Content Validity of the Lee Chronic Graft-versus-Host Disease Symptom Scale as Assessed by Cognitive Interviews

Emily C. Merkel1, Sandra A. Mitchell2, Stephanie J. Lee3,*

1 Department of Medicine, School of Medicine, University of Washington, Seattle, Washington
2 Division of Cancer Control and Population Sciences, Healthcare Delivery Research Program, Outcomes Research Branch, National Cancer Institute, Bethesda, Maryland
3 Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington

Article history:
Received 20 November 2015
Accepted 29 December 2015

Key Words:
Allogeneic hematopoietic cell transplantation
Chronic graft-versus-host disease
Patient-reported outcome
Symptoms
FDA qualification
Cognitive interview

A B S T R A C T
The Lee Chronic Graft-versus-Host Disease (cGVHD) Symptom Scale has been recommended for use by the 2005 and 2014 National Institutes of Health (NIH) Consensus Conferences to capture cGVHD symptoms. Although the cGVHD Symptom Scale was previously validated, this study aims to reexamine the instrument’s content validity by exploring the clarity, comprehensibility, relevance, and ease of use in a contemporary cGVHD sample, toward Food and Drug Administration (FDA) qualification of this patient-reported outcomes (PRO) instrument as a drug development tool. Attaining FDA qualification means that an instrument has been judged to be a reliable and valid measure of clinical benefit. Twenty adult patients with a median age of 58 year (range, 31 to 79 years) participated. The median duration of cGVHD was 33 months (range, 0 to 134.4 months), and current NIH severity score was mild in 1 patient, moderate in 10 patients, and severe in 9 patients, with a median of 5.5 treatments (range, 0 to 14) ever used for cGVHD. The median summary score was 23 (range, 8 to 51), and the median time to complete the scale was 2 minutes, 7 seconds (range, 1 minute, 8 seconds to 4 minutes). Symptoms of cGVHD were well captured on the Lee cGVHD Symptom Scale, although 4 additional symptoms/signs were mentioned by 15% of the participants. Participants mostly reported that item wording was clear and provided accurate definitions of specific terminologies; however, 7 participants (35%) reported finding 1 or more items in the skin domain unclear, reporting, for example, that rashes and itchy skin seemed synonymous. Two of 19 participants (10.5%) described how their answers would have changed had they been asked about their symptoms within the past month instead of within the past week, owing to recently resolved symptoms. All participants were able to accurately explain the concept of “bother” in their own words and distinguish it from symptom severity or other related symptom attributes. In summary, participants found the Lee GVHD Symptom Scale to be a comprehensive and understandable way to report their cGVHD symptom experience. Future work will focus on options for the recall period, the phrasing of skin items, and whether some very rare symptoms (eg, feeding tube, use of oxygen) should continue to be a part of the scale.

© 2016 American Society for Blood and Marrow Transplantation.

INTRODUCTION
Chronic graft-versus-host disease (cGVHD) is a major complication of allogeneic hematopoietic stem cell transplantation (HCT) that is associated with decreased quality of life, impaired functional status, continued need for immunosuppressive medication, and increased mortality [1,2].

A recent National Institutes of Health (NIH) Consensus Conference proposed various tools for standardizing diagnosis, scoring, histopathology, biomarker assays, response assessment, and conduct of clinical trials; patient-reported measures are included in these recommendations [3,4]. In 2002, the development and validation of the Lee cGVHD Symptom Scale to measure symptoms in outpatient age > 18 years with cGVHD was reported [5]. This instrument, recommended by the 2005 and 2014 NIH Consensus Conferences [4,6], is now commonly used to evaluate symptoms in clinical practice and in trials of new therapies for cGVHD. The scale contains 30 items in 7 subscales (skin, eye, mouth, lung, nutrition, energy, and psychological). Patients report

Financial disclosure: See Acknowledgments on page 757.
* Correspondence and reprint requests: Stephanie J. Lee, MD, MPH, Clinical Research Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, DS-200, Seattle, WA 98109.
E-mail address: sjlee@fredhutch.org (S.J. Lee).
their level of symptom “bother” over the previous month on a 5-point Likert scale: not at all, slightly, moderately, quite a bit, or extremely (Appendix 1). Subscale scores and the summary score range from 0 to 100, with a higher score indicating worse symptoms. A clinically meaningful difference of 6 to 7 points has been suggested for the summary score [1].

