Defining changes in physical limitation from the patient perspective: insights from the VITALITY-HFpEF randomized trial

Javed Butler[1]*, John A. Spertus[2], Luke Bamber[3], Muhammad Shahzeb Khan[4], Lothar Roessig[3], Vanja Vlajnic[5], Josephine M. Norquist[6], Kevin J. Anstrom[7], Robert O. Blaustein[6], Carolyn S.P. Lam[8], and Paul W. Armstrong[9], VITALITY-HFpEF Study Group

[1]Department of Medicine, University of Mississippi, Jackson, MS, USA; [2]Saint Luke's Mid America Heart Institute, University of Missouri, Kansas City, MO, USA; [3]Bayer AG, Wuppertal, Germany; [4]Division of Cardiology, Duke University Medical Center, Durham, NC, USA; [5]Bayer U.S., LLC, Whippany, NJ, USA; [6]Cardiovascular Clinical Research, Merck & Co, Inc, Kenilworth, NJ, USA; [7]Duke Clinical Research Institute, Durham, NC, USA; [8]National Heart Centre Singapore, Duke-National University of Singapore, Singapore; and [9]Division of Cardiology, Canadian VIGOUR Centre, University of Alberta, Edmonton, Canada

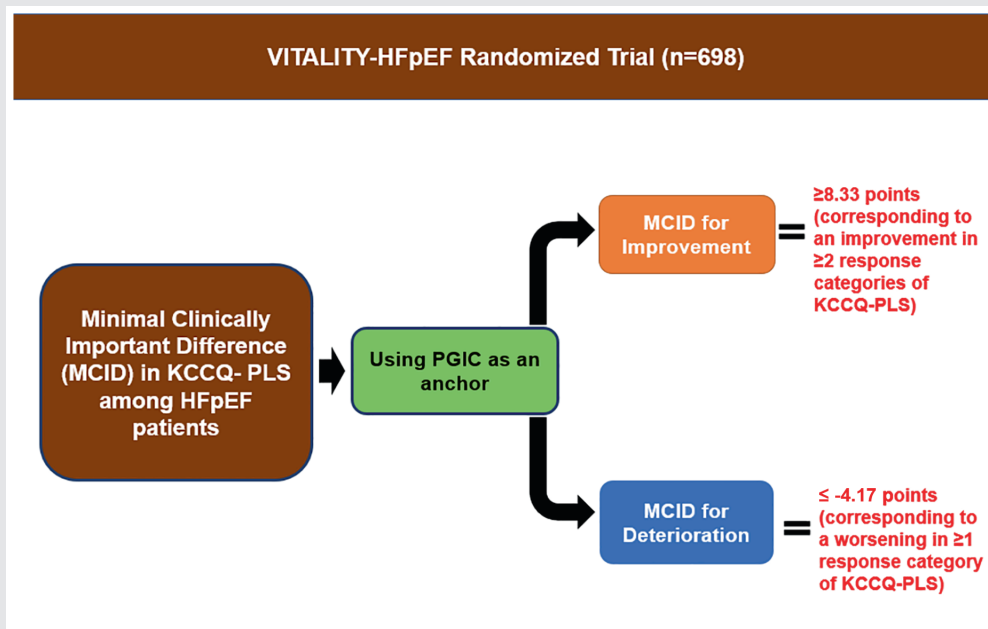|	
Aims	Clinically important thresholds in patient-reported outcomes measures like the Kansas City Cardiomyopathy Questionnaire (KCCQ) have not been defined for patients with heart failure and preserved ejection fraction (HFpEF). The aim of this study was to estimate meaningful thresholds for improvement or worsening in the KCCQ physical limitation score (PLS) in patients with HFpEF.
Methods and results	In this pre-specified analysis from VITALITY-HFpEF, anchor- and distribution-based approaches were used to estimate thresholds for improvement or worsening in the KCCQ-PLS using Patient Global Impression of Change (PGIC) as an anchor. The KCCQ-PLS contains six elements, with each increment in response resulting in a change of 4.17 points when converted to a 0–100 scale. The mean change in KCCQ-PLS from baseline to week 12 was calculated for each PGIC group to estimate a meaningful within-patient change. Of 789 patients enrolled, 698 had complete KCCQ-PLS and PGIC data at week 12. The mean ($\pm$ standard deviation) changes in KCCQ-PLS corresponding to PGIC changes of 'a little better', 'better', and 'much better' were 5.7 ± 18.6, 11.6 ± 19.3, and 18.4 ± 25.3 points, respectively. The scores of patients who responded 'a little better' ($n = 177$) overlapped substantially with those who reported 'no change' ($n = 193$; mean change 2.8 ± 18.9). The mean change in KCCQ-PLS for patients responding 'a little worse' ($n = 32$) was -2.6 ± 18.0 points. The threshold for meaningful within-patient change in KCCQ-PLS based on distribution-based analyses was 12.3 points. Using area under the curve (AUC) analyses of KCCQ-PLS, the sensitivity and specificity of a 4.17-point change were 0.61 and 0.57, for an 8.33-point change they were 0.49 and 0.64, and for a 12.5-point change they were 0.44 and 0.72 for being at least a little better on the PGIC (AUC = 0.54).
Conclusion	In the VITALITY-HFpEF trial, a change in KCCQ-PLS of ≥ 8.33 points (corresponding to an improvement in ≥ 2 response categories of KCCQ-PLS) may represent the minimal clinically important difference for improvement and a change of ≤ -4.17 points (corresponding to a worsening in ≥ 1 response category of KCCQ-PLS) may suggest deterioration in patients with HFpEF.

