Biol Blood Marrow Transplant 22 (2016) 752–758

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

ASBMT American Society for Blood and Marrow Transplantation

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

Emily C. Merkel [1], Sandra A. Mitchell [2], Stephanie J. Lee [3,*]

[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.

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 outpatients 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, D5-290, Seattle, WA 98109.
E-mail address: sjlee@fredhutch.org (S.J. Lee).

http://dx.doi.org/10.1016/j.bbmt.2015.12.026
1083-8791/

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 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 in-person 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 $\geq$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 $\geq$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