Patient (2016) 9:117–134
DOI 10.1007/s40271-015-0132-1

ORIGINAL RESEARCH ARTICLE

Patient-Centered Research to Support the Development of the Symptoms of Major Depressive Disorder Scale (SMDDS): Initial Qualitative Research

Kelly P. McCarrier[1] · Linda S. Deal[2] · Lucy Abraham[3] · Steven I. Blum[4] · Elizabeth Nicole Bush[5] · Mona L. Martin[1] · Michael E. Thase[6,7] · Stephen Joel Coons[8] · On behalf of the PRO Consortium's Depression Working Group

Published online: 26 June 2015

Abstract

Background Content valid, patient-reported outcome (PRO) measures of major depressive disorder (MDD) symptoms are needed to assess MDD treatment benefit. While a range of questionnaires are currently available to evaluate aspects of depression from the patient's perspective, their comprehensiveness and qualitative development histories are unclear.

Objective The objective of this study was to describe the process and results of the preliminary qualitative development of a new symptom-based PRO measure intended to assess treatment benefit in MDD clinical trials.

Methods Qualitative interviews were conducted with adult MDD patients in the USA who recently experienced a major depressive episode. Experienced interviewers conducted concept elicitation (CE) and cognitive interviews using semi-structured interview guides. The CE interview guide was used to elicit spontaneous reports of symptom experiences along with probing to further explore and confirm concepts. The cognitive interview guide was developed to evaluate concept relevance, understandability, and structure of the draft items, and to facilitate further instrument refinement.

Results Forty patients participated in the CE interviews. A total of 3022 symptom codes, representing 84 different concepts were derived from the transcripts. Data from the CE interviews were considered alongside existing literature and clinical expert opinion during an item-generation process, leading to development of a preliminary version of the Symptoms of Major Depressive Disorder Scale (SMDDS). Fifteen patients participated in three waves of cognitive interviews, during which the SMDDS was further refined.

Conclusions The SMDDS is a 35-item PRO measure intended for use as an endpoint in MDD clinical trials to support medical product labeling. The SMDDS uses a 7-day recall period and verbal rating scales. It was developed in accordance with the US Food and Drug Administration (FDA)'s PRO Guidance and best practices. Qualitative interviews have provided evidence for content validity. Future quantitative studies will confirm the SMDDS's measurement properties and support FDA qualification.

Steven I. Blum was affiliated with Forest Research Institute, Jersey City, NJ, USA (now part of Actavis) at the time of his involvement in the study.

Electronic supplementary material The online version of this article (doi:10.1007/s40271-015-0132-1) contains supplementary material, which is available to authorized users.

✉ Kelly P. McCarrier
mccarrier@hrainc.net

1 Health Research Associates, Inc., 6505 216th St. SW - Suite 105, Mountlake Terrace, WA 98043, USA

2 Clinical Outcomes Assessment, Shire, Wayne, PA, USA

3 Outcomes and Evidence, Pfizer Ltd, Tadworth, Surrey, UK

4 Patient Reported Outcomes, GlaxoSmithKline, King of Prussia, PA, USA

5 PRO Center of Expertise, Eli Lilly and Company, Indianapolis, IN, USA

6 Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA

7 Philadelphia Veterans Affairs Medical Center, Philadelphia, PA, USA

8 Patient-Reported Outcome Consortium, Critical Path Institute, Tucson, AZ, USA

Key Points for Decision Makers

Through a public–private collaboration supported by the US Food and Drug Administration (FDA), the PRO Consortium's Depression Working Group is developing a new patient-reported outcome (PRO) instrument for the assessment of symptoms of major depressive disorder (MDD).

The instrument's development has followed scientific best practices and FDA guidance, and its content validity has been supported by direct patient involvement through qualitative concept elicitation and cognitive interviews.

This process has resulted in a new 35-item PRO questionnaire, the Symptoms of Major Depressive Disorder Scale (SMDDS), which is currently being evaluated and refined through quantitative testing to support FDA qualification as a clinical trial endpoint to assess MDD symptoms.

1 Introduction/Background

Major depressive disorder (MDD) is a severe mental health disorder affecting 16.9 % of the US adult population and nearly 340 million people worldwide [1]. As a leading cause of disability, it is responsible for ≈400 million lost workdays in the USA each year, and the incremental cost of persons with MDD has recently been estimated to exceed US$210 billion annually [2]. MDD is characterized by depressed mood, loss of interest or pleasure, fatigue, poor concentration, associated feelings of worthlessness and guilt, and suicidal thoughts [3], and includes a range of subjectively experienced symptoms. In clinical trials, efficacy of new treatments for MDD is typically evaluated using clinician-reported outcome (ClinRO) measures such as the Hamilton Depression Rating Scale (HAM-D) [4] or Montgomery-Åsberg Depression Rating Scale (MADRS) [5], and a Global Clinical Impression of Change.

While ClinRO measures capture information that is not fully evaluable via self-report (e.g., psychomotor retardation, physical agitation), MDD is primarily a subjective experience, with severity of symptoms directly related to the degree of impairment [6]. Research indicates patient-reported scales contribute more than ClinRO measures in predicting pharmacological treatment outcome for MDD, suggesting patient-report provides clinically important information not accessible through clinician rating [7]. Assessment of depressive symptoms from the patient's perspective is essential to fully evaluate treatment risk–benefit profiles in clinical studies and complements traditional ClinRO measures in assessment of treatment outcomes. Hence, the Critical Path Institute's (C-Path) Patient-Reported Outcome (PRO) Consortium in conjunction with advisors from the US Food and Drug Administration (FDA) identified as a priority the need for a well-defined and reliable patient-reported measure to assess MDD treatment benefit and support product labeling. The PRO Consortium established the Depression Working Group to qualify a PRO instrument for use as a primary endpoint measure in MDD clinical trials [8]. Qualification, as defined by the FDA's Center for Drug Evaluation and Research (CDER), is a formal conclusion that the results obtained from the PRO instrument within a stated context of use can be relied upon to have a specific interpretation and application in drug development and regulatory review [9].

The Depression Working Group comprises representatives from the sponsoring firms and C-Path personnel. Health Research Associates (HRA) was selected, through competitive bidding, to provide contract research services with and on behalf of the working group. The core development team for the Symptoms of Major Depressive Disorder Scale (SMDDS) comprised members of the Depression Working Group and PRO measurement scientists from HRA. An advisory panel of clinical experts provided input at critical stages of the PRO instrument development process. Activities undertaken by the group follow the FDA guidance documents titled Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (referred to here as PRO Guidance) [10] and Qualification Process for Drug Development Tools [9].

This manuscript describes the initial steps to date in the development of the SMDDS: (1) the decision to develop a new PRO measure for MDD rather than select/modify an existing measure; (2) methodological steps/findings from the concept elicitation (CE) interviews, including clinical input and item generation; (3) development of a preliminary version of the SMDDS; and (4) findings from cognitive interviews and resulting modifications to the SMDDS.

2 Methods

2.1 Study Design and Development Steps

Figure 1 shows major study activities in developing the SMDDS. The steps leading to the qualitative development of the preliminary version of the instrument and evidence supporting its content validity are presented below. Further development efforts are currently underway, and will be reported separately when completed.