Given the significant impact of cGVHD on patient symptoms after HCT, it is crucial that new treatments seeking Food and Drug Administration (FDA) approval be evaluated in light of their ability to improve the common symptoms of cGVHD as well as overall patient function [7]. The FDA has not yet qualified a patient-reported outcome (PRO) measure to assess whether potential treatments for cGVHD improve patient symptoms, although helping patients “live better” is one of the criteria for FDA approval, along with “living longer.”

For the FDA to qualify the Lee cGVHD Symptom Scale as a drug development tool, it must perform a rigorous evaluation of whether the Scale is a reliable and valid measure of clinical benefit in a particular context of use [8]. To date, the FDA has qualified only one PRO instrument, a tool recently qualified for measuring exacerbations in chronic obstructive pulmonary disease [9]. Qualifying an instrument means that sufficient quantitative and qualitative evidence concerning the measurement properties has been presented to the FDA such that the measure can be accepted to support a labeling claim [8,10].

Qualitative evidence of content validity is an important component of the evidence dossier for FDA qualification. Evidence in support of the content validity of a measure of cGVHD symptoms should include empiric evidence that the measure is relevant and comprehensible to patients with cGVHD. Although patient input informed the original development of the Lee cGVHD Symptom Scale, this development occurred between 1998 and 2000 [5]. The aim of the present study was to reassess the instrument’s content validity using one-on-one cognitive interviews with 20 patients living with cGVHD.

METHODS
Participants and Data Collection
Between June and August 2015, 2 of the authors (S.J.L. and E.C.M.) conducted one-on-one cognitive interviews with 20 HCT survivors with cGVHD who were attending the Long-Term Follow-Up (LTFU) Clinic at the Seattle Cancer Care Alliance. Inclusion criteria included adults with active cGVHD who could communicate in English. Patients who were eligible and gave written consent participated in a 20- to 30-minute semistructured interview that was audio-recorded.

The scripted, cognitive interview was based on a framework developed by the International Society for Pharmacoeconomics and Outcomes Research evaluating existing PRO instruments and their measures [11-13]. Interview questions and follow-up probes were based on the principles of cognitive interviewing articulated by Willis [14]. Interview questions explored whether the Lee cGVHD Symptom Scale accurately reflected the patient’s experience of cGVHD symptoms and the impact of cGVHD on his or her life, as well as the comprehensibility, relevance, and ease of use of the response choices and recall period (Appendix 2). Recruitment continued until saturation was reached (ie, further interviews did not produce any new relevant themes or categories).

Patients were first asked to describe current and past cGVHD symptoms, to provide researchers with an unbiased list of possible symptoms against which to compare the items in the symptom scale. The participant then completed the Lee cGVHD Symptom Scale during the interview, and was asked to identify any items that were confusing or challenging to answer. Because we were interested in exploring the effects of a shorter recall period, the standard recall period of 1 month was modified in the questionnaire instructions from “the past month” to “the past week.” The interviewers (E.C.M. and S.J.L.) probed the relevance and clarity of the 30 items through open-ended questions. Each respondent was also interviewed in detail about the clarity and relevance of a subset of 5 items. Items were assigned to each interview to ensure that all scale items would receive a detailed evaluation by at least 3 respondents. The semistructured interview questions also addressed comprehensibility, relevance, recall period, and understanding of the concept of “symptom bother.” Interviewers supplemented the cognitive interview script by asking follow-up questions to further probe participant responses. Participants then completed 2 questions to assess their self-perceived overall cGVHD severity, along with a brief demographic questionnaire. The entire process took 15 to 25 minutes.

Following enrollment, participants’ charts were reviewed to collect details about previous treatments and current cGVHD organ involvement using the 2014 NIH consensus criteria for diagnosis and scoring [15].

This study was approved by the Institutional Review Board of the Fred Hutchinson Cancer Research Center. Each participant provided signed written consent before the interview and verbally confirmed consent for the interview to be audio-recorded. The participant was given a $20 gift card after completion of the interview.

Data Analysis
Inductive thematic analysis was performed by 2 of the authors (E.C.M. and S.J.L.) to develop and iteratively modify a codebook. Interviews were transcribed verbatim, and transcripts were then coded line-by-line to facilitate analysis and identification of themes. In this report, illustrative quotes are provided to supplement narrative descriptions, with the participant identification number noted in parentheses.

