



# Investigating gender-based differential item functioning on the Kansas City Cardiomyopathy Questionnaire (KCCQ) using qualitative content analysis

Theresa M. Coles<sup>1,8</sup> · Nicole Lucas<sup>1</sup> · Molly McFatrigh<sup>1</sup> · Debra Henke<sup>1</sup> · Jennifer L. Ridgeway<sup>6</sup> · Emma M. Behnken<sup>6</sup> · Kevin Weinfurt<sup>1</sup> · Bryce B. Reeve<sup>1</sup> · Amy Corneli<sup>1</sup> · Shannon M. Dunlay<sup>6</sup> · John A. Spertus<sup>2</sup> · Li Lin<sup>1</sup> · Ileana L. Piña<sup>3,4,5</sup> · Fraser D. Bocell<sup>5</sup> · Michelle E. Tarver<sup>5</sup> · Heidi Dohse<sup>7</sup> · Anindita Saha<sup>5</sup> · Brittany Caldwell<sup>5</sup>

Accepted: 7 October 2022 / Published online: 2 November 2022  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

## Abstract

**Purpose** The purpose of this study is to evaluate potential gender-based differences in interpreting the Kansas City Cardiomyopathy Questionnaire (KCCQ-23) and to explore if there are aspects of health-related quality of life (HRQOL) not captured by the KCCQ-23 that are important to assess in men and/or women with heart failure (HF).

**Methods** Patients  $\geq 22$  years of age with clinician-diagnosed HF and left ventricular ejection fraction  $\leq 40\%$  were recruited from two academic medical centers to participate in semi-structured concept elicitation and cognitive debriefing interviews. Enrollment was stratified by patient-identified gender (half women/half men). All interviews were conducted over the phone/web and audio recorded. Interviews were transcribed and descriptive qualitative content analysis was used to summarize findings overall and by gender.

**Results** Twenty-five adults (56% women) diagnosed with HF participated. The average age was 67 years (range: 25–88). Women attributed a wider variety of symptoms to HF than men. Some participants had difficulty differentiating whether their experiences were due to HF, side effects of their medications, or age. We found very little evidence that participants interpreted KCCQ-23 items differently based on gender.

**Conclusions** Overall, our findings indicate that interpretation of the KCCQ-23 items were similar in men and women. However, some modifications to items may improve clarity of interpretation for a wide range of patients.

**Keywords** Heart failure · Women · Patient-reported outcomes · Differential item functioning · Qualitative

✉ Theresa M. Coles  
Theresa.Coles@Duke.edu

<sup>1</sup> Department of Population Health Sciences, Center for Health Measurement, Duke University, Durham, NC, USA

<sup>2</sup> Saint Luke's Mid America Heart Institute, University of Missouri Kansas City, Kansas City, MO, USA

<sup>3</sup> Wayne State University, Detroit, Michigan, USA

<sup>4</sup> Central Michigan University, Mount Pleasant, Michigan, USA

<sup>5</sup> Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland, USA

<sup>6</sup> Mayo Clinic, Rochester, NY, USA

<sup>7</sup> Tour de Heart, Hailey, ID, USA

<sup>8</sup> Duke University School of Medicine, 215 Morris Street, Durham, NC 27701, USA

## Plain english summary

This study evaluated potential gender-based differences in interpreting a questionnaire used to assess health-related quality of life for adults diagnosed with heart failure—the Kansas City Cardiomyopathy Questionnaire (KCCQ-23). We also explored if there are aspects of patients' HRQOL experiences not captured by the KCCQ-23 that are important to assess in men and/or women with heart failure. We conducted one-on-one interviews with 25 adults diagnosed with heart failure (56% women). Men and women interpreted KCCQ-23 items similarly. Some KCCQ-23 questions may need minor changes to improve clarity of interpretation for a wide range of patients.

## Introduction

The use of patient-reported outcome measures (PROMs) in clinical trials and clinical care assumes that patients interpret the questions similarly, regardless of their race, ethnicity, gender, age, or other characteristics [1]. When two groups of patients (e.g., men and women) who have the same health status have different probabilities of providing the same PROM response based on the group they belong to, differential item functioning (DIF) is present. DIF can mask true differences by subgroups, by either canceling out the difference in true health status or artificially increasing differences relative to the true difference. Though DIF is typically evaluated with quantitative methods based on empirical data collected by the questionnaire under study, very few studies have used qualitative methods to explore DIF [2, 3].

Understanding whether or not DIF exists is particularly important for PROMs, such as the Kansas City Cardiomyopathy Questionnaire (KCCQ) for heart failure (HF) [4], as the scientific community seeks to understand the underlying mechanisms behind differences in HF outcomes by gender and other key patient characteristics (e.g., race/ethnicity). We posited that DIF may be present on the KCCQ because women have different lived experiences with HF, which could cause women to interpret questions about HF differently [5–8]. Further, there has been a historic lack of representation of women in HF studies [9], which could have excluded some women's perspectives in the development of HF PROMs.

Few studies have evaluated DIF in PROMs developed for evaluating heart failure symptoms and functioning. DIF has been studied in two heart failure-specific PROMs, the Patient-Reported Outcomes Measurement Information System (PROMIS + HF)[10] and Minnesota Living with Heart Failure Questionnaire (MLHFQ)[11, 12]. No statistically significant DIF was identified on the PROMIS + HF, but not all items could be evaluated due to some domains including too few items. DIF of some MLHFQ items were flagged with statistically significant DIF, but of insufficient magnitude to require changes to the MLHFQ. Psychometric properties of the KCCQ were compared by gender [13], showing that validity, reliability and sensitivity to change were similar by gender, but published quantitative DIF studies on the KCCQ are not available. Coles and colleagues have discussed the importance of evaluating DIF in measures used with this population [14]. Similar to the PROMIS + HF measure, the KCCQ includes domains with few items (anchors), limiting the ability to conduct quantitative DIF analyses on all domains.