*Corresponding author. Department of Medicine, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216, USA. Tel: +1 601 984-5600, Fax: +1 601 984-5608, Email: jbutler4@umc.edu

Graphical Abstract

Minimal clinically important difference for improvement and deterioration of Kansas City Cardiomyopathy Questionnaire physical limitation score (KCCQ-PLS) in patients with heart failure with preserved ejection fraction (HFpEF). PGIC, Patient Global Impression of Change.

Keywords   Patient-reported outcomes ● Heart failure with preserved ejection fraction

Introduction

Patients with heart failure (HF) and preserved ejection fraction (HFpEF) face significant impairment in health-related quality of life and physical functioning.[1,2] While researchers continue to seek novel therapies to improve survival among these patients, the therapeutic focus in HFpEF has appropriately broadened to consider the impact of patient-reported outcomes. The Kansas City Cardiomyopathy Questionnaire (KCCQ) is a validated questionnaire that is widely used to assess various health status domains in patients with HF.[3] The KCCQ physical limitation score (KCCQ-PLS) measures patient-reported impairments in physical limitations due to HF. The KCCQ-PLS is one of the domains the Center for Drug Evaluation and Research, part of the US Food and Drug Administration (FDA), has deemed appropriate as a clinical outcome assessment measure for HF drug development.[4] The KCCQ-PLS has been validated against the 6-min walk distance and the New York Heart Association functional class.[3,5,6] Despite its validity, reliability, and prognostic significance, interpreting what magnitude of KCCQ-PLS change is meaningful to patients is ill-defined, particularly in patients with HFpEF.[7] Clarifying the clinical significance of changes in KCCQ-PLS is desirable to enhance interpretation of clinical trials that seek to quantify the impact of novel treatments on patients' physical function impairment.

While prior studies have focused on determining meaningful between-group changes (usually a novel therapeutic and placebo),[6–8] it is equally important to estimate the within-patient difference in scores given the potential for intrinsic variability and what is perceived as meaningful to patients. Hence establishing the definition of responder thresholds that can be used to estimate rates of improvement and determine the number needed to treat in a clinically interpretable manner would be valuable for clinicians and patients alike.[9] The VITALITY-HFpEF (eValuate the effIcacy and safeTy of the orAL sGC stImulator vericiguaT to improve phYsical functioning in activities of daily living in patients with Heart Failure and preserved Ejection Fraction) trial was designed to test whether the novel oral soluble guanylate cyclase (sGC) stimulator vericiguat improved physical functioning in patients with HFpEF. An integral part of the analysis plan was to determine clinically important thresholds of the KCCQ-PLS in patients with HFpEF.

Methods

Study design

VITALITY-HFpEF (NCT03547583) was a randomized parallel-group, placebo-controlled, double-blind, multicentre phase 2b trial to assess the efficacy of the oral sGC stimulator vericiguat in improving physical

functioning limitation as measured by the KCCQ-PLS scale.[10] The detailed eligibility criteria and study design have been published previously.[11] Briefly, patients were enrolled if they presented with chronic HFpEF and a left ventricular ejection fraction (LVEF) ≥45% with New York Heart Association class II–III symptoms within 6 months of a recent decompensation (HF hospitalization or intravenous diuretics for HF without hospitalization), and had elevated natriuretic peptides. There was no run-in phase. Patients were enrolled independent of their baseline KCCQ-PLS. Exclusion criteria included previous LVEF <40%, estimated glomerular filtration rate (eGFR) <30 mL/min/1.73 m^2, symptomatic hypotension, or a resting heart rate <50 or ≥100 bpm. Patients who met the eligibility criteria were randomized 1:1:1 to daily vericiguat 15 mg or 10 mg or placebo. The study protocol was approved by local ethics committees at participating sites; all patients provided written informed consent. The main results of the trial have been published previously; 24-week treatment with vericiguat at either 15 mg or 10 mg daily compared with placebo did not improve the PLS of the KCCQ.[10]