Scores were calculated following the developer’s instructions (Appendix 3). Specifically, the subscale scores (skin, eye, mouth, lung, nutrition, energy, and psychological status) were calculated if ≥50% of items in the subscale were completed. Note that the instrument is formatted for ease of completion, but the order does not exactly match the subscales, which were determined by factor analysis during development. The summary score is the mean of the subscales when ≥50% of the subscales were completed. The theoretical range for each subscale and the summary score was 0 to 100, with higher scores indicating greater symptom bother. Additional quantitative data were drawn from interview transcripts, Lee cGVHD Symptom Scale scores, and chart abstractions. Quantitative analysis was performed using the CORR procedure in SAS version 9.4 (SAS Institute, Cary, NC). Spearman correlation coefficients were calculated between the Lee cGVHD Symptom Scale scores and the cGVHD organ severity scores, derived through chart abstraction and using the 2014 NIH consensus criteria.

RESULTS
Participant Characteristics
A total of 20 patients (11 males [55%]; median age, 58 years; range, 31 to 79 years) were enrolled between June and August 2015. Three participants (15%) were racial or ethnic minorities. Among the participants who self-identified as a racial or ethnic minority, 1 identified as black, 1 identified as Asian, and 1 selected more than 1 race. Sixteen participants (80%) had a college or postgraduate degree, and 16 (80%) were married or living with a spouse or partner. Six participants (30%) were working full time; 3 (15%), part time. Six participants (30%) were retired, and 3 (15%) were disabled and unable to work.

All participants had received a peripheral blood stem cell transplant. Eighteen participants (90%) had a history of acute GVHD, and 18 (90%) had an established diagnosis of cGVHD before the clinic visit at which they were interviewed; 2 participants’ initial diagnosis of cGVHD was confirmed at the clinic visit and before the interview took place. The median duration of cGVHD was 33 months (range, 0 to 134.4), and the global severity of cGVHD using 2014 NIH consensus scoring was mild in 1 participant, moderate in 10 participants, and severe in 9 participants. Participants had received a median of 5.5 treatments (range, 0 to 14) for their cGVHD.

The demographic, clinical, and cGVHD characteristics of the participants are summarized in Tables 1 and 2.

Lee cGVHD Symptom Scale
The duration of the interview was a median of 14 minutes and 48 seconds (range, 6 minutes, 38 seconds to 24 minutes, 18 seconds). The median time to complete the 30 items of the
of each subscale score among symptomatic patients. Endorsing any subscale symptoms and the median and range, the 30 items of the symptom scale could be completed with minimal burden. The median summary score was 23 (range, 6 to 30). Working full time, working part time, in school full time, homemaker, retired, disabled, unable to work, unemployed, not looking for work, other.

Underlying disease, n (%)  
AML  4 (20)  
ALL  5 (25)  
MDS  7 (35)  
HD  1 (5)  
Other  3 (15)

Stage of disease at transplantation, n (%)  
Early  11 (55)  
Intermediate  9 (45)  
Advanced  0

Donor type, n (%)  
Related  9 (45)  
Unrelated  11 (55)  
Matched  15 (75)  
Partially mismatched  5 (25)  
Stem cell source, n (%)  
Peripheral blood stem cells  20 (100)

Characteristics of Table 1 (n = 20)  

Characteristic Value

Male sex, n (%) 11 (55)  
Age, yr, median (range) 58 (31-79)  
Caucasian, n (%) 17 (85)  
Non-white, n (%) 3 (15)  
Married or living with partner, n (%) 16 (80)  
College or postgraduate degree, n (%) 16 (80)  
Employment status, n (%)  
Working full time 6 (30)  
Working part time 2 (10)  
In school full time 1 (5)  
Homemaker 4 (20)  
Retired 6 (30)  
Disabled, unable to work 3 (15)  
Unemployed, not looking for work 1 (5)  
Other 2 (10)

Lee cGVHD Symptom Scale was strongly correlated (r = 0.65; P = .002) with their NIH calculated overall cGVHD severity, making energy the most predictive domain for objective overall cGVHD severity. The Lee cGVHD Symptom Scale domain score with the strongest correlation to the corresponding NIH severity score for that domain was the mouth (r = 0.63; P = .003), and 100% of participants endorsed symptoms in this domain. The eye domain also had a moderately strong correlation between the Lee cGVHD Symptom Scale Score and the NIH domain score (r = 0.52; P = .019), and 18 participants (90%) endorsed eye symptoms.