To investigate how DIF might manifest in the KCCQ-23 by gender, we conducted a qualitative descriptive study. Our study had four objectives. First, we examined

how men and women describe concepts measured by the KCCQ (symptoms, physical function, and social limitations). Second, we explored possible gender differences in how patients interpret concepts, questions, and response choices on the KCCQ-23, with a focus on KCCQ-23 items that were indicated, a priori, by clinicians to potentially exhibit DIF. Third, we investigated whether there were impacts of HF on HRQOL that are not currently captured by the KCCQ-23, but that may be important to assess in men and women with HF. Fourth, we explored whether answering KCCQ-23 questions amid the COVID-19 pandemic may influence how patients chose responses.

## Methods

### DIF supposition generation

To inform qualitative interview questions so that focus could be directed to those items that were most likely to be interpreted differently in men and women, we invited 14 clinicians (using non-probability sampling) to complete a worksheet with their impressions of possible interpretational differences by men and women on the KCCQ-23. The worksheet first provided a definition and example of DIF. Next, the worksheet presented KCCQ-23 item stems and response choices and asked clinicians to provide their own suppositions for DIF (if any) for each KCCQ-23 item (Supplemental Appendix 1). Results were tabulated by KCCQ-23 item. Open-ended responses were reviewed and common themes to explore in the patient interviews were identified by comparing across clinicians by item. KCCQ-23 items were prioritized for inclusion in the qualitative interviews if 5 or more clinician participants agreed on DIF suppositions per item.

### Qualitative interview participation selection and recruitment

Within a constructivist research paradigm [15, 16], one-on-one qualitative interviews were conducted with adults  $\geq 22$  years of age with clinician-diagnosed New York Heart Association (NYHA) functional class I–IV HF and left ventricular ejection fraction (LVEF)  $\leq 40\%$  on their most recent cardiac imaging for those who received care at Duke University (Durham, NC) or Mayo Clinic (Rochester, MN). Exclusion criteria included significant cognitive or memory impairment, hospitalization or referral to hospice at the time of the interview, individuals who had a left ventricular assist device (LVAD) or were planning to have surgery to implant an LVAD within 2 weeks of screening, and individuals who had received a cardiac transplant.

Purposive sampling was used to ensure diversity of patient experiences. To compare responses by gender, the primary sampling objective was stratification by patient-defined gender to include approximately equal numbers of women and men. Within each gender category, we aimed for diversity by three sampling objectives: (1) HF severity (NYHA: I/II vs III/IV), (2) age ( $\leq 70$  vs  $> 70$  years), and (3) race (Black or African American, White, or other).

Interviews were conducted by phone or web conference (with the video off). Patients at both sites were screened via medical chart review and a screening questionnaire for eligibility. At both centers, study team members reached out to potentially eligible patients via email or phone. For individuals receiving care at Duke, the interviewers conducted verbal consent. For individuals at Mayo Clinic, those interested in participating were sent a signed HIPAA authorization form by mail, and verbal consent was conducted.

## Measures

The KCCQ-23 is a 23-item PROM that assesses 8 HF-related concepts: Physical Limitations (6 items), Symptom Stability (1 item), Symptom Frequency (4 items), Symptom Burden (3 items), Self-efficacy (2 items), Quality of Life (3 items), and Social Limitations (4 items) [4]. The KCCQ-23 has been psychometrically evaluated in multiple studies [4, 17, 18], and has been qualified as a Medical Device Development Tool [19]. All KCCQ-23 domain scores range from 0 to 100, with higher scores reflecting better health status. The KCCQ-23 was completed by patients during the interview on paper.

## Number of interview participants

Guest and colleagues [20] showed that relatively little new information is learned after 12 interviews within a homogeneous sample of interview participants. We anticipated that at least 12 interviews per gender would be needed to achieve sufficient redundancy in information such that new concepts identified in additional interviews would not meaningfully

add to further understanding for objectives of this study [21]. Symptom reports became redundant, suggesting saturation had been achieved, by approximately 12 interviews per gender, and 25 interviews had been conducted (Supplemental Appendix 2).

## Description of interviews

Semi-structured interviews using concept elicitation [16] and cognitive interviewing techniques [22] were conducted using interview guides developed by the research team. Interview questions that informed study conclusions are presented in Supplemental Appendix 3. The interview guide was developed and refined collaboratively by the study team including researchers, clinicians, and a patient representative. Interviews were approximately one hour, and all were conducted in English. Each interview was conducted in three sections, representing the three study objectives (Fig. 1).

Interviews were conducted via phone or web from July 2020 through December 2020. All interviews were audio recorded and transcribed verbatim for analysis purposes. Interviewers completed debriefing forms to summarize their findings and monitor data saturation. Participants received remuneration of \$50. The study was approved by Duke University (coordinating center) and Mayo Clinic institutional review boards.

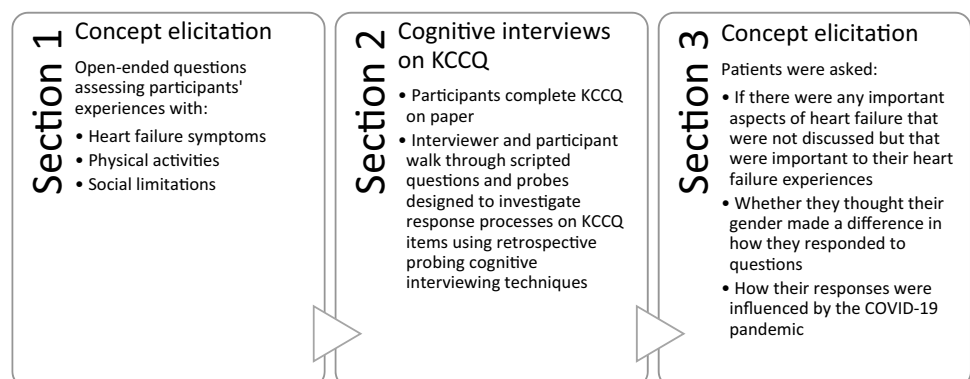
All interviewers from Duke and Mayo Clinic were women. Interviews were conducted throughout the COVID-19 pandemic.

## Analytic methods

Descriptive statistics were calculated to summarize characteristics of the study participants, as well as KCCQ-23 domain responses using SAS 9.4 software (© 2016, SAS Institute Inc., Cary, NC, USA). Sample sizes were not sufficient to conduct quantitative DIF testing.