Health-related quality of life assessment

The KCCQ is a 23-item, disease-specific measure intended for the assessment of HF patients' perspectives of how their disease impacts their health status. The KCCQ has been shown to be valid, reliable, sensitive to clinical changes, and associated with death, hospitalization, and healthcare costs.[12–19] The KCCQ measures four clinical domains: symptoms frequency and severity, physical limitation, quality of life, and social limitation. As physical functioning was considered an important measure of treatment benefit with interventions in HFpEF, the change from baseline to week 24 in the KCCQ-PLS was chosen as the primary endpoint in the VITALITY HFpEF trial.[11] The KCCQ-PLS measures the limitations imposed by HF that a patient experiences in performing activities of daily life. There are six items in the physical limitation domain of the KCCQ, and the response options for each item range from 1 (extremely limited) to 5 (not at all limited) with an option to also indicate limitations for other reasons or not completing the activity. The algorithm used to calculate the PLS applies equal weighting to each of the items; the raw PLS, which ranges from 6−30 is then transformed to a 0−100 scale, with higher scores reflecting less physical function limitation.[3] Given this conversion, a 1-response category change on a single item in the PLS results in a change of 4.17 points on the 0−100 scale (100 total points/24).

Patient global impression of change and severity

To best determine the clinical meaningfulness of a change in KCCQ-PLS, the change needs to be correlated with an assessment of the degree to which a patient's health status has changed. A common approach is to administer questions that ask patients how they believe that their health has changed as compared with a prior time period. Accordingly, Patient Global Impression of Change (PGIC) questions were administered to assess the degree of change in physical limitation due to HF when compared with their limitation at the start of the treatment. Response options included 'much better', 'better', 'a little better', 'the same', 'a little worse', 'worse', or 'much worse'. A subsequent question then asked the patient to indicate whether they felt the degree of change reported was important or not, with response options of yes or no. The PGIC was administered at weeks 2, 6, 12, 18, and 24 post-randomization. In addition to PGIC,

a Patient Global Impression of Severity (PGIS) questionnaire was administered to assess the current severity of physical limitations due to HF, with response options of 'no limitations', 'mild', 'moderate', 'severe', or 'very severe'. The PGIS was administered at baseline and at all post-randomization visits. The KCCQ, PGIC, and PGIS were completed electronically, without knowledge of treatment assignment, before all other assessments during the clinic visits.

Statistical analyses

Meaningful within-patient changes for KCCQ-PLS were estimated using an anchor-based approach.[20,21] Although anchor-based analyses of week 12 data were considered primary, week 24 data were also examined to provide support to week 12 data and to assess the stability of the estimates. KCCQ-PLS mean change scores from baseline to week 12 were calculated for each PGIC response category to estimate the average change of participants reporting different magnitudes of change. Additional anchor-based analyses were conducted using the PGIS as an additional anchor. The degree of change from baseline to follow-up on the PGIS was used as an anchor (e.g. improved by 1 response category, declined by 1 response category). Although the primary endpoint in the VITALITY-HFpEF trial was the change in KCCQ-PLS from baseline to week 24, the primary time point chosen for the current analysis was pre-specified to be 12 weeks. This strategy allowed for the use of blinded interim data from the VITALITY-HFpEF trial prior to database lock to estimate the thresholds for improvement and worsening on the primary endpoint, change in KCCQ-PLS.

Distribution-based approaches were also performed to supplement the anchor-based approaches. Graphical displays via empirical cumulative distribution functions (CDFs) and probability density functions (PDFs) were used to illustrate the percent of patients who experienced different levels of change on the KCCQ-PLS across PGIC categories. The distribution of change scores via empirical CDF and PDF plots provides supporting evidence for the identified responder threshold.[22] It has been suggested that 0.2, 0.5, and 0.8 times the standard deviation (SD) of baseline scores represents small, moderate, and large changes.[23] For these analyses, 0.5 SD and 1 standard error of measurement (SEM) of the KCCQ-PLS at baseline were selected to represent a moderately large change.[23] The SEM was calculated as the SD at baseline multiplied by the square root of 1 minus the reliability of the KCCQ-PLS. The reliability coefficient was estimated by calculating the correlation (intra-class correlation coefficient [ICC]) between the baseline and week 6 KCCQ-PLS scores among stable patients.[24] Stable patients were defined as patients who responded 'the same' to the PGIC at week 6.

All analyses were pre-specified and blinded, with all treatments combined. Correlations between KCCQ-PLS and anchors were calculated using Spearman's rank correlation coefficient to assess the level of confidence in the interpretation of results. Correlations >0.3 are generally considered acceptable.[25] Because each 'shift' in response to the KCCQ-PLS results in a change of 4.17 points, the definitions of meaningful thresholds for KCCQ-PLS were made in reference to this scoring interval, although the exact values can reside anywhere between 0−100.

Receiver operating characteristic (ROC) curve analyses were used to assess the ability of change in KCCQ-PLS to discriminate between patients responding 'a little better', 'better', or 'much better' on the PGIC questionnaire at week 12 and those with other responses. The point(s) on the curve that optimize sensitivity and/or specificity were considered the optimal threshold in differentiating between the two

groups. Area under the curve (AUC) estimates of approximately 0.5 indicate no ability to discriminate between categories.