Inclusiveness of Symptom Scale Items

Before completing the Lee cGVHD Symptom Scale, participants were asked to describe their current and past GVHD symptoms. This information provided an unbiased list of possible symptoms against which to compare the items in the symptom scale. The cGVHD symptoms that participants reported spontaneously were well captured on the symptom scale, with mouth and eye symptoms the most discussed. However, edema/swelling, vaginal, liver, and fingernail symptoms (items not on the scale) were mentioned as symptoms by 3 participants (15%) each. In the interviews, several participants highlighted that vaginal symptoms are often left out of the conversation about cGVHD, and expressed a desire to see vaginal symptoms better addressed:

“I’ve had vaginal GVH since early on and it’s just...like I know the team would talk about there was somebody else to deal with girl parts, and I’m a nurse, and I feel like that needs to be talked about more openly. Like it’s just kind of left out, so unless I bring it up, it gets skipped over. And so I’m okay with bringing it up, but I feel like a lot of people aren’t, and so having stuff going on in your sexual or intimate life can be a really big deal, and it just might mean that people aren’t as lucky as I am to have people to talk to about it” (11).

Participants noted that although elevated liver function test results due to cGVHD were not included in the symptom scale, these abnormalities did not have a significant impact on their quality of life, because they caused them minimal to no bother.

“Liver function is the one that jumps out at me. Um, you know it was never annoying but something that showed up in the labs that they felt the need to address with medication. I mean, it was another pill, which I didn’t love, but that was essentially a nonissue” (7).

Clarity and Relevance of Symptom Scale Items

Skin domain

During the course of the interview or immediately after completing the symptom scale, seven participants (35%) indicated that they found 1 or more items in the skin domain unclear. For example, 2 participants (10%) believed that rashes and itchy skin were synonymous. Five participants (25%) spoke of difficulties in assigning their skin symptoms to only 1 of the available categories of cGVHD skin characteristics, specifically abnormal skin color, rashes, thickened skin, sores on skin, or itchy skin.

Several patterns emerged when comparing the Lee cGVHD Symptom Scale domain scores with the NIH consensus scoring system based on chart abstraction for the same domains, participants’ scoring on the energy domain on the Lee cGVHD Symptom Scale was strongly correlated (r = 0.65; P = .002) with their NIH calculated overall cGVHD severity, making energy the most predictive domain for objective overall cGVHD severity. The Lee cGVHD Symptom Scale domain score with the strongest correlation to the corresponding NIH severity score for that domain was the mouth (r = 0.63; P = .003), and 100% of participants endorsed symptoms in this domain. The eye domain also had a moderately strong correlation between the Lee cGVHD Symptom Scale Score and the NIH domain score (r = 0.52; P = .019), and 18 participants (90%) endorsed eye symptoms.

Inclusiveness of Symptom Scale Items

Before completing the Lee cGVHD Symptom Scale, participants were asked to describe their current and past GVHD symptoms. This information provided an unbiased list of possible symptoms against which to compare the items in the symptom scale. The cGVHD symptoms that participants reported spontaneously were well captured on the symptom scale, with mouth and eye symptoms the most discussed. However, edema/swelling, vaginal, liver, and fingernail symptoms (items not on the scale) were mentioned as symptoms by 3 participants (15%) each. In the interviews, several participants highlighted that vaginal symptoms are often left out of the conversation about cGVHD, and expressed a desire to see vaginal symptoms better addressed:

“I’ve had vaginal GVH since early on and it’s just...like I know the team would talk about there was somebody else to deal with girl parts, and I’m a nurse, and I feel like that needs to be talked about more openly. Like it’s just kind of left out, so unless I bring it up, it gets skipped over. And so I’m okay with bringing it up, but I feel like a lot of people aren’t, and so having stuff going on in your sexual or intimate life can be a really big deal, and it just might mean that people aren’t as lucky as I am to have people to talk to about it” (11).

Participants noted that although elevated liver function test results due to cGVHD were not included in the symptom scale, these abnormalities did not have a significant impact on their quality of life, because they caused them minimal to no bother.

“Liver function is the one that jumps out at me. Um, you know it was never annoying but something that showed up in the labs that they felt the need to address with medication. I mean, it was another pill, which I didn’t love, but that was essentially a nonissue” (7).