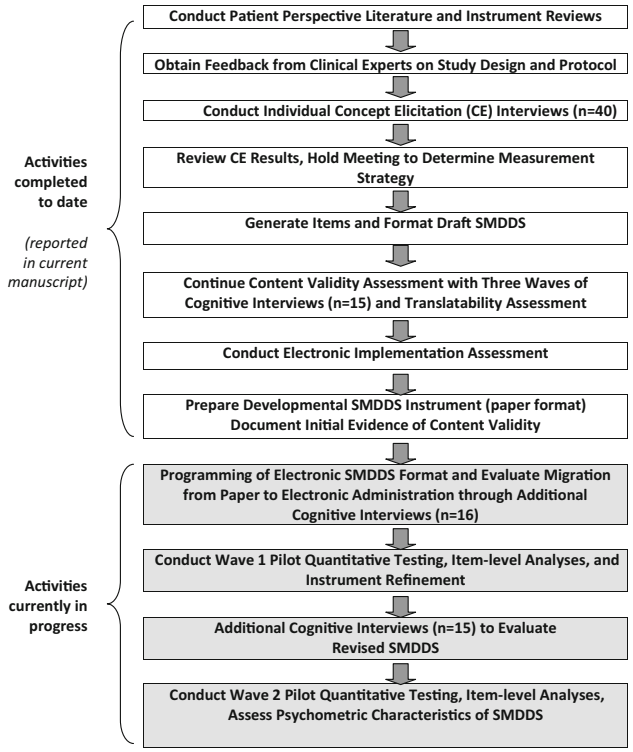

Fig. 1 Chronology of Symptoms of Major Depressive Disorder Scale (SMDDS) development activities. *CE* concept elicitation

In the earliest developmental step, two separate systematic literature reviews were conducted. The first examined previously published qualitative research in MDD to identify relevant symptom concepts and domains from the patient perspective. It identified a predominant focus of patient experience on emotions rather than clinical symptoms (cognitive and executive functioning symptoms) [11].

The second review evaluated published evidence for existing MDD-focused outcome assessment instruments potentially suitable for FDA qualification, or to identify potential items that could be used in development of a new measure. The outcome measures identified varied in concepts measured, instrument length, response options, anchoring, scoring algorithms, and recall period. However, limited information was available on how these tools were developed, in particular the extent of direct patient input [12]. Therefore, existing instruments were considered unlikely to satisfy requirements of the FDA's PRO Guidance [9]. Further qualitative work with MDD patients was deemed necessary to determine if any existing scale could be shown to have sufficient content validity.

The literature and instrument reviews informed the development of a qualitative study protocol and CE interview guide for qualitative work that met the methodological expectations outlined in the PRO Guidance. The qualitative study protocol and recruitment and interview forms were reviewed and approved by Quorum Review IRB (Seattle, WA, USA). All study participants provided written informed consent prior to participation in study activities. The study was conducted in compliance with the Declaration of Helsinki. Individual qualitative CE interviews were conducted to identify MDD-related symptom concepts relevant to MDD patients, documenting language patients used when describing symptoms, and exploring symptom characteristics such as severity, frequency, and duration.

After qualitative coding and analysis of interview transcripts, the working group determined that development of a new PRO measure for MDD was merited. Some existing measures included many of the identified concepts, but no single instrument was found suitable for the current purposes. All would require some degree of modification; therefore, the qualitative data were used to support an item generation process resulting in a preliminary draft of the SMDDS.

The preliminary version of the SMDDS was further tested and refined through three iterative waves of cognitive interviews conducted to evaluate understanding, acceptability of formatting, appropriateness of the instructions, recall period, and response options. Concurrent with the series of cognitive interviews, the SMDDS was also evaluated for translatability into five different languages. Additionally, refinements to support the implementation of the SMDDS across all currently available electronic PRO (ePRO) data capture platforms were performed before the preliminary SMDDS was finalized for ePRO programming and quantitative testing.

2.2 Study Participants

The qualitative study targeted a diverse sample of interview participants similar to those who would be completing the PRO instrument in future MDD clinical trials. Identical eligibility criteria were employed for CE and cognitive interview participants, reflecting common entry criteria for clinical trials designed to evaluate MDD treatment benefit. The study included male and female participants aged 18–65 years meeting Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition, Text Revision) DSM-IV-TR [13] and Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) DSM-5 [3] criteria for MDD, who were being treated for MDD on an outpatient basis. All had experienced a major depressive episode (MDE) within 6 months of screening and had a HAM-D score >18 at screening. Participants were required to read, write, and speak English at a level allowing them to provide written informed consent and actively contribute in an interview.

To ensure concepts elicited were about MDD, individuals with a current or past history of a personality disorder, bipolar disorder, schizophrenia or other psychotic disorder, obsessive–compulsive disorder, post-traumatic stress disorder, organic mental disorders, or mental disorders due to general medical conditions were excluded from the study. Other exclusion criteria included a recent history (12 months) of significant drug or alcohol abuse; a positive urine screen for drugs of abuse at time of enrollment; significant risk of suicide (determined by investigator or Columbia-Suicide Severity Rating Scale); or history of electroconvulsive therapy, vagal nerve stimulation, or deep-brain stimulation treatment for MDD.

Subjects were recruited from clinical sites in six US states (Connecticut, Florida, Illinois, New York, Oklahoma, and Washington) between March and April 2012. While no formal diversity quotas were employed, each clinical site strove to recruit patients with varying MDD treatment history and disease severity, and broad demographic characteristics (age, sex, race/ethnicity, marital status, educational attainment, and employment status).

2.3 Concept Elicitation (CE) Interviews

The study protocol and qualitative interview guide were developed prior to CE recruitment, based on a hypothesized conceptual framework, the results from the literature review, and clinician expert input (see the Acknowledgments section). CE interviews were conducted by trained qualitative research staff in private interview rooms at each participating clinic or at a nearby market research facility. In total, four different interviewers were involved in qualitative data collection, with four interviewers conducting the CE interviews, and three of those conducting cognitive interviews. Interviewers had between 3 and 25 years of experience with interviewing techniques for PRO measurement development and all were experienced in conducting individual patient interviews in mental health settings.

Interviews followed the semi-structured CE interview guide, employing open-ended questions and day-reconstruction exercises to elicit spontaneous reports of MDD-related symptom concepts (see the Electronic Supplementary Material). Open-ended questions were followed by probing, to assess concepts not spontaneously reported by study participants. The interview probe content was based on concepts identified in the systematic review of MDD literature. For each symptom they reported, interview participants were asked to rate the severity and level of bother or difficulty.

2.4 Analysis of Qualitative Data

Interviews were audio recorded and transcribed. Transcripts were reviewed independently by trained coders to identify patient-expressed concepts. Code assignment facilitated grouping of concepts with other codes of similar content. ATLAS.ti^{TM} software [14] assisted coders in tagging concepts and in cataloging assigned codes using an iterative framework (Fig. 2). Concepts were grouped by similarity of content and analyzed to identify the most relevant expressions and most common language used by patients.