Qualitative content analysis was used to analyze participant narratives [23]. Study team members, including interviewers, clinicians, and a patient representative,

**Fig. 1** Interview content



met regularly to discuss the analysis plan, coding application, and interpretation of results. A codebook was developed collaboratively among investigators at both sites (Supplemental Appendix 4). Duke team members used NVivo qualitative data analysis software (Version 12 for Windows) to apply codes to transcripts. To assess coder agreement for the first transcript, three Duke analysts independently coded a transcript and then met to reconcile codes. Discrepancies were discussed, documented, and reconciled, and the codebook was reviewed and revised. The analysts then divided the remaining transcripts and independently coded them. Study team members met regularly throughout analysis to discuss questions and ensure consistency of code application and interpretation, and questions that arose about the applicability of codes to transcript sections. Throughout coding, if revisions were needed, new codes were documented and dated, and applied to all transcripts. After coding was complete, analysts identified themes by reviewing the code reports categorized by patients' self-reported gender. Themes were identified when multiple patients reported similar experiences or interpretations of an item. Next, themes were compared descriptively by gender, and analysis identified themes across gender. Other participant characteristics (age, NYHA status) were also associated with each transcript, allowing analysts to provide context to salient themes. Illustrative quotes were identified. Results were compared with clinician DIF suppositions.

Percentages were reported only when the analysis team could calculate a denominator based on the number of participants who were asked a question. It is important to note that this study design posits that topics most salient to patients are reported by patients; thus even if a topic was not discussed, it does not mean it was not experienced by a participant. Results.

## Results

### Results of the DIF supposition generation

Of the 14 invited clinicians, 12 provided responses to DIF worksheets. Of the 12 clinicians who responded, 7 were cardiologists, 2 were transplant specialists, 1 was a cardiothoracic surgeon, 1 was a nurse practitioner, and 1 did not provide demographic information. The average years in practice was approximately 23 years (range: 12–40 years). Figure 2 presents an overview of the item-level gender DIF suppositions shared by at least five clinicians (approximately 40% of participants). KCCQ-23 items were prioritized for inclusion in the qualitative interviews based on clinicians' agreement on DIF suppositions.

### Description of the qualitative sample

Twenty-five patients with HF (13 women, 12 men) were included in the qualitative study (Table 1). Most participants were white (76%) and had an average LVEF of 30.3%. Additional patient characteristics, such as comorbid conditions and HF etiology, are presented in Supplemental Appendix 5.

In general, men had higher scores in almost all KCCQ-23 domains compared to women, representing better health status (Supplemental Appendix 6).

### Objective 1 results: KCCQ-23 concepts by gender

#### Symptoms

Fatigue ( $n = 19$ ), shortness of breath ( $n = 15$ ), and swelling ( $n = 7$ ) were the most frequently mentioned symptoms participants attributed to their HF. Participants used a range of synonyms to describe symptoms. For example, for fatigue, patients used the term “sluggishness,” “needing to take naps,” “feeling tired,” and “lack of energy.” Women in this study described a wider range of symptoms than men (Supplemental Appendix 2), and only women spontaneously reported chest pain ( $n = 4$ ). Some patients had trouble differentiating if their symptom was due to HF or a side effect of their medication.

#### Physical limitations

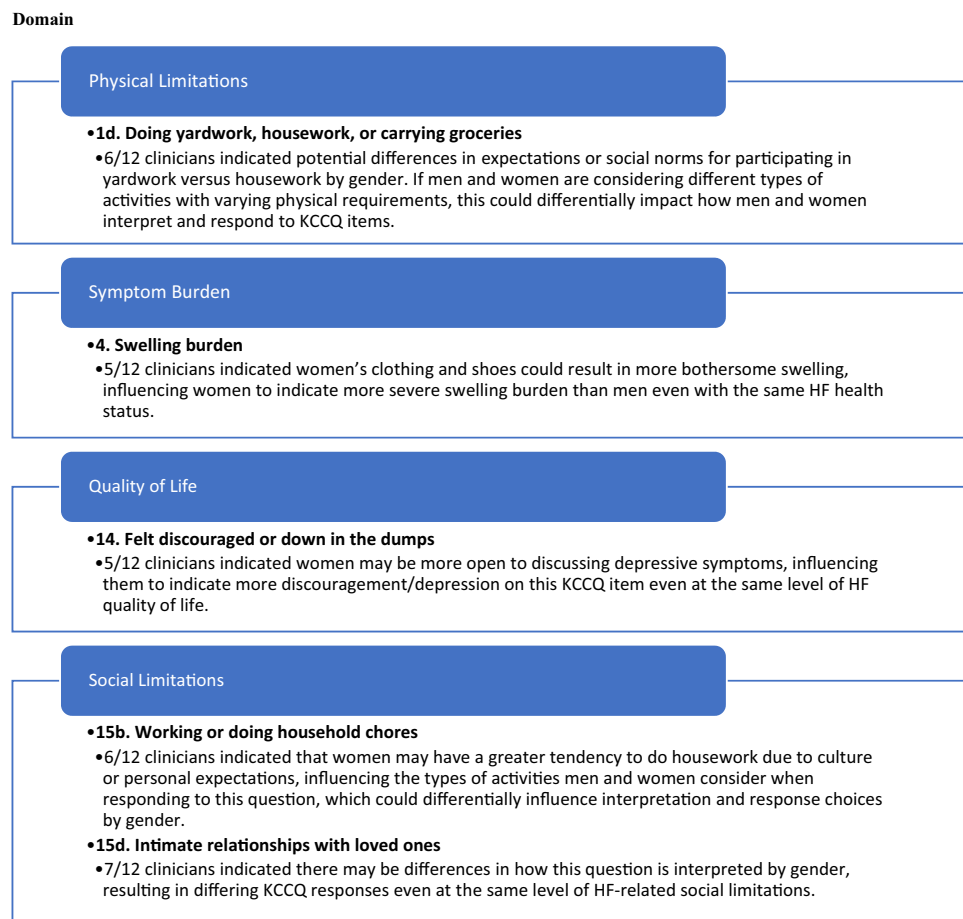
On the question “How does having heart failure impact your day-to-day physical activities,” four participants did not specify any physical limitations due to HF (2 women, 2 men). All of these patients were under 70 and were classified as NYHA I or II. Approximately a third of participants ( $n = 8$ ) indicated they experienced shortness of breath with activities such as walking, doing chores, or going up stairs. Half of all participants in the study ( $n = 14$ , 56%) indicated that they avoid some activities altogether or needed to take breaks to complete the activity:

