

*Open*

# Development of a Symptom-Focused Patient-Reported Outcome Measure for Functional Dyspepsia: The Functional Dyspepsia Symptom Diary (FDSD)

Fiona Taylor, MBiochem<sup>1</sup>, Sophie Higgins, MPH<sup>1</sup>, Robyn T. Carson, MPH<sup>2</sup>, Sonya Eremenco, MA<sup>3</sup>, Catherine Foley, MPH, MA<sup>1</sup>, Brian E. Lacy, MD, PhD<sup>4</sup>, Henry P. Parkman, MD<sup>5</sup>, David S. Reasner, PhD<sup>6</sup>, Alan L. Shields, PhD<sup>1</sup>, Jan Tack, MD, PhD<sup>7</sup> and Nicholas J. Talley, MD, PhD<sup>8</sup> on behalf of the Patient-Reported Outcome Consortium's Functional Dyspepsia Working Group

- OBJECTIVES:** The Functional Dyspepsia Symptom Diary (FDSD) was developed to address the lack of symptom-focused, patient-reported outcome (PRO) measures designed for use in functional dyspepsia (FD) patients and meeting Food and Drug Administration recommendations for PRO instrument development.
- METHODS:** Concept elicitation interviews were conducted with FD participants to identify symptoms important and relevant to FD patients. A preliminary version of the FDSD was constructed, then completed by FD participants on an electronic device in cognitive interviews to evaluate the readability, comprehensibility, relevance, and comprehensiveness of the FDSD, and to preliminarily evaluate its measurement properties.
- RESULTS:** During concept elicitation interviews, 45 participants spontaneously reported 19 symptom concepts. Of those, seven symptoms were selected for assessment by the eight-item FDSD. Cognitive interviews with 57 participants confirmed that participants were able to comprehend and provide meaningful responses to the FDSD, and that the handheld electronic FDSD format was suitable for use in the target population. Scores of the FDSD were well-distributed among response options, item discrimination indices suggested that the FDSD items differentiate among patients with varying degrees of FD severity, and inter-item correlations suggested that no items of the FDSD were capturing redundant information. Internal consistency estimates (0.87) and construct-related validity estimates using known-groups methods were within acceptable ranges.
- CONCLUSIONS:** The FDSD is a content-valid PRO measure, with preliminary psychometric evidence providing support for the FDSD's items and total score. Further psychometric evaluations are recommended to more fully test the FDSD's score performance and other measurement properties in the target patient population.

**SUPPLEMENTARY MATERIAL** is linked to the online version of the paper at <http://www.nature.com/ajg>

*Am J Gastroenterol* 2018; 113:39–48; doi:10.1038/ajg.2017.265; published online 19 September 2017

## INTRODUCTION

Functional dyspepsia (FD) is a common functional gastrointestinal disorder characterized by heterogeneous symptoms thought to originate in the gastroduodenal region, including postprandial fullness, early satiety, and epigastric pain and burning (1). FD is

further subdivided into two, potentially co-existing, diagnostic categories as follows: (i) postprandial distress syndrome (PDS), characterized by postprandial fullness and early satiation, and (ii) epigastric pain syndrome (EPS), characterized by epigastric pain and burning (2). Upon routine diagnostic investigation, FD

<sup>1</sup>Adelphi Values, Patient-Centered Outcomes, Boston, Massachusetts, USA; <sup>2</sup>Allergan Plc., Jersey City, New Jersey, USA; <sup>3</sup>Critical Path Institute, Tucson, Arizona, USA; <sup>4</sup>Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA; <sup>5</sup>Temple University School of Medicine, Philadelphia, Pennsylvania, USA; <sup>6</sup>Ironwood Pharmaceuticals, Inc., Cambridge, Massachusetts, USA; <sup>7</sup>University of Leuven, Leuven, Belgium; <sup>8</sup>University of Newcastle, Callaghan, New South Wales, Australia. **Correspondence:** Fiona Taylor, MBiochem, Adelphi Values, Patient-Centered Outcomes, 290 Congress Street, 7th Floor, Boston, Massachusetts 02210, USA. E-mail: [fiona.taylor@adelphivalues.com](mailto:fiona.taylor@adelphivalues.com)

Received 24 April 2017; accepted 11 July 2017

symptoms appear to lack a structural or metabolic cause to readily explain their presence, making diagnosis, treatment, and the establishment of treatment outcomes difficult (1). Although the symptoms of FD overlap with other gastrointestinal disorders, such as gastroesophageal reflux disease, irritable bowel syndrome, and gastroparesis (1,2), FD is considered a distinct disorder with considerable impact on the quality of life of those diagnosed with it (3).

Patient assessment is critical in FD, because, lacking a clear organic origin, it is considered a symptom-defined disorder. Although patient-reported outcome (PRO) questionnaires for gastrointestinal disorders exist, including for FD (e.g., Dyspepsia Symptom Severity Index (4) and Nepean Dyspepsia Index (5)), until recently it had been unclear to what extent the development of these existing questionnaires was consistent with the US Food and Drug Administration's (FDA) Guidance for Industry—Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (6) (hereafter referred to as “FDA PRO Guidance”) and, therefore, to what extent the questionnaires were suitable for use in regulated clinical trials to evaluate new product treatment claims. Broadly, the FDA PRO Guidance explains that PRO measure development should be informed by rigorous and well-documented qualitative research, in order to ensure that the tool assesses concepts that are (i) relevant to the disease or condition, (ii) important to individuals with the disease or condition, and (iii) understandable to respondents (that is, the patient). In particular, the FDA PRO Guidance assigns a premium to direct patient input toward the development of PRO measures intended for use to support product approval and labeling.

In recognition of the importance of scientifically defensible PRO measures for use in clinical trials, the Critical Path Institute's (C-Path) PRO Consortium (7), through its FD Working Group and in conjunction with advisors from the FDA, aimed to make publicly available an FD symptom-focused PRO measure that could be used to support primary endpoints in regulated FD clinical trials and to submit the measure for qualification under the FDA's Drug Development Tool Qualification Program (8). As an initial step, the FD Working Group documented the primary symptoms of FD from the literature and evaluated the extent to which existing questionnaires target those symptoms and were defensible for use in regulated clinical trials to assess treatment efficacy claims intended for product labeling (9). In this study, a total of 56 articles and 16 instruments assessing FD symptoms were reviewed. Concepts listed in the Rome III criteria for FD ( $n=7$ ), those assessed by existing FD instruments ( $n=34$ ), and symptoms reported by patients in published qualitative research ( $n=6$ ) were summarized in an FD conceptual model (reproduced in **Figure 1**). Of note, each of the symptoms described in the published qualitative research reports was also specified in the Rome III criteria, with the exception of vomiting.