Clarity and Relevance of Symptom Scale Items

Skin domain

During the course of the interview or immediately after completing the symptom scale, seven participants (35%) indicated that they found 1 or more items in the skin domain unclear. For example, 2 participants (10%) believed that rashes and itchy skin were synonymous. Five participants (25%) spoke of difficulties in assigning their skin symptoms to only 1 of the available categories of cGVHD skin characteristics, specifically abnormal skin color, rashes, thickened skin, sores on skin, or itchy skin.
“I don’t really know where my leg lesions fall in. Maybe they fall under one of these; maybe it’s here [gestures to skin section of survey]. They’re not really sores, though… maybe abnormal skin color” (8).

“I have a couple things on my skin that aren’t really sores. You know, they were…kind of like little spots, and now they’ve turned into sort of scaly little lumps, so they kind of don’t fit into any of these, but they’re there” (11).

“If I were to look at this, I kind of equate this with this [points to rashes and itchy skin]. Yeah, so there’s some bleeding over there. But no, I guess you could have a rash that wasn’t itchy. I don’t know” (15).

Need to use oxygen and receiving nutrition from an intravenous line or feeding tube
In addition, 2 (10%) participants mentioned they found the terms “need to use oxygen” and “receiving nutrition from an intravenous line or feeding tube” not meaningful, citing their extreme nature.

“Receiving nutrition from an intravenous line or feeding tube. I don’t think that’s that; that’s kind of way out there. It seems like it’s not part of—I mean, these are things that…just seems like a real—like you’re in the hospital doing that” (1).

“I don’t know about the need to use oxygen. That seems really extreme” (4).

Psychological domain
Three participants (15%) noted that answering the mental and emotional questions regarding symptoms of depression and anxiety was more challenging for them because those symptoms might have non-GVHD causes or because of difficulty pondering or evaluating these particular emotions. All participants understood what was meant by the questions, however.

“I think that mental and emotional stuff is harder to measure, harder to evaluate” (6).

“So I’m saying no to these 3 things [points to depression, anxiety, and difficulty sleeping] as a consequence of GVHD. I have—there are other reasons I might be having them, okay?” (18).

“I just don’t like the mental and emotional. Sometimes I don’t like to answer those…it’s perfectly worded” (20).

Overall clarity and relevance
A majority of the participants reported that item phrasing, including the response choices, was clear for all items, and they were able to give accurate explanations of the symptom terms on which they had received in-depth instruction. In the interviews, several participants noted the comprehensiveness and relevance of the items included in the scale and recognized that the items on the scale were common cGVHD symptoms even if they had never personally experienced those symptoms.

“I’m not experiencing problems with breathing, but just because I’m not doesn’t mean someone else isn’t” (5).

“Well, some of them don’t apply to me. Loss of sleep—I don’t have loss of sleep, so I just marked zero on them, but I take it they’re all common symptoms of GVHD. I just don’t have all of them” (13).

Recall Period
Seventeen of 19 participants (89.5%) said that their answers would not change if asked about their symptoms within the past month instead of within the past week.
Participants comprehended that the recall period “within the past week” was included in the recall of the past month.

“It probably wouldn’t change, because the past week is included in the past month, isn’t it? So, I guess no, probably not” (4).

Two of 19 participants (10.5%) said that their answers would change if asked about their symptoms within the past month instead of within the past week. When asked to clarify why their answers would change and what specific symptoms would change, both noted recently resolved symptoms.

“Well, I can mainly think of the cough, because that was—dealing with the cough a month ago. Yeah, that would be the main one” (18).

**Bother**

Participants did not report any confusion about the concept of bother, defining it within 4 general categories: an irritating or annoyance, an interruption, a deviation from the norm, or a discomfort.

“I’m still struggling to find normal—normalcy, what that means to me now, and just think I’m on that path to kind of a routine and something happens that I don’t expect or can’t predict. That’s what I would consider bothersome. Usually I can anticipate that things are going to happen...I kind of deal with it a little differently, but when it’s unpredictable, it happens and it’s just kind of a surprise, it’s a bit of a bother—a little irritating.” (5)

“It’s not anything that’s life-threatening, but it’s just annoying” (7).

“Just that you notice it to a point where you feel like it’s kind of an annoyance or its abnormal or, you know, not what you were you used to prior to...the transplant” (8).