No imputation was used for missing KCCQ scores and only patients with complete data were analysed. Analyses were performed using SAS version 9.4 (SAS Institute, Inc., Cary, NC, USA) and a *p*-value <0.05 was considered statistically significant.

Results

Patient characteristics

Between 15 June 2018 and 23 April 2019, 789 patients were randomized at 168 sites in 21 countries. Of these, 761 (96%) patients had non-missing data for the KCCQ-PLS at baseline. Of these 761 patients, 695 (91%) at week 12 and 643 (84%) at week 24 had non-missing data for the KCCQ-PLS, PGIC, and PGIS measures. The mean age of the cohort was 72.7 ± 9.4 years and 48.8% were female. Online *Table S1* shows the baseline characteristics of the patients enrolled.

KCCQ-PLS scores and PGIC

The mean (SD) baseline KCCQ-PLS score was 58.8 ± 24.5 corresponding to a moderate limitation in physical activities. There was substantial variation among patients at baseline (range 0–100). On average, study participants improved over time with KCCQ-PLS values increasing from baseline at all visits (change from baseline to week 6: 7.3 ± 18.4; week 12: 6.9 ± 20.8; week 18: 7.8 ± 20.9; week 24: 7.6 ± 20.3). Individual change scores ranged from −100 to +87.5.

Most participants reported improving over the first 12 weeks, as assessed by the PGIC. Overall, the evolution in changes was distributed as follows: 1.6% (*n* = 11), 28% (*n* = 193), 30% (*n* = 211), 21% (*n* = 144), and 13% (*n* = 91) reported being worse, the same, a little better, better, and much better, respectively. Similarly, at week 24 changes in PGIC revealed that 3% (*n* = 18), 26% (*n* = 163), 25% (*n* = 163), 23% (*n* = 145), and 18% (*n* = 116) of patients reported being worse, the same, a little better, better, and much better, respectively. The Spearman correlations between the PGIC

and the change in KCCQ-PLS scores were 0.28 at week 12 and 0.31 at week 24.

Meaningful within-patient change in KCCQ-PLS using PGIC as anchor

At week 12, the mean (SD) change in KCCQ-PLS was 5.7 ± 18.6 points for patients responding 'a little better', 11.6 ± 19.3 points for those responding 'better', and 18.4 ± 25.3 points for those responding 'much better' (*Table 1*). At week 24, the corresponding mean (SD) changes in KCCQ-PLS were 8.2 ± 17.7 points, 12.4 ± 20.1, and 16.1 ± 21.7 points. The mean change in KCCQ-PLS for the patient responding 'a little worse' was −2.6 ± 18.0 at week 12 and −1.4 ± 22.2 at week 24 (*Table 2*). The mean change in KCCQ-PLS for patients who reported any level of improvement on the PGIC (much better, better, or little better) was 10.5 ± 21.0 and 12.0 ± 20.0 at weeks 12 and 24, respectively. The mean KCCQ-PLS change for patients reporting any level of worsening (much worse, worse, or a little worse) was −5.4 ± 18.7 points at week 12 and −9.2 ± 23.5 points at week 24.

Empirical cumulative distribution function, probability density function, and area under the curve

The CDFs showed a clear distinction between the improvement and worsening groups overall (*Figure 1*). Within the improvement subcategories, the 'a little better' (*n* = 211) curve overlapped substantially with the 'no change' (*n* = 193) category. The 'better' (*n* = 144) and 'much better' (*n* = 91) curves were more clearly differentiated from the 'no change' group. The median values on the 50th percentile line for the 'better' and 'much better' groups were ~10–15 points on the KCCQ-PLS (*Figure 2*). The worsening categories on the PGIC were differentiated from the 'no change' group and from each other. The median values for the 'a little worse' and 'worse' groups were −5 and −15 points, respectively. *Figure 2* shows ~35% of patients in the 'a little worse' category and

Table 1 Summary of Kansas City Cardiomyopathy Questionnaire physical limitation score meaningful improvement thresholds by Patient Global Impression of Change as anchor

PGIC anchor	Meaningful change estimate					
	Week 12			Week 24		
	n	Mean KCCQ-PLS change (SD)	Median KCCQ-PLS change (25th, 75th)	*n*	Mean KCCQ-PLS change (SD)	Median KCCQ-PLS change (25th, 75th)
The same	193	2.77 (18.88)	4.17 (−4.17, 12.50)	163	3.43 (16.14)	4.17 (−8.33,14.17)
PGIC: A little better and 'important'	177	5.73 (18.60)	4.17 (−4.17, 16.67)	140	8.21 (17.73)	8.33 (−0.83, 17.08)
PGIC: Better and 'important'	136	11.56 (19.30)	9.58 (0.00, 25.00)	142	12.40 (20.11)	12.50 (0.00, 25.00)
PGIC: Much better and 'important'	88	18.44 (25.29)	16.67 (0.63, 32.08)	115	16.14 (21.68)	12.50 (4.17, 29.17)

KCCQ-PLS, Kansas City Cardiomyopathy Questionnaire physical limitation score; PGIC, Patient Global Impression of Change; SD, standard deviation.
The Spearman correlation between PGIC and KCCQ-PLS was 0.28 and 0.31 at 12 and 24 weeks, respectively.

