

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

Fiona Taylor¹ · David S. Reasner² · Robyn T. Carson³ · Linda S. Deal⁴ · Catherine Foley¹ · Ramon Iovin¹ · J. Jason Lundy⁵ · Farrah Pompilus¹ · Alan L. Shields¹ · Debra G. Silberg⁶

Published online: 28 March 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Objectives The aim was to document, from the perspective of the empirical literature, the primary symptoms of functional dyspepsia (FD), evaluate the extent to which existing questionnaires target those symptoms, and, finally, identify any missing evidence that would impact the questionnaires' use in regulated clinical trials to assess treatment efficacy claims intended for product labeling.

Methods A literature review was conducted to identify the primary symptoms of FD and existing symptom-based FD patient-reported outcome (PRO) instruments. Following a database search, abstracts were screened and articles were retrieved for review. The primary symptoms of FD were organized into a conceptual model and the PRO instruments were evaluated for conceptual coverage as well as compared against evidentiary requirements presented in the FDA's PRO Guidance for Industry.

Results Fifty-six 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 the FD conceptual model. Except for vomiting, all of the identified symptoms from the published qualitative research reports were also specified in the Rome III criteria. Only three of the 16 instruments, the Dyspepsia Symptom Severity Index (DSSI), Nepean Dyspepsia Index (NDI), and Short-Form Nepean Dyspepsia Index (SF-NDI), measure all seven FD symptoms defined by the Rome III criteria. Among these three, each utilizes a 2-week recall period and 5-point Likert-type scale, and had evidence of patient involvement in development. Despite their coverage, when these instruments were evaluated in light of regulatory expectations, several issues jeopardized their potential qualification for substantiation of a labeling claim.

Conclusions No existing PRO instruments that measured all seven symptoms adhered to the regulatory principles necessary to support product labeling. As such, the development of a new FD symptom PRO instrument is supported.

On behalf of the Patient-Reported Outcome Consortium's Functional Dyspepsia Working Group.

✉ David S. Reasner
DReasner@IronwoodPharma.com

¹ Adelphi Values, Boston, MA, USA

² Study Endpoints, Ironwood Pharmaceuticals, Inc., 301 Binney Street, Cambridge, MA 02142, USA

³ Allergan plc., Jersey City, NJ, USA

⁴ Pfizer, Collegeville, PA, USA

⁵ Outcometrix, Tucson, AZ, USA

⁶ Shire, Zug, Switzerland

Key Points for Decision Makers

There are no functional dyspepsia-specific clinical outcome assessments that have adequate documentation to support their use for US labeling purposes in regulated clinical trials.

There is a need for the development of a new clinical outcome assessment for patients with functional dyspepsia that reflects the preliminary conceptual model proposed by the authors.

1 Introduction

Functional dyspepsia (FD) is a complex disorder defined by upper gastrointestinal symptoms, including epigastric pain and burning, postprandial fullness, early satiation, bloating, belching, nausea, and vomiting, outside of any evident structural disease [1]. The definition of FD will likely continue to evolve with the publication of the revised Rome IV diagnostic criteria in 2016 [2]; however, the current definition, formulated by the Rome task force in 2006 as part of Rome III, includes two subtypes, postprandial distress syndrome and epigastric pain syndrome, that may overlap to varying degrees [3]. Postprandial distress syndrome is characterized by postprandial fullness and early satiation, and epigastric pain syndrome is characterized by epigastric pain and burning. Given the current definition, the prevalence of FD as defined in Rome III is difficult to determine because, based on the diagnostic criteria, an upper endoscopy must be performed to rule out the presence of any structural disease [3]. As a result, the literature commonly reports the prevalence of “uninvestigated dyspepsia,” with estimates ranging from 5 to 15 % depending on the population and definition used within individual studies [1].

The Critical Path Institute’s Patient-Reported Outcome (PRO) Consortium [4], in conjunction with advisors from the US Food and Drug Administration (FDA), has identified the need for a well-defined and reliable patient-reported instrument to assess treatment benefit in FD clinical trials. Although self-reporting is central to identifying and evaluating treatments for this symptom-defined condition, it is unclear to what extent the development of existing FD symptom PRO questionnaires was consistent with the FDA’s Guidance for Industry, “Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims” (hereafter referred to as “FDA PRO Guidance”) [5]. Therefore, the PRO Consortium established the Functional Dyspepsia Working Group (WG) to develop a PRO instrument for use as a primary endpoint assessment in FD clinical trials, and to submit the instrument for qualification under the FDA’s Drug Development Tools (DDT) Qualification Program [6].

As a first step, the WG set out to develop the evidentiary basis for the construction of a new PRO instrument to evaluate FD treatment efficacy claims. To this end, both a symptoms review and an instrument review were conducted. In keeping with the FDA PRO Guidance [5], the objectives of these activities were to document, from the perspective of the empirical literature, the primary symptoms of FD, evaluate the extent to which existing questionnaires target those symptoms, and, finally, identify any missing evidence that would impact the questionnaires’ use

in regulated clinical trials to assess treatment efficacy claims intended for product labeling. To the extent that existing instruments did not measure the concepts of interest (i.e., the primary symptoms of FD) or failed to meet the FDA’s evidentiary standards, decisions could be informed with respect to developing a new instrument or repurposing an existing one. In that light, it is important to note that the FDA categorizes a “questionnaire” as a set of questions or items shown to a respondent to obtain answers for research purposes, while an “instrument” consists of both a means to capture data (i.e., a questionnaire) plus all the information and documentation that supports its use (e.g., methods and instructions for administration or responding, a standard format for data collection, and methods for scoring, analysis, and interpretation of results) [5].