*“I’ve learned to adjust my expectations. I don’t like making that adjustment because I don’t like spreading out a chore over days, but, you know, it’s fine. It’s better than not being able to do it at all.”* [Age 61, female, White, NYHA III]

Three participants noted that each day could be different, labeling some days as good and some as bad. For example, one participant said:

*“On a really bad day I’ve had to sit down before I even complete dressing [...] those days [are] pretty much [...] a wash because I do not even think about*

**Fig. 2** Clinician-generated gender-based Differential Item Functioning (DIF) suppositions by KCCQ-23 domain



*going to get groceries. I do not even really think about sweeping, mopping, vacuuming... Now on good days I can wake up, go ahead, get my bath, feel pretty good. I can go get a few groceries definitely before the heat ...Come back, put the groceries up, fix meals. I've even put up stuff in the freezer, sweep, mop, vacuum.”* [Age 57, female, White, NYHA III]

Age was mentioned by a few participants as a possible contributor to their physical limitations. It was difficult for them to determine whether some limitations were due to HF or their age exclusively:

*“I think it might be related to [...] heart failure. And I think now that age has played a role in that too... I think it comes from both of those things. I don't know which has more impact...”* [Age 75, male, Black, NYHA I]

No participant mentioned the most effortful physical activity covered by the KCCQ-23 Physical Limitations domain: running or hurrying (as if to catch a bus). Overall, men and women reported similar activities and physical limitations.

### Social limitations

Fifteen participants said that HF does not impact or mostly does not impact their relationships or social activities with family or friends (7 women, 8 men; NYHA: I=2; II=4, III=9). Seven of these participants were under 70 years old and 9 were older. Four participants reported a change in their friendships, with friends treating them differently because of their HF:

*“I'm just seen as [...] the poor person with a heart condition... I feel like a sad story sometimes, but I guess it makes it, [...]hard, and then I just try not to [...]talk about it very much with people because I know it makes them really uncomfortable.”* [Age 25, female, White, NYHA II]

Approximately a third of patients mentioned their HF affected relationships with family members. Three participants said they were not able to keep up with family activities or being “sidelined” for certain activities. For



**Table 1** Qualitative study patients characteristics overall and by gender

Level or Statistic	Overall (N=25)	Male (N=12)	Female (N=13)
<b>NYHA functional class</b>			
I	6 (24.0%)	5 (41.7%)	1 (7.7%)
II	7 (28.0%)	4 (33.3%)	3 (23.1%)
III	11 (44.0%)	3 (25.0%)	8 (61.5%)
IV	1 (4.0%)		1 (7.7%)
<b>Age</b>			
Median	68.0	68.5	68.0
Mean	67.0	68.6	65.5
Min	25.0	52.0	25.0
Max	88.0	81.0	88.0
<b>Ejection fraction</b>			
Median	31.0	32.0	30.0
Mean	30.3	30.4	30.2
Min	15.0	15.0	20.0
Max	40.0	40.0	40.0
<b>Ethnicity</b>			
Hispanic	1 (4.0%)	1 (8.3%)	
<b>Race</b>			
White	19 (76.0%)	10 (83.3%)	9 (69.2%)
Black or African American	6 (24.0%)	2 (16.7%)	4 (30.8%)
<b>Education</b>			
Less than high school	1 (4.0%)		1 (7.7%)
High school graduate or GED	1 (4.0%)		1 (7.7%)
Completed some college; no degree	10 (40.0%)	5 (41.7%)	5 (38.5%)
Completed associate degree	2 (8.0%)		2 (15.4%)
College graduate	5 (20.0%)	2 (16.7%)	3 (23.1%)
Graduate school	6 (24.0%)	5 (41.7%)	1 (7.7%)
<b>Insurance</b>			
Medicare	18 (72.0%)	9 (75.0%)	9 (69.2%)
Medicaid	2 (8.0%)	1 (8.3%)	1 (7.7%)
Private plan	16 (64.0%)	7 (58.3%)	9 (69.2%)
Uninsured	1 (4.0%)	1 (8.3%)	
Other	1 (4.0%)		1 (7.7%)

example, at the pool with grandchildren, one woman indicated that she was limited to floating in the water rather than playing with the children. Five participants mentioned that romantic relationships were impacted by HF (3 women, 2 men), including limiting physical/sexual aspects due to lack of energy, or being emotionally affected by HF—feeling like a burden to their partner or being in a bad mood. Erectile dysfunction due to HF was mentioned by a few men. Though a wide variety of social impacts were mentioned, themes were consistent by gender for the types of activities mentioned or how social activities were affected.

## Objective 2 results: response processes for KCCQ-23 Items by gender

### KCCQ1d: doing yardwork, housework, or carrying groceries

A wide range of physical effort was noted by patients when reflecting on this question, including weeding, cutting tree limbs with a chainsaw, carrying groceries, cleaning bathrooms, cutting grass, washing dishes, and sweeping. Several participants reported they needed to do chores in small time allotments because of HF. Willingness to take breaks did not vary by gender. All participants described helping with some aspect of yardwork, housework, and groceries. No one stated that they did not do a task because of how labor was divided by gender in their household.

During interviews, participants with limitations tended to rate only the concepts in the question that they were able to do and ignored those they were unable to do:

“I [responded] Moderately [Limited] because the things that are grouped together there, I can do some of them better than others. I don't have any problem carrying groceries and I can do some housework. I do have some help every other week, so that does take a stress off of [me] a little bit. But yard work would probably be lower rated than that because that was the one that I'm not moderate in ...” [Age 71, female, White, NYHA IV] Themes were consistent across genders in how patients chose which tasks they focused on when answering the question.

