege AA, Naidu SS, et al. Development of the hypertrophic cardiomyopathy symptom questionnaire (HCMSQ): a new patient-reported outcome (PRO) instrument. Pharmacoeconomics - Open. https://doi.org/10.1007/s41669-020-00335-5.

17. Olivotto I, Oreziak A, Barriales-Villa R, Abraham TP, Masri A, Garcia-Pavia P, 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.

18. Spertus JA, Fine JT, Elliott P, Ho CY, Olivotto I, Saberi S, et al. Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): health status analysis of a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet. 2021. https://doi.org/10.1016/S0140-6736(21)00763-7.

19. Saberi S, Cardim N, Yamani M, Schulz-Menger J, Li W, Florea V, et al. Mavacamten favorably impacts cardiac structure in obstructive hypertrophic cardiomyopathy. Circulation. 2021;143(6):606–8. https://doi.org/10.1161/CIRCULATONNAHA.120.052359.

20. Ho CY, Olivotto I, Jacoby D, Lester SJ, Roe M, Wang A, et al. Study design and rationale of EXPLORER-HCM. Circ Heart Fail. 2020;13(6):e006853. https://doi.org/10.1161/CIRCHEARTFAL-URE.120.006853.

21. 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.

22. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36. https://doi.org/10.1007/s11136-011-9903-x.

23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

24. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

25. Cohen J. Statistical power analysis for behavioral sciences. 2nd edn. Hillsdale; 1988.

26. FDA. Discussion document for patient-focused drug development public workshop on guidance 3—methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments. 2018. https://www.fda.gov/media/116277/download. Accessed 21 June 2021.

27. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. Health Qual Life Outcomes. 2006;4(1):70. https://doi.org/10.1186/1477-7525-4-70.

28. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ. 2020;369: m1714. https://doi.org/10.1136/bmj.m1714.