“It just interferes with my daily life a lot...And then the other thing is just related to GVHD and immune suppression—feeling like I can’t do, you know, a lot of things I’d normally do” (11).

“if it makes me uncomfortable...That’s what it means to me, if it makes me uncomfortable” (3).

**Overall Impressions of the Scale**

All participants (100%) described a favorable overall impression of the scale, citing its brevity, clarity, and ease of completion. A majority of respondents explained that the scale, response choices, and directions for completion were already as clear as possible and were unable to offer any suggestions for change.

“It felt really reasonable compared to a lot of the surveys I fill out. Pretty quick” (7).

“It was good. It was short” (4).

“I really don’t know of any way you could make it easier” (3).

“A piece of cake...it’s very easy” (20).

“The directions are very clear, and the selection is in terms of assessing how you feel about each item is pretty spot on” (5).

Three participants (15%) noted that they particularly liked the verbal descriptors for the response options and the 5-point Likert scale, highlighting that more than 5 points would have been too challenging and that qualitative descriptors are preferable to numbers. The exception to this was 1 participant who was interested in knowing how the investigators wanted him to map his symptoms to the qualitative scale.

“Well, I think this scale is not bad because it spells out things rather than putting them on a scale of 1 to 10 or whatever. I find those not very helpful, because the numbers are relative. How would you know what the end of the scale is if you haven’t experienced it, so you don’t know really where you are on the scale. But if you say not at all or slightly or moderately, I think that’s easier to answer those questions” (13).

“I think the 4, well 5 columns actually, are really good. I wouldn’t do any more...This seems to break it down, so I particularly like that, and I thought the categories were good” (16).

“I think there’s going to be a level of subjectivity in these categories, just because I think some people by nature underestimate the impact of things and some people overestimate maybe... I don’t know how you’d do this—but I could see just a sentence or two describing each one of these. You know, if it was more—here’s the engineer in me coming out—but if it was more quantitative, you know, if moderately meant, you know, you think about it 3 times a day and extremely meant that it has a significant impact on your daily activity or something like that. That kind of a definition might be in making sure that I was aligned with how these terms are generally thought of” (7).

Participants also appreciated being asked to provide their own perspectives of the cGVHD symptom experience.

“The subtext for all of this stuff is, well, you’re still alive, but when you go into this, you don’t know how it’s going to be afterward, and it’s a pretty big deal” (11).

When asked if there was anything that could be done to make the Lee GVHD Symptom Scale easier to complete, 2 participants (10%) suggested offering the symptom scale online, and 2 participants (10%) suggested adding a section for participants to write in additional free text if they wanted to qualify any of their responses or indicate any unsolicited symptoms.

**DISCUSSION**

The participants’ responses to the Lee cGVHD Symptom Scale were overwhelmingly positive, highlighting the scale’s clarity, relevance, and acceptability. Participants expressed interest in and gratitude for being asked to provide their own assessment of their cGVHD symptoms and how those symptoms impact their lives.

Participants were consistent in their understanding of bother when prompted for a definition before seeing the Lee cGVHD Symptom Scale. This suggests a universal, culturally shared understanding of this concept, at least among English speakers. Participants also reported that it was a reasonable measure against which to consider the impact of cGVHD symptoms on their day-to-day lives. However, whether symptoms should be scaled for intensity, frequency,
interference, bother, or all of these dimensions, remains a topic of debate in the literature [16-19].

The majority of patients clearly understood the recall period ("within the past week") and said that their answers would not change if asked about their symptoms within the past month instead of within the past week. Although the scale was originally validated with a 1-month recall period, the participants' responses support the acceptability of the shorter recall period.

When the Lee cGVHD Symptom Scale was first developed and validated in the late 1990s, severe manifestations of cGVHD were more common. Two of the symptom scale items, "need to use oxygen" and "receiving nutrition from an intravenous line or feeding tube," reflect that reality. As transplantation regimens and GVHD prophylaxis and treatment continue to evolve, these manifestations are becoming increasingly rare. Two participants noted that these symptoms seemed to be extreme compared with the remainder of the scale, and it is worth evaluating whether or not these symptoms are sufficiently common so that they should continue on the scale moving forward. Although our sample was small, none of our study participants endorsed these problems. Using larger populations, we plan to examine the presence of floor effects for these items and consider the potential impact of removing them from the scale.