2 Methodology

2.1 Symptoms and Instrument Literature Searches

Literature searches were conducted in the Embase[®], PsycINFO[®], and MEDLINE[®] databases, using the OvidSP platform. Searches were limited to English-language studies in humans. To prevent the aggregation of unrelated or previously misclassified diseases, search results were also limited to those conducted from January 2006 onward, the year the Rome III criteria [3] for FD were published; and the search results closed May 2013, after which our instrument development transitioned to qualitative research. Search terms included “functional dyspepsia,” “nonulcer dyspepsia,” “idiopathic dyspepsia,” “essential dyspepsia,” “postprandial distress syndrome,” and “epigastric pain syndrome”; these were required to be present in the abstract of the article. In addition, abstracts were required to include at least one search term related to symptoms, questionnaires, or qualitative research methods. A manual inspection of the reference lists of all retrieved articles was also performed to identify any relevant citations not captured with the specified search criteria.

Search results were screened by abstract for relevancy and retrieved for full review according to pre-defined eligibility criteria. Specifically, articles were selected for review if they were specific to FD in adults, referenced symptoms of FD as experienced by patients, or referenced FD symptom-specific questionnaires or published qualitative research with patients (i.e., referenced qualitative design/methodology/analysis, such as interviews, focus groups, or patient reports); articles were excluded if they concerned FD in a pediatric population, concerned symptom assessment in a mixed population (i.e., population was not exclusive to patients with FD and included patients

with other gastrointestinal conditions), reported data on patients recruited into a study based on pre-Rome II criteria, reported FD symptoms without specifying a qualitative methodology or naming a specific questionnaire by which they were collected, focused on questionnaires used for diagnostic purposes only (not study endpoints), or were derived from non-peer-reviewed research (e.g., conference proceedings or dissertations). From each article selected for full-text review, data pertaining to the study aim, sample demographics, methodology, and results were extracted by populating a data extraction spreadsheet with these pre-specified variables.

In addition to the review of the literature, the patient-reported outcome and quality of life instruments database (PROQOLID) was searched to identify existing PRO instruments used to measure FD symptoms. PROQOLID is a database of PRO instruments managed by the Mapi Research Trust [7]. It includes information on psychometric properties, development and validation, available translations, and conditions of use from over 1000 reports on PRO instruments. Questionnaires focusing on or comprising domains focusing on the assessment of the patient experience of FD symptoms were considered for review. Questionnaires evaluating the impacts of FD, or health-related quality of life or quality of life, were excluded, as were any questionnaires not available in English.

2.2 Conceptual Model Development

Collectively, symptom concepts (e.g., epigastric pain and burning) identified in each of the reviewed articles were used to construct a preliminary conceptual model of the symptom experience of FD [8]. For comprehensiveness, the conceptual model included all concepts reported from each of the three sources, including existing PRO instruments, published qualitative research, and the Rome III criteria [8–10]. The published qualitative research included documented instances of symptom concepts spontaneously reported by patients with FD. Published interviews, focus groups, or written reports in which patients were allowed to freely expound outside the confines of a previously described or developed instrument were all considered eligible for analysis.

Proposed by Wilson and Cleary [8], a conceptual model is a heuristic classification scheme that links a specified disease state or condition to its proximal and increasingly distal health outcomes. In general, proximal concepts tend to be uni-dimensional, have a direct relationship with the condition and effects of treatment, and have greater potential to be characterized by primary trial outcomes. Distal concepts tend to be multi-dimensional, indirectly related to the condition and effects of treatment, and have less potential to be characterized by primary trial outcomes

[9, 10]. In the present context, the preliminary conceptual model was intended to capture only proximal, condition-level concepts (i.e., symptoms of FD).

2.3 Instrument Evaluation

Following the instrument literature review, existing PRO instruments were evaluated using a two-step approach. First, the adequacy of concept coverage was evaluated by comparing the concepts assessed by the reviewed instruments to the primary FD symptoms specified in the Rome III criteria (postprandial fullness, early satiation, postprandial nausea, excessive belching, epigastric pain, epigastric burning, and upper abdominal bloating) [3]. Second, instruments deemed to have adequate conceptual coverage were further evaluated against the evidentiary requirements set forth in the FDA PRO Guidance [5], and Clinical Outcome Assessment (COA) criteria established by the FDA's Study Endpoints and Label Development (SEALD) group [11].

3 Results

3.1 Symptoms Literature Search

The literature search identified a total of 787 articles. Through the initial database search (Embase[®], PsycINFO[®], and MEDLINE[®]), 780 potentially relevant articles were identified. Subsequently, seven additional articles were added from the article reference lists; 56 articles were included in the final review (Fig. 1).