Table 2 Summary of Kansas City Cardiomyopathy Questionnaire physical limitation score meaningful deterioration thresholds by Patient Global Impression of Change as anchor

PGIC anchor	Meaningful change estimate					
	Week 12			Week 24		
	n	Mean KCCQ-PLS change (SD)	Median KCCQ-PLS change (25th, 75th)	n	Mean KCCQ-PLS change (SD)	Median KCCQ-PLS change (25th, 75th)
The same	193	2.77 (18.88)	4.17 (−4.17, 12.50)	163	3.43 (16.14)	4.17 (−8.33, 14.17)
PGIC: A little worse and 'important'	32	−2.59 (18.04)	−2.08 (−12.50, 8.33)	32	−1.35 (22.20)	0.00 (−10.42, 12.50)
PGIC: Much worse, worse, or a little worse and 'important'	47	−5.36 (18.65)	−4.17 (−16.67, 6.25)	52	−9.16 (23.50)	−4.17 (−21.67, 6.25)

KCCQ-PLS, Kansas City Cardiomyopathy Questionnaire physical limitation score; PGIC, Patient Global Impression of Change; SD, standard deviation.

20% in the 'worse' category reported an improvement on their KCCQ-PLS at 12 weeks.

The AUC analyses demonstrated minimal ability for the KCCQ-PLS to discriminate between patients' PGIC ratings for any improvement versus no change or worsening by 0.54 at week 12 and 0.56 at week 24. Sensitivity and specificity >0.70 were not achieved in any scenario. Although it would have been ideal to balance high specificity with high-sensitivity estimates, an imbalance in favour of high specificity was considered preferable in this specific context. This would reduce the likelihood that failure to experience a meaningful change is categorized as having experienced meaningful change (i.e. a false positive result). Within this framework, increases in the KCCQ-PLS of 8.3 to 12.5 provided a sensitivity >0.40 and a specificity >0.70. The sensitivity and specificity of a 4.17-point change were 0.61 and 0.57; for an 8.33-point change they were 0.49 and 0.64, and for a 12.5-point change were 0.44 and 0.72 for being at least a little better on the PGIC (AUC = 0.54). The threshold for meaningful within-patient change in KCCQ-PLS from distribution-based methods corresponded to 12.2 for 0.5 SD at baseline, and 10.8 from the 1 SEM based on ICC = 0.80, which indicates an acceptable level of test–retest reliability for the KCCQ-PLS.

Meaningful within-patient change in KCCQ-PLS using PGIS as anchor

The Spearman correlations between the PGIS and change in KCCQ-PLS were 0.31 at week 12 and 0.41 at week 24. The thresholds of KCCQ-PLS scores associated with deterioration or improvement by 1 or 2 response categories from baseline are shown in online supplementary *Table S2*.

Discussion

As patient-reported outcomes are increasingly used in clinical trials, it is important to understand how best to determine the amount of change in a patient-reported outcome score that is considered meaningful to healthcare providers and patients. In this pre-specified and blinded pooled analysis of VITALITY-HFpEF data, several key findings are evident. First, anchor-based analyses suggest that 8.33 points, on a 0−100 scale, represents a meaningful within-patient improvement threshold for KCCQ-PLS in patients with HFpEF, with 12.5 points representing a clearly important change. This was also supported by the results from distribution-based methods, where the threshold for meaningful within-patient change in KCCQ-PLS corresponded to 12.2 for 0.5 SD at baseline. Second, the anchor-based analyses suggest that any degree of worsening from baseline on the KCCQ-PLS is considered important to patients. Since the construct of the KCCQ requires that each 'shift' in the KCCQ-PLS corresponds to a change of 4.17 points, for an individual patient this implies that a 4.17 change represents the threshold for meaningful worsening in KCCQ-PLS (*Graphical Abstract*). These data have important clinical implications as these thresholds for meaningful within-patient improvement and worsening on the KCCQ-PLS may aid in assessment of treatment efficacy for therapy in HFpEF and the design of future trials.