With respect to the 16 instruments found in the literature review, three (the Dyspepsia Symptom Severity Index (4), Nepean Dyspepsia Index (5), and Short-Form Nepean Dyspepsia Index (10)) assessed all seven FD symptoms listed in the Rome III criteria (i.e., early satiation, epigastric burning, postprandial

fullness, postprandial nausea, excessive belching, epigastric pain, and upper abdomen bloating). Despite their strong FD symptom coverage (and evidence of patient involvement in development), their potential qualification for use to substantiate product labeling goals was questionable in light of concerns regarding regulatory expectations around specified recall periods and response options (9).

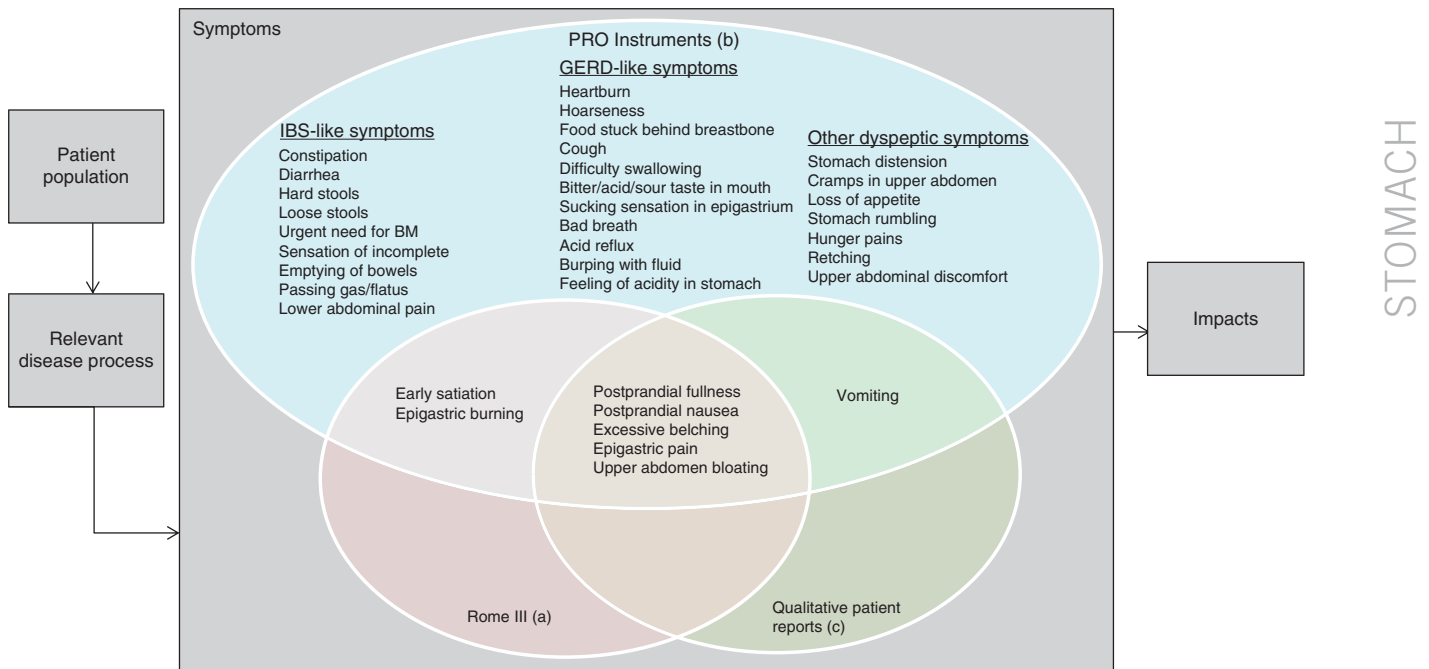
Given the conclusion that none of the existing PRO measures assessing all seven FD symptoms listed in the Rome III criteria adhered to the regulatory principles necessary to support product labeling, the FD Working Group initiated the development of a novel tool, the Functional Dyspepsia Symptom Diary (FDSD). The purpose of this study is to describe the research and evidentiary basis that contributed to and supported the development of the FDSD. The specific development activities included the following: (i) identification and documentation of FD symptoms from the patient perspective; (ii) selection of the FD symptom concepts to be targeted for assessment by the newly created measure as well as the creation of the measure itself; (iii) evaluation of the content of the FDSD among participants with FD; and (iv) preliminary evaluation of the measurement properties of the items and scores produced by the FDSD when completed by participants with FD.

## METHODS

### Concept elicitation interviews with participants

A total of 45 face-to-face, 60 min concept elicitation interviews were conducted to identify and document the symptoms of FD from the perspective of individuals with the condition. Study participants were recruited from four sites in the United States, based on the study's inclusion and exclusion criteria (**Supplementary Table 1**), and interviewed between March and October 2014. Broadly, participants were adults  $\geq 18$  years old, who met the Rome III diagnostic criteria (2) for FD (diagnosis and subtype categorization were determined by the recruiting clinician) without other gastrointestinal disorders (e.g., active constipation, irritable bowel syndrome, or gastroesophageal reflux disease).

Approval to execute the study was received from Copernicus Group Independent Review Board on 1 November 2013 and from Mary Hitchcock Memorial Hospital on 14 March 2014. All participants provided written informed consent before participation and, once enrolled, completed a Demographic Health and Information Form. A clinician-completed Case Report Form was also used to collect clinical data. During the interviews, interviewers followed a semi-structured interview guide to elicit symptom concepts from participants. When addressing a symptom, the interviewer probed using follow-up questions in order to collect data on dimensions of the symptom, including duration, frequency, and severity. All interviews were audio-recorded (with participant consent), anonymized, and imported into ATLAS.ti, a computerized qualitative data analysis package (Berlin, Germany), to facilitate content analysis. To evaluate saturation (i.e., the point at which no new or relevant information is gained from additional interviews), concept emergence was documented across sets of successive interviews, according to established methods (11–13). Unique



**Figure 1.** Literature-based conceptual model for functional dyspepsia symptoms (9). BM, bowel movement; GERD, gastroesophageal reflux disease; IBS, irritable bowel syndrome; PRO, patient-reported outcome. (a) Concepts in Rome III criteria for FD; (b) concepts assessed by existing FD instruments; (c) concepts reported by patients in published qualitative research. This figure was originally published in Taylor *et al.* (9).

concepts (e.g., bloating) were tabulated for frequency and mean bothersome ratings (on a scale of 0 (No bother) to 10 (Most bothersome)) were calculated for each symptom. In addition, separate from the bothersome rating exercise, participants were asked to rank the symptoms in which they would most like to see improvement with treatment.