3.2 Conceptual Model for Functional Dyspepsia Symptoms

The FD symptom conceptual model derived from the symptoms literature review is presented in Fig. 2 (depicted as a Venn diagram). The bottom left circle (a) of the model contains those concepts listed in the Rome III criteria for FD [3]. The top circle (b) contains the additional concepts assessed in existing questionnaires, which largely fall into one of two distinct categories: irritable bowel syndrome (IBS)-like symptoms [3] and heartburn or gastroesophageal reflux disease (GERD)-like symptoms [12]. Other dyspeptic symptoms that did not fit into one of these two categories are listed separately. Heartburn or GERD-like symptoms have been shown to co-occur with dyspeptic symptoms and are deemed interconnected to the key diagnostic symptoms of FD [1]. Further, patients with FD frequently experience IBS-like symptoms, making it difficult to separate IBS from the symptoms used as diagnostic criteria of FD.

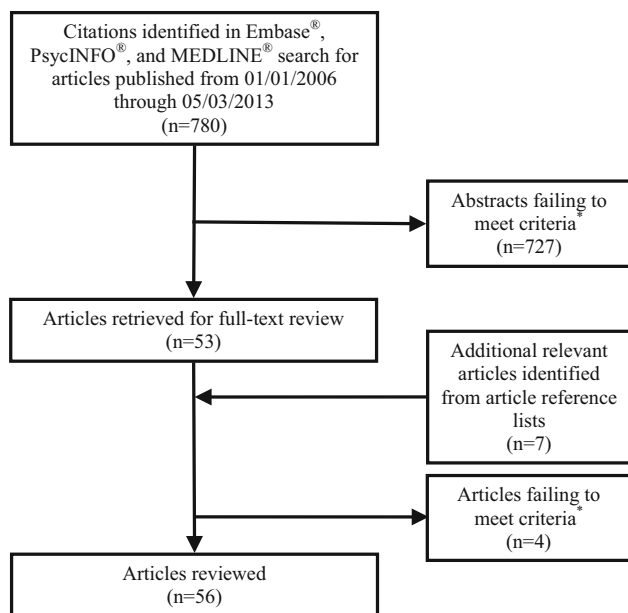


Fig. 1 Symptoms literature search flow diagram. *Asterisk* articles were excluded if they concerned FD in a pediatric population, a mixed population (i.e., not exclusive to FD), reported data on patients recruited into a study based on pre-Rome II criteria, reported FD symptoms without specifying a qualitative methodology or naming a specific questionnaire by which they were collected, focused on questionnaires used for diagnostic purposes only, or were derived from non-peer-reviewed research. *FD* functional dyspepsia

Finally, the bottom right circle (c) contains the symptoms reported by patients with FD in published qualitative research (non-instrument-based). These concepts include five of the diagnostic criteria from Rome III [3]—postprandial fullness, postprandial nausea, excessive belching, epigastric pain, and upper abdominal bloating—as well as vomiting. Excluding the articles describing the development of FD instruments, only one article was identified that describes spontaneous patient reports of FD symptoms [13]. In this study, which examined the relationship between FD symptoms and dietary patterns, 41 patients [$n = 20$ with FD (17 women; mean age = 45 ± 3 years; age range = 23–73 years) and $n = 21$ healthy controls (18 women; mean age = 40 ± 4 years; age range = 20–74 years)] completed symptom diaries in which they recorded any symptoms experienced, their severity, and the time at which they occurred. Subjects reported a total of 612 symptoms, which were divided into meal-related, meal-unrelated, or other symptoms on the basis of the time of their occurrence. Meal-related symptoms were reported as bloating, nausea, upper-abdominal pain, belching, epigastric pain, fullness, vomiting, and discomfort. Information on other symptom categories was not reported. With the exception of vomiting, all meal-related symptoms were redundant with the Rome III diagnostic criteria [3].

3.3 Instrument Evaluation: Conceptual Coverage of Functional Dyspepsia Instruments

A total of 16 PRO instruments assessing symptoms of FD were identified. The conceptual coverages of these instruments were evaluated against the seven primary symptom concepts defined by the Rome III criteria [3]. A total of three instruments were found to measure all seven symptoms: the Dyspepsia Symptom Severity Index (DSSI), Nepean Dyspepsia Index (NDI), and Short-Form Nepean Dyspepsia Index (SF-NDI). A summary of the conceptual coverages of the identified FD instruments is provided in Table 1.

Of the seven primary symptoms of FD, the most commonly measured by existing FD instruments were postprandial nausea/nausea and epigastric pain/pain, with 13 out of 16 instruments containing items assessing these two symptoms. Eleven of the 16 instruments also had items measuring “upper abdominal bloating” or “bloating.” Among the 13 instruments with incomplete conceptual coverage of the seven primary FD symptoms, “epigastric burning” and “postprandial fullness” were the symptoms most often omitted from assessment (omitted in 11 and ten instruments, respectively).

3.4 Instrument Evaluation: Description and Developmental History of Functional Dyspepsia Instruments

Instruments concluded to have adequate conceptual coverage—the DSSI, NDI, and SF-NDI—were further evaluated with regard to the instruments’ recall periods, item wording, response options, scoring, and development histories (Table 2).