Defining clinically important differences in patient-reported outcomes is challenging. The critical issue is determining whether or not observed within-patient changes from baseline in a patient-reported outcome endpoint score are meaningful. Whereas the KCCQ asks patients to describe their health status at one point in time, the PGIC questions prompt patients to reflect on how their health has changed over time. In this study, the correlation between the PGIC and KCCQ-PLS of 0.28 is slightly below the recommended threshold of 0.30, indicating that the PGIC may not be a reliable anchor upon which to quantify a change. The reasons for this disparity are unclear but may reflect patients' expectation for improvement after volunteering for a clinical trial thereby potentially biasing the accuracy of their recollection of physical impairment. Nevertheless, this anchor-based approach represents the best strategy currently available and one that has been endorsed by the FDA.[26] The thresholds we have defined correspond to a modest but clinically important change of 8.3 points and require (on average) an improvement of two response categories in KCCQ-PLS. This provides strong face validity when reflecting on the actual KCCQ-PLS composition and is well aligned with prior estimates of a 5-point improvement in KCCQ-PLS score being
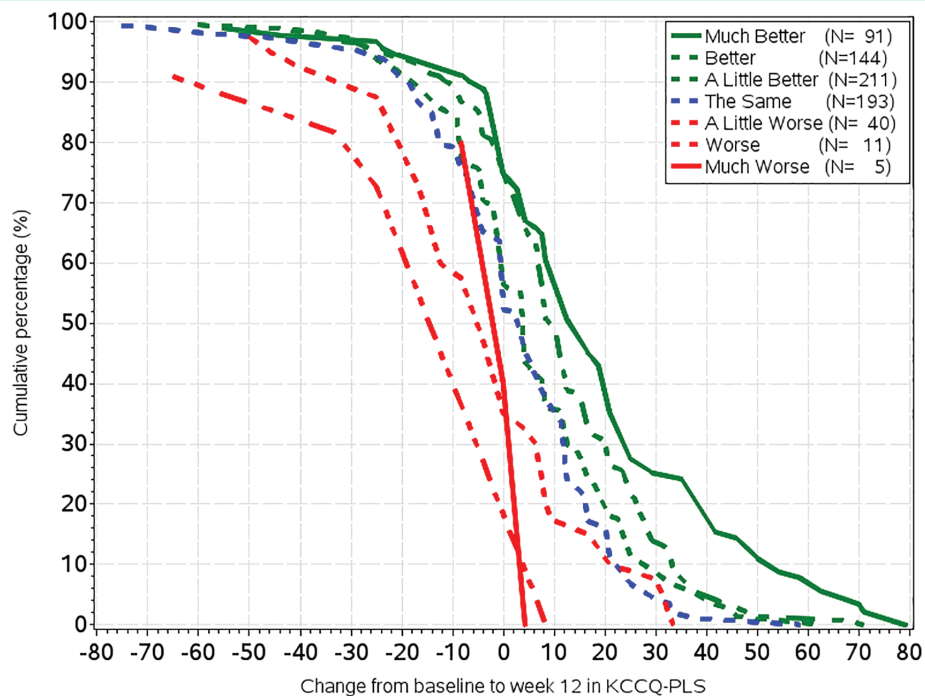
Figure 1 Empirical cumulative distribution functions of the change from baseline to week 12 in the Kansas City Cardiomyopathy Questionnaire physical limitation score (KCCQ-PLS) for each anchor category of Patient Global Impression of Change.

clinically important in HF with reduced ejection fraction.[6–8] As PLS is scaled to 100, a shift of 8.33 on a 0–100 scale corresponds to a change of ~5 on the KCCQ raw score. The 5-point threshold has been associated with a 6% (3%–9%) reduction in cardiovascular death and hospitalization in the TOPCAT trial.[13] Larger changes in patients' perceived improvement were associated with proportionally larger changes in the KCCQ-PLS that were also consistent with prior reports of 10 and 20-point changes representing moderate-to-large and large-to-very large clinical changes.[13]

Beyond extending the prior work in HF with reduced ejection fraction populations, the novel data from the current study provide a useful foundation for interpreting future studies that use KCCQ as an endpoint. While clinical trials have traditionally reported the mean differences between treatment groups, these are difficult to interpret as they average the experiences of patients within the population. An advantage of determining clinically meaningful within-patient thresholds as described herein is the enabling of comparisons between patient groups of the proportions who reported feeling better, were unchanged, or got worse. This responder analysis methodology facilitates estimating the number needed to treat to achieve a benefit of a certain magnitude. One recently reported trial used this approach to report a significant 3.7-point better 12-week KCCQ overall summary score in patients treated with dapagliflozin, and a number needed to treat of ~10 for experiencing a very large improvement in overall health status.[27] It is noteworthy that the magnitude of meaningful within-patient estimates for improvement and deterioration differed in our study. This asymmetry has been observed in other studies as well and suggests that a relatively smaller change in KCCQ-PLS can indicate patient worsening.[28,29]