### Concept selection and preliminary FDS construction

The FD Working Group held a 1-day meeting with an expert panel to identify and select target measurement concepts for inclusion in the preliminary FDS. Together with the FDA PRO Guidance (6), the following criteria were considered when selecting potential concepts: concept's frequency of report by participants, bothersome rating (i.e., participants' level of subjective bother associated with the symptom), inclusion in the Rome III Diagnostic Criteria (2), documentation in the empirical literature, and applicability to all participants, regardless of FD subtype. Following concept selection, the preliminary FDS was constructed in the context of defining the following: the context of use, mode of data collection, recall period, instructions, items, and response options.

Upon creation of the preliminary measure, a translatability assessment of the FDS was conducted by having linguistic validation experts review the tool, comment on any text or concepts that may present difficulty in future translation efforts, and provide solutions to mitigate concerns. In this regard, the following languages were included in the translatability assessment: German (Germany), Italian (Italy), Russian (Russia), Hindi (India), Japanese (Japan), French (France), Spanish (Mexico), Arabic (Egypt), Chinese (China), and Korean (Korea). In addition, an electronic

implementation assessment was conducted by the Electronic Patient-Reported Outcome (ePRO) Consortium's Instrument Migration Subcommittee to determine the suitability of the preliminary FDS for data collection on a handheld ePRO device.

### Cognitive interviews

Following the FDS's construction and its implementation on a handheld ePRO device (LG Nexus 5 smartphone, programmed by Biomedical Systems (Maryland Heights, MO), and hereafter referred to as "device"), a total of 57 face-to-face, 60 min cognitive interviews were conducted with participants with FD in two waves, to collect qualitative data on the readability, comprehensibility, relevance, and comprehensiveness of the FDS. Eight participants in the initial wave completed paper-based screenshots of the electronic FDS to test preliminary wording before device programming. A second wave of 49 participants completed the FDS on the device. All participants first completed the FDS without interruption and then were probed in a structured manner to evaluate the readability, comprehensibility, and relevance of the FDS. Participants in the second wave were also questioned on the usability of the FDS on the device. All interviews were conducted in the United States between June 2015 and September 2016; due to minimal revisions to the FDS between waves, data collected in both waves were pooled for the qualitative and quantitative analyses reported here.

To facilitate recruitment of a sample that reflects the real-world FD population (14), those with active irritable bowel syndrome or chronic constipation were eligible to participate in the cognitive interviews. The inclusion and exclusion criteria were amended

and protocol amendments were approved by the aforementioned IRBs before the start of recruitment. Participant clinical and sociodemographic data were collected using the Demographic Health and Information Form and Case Report Form. Content analysis compared participants' interpretations of the instructions, items, and response options to the developer definitions that were drafted following the initial development of the FDS. In addition, the second wave participants' responses to questions regarding their overall opinion on the usability of the FDS on the device were analyzed.

#### Preliminary evaluation of the FDS's measurement properties

Data collected from participants' initial completion of the FDS (i.e., the numeric responses to the FDS provided by participants) were analyzed to gain a preliminary understanding of the performance of the items and proposed scale (i.e., total symptom) score produced by the FDS. Specifically, using SAS version 9.4 (SAS Institute, Cary, NC), the following properties were evaluated: missingness, score distributions, floor and ceiling effects (defined as  $\geq 25.0\%$  of participants selecting the response that reflects the worst or best possible state, respectively), item discrimination, inter-item correlations, internal consistency reliability, and construct-related validity using known-groups methods. The cross-sectional study design, while considered sufficient for the stated research goals, precluded the ability to generate results related to other indicators of psychometric performance, including test-retest reliability and sensitivity to change.

## RESULTS

### Concept elicitation interviews with participants

A total of 45 interviews were conducted (Table 1) and participants spontaneously reported 19 symptom concepts, 95.7% ( $n=18$ ) of which were elicited in the first 75% of interviews. Within each FD subtype, a similar downward trend in the elicitation of new concepts was observed, with no new relevant concepts emerging in the last 25% of interviews, thus providing evidence that conceptual saturation was reached. The 19 identified concepts are listed in Figure 2 for the full sample and in a graph (Figure 3) by FD subtype. Across all 19 symptoms, mean participant-reported bothersome ratings ranged from 4.0 to 9.0, whereas mean bothersome ratings for symptoms rated by at least two-thirds of participants each (i.e., bloating, early satiety, stomach pain, and nausea) ranged from 6.3 to 7.5. The two symptoms ranked by participants as most important to improve if an effective treatment were available were bloating and stomach pain. Based on these results, the following seven symptom concepts were identified for inclusion in the preliminary measure: stomach pain, upper abdominal burning, nausea, bloating, postprandial fullness, early satiety, and burping/belching.

### Development of the preliminary FDS

The preliminary FDS was an eight-item measure, intended for daily administration on a handheld ePRO device, to assess symptoms of FD in the context of a clinical trial. Items 1 through 6

assessed the severity of stomach pain, upper abdominal burning, nausea, bloating, postprandial fullness, and early satiety, respectively. Items 7 and 8 assessed burping/belching in terms of the level of bother and severity, respectively. A diagram showing the location of the stomach was included at the beginning of the FDS to instruct respondents to think only about symptoms in this area. In addition, respondents were asked to reflect over the past 24 h while responding to the FDS and responses were scored on an 11-point numeric rating scale from 0 (no (concept)) to 10 (worst imaginable (concept)). A 24 h recall period was deemed appropriate, as the FDS concepts of measurement can be variable both between days and within a day. The selection of an 11-point numeric rating scale is consistent with suggestions that the scale has relative advantages in minimizing missing data, patient preference, ease of recording, and ease of implementation in clinical trials (15).