Recall period An instrument’s recall period is defined as the time period that respondents are asked to consider when replying to items. The DSSI, NDI, and SF-NDI all utilize recall periods of the 2 weeks prior to assessment. The FDA has indicated, particularly for symptom assessment (i.e., concepts that vary over short periods of time), that recall periods that ask respondents to focus on their current or recent state (e.g., the past 24 h) are preferable to longer intervals [5]. If FD symptoms vary within a day and/or between days, the instruments reviewed here may raise concern due to potential recall bias and an inability to account for daily variation in FD symptoms.

Item wording An item was considered well-constructed based on the following criteria: contains no medical jargon or slang, double negatives, or overlapping or unbalanced responses; assesses only one concept (i.e., not double-barreled); and does not assess vague or ambiguous concepts. Item wording was deemed adequate in both the full

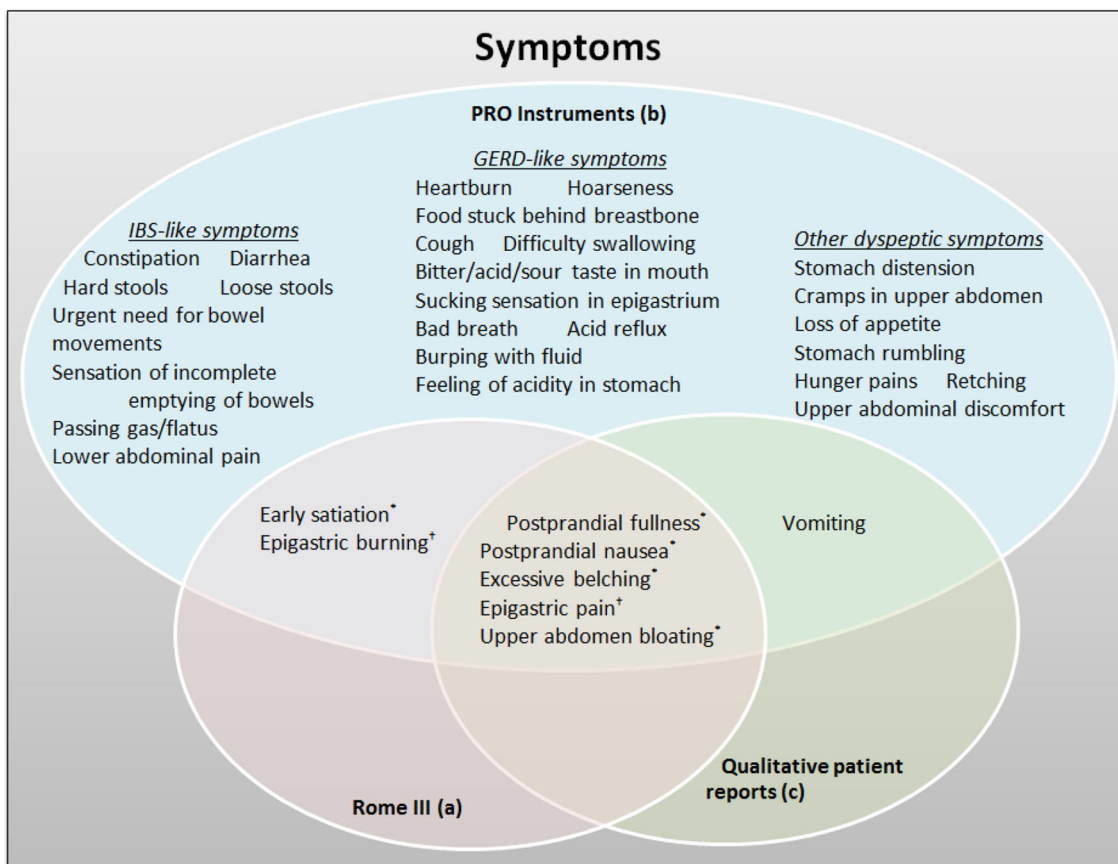


Fig. 2 Conceptual model for FD symptoms. Qualitative patient reports from Pilichiewicz et al. [13]. *Asterisk* postprandial distress syndrome symptoms, as defined by Rome III criteria. *Dagger*

epigastric pain syndrome symptoms, as defined by Rome III criteria. *FD* functional dyspepsia, *GERD* gastroesophageal reflux disease, *IBS* irritable bowel syndrome, *PRO* patient-reported outcome

and short forms of the NDI; however, the DSSI includes an item assessing “discomfort,” which may be best measured with multiple items assessing various dimensions of “discomfort.”

Response options Response options comprise the range of choices a questionnaire provides to respondents in replying to individual items. The DSSI, NDI, and SF-NDI all use 5-point Likert-type scales. For the DSSI, the wording of response options was considered clear and appropriate, and response options were appropriately ordered, with strong distinctions offered between choices.

The NDI, however, uses a “bother” scale to assess symptoms, in addition to frequency and intensity scales, and scoring involves the summation of frequency, intensity, and bother scores. Incorporating bother scores may raise concerns because bother is a complex evaluative concept that may encompass aspects of symptom frequency, severity, and associated impact. Additionally, greater opportunity for individual variation exists among participants responding to a measure of symptom bother than for items assessing symptom frequency or severity.