There are some limitations to this study. Developing a measure, independent of the KCCQ, to define clinical change is subject to misclassification. The correlations between KCCQ-PLS change and the PGIC were lower than anticipated and below a threshold for which these analyses are usually performed. Moreover, a sensitivity and specificity of >0.7 together was not achieved to discriminate between patients' PGIC ratings for any improvement versus no change or worsening. This may be due, in part, to a restricted range of PGIC values, with relatively few patients reporting worsening, or it could be that a 3-month recall period in the context of a HFpEF clinical trial was too long for patients to accurately remember how they were feeling at the start of the trial. Nevertheless, our results were similar with distribution-based analyses. Quality of life in patients with HF may deteriorate over time and thus meaningful within-patient changes may be dependent, at least in part, on duration of follow-up. However, assessing these changes beyond a 3-month duration may be difficult due to recall bias. Furthermore, the generalizability of these results is limited to patients with HFpEF with recent decompensation. Because few patients reported worsening on the PGIC, we were limited in our ability to define clinically important thresholds for deterioration. Overall, patients largely improved in all treatment groups, probably related in part to the placebo effect and intensification of background therapy. This improvement could have affected the estimate for threshold of deterioration. Lastly, while patient-reported outcomes are meaningful as endpoints in themselves, their association
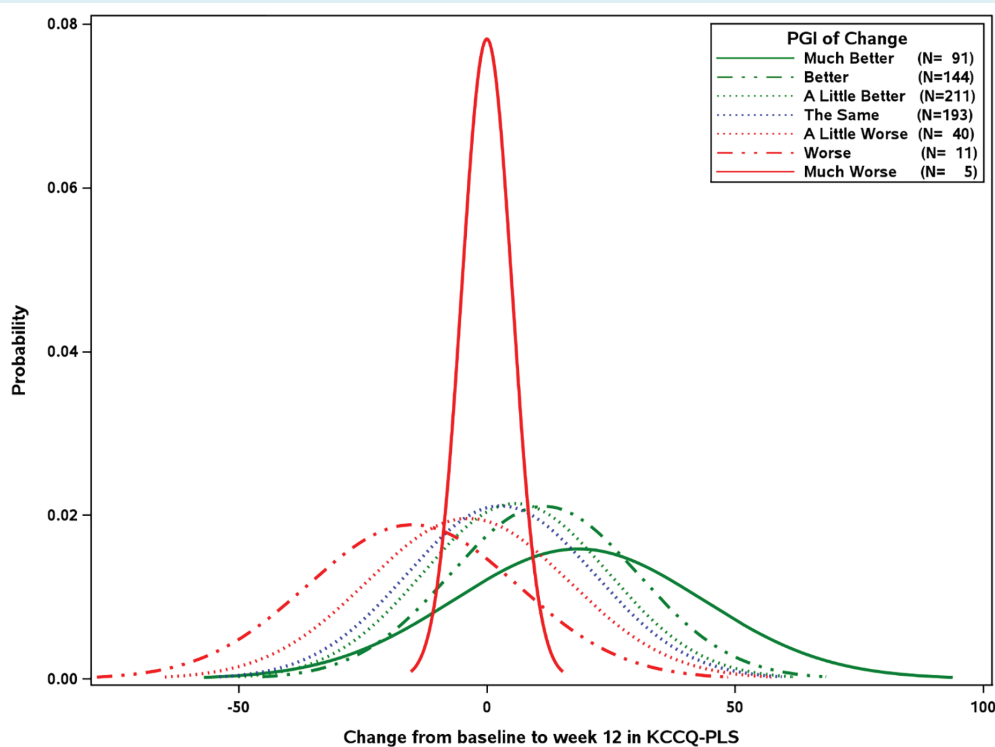
Figure 2 Empirical probability distribution functions of the change from baseline to week 12 in the Kansas City Cardiomyopathy Questionnaire physical limitation score (KCCQ-PLS) for each anchor category of Patient Global Impression (PGI) of Change.

with clinical outcomes is also important. Future studies should assess if the proposed thresholds of the KCCQ-PLS are associated with subsequent morbidity and mortality.

In conclusion, we found that an improvement of 8.3 points on a 0−100 scale appears clinically important to patients, as was any level of deterioration in KCCQ-PLS. Using these thresholds for meaningful within-patient improvement, future trials can perform responder analyses to better communicate the number needed to treat to improve patients' health status, thus improving the interpretability of mean differences in scores observed in HFpEF clinical trials.