### KCCQ4: swelling burden

Almost half of participants ( $n = 11$ ; 44%) said they had not experienced swelling in the past 2 weeks (4 women, 7 men). Four patients chose the response indicating that swelling was “Not at all bothersome.” They said they had swelling in their ankles or feet, but it did not interfere with daily activities. Those who selected “slightly bothersome” (5 women, 1 man) did so for several reasons. Two female participants said they felt bothered by how “heavy” their legs felt when they experienced swelling (makes walking feel strenuous). One participant said:

“It's the heaviness. It's like carrying another human on you [...]. If I'm walking, it's more heavy. It's more tedious and more strenuous than most days.” [Age 48, female, Black, NYHA I]

Overall, regardless of response choice, participants discussed taking medications (diuretics), monitoring their salt and fluid intake as well as their weight, or using compression socks to prevent swelling or maintain it at a reasonable level.

Women were more likely to report that they were “moderately,” “somewhat,” or “slightly” bothered by swelling ( $n = 9$ ; 7 women, 2 men). Both men and women were bothered by their clothes, but more women mentioned that swelling bothered them due to how their clothes/shoes fit. Swelling in ankles and feet caused women to have to change what shoes they could wear. Two women said they were limited to only wearing sandals when their feet/ankles were swollen. One said she wears panty hose and special shoes for swelling, and the other said that shirts might get tighter if she has swelling in her arms or her bra might fit tighter. On the other hand, not all participants wore tight clothing that caused trouble due to swelling:

*“How do my clothes affect the swelling of my feet? Pretty much, they don’t. I don’t wear tight socks, I don’t wear tight shoes, I wear shorts most of the time because I live in North Carolina. So, my clothes have very little effect.”* [Age 75, male, White, NYHA I]

#### KCCQ14: felt discouraged or down in the dumps

When asked to define “discouraged or down in the dumps,” patients mentioned sadness, fear, “*not wanting to do anything*,” dissatisfaction, frustration, unhappiness, anger, or stress. A few used terms implying self-judgment such as “*self-pity*” ( $n = 2$ ), “*woe is me*” ( $n = 3$ ), or “*feeling sorry for yourself*.” A few stated they were more likely to feel discouraged when their HF symptoms were worse. When some were asked explicitly if they would have trouble owning up to their feelings, two men and one woman indicated that they may have difficulty describing their feelings of sadness or depression. Themes were consistent across genders in response processes, including how patients interpreted the intent of the question and how they got to their response.

The KCCQ-23 asks patients to consider only their HF-related discouragement when answering KCCQ14. Some participants had trouble limiting their responses to HF-specific discouragement:

*“I have occasionally, and some of that doesn’t have anything to do directly with heart failure, but it has directly to do with my mindset, given the virus and the social issues, and a president that has run amok, and just a lot of things that have gone on. I have [...] had some issues with PTSD, so there will be occasions in which [I] fall into that space. [I do not] plan to go there, but [I] just drop off the vine for a couple of minutes.”* [Age 75, male, Black, NYHA I]

#### KCCQ15d: working or doing household chores

Almost one-third (7/24) of participants said they were not limited at all. Those who said they were limited quite a bit

or severely limited were all women, whereas the majority of men (8/12) said they were not limited at all or only slightly limited. Sweeping, mopping, and vacuuming were mentioned as the most physically demanding chores by both genders. Patients who mentioned these tasks answered “moderately” or “limited quite a bit.” Some avoided these tasks altogether because they caused shortness of breath and fatigue. These individuals were not able to stand for long periods of time and had to sit down to do chores (folding laundry, doing dishes, and vacuuming). Men and women provided similar reasons for why they chose response options. Many participants said they feel limited because of the breaks they have to take, making them feel they cannot complete work or chores at the same pace they used to before HF:

*“But after I have finished with this floor, it’s time for me to sit down. And then, when I get back up, I can go do another part of the project, maybe the floor in the back room or something.”* [Age 75, male, Black, NYHA I]

Not all participants were bothered by chores taking longer due to needing to take breaks. There were no differences in willingness to take breaks by gender that could influence how patients understood the question or got to their response.

Patients who were more physically active before their HF diagnosis may have felt more limited due to HF than patients who were not as active. For example, one man said:

*“Well normally, I’d be out there with my [...] chainsaw cutting trees down and taking care of that, and I can’t do that, so I’m limited.”* [Age 66, male, White, NYHA II]

#### KCCQ15d: intimate relationships with loved ones

Forty percent of participants (10/25) reported that they did not experience limitations in their intimate relationships due to HF. Overall, women chose the more-severe response choices, “moderate” or “severe limitations,” compared with men (7 women, 3 men). Most participants ( $n = 18$ ; 72%) cited that intimate relationships referred to sexual relationships, “in the bedroom,” with spouses or partners (including sex, as well as physical touch like caressing, hugging, or kissing a partner):

*“Intimate relationships could be closeness. You know, emotional. Hugging, touching, loving. I do all of that part, but I’m not interested in the extracurricular activities [sexual penetration, oral sex, etc.] when it comes to intimacy.”* [Age 68, female, Black, NYHA II]

There was an even split in how participants defined “intimate relationships” by gender:

- Sex only (5 men, 6 women)
- Sex and family/friends (4 men, 3 women)
- Family/friends only (3 men, 4 women)

A response process challenge theme emerged, with some patients responding differently depending on their definition of intimate relationships:

*“Well, I didn’t know [...] if we’re talking about erectile dysfunction [...]. But if we’re talking about intimate, then I’m limited quite a bit. If you’re talking about loving my family and enjoying my family and them being around, well then I’m not limited from that standpoint.”* [Age 67, male, White, NYHA II]

Participants under 70 years of age also endorsed more limitations with intimate relationships than those over 70. Some participants over age 70 mentioned they and their spouse are no longer having sex, and that their definition of “intimate relationships” has changed over time.