Moreover, how individual patients arrive at different responses is potentially unclear. The use of “bother” may also prove difficult to implement in global trials, where translation of the instrument will likely be required [31]. Finally, incorporating multiple scales in one instrument may raise concerns, as certain symptoms lend themselves to assessment along particular dimensions better than others. For example, patients may find it more difficult to assess vomiting in terms of intensity rather than frequency; therefore, the assessment of both frequency and intensity in the NDI may be problematic for some symptoms.

Development history An instrument’s development history consists of the methods used to inform its construction. The DSSI, NDI, and SF-NDI provided some evidence indicating patient involvement in their development. However, few details were provided concerning the characteristics of the participant populations used and the way in which data generated from these activities informed item development. Furthermore, no reference to “concept saturation” (i.e., the point at which no new information would likely be generated upon conduct of additional interviews, a measure

Table 1 Coverage of core concepts from Rome III FD diagnostic criteria in existing instruments

Name of instrument (abbreviation)	Items (total)	Items (symptoms)	Core concepts from Rome III							Total (N = 7)
			Postprandial fullness	Early satiation	Postprandial nausea	Excessive belching	Epigastric pain	Epigastric burning	Upper abdominal bloating	
Dyspepsia Symptom Severity Index (DSSI) [14]	19	19	✓	✓	✓	✓	✓	✓	✓	7
Gastrointestinal Symptom Rating Scale, interviewer-administered (GIRS) [15]	15	15	✓	✓	✓	✓	✓	✓	✓	4
Gastrointestinal Symptom Rating Scale, self-administered (GIRS-self) [16]	15	15	✓	✓	✓	✓	✓	✓	✓	4
Gastrointestinal Symptom Score (GIS) [17]	10	10	✓	✓	✓	✓	✓	✓	✓	4
Glasgow Dyspepsia Severity Score (GDSS) [18]	8	1								0
Global Overall Symptoms (GOS) scale, including specific symptom subtypes (SSS) [19]	11	11	✓	✓	✓	✓	✓	✓	✓	6
Hong Kong Index of Dyspepsia (HKID)/Hong Kong Index of Severity (HKIS) [20]	12	12	✓	✓	✓	✓	✓	✓	✓	4
Leeds Dyspepsia Questionnaire (LDQ) [21]	15	15	✓	✓	✓	✓	✓	✓	✓	3
Nepean Dyspepsia Index (NDI) [22]	42	15	✓	✓	✓	✓	✓	✓	✓	7
Short-Form Nepean Dyspepsia Index (SF-NDI) [23]	25	15	✓	✓	✓	✓	✓	✓	✓	7
Patient Assessment of Upper Gastrointestinal Symptom Severity Index (PAGI-SYM) [24]	20	20	✓	✓	✓	✓	✓	✓	✓	5
Porto Alegre Dyspeptic Symptoms Questionnaire (PADYQ) [25]	11	11	✓	✓	✓	✓	✓	✓	✓	4
Reflux Disease Questionnaire (RDQ) [26]	26	26				✓	✓	✓	✓	3
Severity of Dyspepsia Assessment (SODA) [27, 28]	17	13		✓	✓	✓	✓	✓	✓	3
Short-Form Leeds Dyspepsia Questionnaire (SF-LDQ) [29]	5	5		✓	✓					1
Symptom Index of Dyspepsia (SID) [30]	4	4	✓					✓	✓	4

Blank = concept not assessed by instrument; ✓ = concept assessed by instrument
 FD functional dyspepsia

Table 2 Assessment of instrument face validity and developmental history

Name of instrument (abbreviation)	Recall period	Item wording	Response options	Scoring	Developmental history
Dyspepsia Symptom Severity Index (DSSI) [14]	× 2 weeks	× Use of term “discomfort” “Discomfort” is a multidimensional concept	✓ 5-Point Likert-type scale	Scoring appears adequate: Total scale score represented by the mean across the three subscales Higher scores indicate more severe symptoms Subscale scores (representing reflux-, ulcer-, and dysmotility-like symptoms) can be computed by taking the mean value across all symptoms composing the subscale	Some evidence of patient involvement during development: Focus groups (n = 4, 90 min, 8–10 patients with dyspepsia symptoms in each) were conducted to identify the words and phrases patients themselves use to describe dyspeptic symptoms A bank of 21 items and an overall assessment item were generated on the basis of the literature review and patient focus groups Instrument was not cognitively debriefed with patients, as the instrument was tested with gastroenterologists
Nepean Dyspepsia Index (NDI) [22]	× 2 weeks	✓	× 5-Point Likert-type scale measuring either frequency, intensity, or bother (possibly not appropriate to assess symptoms such as vomiting in terms of intensity)	Issues with scoring: Although scored separately from HRQoL, symptom items are scored by the additive model (score = frequency + intensity + bothersomeness)	Some evidence of patient involvement during development: Initial items developed through consultations with patients and local clinicians Items pre-tested among 55 FD patients in four countries to identify relevant disease-specific items, using focus groups and face-to-face interviews Items selected by applying the criterion that if ≥60 % of subjects responded “not at all” or “not applicable,” then the item was not relevant. Twenty-five items remained after applying this criterion
Short-Form Nepean Dyspepsia Index (SF-NDI) [23]	× 2 weeks	✓	× 5-Point Likert-type scale measuring either frequency, intensity, or bother (possibly not appropriate to assess symptoms such as vomiting in terms of intensity)	Issues with scoring: Although scored separately from HRQoL, symptom items are scored by the additive model (score = frequency + intensity + bothersomeness)	Some evidence of patient involvement during development: As above for NDI The two items with the highest factor loadings on each of the five subscales representing the relevant areas were selected to be included in the new, ten-item SF-NDI