Supplementary Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Funding

References

1. Dunlay SM, Roger VL, Redfield MM. Epidemiology of heart failure with preserved ejection fraction. *Nat Rev Cardiol*. 2017;**14**:591−602.
2. Shah SJ, Heitner JF, Sweitzer NK, Anand IS, Kim HY, Harty B, et al. Baseline characteristics of patients in the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist trial. *Circ Heart Fail*. 2013;**6**:184−92.
3. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol*. 2000;**35**:1245−55.
4. Clinical Outcome Assessments (COA) Qualification Submissions Office of Cardiology Hematology, Endocrinology, and Nephrology (OCEHM) Division of Cardiovascular and Nephrology (DCN). DDT COA #000084: Kansas City Cardiomyopathy Questionnaire (KCCQ). April 16, 2020. https://www.fda.gov/drugs/clinical-outcome-assessment-coa-qualification-program/ddt-coa-000084-kansas-city-cardiomyopathy-questionnaire-kccq. (14 March 2022).
5. Flynn KE, Piña IL, Whellan DJ, Lin L, Blumenthal JA, Ellis SJ, et al.; HF-ACTION Investigators. Effects of exercise training on health status in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA*. 2009;**301**:1451−9.
6. Spertus J, Peterson E, Conrad MW, Heidenreich PA, Krumholz HM, Jones P, et al. Monitoring clinical changes in patients with heart failure: a comparison of methods. *Am Heart J*. 2005;**150**:707−15.
7. Dreyer RP, Jones PG, Kutty S, Spertus JA. Quantifying clinical change: discrepancies between patients' and providers' perspectives. *Qual Life Res*. 2016;**25**:2213−20.
8. Butler J, Khan MS, Mori C, Filippatos GS, Ponikowski P, Comin-Colet J, et al. Minimal clinically important difference in quality of life scores for patients with heart failure and reduced ejection fraction. *Eur J Heart Fail*. 2020;**22**:999−1005.
9. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ*. 1995;**310**:452−4.
10. Armstrong PW, Lam CSP, Anstrom KJ, Ezekowitz J, Hernandez AF, O'Connor CM, et al.; VITALITY-HFpEF Study Group. Effect of vericiguat vs placebo on quality of life in patients with heart failure and preserved ejection fraction: the VITALITY-HFpEF randomized clinical trial. *JAMA*. 2020;**324**:1512−21.
11. Butler J, Lam CSP, Anstrom KJ, Ezekowitz J, Hernandez AF, O'Connor CM, et al. Rationale and design of the VITALITY-HFpEF trial. *Circ Heart Fail*. 2019;**12**:e005998.
12. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. *Circ Cardiovasc Qual Outcomes*. 2015;**8**:469−76.
13. Pokharel Y, Khariton Y, Tang Y, Nassif ME, Chan PS, Arnold SV, et al. Association of Serial Kansas City Cardiomyopathy Questionnaire assessments with death and hospitalization in patients with heart failure with preserved and reduced ejection fraction: a secondary analysis of 2 randomized clinical trials. *JAMA Cardiol*. 2017;**2**:1315−21.
14. Joseph SM, Novak E, Arnold SV, Jones PG, Khattak H, Platts AE, et al. Comparable performance of the Kansas City Cardiomyopathy Questionnaire in patients with heart failure with preserved and reduced ejection fraction. *Circ Heart Fail*. 2013;**6**:1139−46.
15. Arnold SV, Spertus JA, Lei Y, Allen KB, Chhatriwalla AK, Leon MB, et al. Use of the Kansas City Cardiomyopathy Questionnaire for monitoring health status in patients with aortic stenosis. *Circ Heart Fail*. 2013;**6**:61−7.
16. Dunlay SM, Gheorghiade M, Reid KJ, Allen LA, Chan PS, Hauptman PJ, et al. Critical elements of clinical follow-up after hospital discharge for heart failure: insights from the EVEREST trial. *Eur J Heart Fail*. 2010;**12**:367−74.
17. Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS, et al. Health status identifies heart failure outpatients at risk for hospitalization or death. *J Am Coll Cardiol*. 2006;**47**:752−6.
18. Arnold SV, Spertus JA, Vemulapalli S, Dai D, O'Brien SM, Baron SJ, et al. Association of patient-reported health status with long-term mortality after transcatheter aortic valve replacement: report from the STS/ACC TVT registry. *Circ Cardiovasc Interv*. 2015;**8**:e002875.
19. Chan PS, Soto G, Jones PG, Nallamothu BK, Zhang Z, Weintraub WS, et al. Patient health status and costs in heart failure: insights from the Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study (EPHESUS). *Circulation*. 2009;**119**:398−407.
20. Lydick F, Epstein RS. Interpretation of quality of life changes. *Qual Life Res*. 1993;**2**:221−6.
21. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;**10**:407−15.
22. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Qual Life Res*. 2018;**27**:33−40.
23. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;**41**:582−92.
24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;**86**:420−8.
25. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;**61**:102−9.
26. Patient-Focused Drug Development Guidance Public Workshop. Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making. US FDA Patient-Focused Drug Development Guidance Public Workshop December 6, 2019. https://www.fda.gov/media/132505/download (14 March 2022).
27. Nassif ME, Windsor SL, Tang F, Khariton Y, Husain M, Inzucchi SE, et al. Dapagliflozin effects on biomarkers, symptoms, and functional status in patients with heart failure with reduced ejection fraction: the DEFINE-HF trial. *Circulation*. 2019;**140**:1463−16.
28. Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA. Interpreting clinically significant changes in patient-reported outcomes. *Cancer*. 2007;**110**:196−202.
29. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res*. 2002;**11**:207−21.