*“Like I said, my wife and I get along well. There’s some things that we used to do that we don’t do any more, and I think you know what I’m talking about [sex], but it’s okay...right now, [our] intimate relationship is just being able to maybe go to the movie or going out to supper or something with my wife and just being with us two together and enjoying one another.”* [Age 75, male, White, NYHA II]

Participants were asked explicitly if they believed gender influenced how they responded to or thought through answering questions on the KCCQ. Ten participants (4 women, 10 men) did not think gender influenced how they answered the questions. Four women stated they did more housework than their male domestic partner. Other hypotheses cited included erectile dysfunction being a male-only experience ( $n=3$ ), women experience different symptoms of heart failure than men ( $n=2$ ), and that women may be more likely to modify their behaviors to control their heart failure symptoms ( $n=1$ ).

Item-specific response process challenges are described above, and there were two response process challenges that were present across multiple items and genders. Fifteen patients thought about a different condition or their age in addition to HF or instead of HF when answering some KCCQ-23 items. For KCCQ-23 items referring to more than one activity, participants tended to only think about and rate the concepts in the question that they were able to do.

### Objective 3 results: examine potential impacts of HF on HRQOL not captured in KCCQ-23

At the end of the interview, patients were asked if there were any other important aspects of HF that were not discussed during the interview but that were important to them. Many participants indicated that the interview already covered the important aspects of HF. For patients who provided answers to this question, none of the concepts were health status related. A description of these non-health-status-related results are provided in Supplemental Appendix 7.

### Objective 4 results: influence of COVID-19 pandemic on responses.

Participants were asked how the COVID-19 pandemic influenced their social relationships and activities that they would normally be involved in. Answers for this question varied with some saying no impact, while others said it has had a large impact. Participants mentioned not being able to do activities they usually enjoy due to infection risks (e.g., attending concerts, going to the gym, traveling). On the other hand, participants indicated they were still able to do a number of activities amid the pandemic such as dog training, getting a haircut, going to the store, or going outside for functions. Some patients experienced a positive impact on their lives from the pandemic, with two individuals noting that they were able to see their family more because of COVID-19. Thinking about the pandemic influenced how some participants managed the 2-week recall period of the KCCQ, with a few participants referencing times outside of the two-week recall period, “before COVID”, when choosing a response. However, this recall response challenge was only observed for some items and only by some participants. For example, when asked about hobbies and recreational activities, one participant referenced things they were unable to do before the pandemic began and before they had heart failure. This participant used a two-week recall period for the other KCCQ items. No themes emerged by gender related to how participants thought through their responses on the KCCQ due to COVID-19.

## Discussion

In this descriptive qualitative study, we explored four primary research questions. For objectives 1 and 3, we explored how women and men described concepts measured by the KCCQ-23 (objective 1) and whether there were any impacts of HF on HRQOL that were not captured by the KCCQ-23 that would be important to capture (objective 3). We



began by investigating symptoms. Overall, men and women indicated experiencing a similar set of symptoms—fatigue, shortness of breath, and swelling—all covered by the KCCQ-23. A few women described experiencing a slightly wider variety of HF symptoms including chest pain or joint pain. Further clinical investigation is needed to determine whether these symptoms were due to HF. For physical limitations, women and men experienced similar limitations. The types of activities spontaneously reported by patients were represented on the KCCQ-23, such as bathing, walking, or housework. The KCCQ-23 also includes questions about running/hurrying or climbing stairs, which were not concepts spontaneously reported by participants in this qualitative study. For social limitations, emotional impacts of the social limitation—burdening others with their HF limitations or being “sidelined” when trying to participate in social activities—were frequently described. During concept elicitation, participants mentioned all of the concepts represented on the KCCQ-23 Social Limitation domain except for working or doing household chores. However, participants did mention limitations with household chores during other parts of the interview, so it is likely that participants did not consider work or household chores to fall under the category of social limitations related to HF.

Regarding objective 2, for KCCQ-23 items thought by clinicians to potentially exhibit DIF, we found very few indications that participants interpreted the items or generated their responses in a different way based on gender. For KCCQ1d (doing yardwork, housework, or carrying groceries), some clinicians indicated that social norms may influence which activities patients participated in. However, all participants described helping with aspects of yardwork, housework, and groceries. Clinicians suggested that swelling might be more bothersome for women because women sometimes wear tighter clothes or shoes than men. This supposition had some support from our cognitive interview results on KCCQ4 (swelling burden); women and men were bothered by swelling, but more women referenced their clothes or shoes bothering them when they experienced swelling. When responding to the KCCQ, it is possible that women would respond indicating that they were more bothered by swelling than men with the same amount of swelling because women are more likely to wear tighter-fitting clothes and shoes. Not all patients wore tight clothing, however, making it challenging to compare these experiences by gender.

Clinicians proposed that men may be less likely to acknowledge feeling depressive symptoms (KCCQ14). This was a challenging supposition to address because individuals who do not acknowledge their feelings may not recognize they are not acknowledging their depressive symptoms. However, we did ask participants to define the term “discouraged or down in the dumps,” and interpretation of KCCQ14 was consistent across genders. For KCCQ15

(“working or doing household chores”), clinicians indicated that women may have a tendency to engage in housework due to cultural or personal expectations, which could influence how they interpret or respond to KCCQ15b. However, the level of effort or activities men and women were thinking about when answering this question were similar by gender. Participants reported a wide range of definitions for “intimate relationships” on KCCQ15d (“intimate relationships with loved ones”); however, there were no differences in interpretation by gender.

For objective 4, we explored whether participants’ response processes were influenced by the COVID-19 pandemic. While the pandemic influenced patients’ ability to conduct some regular activities, it also opened up some opportunities for more social interaction with family members. There were no themes observed by gender in how the pandemic influenced participants’ response processes. The two-week recall period was not used by all participants, as some participants talked about life “before COVID”.

Across KCCQ-23 items, we detected some response process challenges. First, the KCCQ-23 is a disease-specific measure that asks patients to consider outcomes only related to their HF. Patients are not always able to differentiate impacts of their condition from other comorbid conditions or medications [24]. This specific response challenge can contribute to “noise” in patients’ responses when assessing health status in multi-morbid populations. Another response process challenge was related to items listing more than one activity. Participants tended to only consider and rate the concepts in the question that they were able to or did do rather than those they were unable to do or did not do. Though our cognitive interviews did not identify gender-specific response process challenges, response process challenges may limit how deeply we can examine interpretational differences by gender that could indicate DIF.

