VALUE IN HEALTH 21 (2018) 1115–1123

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Patient-Reported Outcomes

Evaluating the Content Validity of Four Performance Outcome Measures in Patients with Elective Hip Replacements and Hip Fractures

Rachel Ballinger, PhD[1,*], *Cicely Kerr, PhD*[2], *Fiona Mowbray, PhD*[2], *Elizabeth Nicole Bush, MHS*[3]

[1]*ICON Clinical Research, Abingdon, Oxfordshire, UK;* [2]*Formerly ICON Clinical Research, Oxford, Oxfordshire, UK;* [3]*Patient-Focused Outcomes Center of Expertise, Eli Lilly and Co., Indianapolis, IN, USA*

ABSTRACT

Objectives: To assess the content validity of performance outcome (PerfO) measures for use with patients undergoing hip fracture (HF) surgery and elective total hip replacement (eTHR). **Methods:** This study was a substudy of a broader evaluation of measurement properties of PerfO measures. The PerfO measures assessed were timed up and go (TUG), four-step stair climb (4SC), long stair climb (LSC), and repeated chair stand (RCS). For this substudy, HF and eTHR participants were interviewed to evaluate the relevance and difficulty of each PerfO measure. Qualitative analysis was conducted on interview transcripts, and summaries of coded data were produced to assess saturation. **Results:** All 18 HF participants related the PerfO measures (TUG, 4SC, and RSC) to activities they completed in daily life, with slight variations in some specific aspects. For the eight eTHR participants, the correspondence between the PerfO measures (TUG, 4SC, and LSC) and activities in daily life varied: all participants saw similarity in the movements for the TUG; most undertook short stair climbs in daily life, but most did not regularly undertake LSC in daily life. Nevertheless, all HF and eTHR participants reported that the PerfO measures were relevant and had a level of difficulty similar to daily activities. **Conclusions:** This study contributes novel methods that adapt US regulatory guidance for patient-reported outcome measures to the evaluation of PerfO measures. A structured approach was used to explore specific details of each measure and correspondence to everyday life. This study demonstrates how content validity of PerfO measures can be meaningfully assessed.

Keywords: content validation, hip fracture, hip replacement, performance outcomes.

Introduction

Performance outcome (PerfO) measures are a type of clinical outcome assessment (COA) outlined by the US Food and Drug Administration (FDA). A PerfO measure is based on tasks performed by a patient according to instructions; it is administered by a health care professional but requires patient cooperation and motivation [1]. PerfO measures can provide specific information about functional status and mitigate variance introduced by perception of functional ability [2]. Specifically in clinical trial evaluation of orthopedic treatment, PerfO measures have been used to assess functioning, including timed up and go, stair climb, chair stand, fast-paced walk, and 6-minute walk tests [3,4].

As with all types of COAs used in clinical trials of medical products, PerfO measures should reflect the health experiences of patients in terms of how they feel or function in everyday life [1,5]. Nevertheless, although some PerfO measures assess abilities and actions that closely simulate how a patient functions in typical life, others assess concepts of interest for which the connection to everyday activities is less clear, such as supine quadriceps isometric strength [6]. In the regulatory context, the degree of correspondence between the COA measurement concept and how patients feel or function in everyday life is considered a key element of content validity. The FDA has provided guidance for the assessment of content validity of patient-reported outcome (PRO) measures and defined *content validity* as the extent to which the PRO instrument measures the concept of interest [5]. Furthermore, qualitative evidence is required to demonstrate that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use [5], and it must be based on direct input from an adequate sample of patients from the targeted clinical study population [7]. Such specific guidance in relation to content validation of PerfO measures is not available. In addition, PRO measures are designed to directly capture patient experience, for example,

* *Address correspondence to:* Rachel Ballinger, ICON Clinical Research UK Ltd., 100 Park Drive, Milton Park, Abingdon, Oxon OX14 4RY, UK.

E-mail: rachel.ballinger@iconplc.com

1098-3015$36.00 – see front matter https://doi.org/10.1016/j.jval.2018.02.005

how patients feel or function, and hence PRO measurement concepts link closely to meaningfulness to patients. For other types of COAs, the meaningfulness of the measurement concept to patients' everyday life may need to be considered differently or separately.

On the basis of literature review and expert clinical opinion, four PerfO measures were selected to assess performance in three study populations as part of the main evaluation study (reported elsewhere [8]). This substudy specifically assessed the content validity of four PerfO measures: timed up and go (TUG), four-step stair climb (4SC), long stair climb (LSC), and repeated chair stand (RCS). Assessment of content validity was based on FDA's 2009 industry guidance titled "Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims" (henceforth, PRO guidance) [5] to the extent that it could be applied to PerfO measures. In addition, specific feedback was received that indicated that the FDA was interested in knowing the relevance of the measures and how the measures' level of difficulty related to everyday functioning.