Participants generally understood the meaning of the item phrasing in the skin domain; however, several participants found it difficult to report their own skin symptoms using the available items, explaining that their skin symptoms could have been described using various terminologies. It is interesting to note that when prompted to define the items in the skin domain, all participants were able to do so without hesitation. Options that could be explored include providing definitions (eg, "rashes, including bumps, scaling, roughness or other changes in skin texture/feel") or pictures of skin manifestations or expanding the skin domain to include additional specific or nonspecific cutaneous symptoms. Another possibility is to incorporate free text fields into the Lee cGVHD Symptom Scale as single items, thus inviting respondents to list any cGVHD symptoms they feel are not adequately captured by the scale and grade their associated degree of bother. These free text descriptions, using their own words, would then be presented each time they completed the symptom scale, although this could complicate scoring and interpretation of the summary score.

Although all cGVHD symptoms were generally well represented on the Lee cGVHD Symptom Scale, edema/swelling, and vaginal, liver, and fingernail symptoms were each mentioned by 3 participants as cGVHD symptoms that did not appear among the scale items. The symptom of edema and swelling was mentioned by 3 different participants, who described 3 different phenomena that may or may not have been related to cGVHD. This inconsistency in definition requires further exploration before being considered for inclusion in the symptom scale.

Several women mentioned vulvovaginal symptoms without prompting, and several more endorsed vaginal symptoms when asked. In addition, the assessment of vaginal cGVHD symptoms is now included in the NIH consensus scoring and evaluation as an exploratory measure, supporting the addition of such symptoms to the Lee cGVHD Symptom Scale. More qualitative work is needed to assess whether women are able to discern vaginal symptoms from cGVHD as opposed to vaginal symptoms from menopause and hormonal changes, considering that some manifestations may overlap. Vulvovaginal symptoms caused by infections or estrogen deficiency would confound the assessment of cGVHD treatment. No men spontaneously mentioned genital symptoms.

Elevated liver function test results are a common complication of cGVHD. Several participants noted that they had had elevated liver function test values, but participants acknowledged that this did not bother them. Therefore, liver abnormalities are not a good candidate for addition to the Lee symptom scale. Similarly, fingernail symptoms were mentioned by several participants, but were not described as bothersome.

Limitations of this study include the small cohort and a sample drawn from a relatively homogenous population of outpatients at a single center. We enrolled only 20 participants because we reached saturation about interview topics. Racial/ethnic minorities were underrepresented, and participants were very well educated. As with all PRO measures, our instrument was designed for use in patients who are capable of meaningfully self-reporting. The scale was not intended for cognitively impaired patients, children age <12 years, or proxy reporting of symptoms. In-trial guidance has been published to address situations in which study participants have a PRO endpoint due for collection and have cognitive impairment [20].

In summary, the results of this study support the content validity of the Lee cGVHD Symptom Scale, a PRO measure of cGVHD symptom bother. Study participants believed that the scale captured almost all of their symptoms and was generally clear. They also appreciated the brevity, and offered few suggestions for improvement. Future work will focus on additional evaluation of construct validity and responsiveness to change. Although this testing was previously done during development, it has been almost 15 years since this initial work was completed. Evaluation in larger samples and using modern measurement theory will provide additional evidence of the measurement properties of this instrument and support FDA qualification. Because cGVHD has prominent effects on symptom burden and quality of life, and because symptom improvement is part of the criteria for defining therapeutic response, it is crucial that we have available valid, reliable, and responsive measure of cGVHD bother for use in trials of new therapies for cGVHD.

ACKNOWLEDGMENTS

The authors thank Nancy Kline Leidy, PhD, RN, for her contributions to the design of the study, and Tracey Godlewski, PA-C, Carina Moravec, ARNP, Kristen Kindsvogel, ARNP, and Mary Flowers, MD, for their help with enrollment.

Financial disclosure: Funding for this study was provided by the American Cancer Society (ECM) and CA118953 and CA163438 (SJL). The Chronic GVHD Consortium U54 CA163438 is part of the NCATS Rare Diseases Clinical Research Network (RDCRN). RDCRN is an initiative of the Office of Rare Diseases Research (ORDR), NCATS, funded through a collaboration between NCATS and the National Cancer Institute.

Conflict of interest statement: The authors have no conflicts of interest to report.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.bbmt.2015.12.026.
REFERENCES