FDS item-level scores and a Total Symptom Score (TSS) were calculated. The TSS comprised Items 1, 2, 4, 5, and 6 of the FDS; Items 3, 7, and 8 were considered supplementary items (symptoms relevant to 68.9–73.3% of participants but not considered cardinal symptoms of FD by the expert panel) and were not included in the TSS. The FDS TSS ranged from 0 to 50, with higher scores indicating greater symptom burden. Although some minor revisions were suggested to the FDS following the translatability and electronic implementation assessments, it was decided that no revisions to the FDS would be made before the cognitive interviews. Developer's definitions were agreed upon for each item to help ensure conceptual equivalence when translating the FDS into other languages.

### Cognitive interviews with participants

A total of 57 interviews were conducted with participants with FD (Table 1). Although individuals with irritable bowel syndrome or chronic constipation were eligible for participation in the cognitive interviews, very few participants with these comorbidities were recruited ( $\leq 5$  for each).

Participants who provided an interpretable response interpreted the FDS instructions (Part 1: 94.5%,  $n=52/55$ ; Part 2: 98.2%,  $n=55/56$ ), diagram (96.4%,  $n=54/56$ ), response anchors ( $\geq 92.0\%$  for each item), and recall period of the past 24 h (94.4%,  $n=51/54$ ) as intended. The majority of participants in the first wave (62.5%,  $n=5/8$ ) reported a preference for the recall period at the beginning of each item (i.e., "over the past 24 h..."), and all items were revised to this format for the second wave.

Overall, participants were able to read, understand, and provide meaningful responses to all eight items of the FDS. Specifically, for Item 1 and Items 3–8, at least 81.8% of participants interpreted the item as intended. Interpretation issues included attribution of the concept to an incorrect location or item interpretations that did not align with developers' definitions. For example, when interpreting Item 2 (burning in the stomach), all but 1 of the 11 participants (21.6%) who did not interpret the item as intended were incorrectly thinking of either heartburn or burning in the throat/esophagus or chest. In addition, Item 4 (bloating) and Item 5 (stomach fullness) were not interpreted as intended by 14.3 and

**Table 1. Clinical and sociodemographic characteristics of participants**

	Concept elicitation interviews (N=45)		Cognitive interviews (N=57)	
<b>Age</b>				
Mean, s.d. (range)	46.2, 13.0 (21.0–73.5)		42.6, 14.7 (22.1–69.8)	
<b>Sex (n (%))</b>				
Female	32 (71.1%)		45 (78.9%)	
Male	13 (28.9%)		12 (21.1%)	
<b>Hispanic or Latino ethnicity (n (%))</b>				
Yes	12 (26.7%)		13 (22.8%)	
No	33 (73.3%)		44 (77.2%)	
<b>Race (n (%))</b>				
White	31 (68.9%)		45 (78.9%)	
Asian	2 (4.4%)		0 (0.0%)	
Black or African American	1 (2.2%)		7 (12.3%)	
Native Hawaiian or other Pacific Islander	1 (2.2%)		0 (0.0%)	
Other <sup>a</sup>	6 (13.3%)		4 (7.0%)	
Not answered	4 (8.9%)		1 (1.8%)	
<b>Highest level of education (n (%))</b>				
Some college or certificate program	11 (24.4%)		22 (38.6%)	
College or university degree (2 or 4 year)	18 (40.0%)		19 (33.3%)	
High school diploma (or GED) or less	12 (26.7%)		10 (17.5%)	
Graduate degree	4 (8.9%)		5 (8.8%)	
Currently in college	0 (0.0%)		1 (1.8%)	
<b>Clinician-reported FD subtype (n (%))<sup>b</sup></b>				
EPS	14 (31.1%)		20 (35.1%)	
PDS	14 (31.1%)		20 (35.1%)	
Co-existing EPS and PDS	17 (37.8%)		17 (29.8%)	
<b>FD severity (n (%))<sup>c</sup></b>				
	<i>Clinician-reported</i>	<i>Participant-reported</i>	<i>Clinician-reported</i>	<i>Participant-reported</i>
Mild	11 (24.4%)	7 (15.6%)	17 (29.8%)	10 (17.5%)
Moderate	25 (55.6%)	26 (57.8%)	20 (35.1%)	34 (59.6%)
Severe	9 (20.0%)	12 (26.7%)	20 (35.1%)	13 (22.8%)

EPS, epigastric pain syndrome; PDS, postprandial distress syndrome.

<sup>a</sup>For concept elicitation interviews, the following responses were marked as “Other” by participants: Hispanic ( $n=6$ , 13.3%). For cognitive interviews, the following responses were marked as “Other” by participants: Hispanic ( $n=2$ , 3.5%), Mexican/Puerto Rican ( $n=1$ , 1.8%), and Spanish ( $n=1$ , 1.8%).

<sup>b</sup>As determined by the Rome III Diagnostic Criteria(2) at time of participant screening.

<sup>c</sup>Determined by participants’ responses on the Demographic Health and Information Form and clinicians’ responses on the Case Report Form.

14.5% of participants, respectively ( $n=8/56$  and  $n=8/55$ ). Participants who misinterpreted these items were most commonly thinking about bloating as a sensation of being full of food (without mention of air/gas), and stomach fullness as the feeling of satisfaction or contentment following completion of a meal (rather than an uncomfortable fullness).

Overall, all items of the FDSD were relevant to the target population, with participants reporting that they were currently experiencing or had experienced the symptoms assessed by the measure. For Items 1–7, at least 90.7% of participants reported experiencing

the concept being evaluated by the FDSD, either within or before the 24 h recall period. For Item 8, most participants had experienced being bothered by burping/belching at some time (84.2%,  $n=48/57$ ), but 15.8% of participants ( $n=9/57$ ) reported never being bothered by burping/belching due to FD.

All participants who completed the FDSD on the device reported that it was easy to read the items on the screen of the device (100%,  $n=48/48$ ) and had an overall positive opinion of using the device to complete the FDSD (100%,  $n=49/49$ ). Median time for FDSD completion was 1 min 35 s (range=44 s to 6 min 43 s). Additional results

regarding the usability assessment of the device are provided as **Supplementary Material (Supplementary Table 2)**.