2.5 Methods for Data Quality Assurance

Inter-rater agreement (IRA) analyses assessed consistency in how coders assigned concept codes. Ten percent of transcripts (five of 40) were randomly selected for independent dual-coding by two members of the coding team and compared to evaluate differences in code assignment. Consistency of coding was characterized by agreement in (1) the identification of concepts; and (2) assignment of codes to each identified concept.

To assess saturation (the point at which no more new information is provided), transcripts were ordered chronologically into groups of eight each. Codes reported for each

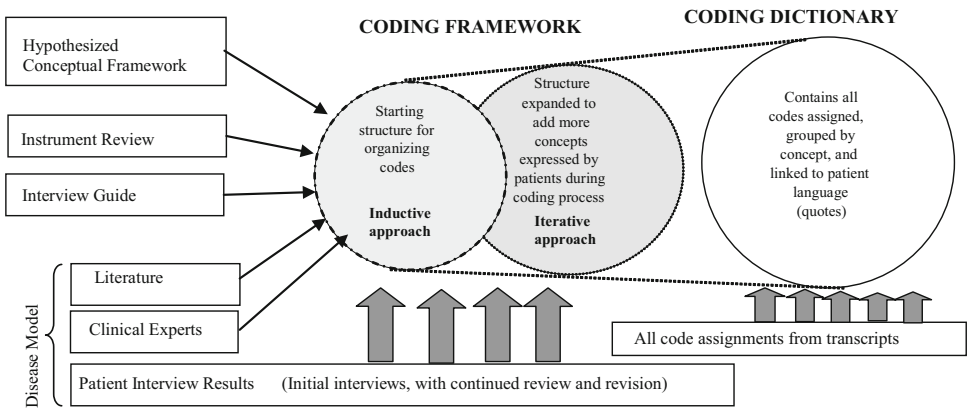

Fig. 2 Developing the coding framework and coding dictionary

subsequent transcript group were compared to codes for the preceding group until no new codes (thus, no new information) appeared. The study used data collection and analysis techniques based on current best practice recommendations for establishing content validity of PRO instruments for medical product evaluation [15, 16].

2.6 Determination of Measurement Strategy and Process for Item Generation

The core development team and the panel of MDD clinical experts met to review the CE results, and determine whether to select an existing PRO instrument or opt to develop a new measure. Tabulated results from the qualitative CE interviews were reviewed alongside concepts identified from published literature and existing instruments, to guide the group's selection of concepts to include in PRO measurement. Data were reviewed within the context of the overall goals of the final measure to accurately reflect treatment-related changes in MDD symptoms that are relevant and meaningful to the patient and important in clinical assessment, with sufficient evidence to allow use of the measure to support medical product label claims in the USA. The agreed list of target concepts was cross-referenced against the content coverage of instruments evaluated during the systematic review to determine if an existing instrument would meet the PRO measurement needs of the working group.

Words and phrases from CE interview data were used to construct the wording of preliminary scale items for each targeted concept. When selecting concepts and drafting item language, the development team determined appropriateness of each potential item against the following criteria: (1) relevance to patients with MDD, as determined by the frequency with which the item was mentioned by patients, particularly when mentioned spontaneously; ratings of bother or importance by patients; and/or other sources of support from qualitative work with patients that indicated relevance; (2) item represents a single, not multidimensional symptom; (3) item is written with vocabulary and phrases commonly used and understood by people with MDD, as informed by the coding dictionary and coding summaries from the qualitative transcript data; (4) clinical expert panel and core development team agree the item is likely to change with successful treatment of MDD; (5) item is unlikely to be vulnerable to ceiling or floor effects in individuals with MDD; (6) item is likely to have semantic or conceptual equivalence in other languages; and (7) item is likely to measure change within the timeframe used for recall in the new measure. During subsequent review, these targeted concepts and preliminary items were further refined, addressing any synonymous/duplicative concepts. A formatted version of the questionnaire was prepared for evaluation in cognitive interviews, translatability assessment, and electronic implementation assessment.

2.7 Cognitive Interviews and Instrument Refinement

Cognitive interviews evaluated relevance, understandability, clarity of language, and structure of preliminary items and their instructions to facilitate further instrument refinement. For the cognitive interviews, new patients naïve to the study were recruited from three of the participating clinical sites (Connecticut, Illinois, and Oklahoma). Each cognitive interview was with one participant in a face-to-face session lasting 60–90 min. During this interview, participants were asked to first self-administer the SMDDS, and then answer a series of interview questions designed to understand their cognitive process with each of the items.

The semi-structured cognitive interview guide standardized the interview and followed a think-aloud process to evaluate each item. During the interview, questions were asked about comprehension and relevance of the individual items; fit of the response scales; appropriateness of the recall period and item wording; and any lack of clarity of items, terminology, instructions, or sentence structure. In some cases, participants were asked to reflect on alternate wording of the item stems with regard to the recall period references and most appropriate phrasing of the symptom concept.

Interviews were audio-recorded, transcriptions were summarized, and participant quotes were organized by like items to facilitate evaluation. Three iterative waves of five interviews each were conducted. Following each wave, the core development team reviewed interviewer notes and refined the instrument based on interview results. An item-tracking matrix recorded changes made to each item during the refinement process.

Parallel with cognitive interviews, experienced PRO linguistics consultants conducted a translatability assessment in five languages (German, Spanish, French, Russian, and Chinese) to identify potential difficulty in translating items while maintaining concept equivalence. The linguists rated the English text for each item regarding level of difficulty in finding a suitable translation that would maintain concept equivalence. Difficulties were rated on a 5-point scale (1 = not difficult at all, 2 = slightly difficult, 3 = moderately difficult, 4 = very difficult, and 5 = extremely difficult). The consultants also provided suggestions and explanations for ways to maintain concept equivalence if translations were possible. Findings were used to make revisions to selected items prior to the closure of the cognitive interview process.

Following instrument modifications made during the first three waves of cognitive interviews, translatability assessment, and expert input, the SMDDS was further evaluated through a formal electronic implementation assessment. This process assessed the viability of implementing the SMDDS on a broad range of available electronic platforms, and collected structured item-level feedback from ePRO system providers on implementation within currently available electronic platforms (tablet, handheld, interactive voice response, Web, and digital pen). The findings were reviewed by the development team and used to make additional formatting changes to the SMDDS prior to finalization for ePRO programming and quantitative testing.

3 Results

3.1 CE Findings

3.1.1 Demographic and Clinical Characteristics

CE interviews were conducted with 40 participants. The average age of the participants was 46.2 (range 21–63) years, 67.5 % were female, and the average HAM-D total score was 24.4 (range 19–39) at enrollment (Table 1). Participants were White/non-Hispanic (45.0 %), Black/African American (22.5 %), or Hispanic (22.5 %); 77.5 % reported completing at least some college education. Average time since initial MDD diagnosis was almost 8 years.

3.1.2 Content Analysis Results

Analysis of the interview transcripts resulted in 3022 coded symptom expressions, grouped into 84 different concepts based on content and similarity of patient expression, within 11 hypothesized symptom subdomains. Five dual-coded transcripts were assessed for IRA, with 97.5–99.1 % agreement between the two coders regarding concept code assigned to text segments. With approximately 8214 words of narrative text per transcript and 5209 assignable codes, these results suggest high IRA.