Methods

Study Design

This content validation study was a qualitative substudy of a main evaluation study, with a longitudinal design assessing the measurement properties of the same PerfO measures (Fig. 1). The main and substudy protocols were approved by a central institutional review board and all participants provided written informed consent before enrolling. The main PerfO measure evaluation study was conducted at 15 clinical sites in the United States and evaluated select PerfO measures in participants who underwent hip fracture (HF) surgery, elective total hip replacement (eTHR), or elective total knee replacement (eTKR). During each of the three main study visits, PerfO measures were administered by trained health care professionals and included TUG, 4SC, LSC, and RCS (described in Table 1). As a predictor of future falls [10], the RCS was undertaken only with the HF group. The LSC was used only with the eTHR and eTKR groups.

For the content validation substudy, HF and eTHR participants in the main study were invited to complete a telephone interview after one of their scheduled study visits; eTKR participants had completed all three visits at the time the substudy was initiated and so were not included. Interviews were conducted between November 2013 and May 2014. At that time, HF participant recruitment for the main study was ongoing, allowing inclusion of HF participants in the substudy who were at different stages postsurgery; nevertheless, the remaining eTHR participants were all attending their final main study visit 12 weeks (±3 days) postsurgery (Fig. 1).

Structured interview guides for each group were developed. These included a recap of the relevant PerfO measures and instructions given to help focus participants' recall. Questions related to overall experience and specific details of each measure (such as rising to standing and turning), before exploring the relevance of the measures to everyday activities and functioning.

Participants

Inclusion criteria for the content validation substudy were participation in the main evaluation study (eligibility criteria are presented in Table 2) and availability for a telephone interview, ideally within 5 days of a main study visit. With participant permission, selected main study data were made available to the interviewer for reference during the interview (e.g., if a test had not been fully completed).

Interviews

Telephone interviews were conducted by experienced interviewers using the relevant structured interview guide. Interviews were scheduled to last less than 45 minutes and were audio-recorded and transcribed verbatim. Participants were reimbursed for their time.

Analysis

Analysis codes were identified both from the interview guide and from themes that emerged directly from the data. A codebook for each patient group was developed after review of data from the first four interviews in each group; these were reviewed against transcripts by additional members of the study team. The codebooks comprised code names, definitions, and examples to help ensure consistency of coding across interviews. These documents were modified as needed during the coding of subsequent interviews (e.g., to reflect newly identified themes/codes). Qualitative analysis MAXQDA software (Sozialforschung GmbH, Berlin, Germany) was used to code data.

Sample size was limited by the availability of participants remaining on active follow-up in the main study at the time of the substudy data collection, particularly the eTHR group. Nevertheless, data saturation on core themes was thoroughly assessed in both groups to inform level of confidence in the results and conclusions.

Given the structured interview guides with focused exploration of correspondence between specific test movements and everyday life, the standard approach to assessing data saturation in concept elicitation studies [12] was not appropriate for this qualitative study. Instead, data summary grids were completed for each participant group and reviewed for saturation: interview content was summarized by participant and by PerfO measure for each of three core themes, drawing on the content of groups of analysis codes, namely, overall relevance, overall speed (relates to both relevance and difficulty because instructions varied between the PerfO measures assessed: normal walking speed or as fast as safely able), and overall level of difficulty. The content of new details identified from the final eTHR and the last three HF interviews was reviewed to assess the value of any new details identified at that point, to inform consideration of whether additional interviews would yield important additional information. This approach of summarizing new content rather than just indicating application of a new code is similar to that proposed by Brod et al. [13]. A team-based approach was used to develop and check the accuracy of the summary grid content details (as described in Fig. 1).

Results

Participant and Interview Characteristics

The study sample comprised 18 HF participants recruited from three sites (24% of the 75 HF participants at baseline in the main evaluation study) and 8 eTHR participants from five sites (9.5% of the 84 eTHR participants at visit 3 in the main evaluation study). All interviews were conducted within 7 days after the main study visit at which the participants had completed the PerfO measures (mean days after visit: HF, 3; eTHR, 4). HF participants were interviewed 79 to 177 days after surgery and across the three visits of the main evaluation study (following visit 1, n = 4; visit 2, n = 8; visit 3, n = 6). The mean length of the interviews was 36 minutes (range 18–50 minutes).

Participant characteristics are presented in Table 3, including comparison with the main evaluation study sample. This shows that participants were broadly reflective of the main evaluation