**Preliminary evaluation of measurement properties**

There were no missing data recorded for the FDSM, with 100.0% (N=57) of participants providing data for all items. Overall, the responses to items were well distributed among the response

options for the FDSM, indicating that participants were using all levels of the ordinal response scale (see **Supplementary Table 3** for item distribution table). A ceiling effect ( $\geq 25.0\%$ ) was observed for Item 8 (burping/belching bother) and no items demonstrated a floor effect (**Table 2**). Inter-item correlations indicated that no items were capturing redundant information (Pearson's correlation  $r < 0.80$ , in all instances) (16,17) (**Table 3**). The FDSM TSS yielded Cronbach's  $\alpha = 0.87$ , above the *a priori*-identified threshold ( $\alpha \geq 0.70$ ) for acceptable internal consistency reliability (**Table 2**), and remained above threshold following removal of each of the items composing the TSS (Items 1, 2, 4, 5, and 6). Item discrimination index analyses suggest that the FDSM items composing the TSS are able to discriminate among participants with mild/moderate FD and participants with severe FD to varying degrees (**Table 2**). As hypothesized, the FDSM TSS demonstrated an increasing monotonic trend across known severity groups for both participant-reported and clinician-reported FD severity (that is, TSS increased with increasing severity of FD). For participant-reported FD severity, Items 1 (stomach pain), 2 (burning in the stomach), 7 (burping/belching rating), and 8 (burping/belching bother) demonstrated an increasing monotonic trend. For clinician-reported FD severity, all items with the exception of Items 4 (bloating), 7 (burping/belching rating), and 8 (burping/belching bother) demonstrated an increasing monotonic trend across known severity groups (**Table 2**).

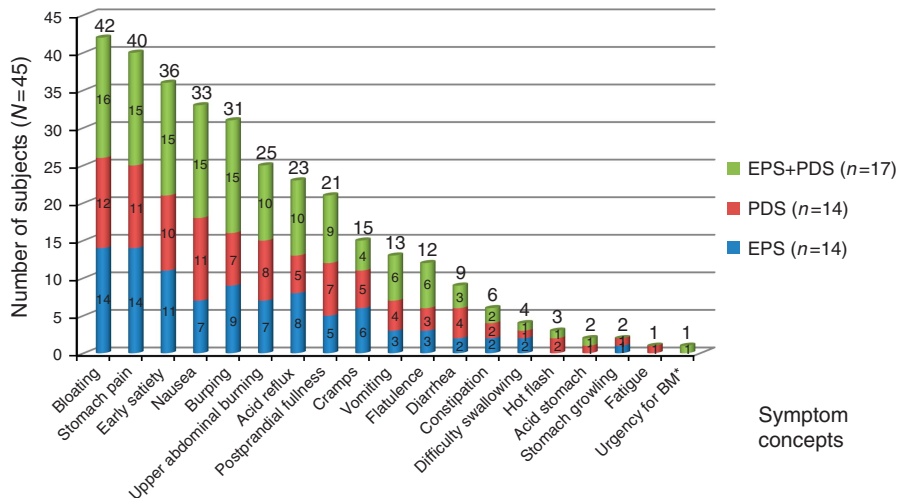
Signs and Symptoms (n=x, %)\*

- Rome III diagnostic or supportive symptoms†
  - Bloating (n=42, 93.3%)
  - Stomach pain (n=40, 88.9%)
  - Early satiety (n=36, 80.0%)
  - Nausea (n=33, 73.3%)
  - Burping (n=31, 68.9%)
  - Upper abdominal burning (n=25, 55.6%)
  - Postprandial fullness (n=21, 46.7%)
- IBS-like lower gastrointestinal symptoms
  - Flatulence (n=12, 26.7%)
  - Diarrhea (n=9, 20.0%)
  - Constipation (n=6, 13.3%)
  - Urgency for bowel movement (n=1, 2.2%)
- GERD-like symptoms
  - Acid reflux (n=23, 51.1%)
  - Difficulty swallowing (n=4, 8.9%)
  - Acid stomach (n=2, 4.4%)
- Other dyspeptic symptoms
  - Cramps (n=15, 33.3%)
  - Vomiting (n=13, 28.9%)
  - Stomach growling (n=2, 4.4%)
- Other symptoms
  - Hot flash (n=3, 6.7%)
  - Fatigue (n=1, 2.2%)

**Figure 2.** List of functional dyspepsia signs and symptoms. \*Percentages represent frequency of participant (N=45) report during concept elicitation interviews † (2).

**Revised FDSM and conceptual framework**

The number of items in the FDSM remained unchanged following analysis of the cognitive interview data; however, revisions were made to item wording and ordering. Following the initial wave of eight cognitive interviews, the recall period for each item was moved from the end of the item to the beginning of the item, based on participant preference. In addition, the order of Item 7 (burping/belching) and Item 8 (burping/belching bother) was switched. Following the second wave of 49 cognitive interviews,



**Figure 3.** Participant-reported functional dyspepsia concepts by subtype. \*BM, bowel movement; EPS, epigastric pain-syndrome; PDS, postprandial distress syndrome.