Saturation of concept was achieved after the fourth of five transcript groups (Table 2). In the first group of eight transcripts, 76 (91 %) of the coded concepts arose. An additional 6 % arose in the second group, 1 % in the third group, and the final 2 % of newly coded concepts appeared in the fourth transcript group. The last remaining group of eight transcripts provided no new information, suggesting additional interviews are unlikely to result in additional concepts being identified, and the sample of 40 interviews was adequate to achieve completeness of concepts from this study population.

3.1.3 Selection of Concepts and Generation of Items for the Symptoms of Major Depressive Disorder Scale

The team reviewed key findings from the qualitative data, literature review, and clinical expert input to identify relevant symptom concepts targeted for inclusion in a PRO instrument. To identify symptom concepts with the strongest support within the qualitative interview data, the development team considered the number of participants expressing each concept, the overall number of coded expressions within the transcripts, severity ratings assigned by participants who reported each symptom, and ratings of how bothersome each symptom was to the participant. This process reduced the initial set of 84 coded concepts to 36 targeted concepts. The key findings from the qualitative interview data and example participant quotes supporting each of these selected concepts are presented in Table 3.

The team then examined the instrument review findings for concepts assessed by existing instruments. While some existing measures included most of the concepts identified through qualitative interviews, no instrument was found suitable for the current purpose, as all would require some degree of modification. The working group decided to develop a new PRO instrument rather than attempt to qualify an existing measure or a modification of an existing measure. Items were drafted for each selected concept to create the preliminary version of the SMDDS, using the coded patient quotations to support the selection of specific wording. The working group decided the measure should assess MDD symptom experience using a 7-day recall period, based on recall period for existing measures, advice of clinical experts, and a reluctance to burden respondents with a daily symptom diary.

The set of newly developed items was reviewed by working group members and clinical experts. Proposed revisions were discussed and adjudicated, and a 36-item preliminary version of the instrument was prepared for use in subsequent cognitive interviews and formal translatability and electric implementation assessments.

3.2 Evaluation and Refinement of the Preliminary Instrument

3.2.1 Cognitive Interviews and Translatability Assessment

Three waves of cognitive interviews included 15 participants. The mean age of the participants was 44.6 years. They were 60.0 % female, 73.3 % White (non-Hispanic),

Table 1 Characteristics of study participants

Characteristic	Concept elicitation ($n = 40$)	Cognitive interviews ($n = 15$)
Age (years)		
Mean (SD)	46.2 (11.8)	44.6 (13.4)
Median	47.0	48.0
Range	21–63	18–59
Sex [*n* (%)]		
Male	13 (32.5)	6 (40.0)
Female	27 (67.5)	9 (60.0)
Marital status [*n* (%)]		
Married	13 (32.5)	6 (40.0)
Living with partner	3 (7.5)	1 (6.7)
Widowed	1 (2.5)	–
Separated	4 (10.0)	2 (13.3)
Divorced	9 (22.5)	3 (20.0)
Never married	10 (25.0)	3 (20.0)
Highest level of education completed [*n* (%)]		
Less than high school	–	–
High school	9 (22.5)	7 (46.7)
Some college	17 (42.5)	5 (33.3)
Bachelor's degree	7 (17.5)	–
Graduate or professional school	7 (17.5)	2 (13.3)
Missing	–	1 (6.7)
Employment [*n* (%)]		
Not employed	18 (45.0)	11 (73.3)
Full-time	14 (35.0)	2 (13.3)
Part-time	7 (17.5)	1 (6.7)
Retired	1 (2.5)	1 (6.7)
Not employed	15 (37.5)	–
Racial and ethnic group [*n* (%)]		
White (non-Hispanic)	19 (47.5)	11 (73.3)
White (Hispanic)	9 (22.5)	2 (13.3)
Black/African American	9 (22.5)	2 (13.3)
Asian	1 (2.5)	–
Other: mixed race	2 (5.0)	–
Household income ($US) [*n* (%)]		
≤9999	9 (22.5)	2 (13.3)
10,000–14,999	2 (5.0)	2 (13.3)
15,000–24,999	3 (7.5)	2 (13.3)
25,000–34,999	5 (12.5)	3 (20.0)
35,000–49,999	6 (15.0)	2 (13.3)
50,000–59,999	4 (10.0)	2 (13.3)
60,000–69,999	4 (10.0)	–
≥70,000 and over	7 (17.5)	2 (13.3)
Self-reported overall health ("How would you rate your overall health?") [*n* (%)]		
Excellent	1 (2.5)	1 (6.7)
Very good	5 (12.5)	2 (13.3)
Good	24 (60.0)	6 (40.0)
Fair	9 (22.5)	5 (33.3)
Poor	1 (2.5)	1 (6.7)

Table 1 continued

Characteristic	Concept elicitation ($n = 40$)	Cognitive interviews ($n = 15$)
No. of years since subject was diagnosed with MDD		
Mean (SD)	7.8 (8.7)	12.3 (12.0)
Median	5.0	7.7
Range	0–40	0.9–42.8
No. of years since onset of most recent major depressive episode		
Mean (SD)	1.0 (1.8)	1.9 (1.5)
Median	0.5	1.4
Range	0–8	0.5–4.8
HAM-D total score[a] at screening		
Mean (SD)	24.4 (4.3)	24.4 (5.3)
Median	23.5	23.0
Range	19–39	19–36

HAM-D Hamilton Rating Scale for Depression, *MDD* major depressive disorder, *SD* standard deviation

[a] The HAM-D total score ranges from a possible 0 to 50. Scores of 14–18 indicate moderate depression, scores of 19–22 indicate severe depression, and scores ≥23 indicate very severe depression

and had an average HAM-D total score of 24.4 at enrollment (Table 1).

During the first wave of five individual interviews, participants expressed confusion with responses for two reverse-scored items addressing the concepts of usefulness and pleasure. They also had difficulty with the transitions between groups of frequency-focused items and those items assessing symptom intensity. Therefore, following wave one, the response scales for the two items were altered to eliminate the need for reverse-scoring and the overall order of items was revised to limit transitions between items assessing different symptom attributes. This reordering resulted in a revised draft for the second wave of interviews in which the 17 intensity-focused items appeared together in the first section of the instrument, followed by the 19 frequency-focused items.

During the second wave, participants lacked clear and distinct interpretation of the exhaustion item, which was subsequently removed from the instrument. Despite the removal of the reverse-scoring for the two items assessing usefulness and pleasure, some participants in the second wave had difficulty with the intensity-focused nature of these two items and felt that frequency was a more natural way to express the concepts. Similarly, information from the translatability assessment indicated that norms of expression in some languages (e.g., French, German) would require phrasing these concepts as frequency-focused items. Based on these cognitive interview findings and confirmation from the translatability assessment, the two items were altered to focus on frequency rather than intensity. Two additional items (assessing cognitive lethargy and fixation on problems) were reworded to focus more directly on the intended concepts. For example, the phrasing of one item stem was changed from "how difficult was it for you to think clearly?" to "how difficult was it for you to think as quickly as you wanted?" to better focus the item on the symptom concept of cognitive lethargy for testing in the third wave.