As with any qualitative cognitive study examining a PROM, an overarching question is: Do we need to make a change to this measure? The results reported here do not provide support for changes to address potential gender-based DIF. However, clarification surrounding items listing more than one activity may facilitate more consistent response processes across individuals with HF. Also, for KCCQ14, while patients understood the intent of the question, there is some evidence that the term “down in the dumps” did not resonate with patients. A potential rewording of the item may provide clarity for participants, especially whether or not they should include HF-related depressive symptoms in their responses or more general depressive symptoms.

This study builds on previous studies that quantitatively evaluated DIF on other HF measures, the PROMIS + HF [10] and MLHFQ [11, 12]. A common theme among these measures and the KCCQ-23 is that some domains include too few items to serve as strong anchors, a prerequisite of

being able to conduct quantitative DIF analyses. This study used qualitative methods to examine DIF when quantitative DIF analyses could not be evaluated, a method that might be useful in examining items that could not be evaluated using quantitative methods.

A strength of this study was that the interviews were conducted at two sites, Duke University and Mayo Clinic, providing important geographic diversity to patient perspectives. Due to the COVID-19 pandemic, all interviews were conducted remotely, which may have made the interview process more accessible to some patients and less accessible to patients without access to computers and internet. Another strength of the study was the engagement with stakeholders, including clinicians and a patient [HD], who provided feedback on data collection instruments, and analysis and interpretation. Diversity of perspectives throughout the study provided the team with an important dimension of understanding participant responses.

A limitation of this study is that responses could be indicative of only one person's experiences or the experiences/interpretation of multiple or many patients within the population. This is also a strength of qualitative analysis in that the purpose is to extract a deeper understanding of individual patient experiences. It is possible that evidence of DIF could be uncovered in other samples, but all attempts were made to ensure that the sample was balanced by gender, race, NYHA class, and age. Another limitation of this study is that only one transcript was double coded to assess consistency of code application. The interview guides were designed to group similar concepts together (structured with scripted probes), thus coding was largely structural in nature and it was easy to agree on code concurrence. This study is also limited by its underrepresentation of patients with lower levels of education, those without health insurance or virtual meeting technology, and non-English-speaking participants within and outside of the United States. All interviewers and analysts are women, which could potentially influence interpretation of the results by gender.

**A note on interpreting the qualitative results:** Due to the open-ended nature of qualitative questions, participants generally are not asked during the interviews if they have or have not experienced all symptoms or other phenomena; rather, they are asked to describe their experiences. The supposition is that topics most salient to patients are reported by patients. Therefore, the number of individuals for whom an experience is not mentioned is not an indication of frequency of that experience in the population. Rather, we looked for general themes across participants and by gender.

The evaluation of DIF can be an important aspect of evaluating the validity of proposed interpretations and uses of PROMs. If DIF unduly influences scores on PROMs, then those scores might not be able to serve their

intended purpose, such as forming an endpoint in a clinical trial [1]. This underscores the importance of exploring potential DIF by gender and other sociodemographic factors, such as age, that might be relevant for understanding a disease or condition. In this qualitative study of the KCCQ-23, we found no evidence that men and women interpret the instrument differently and found overall support for its content validity. This suggests that it can be used equally well in clinical studies and that observed sex-based differences likely reflect the true impact of treatment on patients, rather than being an artifact of differentially measuring the experiences of men and women. Ideally, other PROMs will be evaluated to independently confirm their interpretability in men and women to support their use in disparity research. This study also provides an example of how to use qualitative methods to complement quantitative methods to assess DIF in PROMs.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11136-022-03276-y>.

**Acknowledgements** The authors acknowledge the patients who participated in this study, providing important insight into their experiences with heart failure. We would also like to acknowledge Ashley J. Wilson, a Center for Health Measurement intern for her support with the DIF worksheets, as well as Karen Staman and Liz Wing, who provided medical editing support.

**Author contributions** TMC: designed and is responsible for the study, interpreted results, wrote the draft manuscript, and revised the manuscript; NL, MM, DH: provided feedback on the design and analysis of the study, assisted with protocol preparation and IRB submission, recruited, enrolled, consented and interviewed participants, conducted analyses, drafted sections of the manuscript, interpreted results, and provided feedback on the manuscript; JLR: provided feedback on the design and analysis of the study, interviewed participants, interpreted results, and provided feedback on the manuscript. EMB: assisted with protocol preparation and IRB submission, recruited, enrolled and consented participants, and abstracted medical record data on participants. SMD: provided feedback on the design and analysis of the study, facilitated identification of participants, interpreted results, and provided feedback on the manuscript. JAS, ILP, FDB, MET, AS: interpreted results and provided feedback on the manuscript. LL: conducted analyses and provided feedback on the manuscript. KW, BBR, AC: provided feedback on the design and analysis of the study, interpreted results, and provided feedback on the manuscript. BC: initiated the research study questions and initial design, interpreted results, and provided feedback on the manuscript. HD: provided feedback on the design of the qualitative interview guide, interpretation of results, and provided feedback on the manuscript.

**Funding** This work was supported by the U.S. Food & Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) Office of Women's Health and the Patient Science & Engagement Program within the FDA's Center for Devices and Radiological Health (Contract 75F40119C10080), and as part of a financial assistance award [Center of Excellence in Regulatory Science and Innovation grant to Yale University and Mayo Clinic, U01FD005938] funded by FDA/HHS. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the U.S. Government.

**Data availability** Due to the small sample sizes included in qualitative studies, sharing transcripts or other data could identify an individual. Therefore, these data are not publicly available.