**Table 2.** FDSD acceptability, item discrimination, reliability, and known-groups analysis results (N=57)

FDSD item <sup>a</sup>	Ceiling <sup>b</sup> n (%)	Floor <sup>b</sup> n (%)	Item discrimination index <sup>c</sup>	Cronbach's $\alpha$ (TSS; or on item removal)	Known groups (participant-reported severity) <sup>d</sup> Mean (s.d.)	Known groups (clinician-reported severity) <sup>e</sup> Mean (s.d.)
TSS (0–50)	—	—	—	0.87	Mild: 23.4 (9.86) Moderate: 24.6 (10.57) Severe: 30.8 (14.90)	Mild: 20.1 (9.54) Moderate: 27.4 (11.97) Severe: 29.3 (11.79)
1. Stomach pain (0–10)	3 (5.3%)	3 (5.3%)	0.40	0.84	Mild: 4.1 (2.56) Moderate: 4.5 (2.31) Severe: 6.5 (2.99)	Mild: 3.8 (2.28) Moderate: 4.9 (2.76) Severe: 5.9 (2.51)
2. Burning in the stomach (0–10)	7 (12.3%)	1 (1.8%)	0.38	0.89	Mild: 2.9 (2.69) Moderate: 4.6 (2.09) Severe: 5.8 (3.41)	Mild: 3.3 (2.44) Moderate: 4.9 (1.87) Severe: 5.4 (3.20)
3. Nausea (0–10)	13 (22.8%)	2 (3.5%)	0.08	—	Mild: 3.3 (2.16) Moderate: 4.1 (3.31) Severe: 4.1 (3.38)	Mild: 3.3 (3.12) Moderate: 4.1 (3.19) Severe: 4.4 (3.13)
4. Bloating (0–10)	5 (8.8%)	6 (10.5%)	0.22	0.84	Mild: 5.4 (2.59) Moderate: 5.2 (2.77) Severe: 6.0 (3.87)	Mild: 4.4 (2.47) Moderate: 6.2 (2.73) Severe: 5.6 (3.47)
5. Stomach fullness (0–10)	3 (5.3%)	4 (7.0%)	0.21	0.84	Mild: 6.0 (2.31) Moderate: 5.6 (2.81) Severe: 6.5 (3.31)	Mild: 4.9 (2.18) Moderate: 5.9 (3.23) Severe: 6.6 (2.78)
6. Early satiety (0–10)	8 (14.0%)	3 (5.3%)	0.25	0.83	Mild: 5.0 (2.75) Moderate: 4.8 (3.02) Severe: 6.0 (3.98)	Mild: 3.8 (2.61) Moderate: 5.6 (3.24) Severe: 5.8 (3.41)
7. Burping/belching rating (0–10)	4 (7.0%)	4 (7.0%)	0.23	—	Mild: 3.9 (2.47) Moderate: 4.7 (2.49) Severe: 5.3 (3.82)	Mild: 3.3 (2.44) Moderate: 5.8 (2.36) Severe: 4.9 (3.15)
8. Burping/belching bother (0–10)	15 (26.3%)	3 (5.3%)	0.13	—	Mild: 3.4 (2.80) Moderate: 4.1 (3.17) Severe: 4.7 (3.97)	Mild: 2.8 (2.79) Moderate: 4.8 (3.19) Severe: 4.7 (3.56)

CRF, Case Report Form; DHIF, Demographic Health and Information Form; FD, functional dyspepsia; FDSD, Functional Dyspepsia Symptom Diary; TSS, Total Symptom Score.

<sup>a</sup>FDSD items are scored on an 11-point numeric rating scale from 0 (no (concept)) to 10 (worst imaginable (concept)).

<sup>b</sup>Floor and ceiling effects are represented by the percent of participants responding to the worst possible state/best possible state; in this instance, floor effect refers to a high percentage ( $\geq 25.0\%$ ) of participants selecting the score reflecting a state which cannot get any worse (i.e., 10, worst imaginable (concept)) and ceiling effect refers to a high percentage ( $\geq 25.0\%$ ) of participants selecting the score reflecting a state, which cannot get any better (i.e., 0, no (concept))(20–22).

<sup>c</sup>Item discrimination index=proportion endorsed for severe group—proportion endorsed for mild/moderate group. The item discrimination index assesses the extent to which item responses accurately capture genuine patient experiences and are represented on a scale from 1 to  $-1$ , where negative or zero indices characterize poorly performing items and positive indices characterize well-performing items. Indices were evaluated using the following criterion:  $\leq 0.20$ =poor performance, 0.21 to 0.29=moderate performance, 0.30 to 0.39=good performance,  $\geq 0.40$ =excellent performance(23).

<sup>d</sup>Participant-reported FD severity based on response in DHIF: mild  $n=10$ ; moderate  $n=34$ ; severe  $n=13$ .

<sup>e</sup>Clinician-reported FD severity based on response in CRF: mild  $n=17$ ; moderate  $n=20$ ; severe  $n=20$ .

modifications were made to the instructions and four items. The FDSD instructions were revised to better define the location of FD symptoms, and Item 1 (stomach pain) and Item 2 (burning in the stomach) were reversed, enabling stomach burning to be assessed directly following the diagram. Additional clarification was also added to Item 4 (bloating) and Item 5 (stomach fullness). The revised conceptual framework is shown in **Table 4**. No changes to the provisional scoring of the FDSD were made on the basis of the quantitative analyses.

## DISCUSSION

The FDSD is a novel, content-valid PRO measure that is being developed according to FDA guidance recommendations. Results

from the concept elicitation interviews ( $N=45$ ) suggest that, although participants experience a number of FD-related symptoms, core symptoms of the condition are similar across both FD subtypes and FD severity levels. Using these results, in conjunction with findings from a review of the published literature and expert clinician input, the FD Working Group constructed a preliminary measure: the eight-item FDSD for implementation on a handheld ePRO device. Translatability assessment confirmed that the FDSD's instructions and items should be minimally problematic when future translation and linguistic validation activities are undertaken.

Overall, participants in cognitive interviews were able to read, understand, and provide meaningful responses to all eight items of the preliminary FDSD. Minor changes to the measure were made

**Table 3. Inter-item Pearson's correlations for the FDS items (N=57)**

FDS item*	FDS item number							
	1	2	3	4	5	6	7	8
1. Stomach pain	1.00	—	—	—	—	—	—	—
2. Burning in the stomach	0.66	1.00	—	—	—	—	—	—
3. Nausea	0.48	0.40	1.00	—	—	—	—	—
4. Bloating	0.60	0.40	0.49	1.00	—	—	—	—
5. Stomach fullness	0.61	0.34	0.62	0.69	1.00	—	—	—
6. Early satiety	0.60	0.38	0.59	0.74	0.77	1.00	—	—
7. Burping/belching rating	0.43	0.41	0.55	0.52	0.35	0.38	1.00	—
8. Burping/belching bother	0.30	0.34	0.43	0.45	0.21	0.36	0.78	1.00

FDS, Functional Dyspepsia Symptom Diary.  
 \*FDS items are scored on an 11-point numeric rating scale from 0 (no (concept)) to 10 (worst imaginable (concept)). Pearson's correlation coefficients  $r > 0.80$  indicate items are capturing potentially redundant information (24).