During the third wave, the item modifications from the previous waves were confirmed, and one additional wording modification was made to the item assessing pleasure (specifically, "how much of the time have you looked forward to things with enjoyment?" was simplified to "how much of the time did you feel enjoyment?"). Findings from the cognitive interviews confirmed the relevance of the selected concepts and the appropriateness of the 7-day recall period. Over the three waves, one item was removed (exhaustion), and the content of the four items described above was substantially modified based on cognitive interview and translatability assessment findings.

3.2.2 *Electronic Implementation Assessment*

Additional formatting and wording modifications were made based on the results of the electronic implementation assessment. The tabular format of the instrument was replaced with a layout of self-contained items to facilitate a single-item-per-screen ePRO implementation. Bolded text formatting was removed, and the recall period reference was standardized to "over the past 7 days" in the instructions and all items.

The resulting SMDDS contains 35 items that measure each concept using a 5-point verbal rating scale and a 7-day retrospective recall period. Items in the SMDDS are

Table 2 Saturation of coded symptom concepts

MDD symptom domain	Symptom concept	Group 1 ($n = 8$ transcripts)	Group 2 ($n = 8$ transcripts)	Group 3 ($n = 8$ transcripts)	Group 4 ($n = 8$ transcripts)	Group 5 ($n = 8$ transcripts)
Negative emotions/mood	Anger	X				
	Crying	X				
	Decreased pleasure in things		X			
	Despair	X				
	Empty	X				
	Frustration	X				
	Irritability/hostility	X				
	Less compassion	X				
	Mood swings	X				
	Numbness		X			
	Rage	X				
	Sadness	X				
Negative affect	Feeling lonely	X				
	Focus on negative	X				
	Guilt	X				
	Hopeless/helpless	X				
	Shame	X				
	Worthlessness	X				
Anxiety	Anxiety	X				
	Fear	X				
	Nervousness	X				
	Panic attack	X				
	Stressed	X				
	Worried	X				
Low energy	Drained	X				
	Fatigue/exhaustion	X				
	Lethargic	X				
	No/low energy	X				
	Sleepiness	X				
	Tiredness	X				
	Weakness			X		
Cognition	Cognitive lethargy	X				
	Daydreaming	X				
	Distracted	X				
	Feeling overwhelmed	X				
	Fixation on problems	X				
	Impulsiveness	X				
	Indecisiveness	X				
	Intrusive thoughts	X				
	Memory issues	X				
	Poor comprehension				X	
	Poor concentration	X				
	Racing thoughts	X				
Physical symptoms	Breathing problems	X				
	Chest pressure		X			
	Dizziness	X				

Table 2 continued

MDD symptom domain	Symptom concept	Group 1 ($n = 8$ transcripts)	Group 2 ($n = 8$ transcripts)	Group 3 ($n = 8$ transcripts)	Group 4 ($n = 8$ transcripts)	Group 5 ($n = 8$ transcripts)
	Gastrointestinal problems	X				
	Headaches	X				
	Heart palpitations	X				
	Pain	X				
	Muscle stiffness	X				
	Restlessness	X				
	Stomach discomfort	X				
	Sweat		X			
	Tingling in extremities	X				
Sleep disturbances	Early awakening	X				
	Difficulty falling asleep	X				
	General sleep issues	X				
	Insomnia		X			
	Oversleeping	X				
	Difficulty remaining asleep	X				
Eating behavior	Decreased appetite	X				
	Increased appetite	X				
	Overeating	X				
	Under-eating	X				
	Weight gain	X				
	Weight loss	X				
Low motivation	Desire to be alone	X				
	Lack of drive	X				
	Less/lack of interest	X				
	No interest in activities	X				
	No interest in chores	X				
	No interest in leaving home	X				
	No interest in self-care	X				
	Not wanting to get out of bed	X				
Sense of self	Hate self	X				
	Low self-efficacy	X				
	Low self-esteem	X				
	Self-blame	X				
	Victim	X				
Self-harm/suicide	Better off dead	X				
	Self-harm				X	
	Suicidal ideation	X				
	Thoughts of death	X				
	Number of concepts coded in each group	76	5	1	2	0
	Percentage of relevant symptom concepts coded ($n = 84$)	90.5	6.0	1.2	2.4	0.0

Saturation was calculated across groups of transcripts ordered chronologically by interview date. The first occurrence of each concept is indicated with an 'X'

hypothesized to be organized into 11 domains (Table 3). Based on the findings from the CE and cognitive interviews, frequency and intensity response scale options adequately fit their symptom stems. Sixteen of the items focus on the intensity of symptoms with a rating scale from "not at all" to "extremely." Nineteen items focus on frequency or the amount of time a symptom was experienced, employing a rating scale from "never" to "always." To illustrate the overall question structure and rating scales employed in both the intensity and frequency items, Fig. 3 provides example SMDDS items from the version currently undergoing quantitative testing.

4 Discussion

The SMDDS qualitative development efforts followed the principles outlined in the FDA's PRO Guidance [10] and best practices for establishing PRO instrument content validity described by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) PRO Content Validity Good Research Practices Task Force [15, 17]. The SMDDS is intended for use as a primary efficacy endpoint in clinical trials assessing treatment benefit in adults with MDD. The research described here represents a critical step toward establishing the SMDDS as fit for purpose, providing qualitative evidence of the content validity of the SMDDS.

Content validity refers to the extent to which an instrument appropriately and comprehensively covers all facets of the concept to be measured relative to the intended context of use [10]. Direct input from the target population is essential to ensure content completeness and relevance. Information gathered from the literature review coupled with qualitative evidence, collected from clinicians and CE interviews with participants with MDD, demonstrates achievement of concept saturation. The cognitive interview process led to refinements to item content and instructions so that the SMDDS addresses symptoms relevant to MDD patients, with response scales meaningful to the manner in which patients reflect on their own symptoms. The cognitive interview patient participants considered 7 days a feasible recall period. However, further exploration of a daily recall may be useful. This is especially important in the context of treatment benefit claims that are linked to elements of time (e.g., time to treatment effect and durability of response).

The SMDDS includes one or more items addressing eight of the nine symptomatic criteria used in DSM-IV-TR and DSM-5, including pairs of items covering two of the potentially bidirectional symptoms (oversleeping or insomnia, overeating or undereating). The SMDDS does not include an item addressing psychomotor agitation and has only one item that partly addresses psychomotor retardation, specifically slowed thinking. Both agitation and visible psychomotor retardation are more characteristic of severe depressive episodes (as compared to milder depressive episodes) [18–20]. Moreover, there is evidence that psychomotor agitation is difficult to rate reliably [21] and may not be one of the 'core' symptoms of MDD [20, 22–24]. As such, we felt that it was best to leave such ratings to highly trained clinical evaluators [25]. Nevertheless, we realize that not including an item assessing psychomotor agitation is a potential limitation of this scale.