## Declarations

**Conflicts of interest** JS owns the copyright to the KCCQ and receives license fees for the KCCQ. JS receives consulting fees from Bayer, Imbrija Pharmaceuticals, Merck, Novartis, Bristol Meyers Squibb, Janssen, Terumo, and United Healthcare. JS is supported in grants from Abbott Vascular, Janssen and Bristol Meyers Squibb, and is on the Board of Directors for Blue Cross Blue Shield of Kansas City. RM received research support and honoraria from Abbott, American Regent, Amgen, AstraZeneca, Bayer, Boehringer Ingelheim/Eli Lilly, Boston Scientific, Cytokinetics, Fast BioMedical, Gilead, Innolife, Medtronic, Merck, Novartis, Relypsa, Respicardia, Roche, Sanofi, Vifor, and Windtree Therapeutics. TC received research support from Merck and has a consulting agreement with Regenxbio. LL, KW, BR, IP, FB, MT, AS, DH, and BC have no competing interests to disclose.

**Ethical approval** The study was approved by Duke University (coordinating center) and Mayo Clinic institutional review boards.

**Consent to participate** Participants were verbally consented using a verbal script. A copy of the consent form was sent to patients who enrolled in the study.

**Consent for publication** All authors read and approved the final manuscript. All authors also agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## References

- Weinfurt, K. P. (2021). Constructing arguments for the interpretation and use of patient-reported outcome measures in research: an application of modern validity theory. *Quality of Life Research*, 30(6), 1715–1722.
- Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research*, 16(Suppl 1), 33–42.
- Reeve, B. B., Willis, G., Shariff-Marco, S. N., Breen, N., Williams, D. R., Gee, G. C., Alegría, M., Takeuchi, D. T., & Stapleton, M. (2011). Levin KY comparing cognitive interviewing and psychometric methods to evaluate a Racial/Ethnic discrimination scale. *Field Methods*, 23(4), 397–419.
- Green, C. P., Porter, C. B., Bresnahan, D. R., & Spertus, J. A. (2000). Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *Journal of the American College of Cardiology*, 35(5), 1245–1255.
- Adams, K. F., Dunlap, S. H., Sueta, C. A., Clarke, S. W., Patterson, J. H., Blauwet, M. B., Jensen, L. R., & Tomasko, L. (1996). Koch G Relation between gender, etiology and survival in patients with symptomatic heart failure. *Journal of the American College of Cardiology*, 28(7), 1781–1788.
- Lee, C. S., Hiatt, S. O., Denfeld, Q. E., Chien, C. V., Mudd, J. O., & Gelow, J. M. (2015). Gender-Specific physical symptom biology in heart failure. *The Journal of cardiovascular nursing*, 30(6), 517–521.
- McSweeney, J. C., Cody, M., O’Sullivan, P., Elberson, K., & Moser, D. K. (2003). Garvin BJ women’s early warning symptoms of acute myocardial infarction. *Circulation*, 108(21), 2619–2623.
- Pope, J. H., Aufderheide, T. P., Ruthazer, R., Woolard, R. H., Feldman, J. A., Beshansky, J. R., & Griffith, J. L. (2000). Selker HP missed diagnoses of acute cardiac ischemia in the emergency department. *New England Journal of Medicine*, 342(16), 1163–1170.
- Heiat, A., Gross, C. P., & Krumholz, H. M. (2002). Representation of the elderly, women, and minorities in heart failure clinical trials. *Archives of internal medicine*. <https://doi.org/10.1001/archinte.162.15.1682>
- Ahmad, F. S., Kallen, M. A., Schifferdecker, K. E., Carluzzo, K. L., Yount, S. E., Gelow, J. M., McCullough, P. A., Kimmel, S. E., Fisher, E. S., & Cella, D. (2019). Development and initial validation of the promis (R)-plus-HF profile measure. *Circulation-Heart Failure*, 12(6), e005751.
- Munyombwe, T., Höfer, S., Fitzsimons, D., Thompson, D. R., Lane, D., & Smith, K. (2014). Astin F an evaluation of the Minnesota living with heart failure questionnaire using Rasch analysis. *Quality of Life Research*, 23(6), 1753–1765.
- Sébillle, V., & Auget, J.-L. (2004). Evaluating health-related quality of life: A case-study of differential item functioning analysis in small trials. *Communications in Statistics - Theory and Methods*, 33(6), 1403–1428.
- Hejjaji, V., Tang, Y., Coles, T., Jones, P. G., Reeve, B. B., Mentz, R. J., Spatz, E. S., Dunlay, S. M., Caldwell, B., Saha, A., & Tarver, M. E. (2021). Psychometric evaluation of the Kansas city cardiomyopathy questionnaire in men and women with heart failure. *Circulation Heart Failure*, 14(9), e008284.
- Coles, T. M., Lin, L., Weinfurt, K., Reeve, B. B., Spertus, J. A., Mentz, R. J., Piña, I. L., Bocell, F. D., Tarver, M. E., Henke, D. M., & Saha, A. (2022). Do PRO measures function the same way for all individuals with heart failure? *Journal of Cardiac Failure*. <https://doi.org/10.1016/j.cardfail.2022.05.017>
- Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. *International Journal of higher education*, 6(5), 26–41.
- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., & Molsen, E. (2011). Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrument. *Value in Health*, 14(8), 967–977.
- Spertus, J. A., Jones, P. G., Kim, J., & Globe, D. (2008). Validity, reliability, and responsiveness of the Kansas City Cardiomyopathy Questionnaire in anemic heart failure patients. *Quality of Life Research*, 17(2), 291–298.
- Eurich, D. T., Johnson, J. A., & Reid, K. J. (2006). Spertus JA Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health and Quality of Life Outcomes*, 4, 89.
- U.S. Food and Drug Administration, *Medical Device Development Tool (MDDT) Qualification Decision Summary for Kansas City Cardiomyopathy Questionnaire (KCCQ)*. Accessed October 13, 2021. 2016.
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.
- Corbin, J. M., & Anselm, L. (2015). *Strauss, basics of qualitative research: techniques and procedures for developing grounded theory* (4th ed., p. 431). Los Angeles: SAGE.
- Willis, G. B. (2004). *Cognitive interviewing: a tool for improving questionnaire design*. Sage Publications.

23. Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.
24. Fayers, P. M., & Machin, D. (2016). *Quality of life: the assessment, analysis, and reporting of patient-reported outcomes* (p. 1). John Wiley & Sons Inc.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.