× = unsatisfactory; ✓ = satisfactory

FD functional dyspepsia, HRQoL health-related quality of life

of confidence in qualitative research results) was made in any of the instrument development papers. In light of these findings, further evidence is needed to confirm that the instruments adequately assess all symptoms relevant and important to an FD-specific population.

4 Discussion

Functional dyspepsia is a complex disorder defined by upper gastrointestinal symptoms, including epigastric pain and burning, postprandial fullness, early satiation, bloating, belching, nausea, and vomiting, outside of any evident structural disease. Despite the centrality of patient assessment in properly identifying patients and evaluating treatments for this symptom-defined condition, it is unclear to what extent existing PRO instruments for assessing symptoms of FD were developed according to principles consistent with the approach outlined in the FDA PRO Guidance for supporting product labeling. In keeping with the FDA PRO Guidance approach, the objectives of our symptoms and instrument literature reviews were to document, from the perspective of the empirical literature, the primary symptoms of FD, evaluate the extent to which existing instruments target those symptoms, and, finally, identify any missing evidence that would impact their use in product labeling.

The identified symptoms of FD are summarized in the preliminary conceptual model and include symptoms outside the Rome III criteria (i.e., IBS-like symptoms, GERD-like symptoms, and other less obviously clustered dyspeptic symptoms), as well as one additional symptom identified from the qualitative research reports in the literature (i.e., vomiting). Except for vomiting, all of the identified symptoms from the published qualitative research reports were also specified in the Rome III criteria [3]. Heartburn or GERD-like symptoms identified in the symptoms literature review have been shown to co-occur with dyspeptic symptoms and are deemed interconnected to the key diagnostic symptoms of FD. Further, patients with FD frequently experience IBS-like symptoms, making it difficult to separate IBS from the symptoms used as diagnostic criteria of FD. Therefore, it is likely that future qualitative research soliciting direct patient reports of symptoms in the population of patients with FD will elicit not only the symptoms specified in the Rome III criteria but also a larger set of symptoms as conceptual saturation is achieved. Whether the conceptual framework of a future PRO instrument should adhere to a narrow subset of disease-defining symptoms or attempt to incorporate a broader range of FD-predominant symptoms will be informed by qualitative and quantitative data from interviews with patients with FD.

Sixteen instruments were identified in the literature search for existing PROs used in populations with FD. Of these, three were found to measure all seven core symptoms of FD as defined by the Rome III criteria: the DSSI, NDI, and SF-NDI. However, when these three instruments were evaluated in light of regulatory evidentiary recommendations [5], several issues were identified that could jeopardize qualification of the instruments for substantiation of a labeling claim.

Firstly, all three instruments have a recall period of 2 weeks, which is longer than recommended by the FDA PRO Guidance (Sect. D.1, Content validity: recall period) [5]. Generally, the appropriateness of the recall period should be established for the population, disease state, and application of the instrument and, more specifically, factors such as the variability, duration, frequency, and intensity of the concept measured should be considered. Therefore, given the inherent variability of FD symptoms, the advantage of momentary assessment in reducing recall bias, and the stated objective of regulatory acceptance, a 2-week recall period would not likely be considered acceptable.

Secondly, the NDI and SF-NDI assess all symptoms in terms of frequency, intensity, and bother. Certain symptoms, such as vomiting, lend themselves to assessment by certain dimensions (e.g., frequency) better than others (e.g., intensity); bother, as an evaluative concept, is complex and may encompass aspects of symptom frequency, severity, and impact. Further, qualitative evidence of the content validity of these instruments is unavailable in the empirical literature. In particular, although concepts were elicited from patients during the development of these instruments, no evidence that a conceptual framework was developed or that conceptual saturation was achieved is evident. Lastly, there is a lack of evidence regarding cognitive debriefing across all three instruments. For comparison, a review of the literature conducted by Ang et al. [32] to identify studies of patients with FD that employed clinical outcome assessments similarly concluded that no existing instruments appeared to be sufficiently content valid for use in a Rome III-defined FD trial.

In light of the preliminary conceptual model and the deficiencies of these existing instruments with respect to the approach outlined in the FDA PRO Guidance, fundamental modifications and new documented evidence would be a prerequisite for deploying any of these instruments in a biopharmaceutical development program targeting an FDA-approved product label. Therefore, in consideration of the clear clinical need, the WG recommends the development of a new PRO instrument to assess FD symptoms and, in particular, the assessment of symptom improvement. The conceptual model for FD symptoms derived from our symptoms literature review can serve as a

foundation for the future qualitative research necessary for the development of such an instrument, and will be informed by data from qualitative interviews in patients with FD as well as refined throughout the instrument development process.