The current version of the SMDDS also assesses emotional, psychological, and somatic symptoms commonly experienced in MDD but which do not appear in the DSM-IV-TR or DSM-5 criteria lists. These include anger, frustration, irritability, loneliness, nervousness, shortness of breath, headaches, aches and pains, and self-dislike. It is to be determined if severities of these common associated symptoms co-vary with the core symptoms across longitudinal follow-up, and if the burden of persistent or unremitting associated symptoms can help provide a more finely grained assessment of which patients have truly recovered and which warrant further or additional treatment.

The SMDDS offers much broader coverage of the non-criteria symptoms associated with MDEs than two of the commonly used scales, the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR) and Patient Health Questionnaire-9 (PHQ-9). The SMDDS is closer to the Inventory of Depressive Symptomatology-Self-Report (IDS-SR), which includes broad coverage of associated symptoms [26]. It is the intention of the development team to evaluate the SMDDS for further item reduction after quantitatively assessing its item functioning and identifying redundant or misfit items.

The third widely used self-report is the Beck Depression Inventory (BDI), the first PRO measure in this area to gain wide acceptance. The BDI and its more recent revision (BDI-II) have strong coverage of the so-called cognitive triad of depression (negative thoughts about self, world, and others), but incomplete coverage of the symptom criteria of an MDE. The BDI preceded the Diagnostic and Statistical Manual of Mental Disorders (Third Edition) (DSM-III) [28] by 19 years, however, and its development was led by the same researcher who developed cognitive therapy for depression [27].

The current preliminary version of the SMDDS has 35 items covering 11 hypothesized domains that comprehensively address clinically relevant symptoms of MDD that are important and meaningful to patients. The development of the SMDDS has included direct patient involvement through qualitative interviews to establish content validity for the included concepts, evidence that will be a critical

Table 3 Major depressive disorder symptom concepts included in the Symptoms of Major Depressive Disorder Scale (through the cognitive interview phase)

MDD symptom domain	Symptom concept	Number (%) of patients expressing concept (among $n = 40$ total interviews)	Number of coded expressions of concept[a] (among 3022 total expressions)	Mean severity rating[b]	Mean bothersome rating[c]	Example patient language supporting concepts expressed during CE interviews
Negative emotions/mood	Anger	27 (67.5)	105	7.9	7.5	I get angry for any little thing
						I start to get angry with people
						I get annoyed and angry, … get angry easily
						I am reacting with anger, inside of me
	Frustration	13 (32.5)	35	8.0	7.3	Feeling frustrated … frustration
						I get very frustrated … it frustrates me
						My whole frustration is that I am stuck like this
						I get frustrated in my head
	Crying	14 (35.0)	36	7.8	6.1	Crying about nothing … crying over silly things
						Tears were coming down and I wasn't even aware
						I just start crying … I was crying so much
						Why do I feel like my whole body wants to cry?
	Hopeless/helpless	18 (45.0)	48	8.3	8.0	Feel hopeless because you are in a rut
						Feel helpless, powerless … hopeless and sometimes helpless
						I have a lot of hopelessness feeling
						Nothing is going to change, and nothing I can do about it
	Irritability	29 (72.5)	110	7.5	7.1	I was irritable and didn't want to be there
						Irritability at times … they tell me I am irritable
						At the slightest little thing, will become very irritated
						Everything bothers me and I get irritated
	Sadness	32 (80)	181	8.0	7.6	I don't know why I was sad … I go and be in a sad place
						I just feel sad … general sadness … I do feel sad
						I am so sad … I have a lot of sadness in my life
						This profound sadness that I felt … mostly I'm sad
	Pleasure in doing things	11 (27.5)	47	6.4	7.9	I don't get enjoyment out of anything
						Not being able to enjoy pleasurable things
						Things I loved doing, don't really do anymore
						I don't get the joy out of what I used to
Negative affect	Feeling lonely	28 (70.0)	60	8.1	7.6	I feel like I'm in this alone, … feeling lonely … I am alone
						Feel very alone … you tend to feel lonely
						I feel lonely, nobody cares about me … Loneliness
						I have people but I still feel lonely
	Worthlessness	11 (27.5)	39	9.0	9.0	Being … worthless in society, not worthy
						I don't deserve anything … I don't have a place in the world
						I felt so worthless
						I started feeling that my life wasn't worth anything at all

Table 3 continued

MDD symptom domain	Symptom concept	Number (%) of patients expressing concept (among $n = 40$ total interviews)	Number of coded expressions of concept[a] (among 3022 total expressions)	Mean severity rating[b]	Mean bothersome rating[c]	Example patient language supporting concepts expressed during CE interviews
Anxiety	Feeling overwhelmed	25 (62.5)	87	8.7	8.5	Being overwhelmed, having too much … feeling overwhelmed
						A lot of things hitting me at once … get overwhelmed easily
						Can't deal with it, became overwhelming
						How am I going to do everything, I am just overwhelmed
	Anxiety/ nervousness	28 (70.0)	148	8.1	7.9	I get nervous all the time, don't know why
						Nerves become edgy … I get very nervous
						Anxiety, … anxious, … tense, … a lot of anxiety
						Anxiety about tripping and falling … a sense of tension
	Worry	23 (57.5)	59	8.4	7.1	I get really worried, I always worry, I worry about everything
						Worry too much about little things … worried about myself
						I will be worried and anxious at the same time
						I wish I could be less worried … I am a worrier
Low energy	Tiredness	30 (75.0)	95	7.4	8.0	I can sit on a couch and feel tired, … tiredness, … just tired
						Get home tired, feel physically tired, I'm tired all the time
						I've not done anything for a week, I'm just tired
						just tired, no energy, … no reason for me to feel tired
Cognition	Cognitive lethargy	8 (20.0)	23	8.5	9.0	It's like on slow motion, everything slow motion
						My capacity of thinking or my brain slows up
						My thought process is very slow
						Try to process something and it will take you longer/harder
	Intrusive thoughts	14 (35.0)	32	5.5	7.7	I can't seem to rid that of my mind
						I try to think of nice things but it just goes right back to the bad stuff
						Start dwelling upon yourself (and all the other problems)
						Think negative things, continuous, played in my head
	Poor concentration	22 (55.0)	67	7.0	8.6	Ability to focus is not good, reading is hard, listening to people mind wanders
						Focus and concentration is fragmented
						I am already having a hard time concentrating
						I can't keep my concentration
	Difficulty remembering	16 (40.0)	40	4.8	6.5	A little bit of my memory is not as sharp, … forgetfulness
						I don't remember saying something, I lose a lot of things, misplace a lot of things
						Find myself forgetting to do certain things I should be doing
						There are times when I totally miss them (appointments), I don't remember