**Table 4. Conceptual framework of the FDS TSS<sup>a</sup>**

Domain	Concept	FDS Item
FD symptom severity (TSS)	→ Burning in the stomach	→ Item 1
	→ Stomach pain	→ Item 2
	→ Bloating	→ Item 4
	→ Postprandial fullness	→ Item 5
	→ Early satiety	→ Item 6

FD, functional dyspepsia; FDS, Functional Dyspepsia Symptom Diary; TSS, Total Symptom Score.  
<sup>a</sup>Item 3 (nausea), Item 7 (burping/belching rating), and Item 8 (burping/belching bother) are included in the FDS; however, because they are considered supplementary assessments and are not anticipated to be included in the TSS or used in trial endpoints, they are not included in the conceptual framework (they will instead be scored as individual items).

following both the first and second wave of cognitive interviews. Revisions were implemented to clarify the concept measured by each item and to promote improved understanding for respondents. Although these modifications were intended to improve the items in an evidence-based manner (i.e., revising based on the most frequent misinterpretation), there may be a residual number of participants who still fail to interpret the item as intended; however, further revision to the items was considered a risk to the understandability of the items by the majority of patients. With respect to the implementation of the FDS on the device, participants with FD had, overall, a favorable opinion of using the device, demonstrating its usability within the target patient population.

The preliminary evaluation of the FDS's measurement properties provides support for the performance of the FDS's items and TSS in the assessment of FD symptoms. Responses were well distributed across the 11-point scale and, although a ceiling effect was observed for Item 8 (burping/belching bother) (26.3%), this is considered a supplementary item of the FDS. It is possible that participants in the cognitive interview study experienced relatively

low levels of this concept as part of their symptom experience and this is not indicative of an issue with the performance of the scores produced by the FDS. As this item is not used to calculate the FDS TSS, the observed ceiling effect does not have a bearing on the performance of the overall symptom score.

Item discrimination index analyses generated using participant-reported FD severity level groups suggest that the FDS is able to discriminate among FD patients on the basis of self-reported severity. Inter-item correlations indicate that all items of the FDS correlate positively with one another and no item pairs surpassed the *a priori* threshold ( $r > 0.80$ ) for redundancy, suggesting that, although items of the measure are capturing related information, no items are redundant. Overall internal consistency, as calculated using Cronbach's  $\alpha$  ( $\alpha = 0.87$ ), surpassed the threshold and remained high and in the acceptable range following removal of each of the items composing the TSS (i.e., Items 1, 2, 4, 5, and 6). The known-groups approach to assessing construct validity showed that the FDS is able to distinguish between pre-defined groups based on participant- and clinician-reported levels of FD severity.

One potential limitation to the qualitative research presented was the discrepancy between participant-reported symptoms and the categorization of participants into FD subtypes (EPS, PDS, and co-existing EPS and PDS) by recruiting clinicians. During both concept elicitation and cognitive interviews, participants categorized as EPS nonetheless reported experiencing PDS-specific symptoms (e.g., postprandial fullness and early satiety), and vice versa. Another measure, the Leuven Postprandial Distress Scale (18,19), was recently developed based on results from focus groups and cognitive interviews in a PDS population (as identified by Rome III criteria). Similar to the FDS, the Leuven Postprandial Distress Scale assesses bloating, postprandial fullness, and early satiety, and has been found to be content valid for both PDS and co-existing EPS and PDS subtypes in a controlled treatment trial, although these items have not been tested in a pure EPS population. Future research may focus on FD subtype identification in the PRO measure development space.



Another limitation to the research presented is the relatively small sample size used for the preliminary evaluation of the FDSD's measurement properties. Although the sample sizes used in the concept elicitation and cognitive interview activities ( $N=45$  and  $N=57$ , respectively) are considered sufficient for qualitative analyses, a larger sample would have provided a more robust assessment of the FDSD's measurement properties. Further, it was not possible to evaluate certain psychometric properties, based on the study design of the cognitive interviews. As the FDSD was completed at only one time point, test-retest reliability could not be evaluated and because it was not completed in the context of a clinical trial, sensitivity to change could not be assessed. Convergent and divergent validity evaluations were not conducted owing to a lack of a comparison measure in the study with which to correlate the FDSD.

Future psychometric evaluations are recommended to further evaluate the score performance of the FDSD in the target patient population, including test-retest reliability and sensitivity to change, as well as score interpretation and responder definitions within the context of a treatment outcome trial. The FDSD will be submitted under the FDA's Drug Development Tool Qualification Program and the measure will be made publicly available for use to support primary endpoints in future, regulated FD clinical trials. This support will enable a pathway forward and eliminate barriers to drug development programs in a disease area characterized by unmet need.

## ACKNOWLEDGMENTS

This project is a pre-competitive collaboration that includes pharmaceutical company and Critical Path Institute scientists, academic researchers/clinicians, FDA advisors, and Adelphi Values. We gratefully acknowledge members of FDA's Qualification Review Team for their valuable feedback and advice during the development of the FDSD.

## CONFLICT OF INTEREST

**Guarantor of the article:** Fiona Taylor, MBiochem.

**Specific author contributions:** The authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE), were fully responsible for all aspects of manuscript development, and approved this manuscript for submission. Fiona Taylor and Alan L. Shields participated in planning and executing the study, interpreting the data, and drafting the manuscript. Catherine Foley participated in planning and executing the concept elicitation stage of the study; collecting, analyzing, and interpreting the data; and provided input on the manuscript. Sophie Higgins participated in planning and executing the cognitive interview stage of the study; collecting, analyzing and interpreting the data; and drafting the manuscript. Robyn T. Carson, Sonya Eremenco, and David S. Reasner provided input on the study design, interpretation of results, and manuscript. Brian Lacy, Henry Parkman, Jan Tack, and Nick J. Talley were expert panel members on the study and provided input on the study design, interpretation of results, and manuscript.

**Financial support:** Three PRO Consortium member companies sponsor the Functional Dyspepsia Working Group. These companies

support this effort through financial, in-kind, and intellectual contributions. Funding for this Functional Dyspepsia Working Group research was provided by the following PRO Consortium member firms: Allergan Plc., Ironwood Pharmaceuticals, Inc., and Shire. The Critical Path Institute's PRO Consortium is supported, in part, by grant number U18 FD005320 from the US Food and Drug Administration. A list of the members of the PRO Consortium is available at <http://c-path.org/programs/pro/>.