The next step in the development of a new FD symptom-focused PRO instrument is to conduct concept elicitation interviews among patients with a Rome III-confirmed FD diagnosis to identify and document a comprehensive set of symptoms from the patient perspective. To this end, the WG has developed a semi-structured interview guide based on the present work to facilitate qualitative research in a carefully selected group of patients with FD, which will provide documented evidence suitable for qualification under the FDA's DDT program [33]. As it is evident that a number of symptoms within the conceptual model developed from the literature overlap with those of other functional gastrointestinal disorders, key objectives of future concept elicitation interviews will be to establish whether these overlapping concepts are important and relevant symptoms from the perspective of patients with FD, and to gather evidence that development of a well-defined and reliable patient-reported measure to assess treatment benefit is possible in FD.

5 Limitations

The preliminary conceptual model currently includes concepts assessed by PRO instruments that, although used in studies to assess FD symptoms, were not initially developed in an FD population. The FDA PRO Guidance [5] outlines the importance of ensuring that “the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use”. Therefore, it is possible that qualitative research may demonstrate that some of the concepts currently included in the conceptual model are not important or relevant to patients with FD.

Owing to our chosen eligibility criteria, articles were excluded if they concerned FD in a pediatric population, concerned symptom assessment in a mixed population (i.e., population was not exclusive to patients with FD and included patients with other gastrointestinal conditions), reported data on patients recruited into a study based on pre-Rome II criteria, reported FD symptoms without specifying a qualitative methodology or naming a specific questionnaire by which they were collected, focused on questionnaires used for diagnostic purposes only (not study endpoints), or were derived from non-peer-reviewed research (e.g., conference proceedings or dissertations). Therefore, it is possible that instruments that could have been modified and redeveloped may have been excluded.

However, in light of the results of our instrument evaluation, this possibility appears unlikely.

6 Conclusion

No existing PRO instruments were identified that assessed all seven core Rome III symptoms of FD and adhered to principles consistent with the instrument development approach outlined in the FDA PRO Guidance for supporting product labeling. In light of these findings, the WG recommends the development of a new PRO instrument to measure FD symptoms, and provides a preliminary conceptual model for consideration in the requisite qualitative research.

Acknowledgments This project is a collaboration of the Functional Dyspepsia Working Group within the Critical Path Institute's Patient-Reported Outcome (PRO) Consortium (including pharmaceutical company representatives and selected non-member participants), academic researchers/clinicians, FDA, and Adelphi Values, operating in a pre-competitive environment.

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 by grant No. U01FD003865 from the US FDA.

A list of the members of the PRO Consortium is available at <http://c-path.org/programs/pro/>.