Table 3 continued

MDD symptom domain	Symptom concept	Number (%) of patients expressing concept (among $n = 40$ total interviews)	Number of coded expressions of concept[a] (among 3022 total expressions)	Mean severity rating[b]	Mean bothersome rating[c]	Example patient language supporting concepts expressed during CE interviews
Physical symptoms	Breathing problems	10 (25.0)	16	–	9.0	I don't have any breath
						I feel like somebody is choking me so I lose my breath
						Shortness of breath sometimes
						Feeling like can't breathe
	Headaches	18 (45.0)	54	7.0	5.4	Been having headaches more
						I get a lot of headaches
						I get that extreme headache … my depression induces it
						I get more headaches than I have ever gotten
	Bodily pain	18 (45.0)	42	8.5	8.1	Always getting pains somewhere
						Aches and pains
						I am very achy
						(I get) general bodily pain a lot
	Gastrointestinal problems/ stomach discomfort	18 (45.0)	78	7.3	7.1	A little upset to my stomach
						Digestion wasn't great all of a sudden, stomach was turning all the time
						I definitely know that the stomach aches are related to depression
						The bowel, your stomach is upset
Sleep disturbances	General sleep adequacy	26 (65.0)	121	9.0	7.6	Can't really sleep that much at night
						Hard to get up in morning, only sleep 3 to 5 h a night
						I didn't get a good night's sleep
						I don't get good sleep period
	Oversleeping	15 (37.5)	54	6.8	5.9	I can sleep all day, get up for half an hour, be exhausted, and go back to sleep all day again, 24 h
						I will oversleep, … sleeping so much
						I would sleep for three days at a time
						When I am depressed, I can sleep for 24 h
Eating behaviors	Under-eating	14 (35.0)	27	7.9	7.7	A lot of times if I'm not eating it is a sign I am depressed
						I am not eating as much or as regular
						Sometimes I don't eat anything
						I'll have to force myself to eat something
	Overeating	19 (47.5)	48	7.9	7.7	Eating too much, … I started eating a lot, … I did overeat
						Even if I just ate, I will say I am going to get some crackers
						I am not even hungry and will figure out a reason
						Overeat because It's enjoyable, food becomes medication

Table 3 continued

MDD symptom domain	Symptom concept	Number (%) of patients expressing concept (among $n = 40$ total interviews)	Number of coded expressions of concept[a] (among 3022 total expressions)	Mean severity rating[b]	Mean bothersome rating[c]	Example patient language supporting concepts expressed during CE interviews
Low motivation	Not wanting to get out of bed	23 (57.5)	60	7.1	7.2	Didn't want to get out of bed, … didn't want to wake up
						I just didn't want to get up, … wouldn't get out of bed
						I stay in bed, don't feel like getting up, but can't fall asleep
						Was a struggle to get out of bed
	Less/lack of interest	19 (47.5)	23	7.6	6.6	Have lots of things to do but will just go and lie down
						I just don't feel like doing anything
						Don't take as much interest as did before
						I lost interest in doing a lot around my house
	Lack of drive	25 (62.5)	48	7.5	8.3	Seems like I (have to) push myself to do things now
						Hard to get motivated to go exercise
						I just didn't push myself anymore
						It is lack of interest because I have to force myself to do things I used to do
	No interest in activities	8 (20.0)	18	10.0	8.0	Just want to go back and lay in my bed and do nothing
						Want to lay down, not doing anything I have to do
						Not wanting to do anything
						Used to coach, don't feel like it anymore, don't feel like doing anything
Sense of self	Dislike self	7 (17.5)	14	7.9	7.5	Hate my body, hate myself, don't look in the mirror anymore
						I insult myself, fighting with myself, beating up on myself
						Self-loathing (had when I drank)
						I am boring, don't feel great about myself
	Self-criticism	10 (25.0)	34	9.0	6.8	Start beating up on myself … will criticize myself
						Getting down on myself … feel bad about myself
						Just no pride in what I do … can't do anything right
						I don't have anyone else criticizing me, it's me doing it
	Usefulness	24 (60.0)	85	9.5	8.3	Can I really do this? Can I do anything?
						Felt useless to my kids, felt very out of control
						Why can't I get out of this by myself? Why can't I do better?
						I was crying all the time, I can't do anything
	Self-blame	10 (25.0)	34	9.0	6.8	I think I'm the one who caused it … everything is my fault
						I am blaming myself, something is wrong with me
						It will be my fault if everything crashes and burns
						So many times, when bad things happen I blame myself

Table 3 continued

MDD symptom domain	Symptom concept	Number (%) of patients expressing concept (among $n = 40$ total interviews)	Number of coded expressions of concept[a] (among 3022 total expressions)	Mean severity rating[b]	Mean bothersome rating[c]	Example patient language supporting concepts expressed during CE interviews
Self-harm/suicide	Feeling better off dead	6 (15.0)	11	–	6.0	Feel life is not worth living, … wish I was dead
						If I was dead, I wouldn't have to go through this
						Not wanting to continue to live
						Thought always passes by that I am better off dead
						What's the point of living when I'm depressed and sad all the time?
	Suicidal ideation	5 (12.5)	9	–	7.4	Got to do it in a way that it can be properly faked (life insurance to pay off)
						I think about suicide sometimes and I think it is not such a bad way to go … a possibility
						Used to have thoughts about it … thoughts of suicide
						Occasionally I fixate on methods of suicide
	Thoughts of death	17 (42.5)	31	6.5	10.0	I don't even fear it, once I was sick and thought I would die
						I thought about dying … a sense that I could escape that way
						Thoughts of death, something happening to me
						Sometimes I feel I don't want to be here

Although not listed in this table, "Exhaustion" was included as one of the initial 36 concepts in the draft SMDDS instrument, but was later removed based on results of the cognitive interviews. Conversely, "self-blame" was originally combined with "self-criticism" into a single item, but was split into two separate items in later revisions to the instrument

CE concept elicitation, *MDD* major depressive disorder, *SMDDS* Symptoms of Major Depressive Disorder Scale

[a] In total, 3022 total expressions were coded for the 84 symptom concepts. Each concept was represented by between 2 and 181 coded expressions, with a mean of 36 mentions per concept

[b] For each MDD symptom they experienced, interview participants rated the severity of the symptom "at its worst" on a 0–10 scale anchored with "none" and "very severe." Across all concepts, ratings ranged from 2 to 10, with a mean of 7.6

[c] For each symptom they experienced, participants rated "how much that particular symptom bothers you" on a 0–10 scale anchored with "not bothersome" and "extremely bothersome." Across all concepts, the bothersome ratings given ranged from 0 to 10, with a mean of 7.3

Sample Symptom Intensity Item	*Sample Symptom Frequency Item*
6. Over the past 7 days, how sad have you felt? ☐ Not at All ☐ A Little Bit ☐ Moderately ☐ Quite a Bit ☐ Extremely	30. Over the past 7 days, how much of the time did you feel critical about yourself? ☐ Never ☐ Rarely ☐ Sometimes ☐ Often ☐ Always

Source: Example items are from the Symptoms of Major Depressive Disorder Scale (SMDDS©) and are used with permission of the Critical Path Institute.

Fig. 3 Example items from current developmental version of the Symptoms of Major Depressive Disorder Scale (SMDDS)

component of the FDA qualification process. In qualifying the SMDDS for use in clinical trials, the next steps are to 1) conduct additional cognitive interviews to assess the conceptual equivalence of the paper and ePRO formats, and to 2) collect quantitative evidence to refine and confirm the item content, explore response scale distribution anomalies, and test potential subscale structure. Additionally, quantitative evidence of measurement properties such as internal consistency reliability, reproducibility, construct validity, and responsiveness will need to be gathered. Finally, guidance for interpreting and defining a clinically meaningful change in scores for the SMDDS must be established.