**Potential competing interests:** F.T., S.H., C.F., and A.L.S. are employees of Adelphi Values, which received payment from the sponsors to conduct the research. B.L., H.P.P., J.T., and N.J.T. received payment from the sponsors to participate as expert panel members on this study. R.T.C. is an employee of Allergan and owns stock and stock options in Allergan. S.E. is an employee of the Critical Path Institute and has no competing interests to report. B.L. serves on scientific advisory boards for Ironwood, Salix, and Prometheus. H.P.P. has no further competing interests to report. D.S.R. is a member of Albemarle Scientific Consulting and an employee of Ironwood Pharmaceuticals, and owns stock and stock options in Ironwood. J.T. has provided scientific advice to Abide Therapeutics, Alfa Wassermann, Allergan, Mylan, Novartis, Rhythm, Shire, SK Life Sciences, Takeda, Theravance, Tsumura, Yuhan, and Zeria Pharmaceuticals; has received research support from Abide, Shire, Tsumura, and Zeria; and has served on the speaker bureau for Abbott, Allergan, Shire, Takeda, and Zeria. J.T. was involved in development of the Leuven Postprandial Distress Scale PRO. N.J.T. has received Grant/Research Support from NHMRC, NIH, Rome Foundation, Aus EE, Abbott Pharmaceuticals, Allergan, Datapharm, Pfizer, Salix, Prometheus Laboratories, Commonwealth Laboratories, and Janssen. Consultant/Advisory Boards: GI Therapies, Yuhan, Prometheus Laboratories, and Commonwealth Laboratories. US Patent Holder: Biomarkers of irritable bowel syndrome. N.J.T. developed the Nepean Dyspepsia Index Long and Short forms. Any views expressed in this publication represent the personal opinions of the authors, not those of their respective employers. The authors' respective organizations were given the opportunity to review the manuscript for medical and scientific accuracy, as well as intellectual property considerations.

## Study Highlights

### WHAT IS CURRENT KNOWLEDGE

- ✓ There is a lack of patient-reported outcome (PRO) measures to assess functional dyspepsia symptoms.
- ✓ Existing measures do not meet Food and Drug Administration's (FDA) guidance recommendations.

### WHAT IS NEW HERE

- ✓ The Functional Dyspepsia Symptom Diary (FDSD) is a newly developed, content-valid, PRO measure meeting FDA guidance recommendations.
- ✓ Preliminary psychometric evidence supports FDSD item performance.
- ✓ The FDSD is easy for patients with functional dyspepsia (FD) to complete on a handheld electronic device.

## REFERENCES

- Tack J, Talley NJ, Camilleri M *et al.* Functional gastroduodenal disorders. *Gastroenterology* 2006;130:1466–79.
- Drossman DA, Corazziari E, Delvaux M *et al.* Appendix A: Rome III Diagnostic Criteria for FGIDs. In: Drossman DA, Corazziari E, Delvaux M *et al.* (eds). *Rome III The Functional Gastrointestinal Disorders*, 3rd edn. Degnon Associates, Inc.: McLean, Virginia, 2006, pp. 885–97.
- Aro P, Talley NJ, Agreus L *et al.* Functional dyspepsia impairs quality of life in the adult population. *Aliment Pharmacol Ther* 2011;33:1215–24.
- Leidy NK, Farup C, Rentz AM *et al.* Patient-based assessment in dyspepsia: development and validation of Dyspepsia Symptom Severity Index (DSSI). *Dig Dis Sci* 2000;45:1172–9.
- Talley NJ, Haque M, Wyeth JW *et al.* Development of a new dyspepsia impact scale: the Nepean Dyspepsia Index. *Aliment Pharmacol Ther* 1999;13:225–35.
- US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*, 2009.
- Coons SJ, Kothari S, Monz BU *et al.* The patient-reported outcome (PRO) consortium: filling measurement gaps for PRO end points to support labeling claims. *Clin Pharmacol Ther* 2011;90:743–8.
- US Food and Drug Administration. *Drug Development Tools (DDT) Qualification Programs*, available at <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/default.htm> 2015. Accessed 8 August 2016.
- Taylor F, Reasner DS, Carson RT *et al.* Development of a symptom-based patient-reported outcome instrument for functional dyspepsia: a preliminary conceptual model and an evaluation of the adequacy of existing instruments. *Patient* 2016;9:409–18.
- Talley NJ, Phillips SF, Melton J III *et al.* A patient questionnaire to identify bowel disease. *Ann Intern Med* 1989;111:671–4.
- Lasch KE, Hassan M, Endicott J *et al.* Development and content validity of a patient reported outcomes measure to assess symptoms of major depressive disorder. *BMC Psychiatry* 2012;12:34.
- Charmaz K. *Grounded theory*. In: Smith JA, Harre R, Van Langenhove L (eds). *Rethinking Methods in Psychology*. Sage: London, 1995, pp. 27–49.
- Glaser B, Strauss AL. The constant comparative method of qualitative analysis. In: Glaser B, Strauss AL (eds). *Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine de Gruyter: New York, 1967, pp. 101–16.
- Suzuki H, Hibi T. Overlap syndrome of functional dyspepsia and irritable bowel syndrome—are both diseases mutually exclusive? *J Neurogastroenterol Motil* 2011;17:360–5.
- Dworkin RH, Turk DC, Farrar JT *et al.* Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113:9–19.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- Stevens SS. Mathematics, measurement, and psychophysics. In: Stevens SS (ed). *Handbook of Experimental Psychology*. Wiley: Oxford, 1951, pp. 1–59.
- Carbone F, Holvoet L, Vandenberghe A *et al.* Functional dyspepsia: outcome of focus groups for the development of a questionnaire for symptom assessment in patients suffering from postprandial distress syndrome (PDS). *Neurogastroenterol Motil* 2014;26:1266–74.
- Carbone F, Vandenberghe A, Holvoet L *et al.* Validation of the Leuven Postprandial Distress Scale, a questionnaire for symptom assessment in the functional dyspepsia/postprandial distress syndrome. *Aliment Pharmacol Ther* 2016;44:989–1001.
- Fries JF, Lingala B, Siemons L *et al.* Extending the floor and the ceiling for assessment of physical function. *Arthritis Rheumatol* 2014;66:1378–87.
- Patrick DL, Erickson P. *Health Status and Health Policy: Allocating Resources to Health Care*. Oxford University Press: Oxford, UK, 1993.
- McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
- Ebel RL, Frisbie DA. *Essentials of Educational Measurement*, 4th edn. Prentice-Hall: Englewood Cliffs, NJ, 1986.
- McHorney CA, Ware JE Jr, Lu JF *et al.* The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32:40–66.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2018