Compliance with Ethical Standards The authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE) and were fully responsible for all aspects of manuscript development. Fiona Taylor participated in planning and executing the study, analyzing and interpreting the data, and drafting the manuscript; Catherine Foley and Farrah Pompilus participated in planning and executing the study, collecting, analyzing and interpreting the data, and drafting the manuscript; Ramon Iovin participated in collecting, analyzing, and interpreting the data and drafting the manuscript; Alan L. Shields participated in planning and executing the study, interpreting the data, and drafting the manuscript; David S. Reasner provided input on the study design and drafting of the manuscript; Robyn T. Carson, Linda S. Deal, Debra G. Silberg, and Jason Lundy provided input on the study design and manuscript. FT, CF, and ALS are employees of Adelphi Values, which received payment from the sponsors to conduct the research; RI and FP were employees of Adelphi Values at the time this research was conducted. DSR is a member of Albemarle Scientific Consulting and an employee of Ironwood Pharmaceuticals, and owns stock and stock options in Ironwood. RTC is an employee of Allergan and owns stock and stock options in Allergan. At the time of the research, LSD and DGS were employees of Shire, and have shareholdings in Shire and other pharmaceutical companies through stocks and mutual funds. JJJ was an employee of the Critical Path Institute at the time this research was conducted. Any views expressed in this publication represent the personal opinions of the authors, not those of their respective employer. The authors' respective organizations were given the opportunity to review the manuscript for medical and scientific accuracy as well as intellectual property considerations.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Tack J, Talley NJ. Functional dyspepsia—symptoms, definitions and validity of the Rome III criteria. *Nat Rev Gastroenterol Hepatol.* 2013;10:134–41.
- Rome Foundation. Rome IV. 2013. http://www.romecriteria.org/romeiv_pub/index.cfm. Accessed 19 Jun 2014.
- 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., editors. *Rome III The functional Gastrointestinal Disorders*. 3rd ed. McLean: Degnon Associates, Inc.; 2006. p. 885–97.
- Coons SJ, Kothari S, Monz BU, Burke LB. The patient-reported outcome (PRO) consortium: filling measurement gaps for PRO end points to support labeling claims. *Clin Pharmacol Ther.* 2011;90:743–8.
- U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009.
- U.S. Food and Drug Administration. Drug Development Tools (DDT) Qualification Programs. 2015. <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/default.htm>. Accessed 22 Feb 2016.
- PROQOLID: Patient-Reported Outcome and Quality of Life Instruments Database. 2010. <http://www.proqolid.org/>. Accessed 3 Apr 2012.
- Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA.* 1995;273:59–65.
- Shields AL, Gwaltney C, Tiplady B, et al. Grasping the FDA's PRO Guidance: what the agency requires to support the selection of patient reported outcome instruments. *Applied Clinical Trials.* 2006;15:69–77.
- Marquis P, Caron M, Emery M, et al. The role of health-related quality of life data in the drug approval processes in the US and Europe. *Pharm Med.* 2011;25:147–60.
- U.S. Food and Drug Administration. Clinical Outcome Assessment (COA): Frequently asked questions. 2014. <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370261.htm>. Accessed 6 Jun 2015.
- Vakil N, van Zanten SV, Kahrilas P, et al. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol.* 2006;101:1900–20.
- Pilichiewicz AN, Horowitz M, Holtmann GJ, et al. Relationship between symptoms and dietary patterns in patients with functional dyspepsia. *Clin Gastroenterol Hepatol.* 2009;7:317–22.
- 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.
- Svedlund J, Sjodin I, Dotevall G. GSRS—a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci.* 1988;33:129–34.
- Dimenas E, Glise H, Hallerback B, et al. Quality of life in patients with upper gastrointestinal symptoms. An improved evaluation of treatment regimens? *Scand J Gastroenterol.* 1993;28:681–7.
- Adam B, Liebrechts T, Saadat-Gilani K, et al. Validation of the gastrointestinal symptom score for the assessment of symptoms in patients with functional dyspepsia. *Aliment Pharmacol Ther.* 2005;22:357–63.
- El-Omar EM, Banerjee S, Wirz A, McColl KE. The Glasgow Dyspepsia Severity Score—a tool for the global measurement of dyspepsia. *Eur J Gastroenterol Hepatol.* 1996;8:967–71.
- Veldhuyzen van Zanten SJ, Chiba N, Armstrong D, et al. Validation of a 7-point Global Overall Symptom scale to measure the severity of dyspepsia symptoms in clinical trials. *Aliment Pharmacol Ther.* 2006;23:521–9.
- Hu WH, Lam KF, Wong YH, et al. The Hong Kong Index of Dyspepsia: a validated symptom severity questionnaire for patients with dyspepsia. *J Gastroenterol Hepatol.* 2002;17:545–51.
- Moayyedi P, Duffett S, Braunholtz D, et al. The Leeds Dyspepsia Questionnaire: a valid tool for measuring the presence and severity of dyspepsia. *Aliment Pharmacol Ther.* 1998;12:1257–62.
- Talley NJ, Verlinden M, Jones M. Validity of a new quality of life scale for functional dyspepsia: a United States multicenter trial of the Nepean Dyspepsia Index. *Am J Gastroenterol.* 1999;94:2390–7.
- Talley NJ, Phillips SF, Melton J III, et al. A patient questionnaire to identify bowel disease. *Ann Intern Med.* 1989;111:671–4.
- Rentz AM, Kahrilas P, Stanghellini V, et al. Development and psychometric evaluation of the Patient Assessment of Upper Gastrointestinal Symptom Severity Index (PAGI-SYM) in patients with upper gastrointestinal disorders. *Qual Life Res.* 2004;13:1737–49.
- Sander GB, Mazzoleni LE, Francesconi CF, et al. Development and validation of a cross-cultural questionnaire to evaluate nonulcer dyspepsia: the Porto Alegre Dyspeptic Symptoms Questionnaire (PADYQ). *Dig Dis Sci.* 2004;49:1822–9.
- Shaw M, Dent J, Beebe T, et al. The Reflux Disease Questionnaire: a measure for assessment of treatment response in clinical trials. *Health Qual Life Outcomes.* 2008;6:31.
- Kuykendall DH, Rabeneck L, Campbell CJ, Wray NP. Dyspepsia: how should we measure it? *J Clin Epidemiol.* 1998;51:99–106.
- Rabeneck L. Measuring dyspepsia-related health in randomized trials: the Severity of Dyspepsia Assessment (SODA) and its use in treatment with NSAIDs and COX-2-specific inhibitors. *Rheumatology (Oxford)* 2003;42(Suppl 3):iii32–9.
- Fraser A, Delaney BC, Ford AC, et al. The Short-Form Leeds Dyspepsia Questionnaire validation study. *Aliment Pharmacol Ther.* 2007;25:477–86.
- Zeng F, Qin W, Liang F, et al. Abnormal resting brain activity in patients with functional dyspepsia is related to symptom severity. *Gastroenterology.* 2011;141:499–506.
- Gawlicki MC, McKown SM, Talbert MJ, Brandt BA. Application of Bothers in patient reported outcomes instruments across cultures. *Health Qual Life Outcomes.* 2014;12:18.
- Ang D, Talley NJ, Simren M, et al. Review article: endpoints used in functional dyspepsia drug therapy trials. *Aliment Pharmacol Ther.* 2011;33:634–49.
- U.S. Department of Commerce, Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for Industry and FDA Staff: Qualification Process for Drug Development Tools. 2014.