When finalized and qualified by the FDA, the SMDDS will be made publicly available and is intended to be suitable for implementation on a variety of data collection platforms. The development of the SMDDS for qualification as a drug development tool for assessing treatment benefit from the patient perspective has the potential to support product labeling claims beyond those measured by currently available MDD measures.

Acknowledgments This research was funded by the following members of the PRO Consortium: AbbVie, Bristol-Myers Squibb, Eli Lilly and Company, Forest Research Institute, Janssen, Pfizer, Roche Products Limited, Shire, Sunovion, and Takeda. The PRO Consortium receives support through Grant U01FD003865 from the United States Food and Drug Administration to the Critical Path Institute.

The authors would like to thank Linda Carpenter, MD and Madhukar Trivedi, MD for providing their clinical experience and insight to the development process as members of the clinical expert panel (in addition to Dr. Michael E. Thase, who also participated as an author). We would also like to thank Carla Ascoytia, Cecilia Dedios, and Matthew Wolfe for their valuable contributions to the study's data collection and analysis efforts.

Conflict of interest disclosures Linda S. Deal (Shire: employee, travel support, holder of stock/stock options), Lucy Abraham (Pfizer: employee, travel support, holder of stock/stock options), Steven I. Blum (Actavis/Forest Research Institute: employee, travel support, and stock/stock options), and Elizabeth N. Bush (Eli Lilly: employee, travel support, holder of stock/stock options) disclose their employment and other financial relationships with their respective firms.

Michael E. Thase has provided scientific consultation to Alkermes, AstraZeneca, Bristol-Myers Squibb Company, Dey Pharma, L.P., Eli Lilly and Company, Forest Pharmaceuticals, Inc., Gerson Lehman Group, GlaxoSmithKline, Guidepoint Global, H. Lundbeck A/S, Janssen Pharmaceuticals; MedAvante, Inc., Merck and Co. Inc., Neuronetics, Inc., Novartis, Otsuka, Ortho-McNeil Pharmaceuticals, PamLab, L.L.C., Pfizer (formerly Wyeth-Ayerst Laboratories), Schering-Plough (formerly Organon, Inc.), Shire US Inc., Sunovion Pharmaceuticals, Inc., Takeda (Lundbeck), and Transcept Pharmaceuticals. Dr. Thase receives grant funding from the Agency for Healthcare Research and Quality, Eli Lilly and Company, the National Institute of Mental Health, Otsuka Pharmaceuticals, and Sepracor, Inc. He has equity holdings in MedAvante, Inc. and receives royalty income from the American Psychiatric Foundation, Inc., Guilford Publications, Herald House, Oxford University Press, and W.W. Norton & Company. His wife is employed as the Group Scientific Director for Embryon—formerly Advogent, which does business with BMS and Pfizer/Wyeth.

Kelly P. McCarrier, Mona L. Martin, and Stephen Joel Coons declare no conflicts of interest.

Author contributions All authors were involved in the conception and planning of the work that led to the manuscript, as well as interpretation of the data. Kelly P. McCarrier and Mona L. Martin also led the collection and analyses of the study data. All authors participated in drafting or critical revisions of the manuscript and approval of the final submitted version. Kelly P. McCarrier acts as overall guarantor for this study.

References

1. Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, et al. Prevalence and treatment of mental disorders, 1990 to 2003. N Engl J Med. 2005;352:2515–23.
2. Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry. 2015;76(2):155–62.

3. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington: American Psychiatric Publishing; 2013.
4. Hamilton M. A rating scale for depression. J Neurol Neurosur Psychiatry. 1960;23:56–62.
5. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382–9.
6. Foley DL, Neale MC, Gardner CO, Pickles A, Kendler KS. Major depression and associated impairment: same or different genetic and environmental risk factors? Am J Psychiatry. 2003;160:2128–33.
7. Uher R, Perlis RH, Placentino A, Dernovšek MA, Henigsberg N, Mors O, et al. Self-report and clinician-rated measures of depression severity: can one replace the other? Depress Anxiety. 2012;29:1043–9.
8. Coons SJ, Kothari S, Monz BU, Burke LB. The Patient-Reported Outcome (PRO) Consortium: filling measurement gaps for PRO endpoints to support labeling claims. Clin Pharmacol Ther. 2011;90:743–8.
9. US Food and Drug Administration. Guidance for industry and FDA staff: qualification process for drug development tools. 2014. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf. Accessed 29 Jan 2014.
10. US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed 6 Jun 2013.
11. Ball S, Dedios C, Abraham L. The patient's perspective on major depressive disorder: what do we know? J Ment Health Policy Econ. 2012;15(Suppl 1):S1.
12. Blum SI, Bush EN, Bushnell DM. Systematic review of patient-reported outcome measures used to assess symptoms associated with major depressive disorder. J Ment Health Policy. 2012;15:S2–3.
13. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed., text revision. Washington, DC: American Psychiatric Association; 2000.
14. Friese S. ATLAS.ti 7 User Guide and Reference. Berlin: ATLAS.ti Scientific Software Development GmbH. 2014. http://atlasti.com/wp-content/uploads/2014/05/atlasti_v7_manual_en_201409.pdf. Accessed 10 June 2015.
15. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. Value Health. 2011;14:967–77.
16. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. Qual Life Res. 2009;18:1263–78.
17. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. Value Health. 2011;14:978–88.
18. Parker G, Roy K, Hadzi-Pavlovic D, Wilhelm K, Mitchell P. The differential impact of age on the phenomenology of melancholia. Psychol Med. 2001;31(7):1231–6.
19. Morriss R, Leese M, Chatwin J, Baldwin D, THREAD Study Group. Inter-rater reliability of the Hamilton Depression Rating Scale as a diagnostic and outcome measure of depression in primary care. J Affect Disord. 2008;111(2–3):204–13.
20. Bech P, Fava M, Trivedi MH, Wisniewski SR, Rush AJ. Factor structure and dimensionality of the two depression scales in STAR*D using level 1 datasets. J Affect Disord. 2011;132(3):396–400.
21. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry. 2004;161(12):2163–77.
22. Bech P, Allerup P, Gram LF, Reisby N, Rosenberg R, Jacobsen O, Nagy A. The Hamilton Depression Scale. Evaluation of objectivity using logistic models. Acta Psychiatr Scand. 1981;63(3):290–9.
23. Bech P, Wilson P, Wessel T, Lunde M, Fava M. A validation analysis of two self-reported HAM-D6 versions. Acta Psychiatr Scand. 2009;119(4):298–303.
24. Licht RW, Qvitzau S, Allerup P, Bech P. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? Acta Psychiatr Scand. 2005;111(2):144–9.
25. Hadzi-Pavlovic D, Hickie I, Brodaty H, Boyce P, Mitchell P, Wilhelm K, et al. Inter-rater reliability of a refined index of melancholia: the CORE system. J Affect Disord. 1993;27(3):155–62.
26. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry. 2003;54(5):573–83.
27. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive therapy of depression. New York: Guilford; 1979.
28. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed. Washington, DC: American Psychiatric Association; 1980.