Systematic Review of Patient-Reported Outcome Measures Used to Assess Symptoms Associated with Major Depressive Disorder

Steven I. Blum, MBA
Elizabeth Nicole Bush, MHS
Donald M. Bushnell, MA
on behalf of the Patient-Reported Outcome (PRO) Consortium Depression Working Group

Conflict of Interest and Financial Disclosure

- **Steven I. Blum, MBA**
  - Co-Chair of PRO Consortium Depression Working Group
  - Employee of Forest Research Institute, Inc., a subsidiary of Forest Laboratories, Inc., and a shareholder of Forest Laboratories, Inc.
- **Elizabeth Nicole Bush, MHS**
  - Employee of Eli Lilly and Company
- **Donald M. Bushnell, MA**
  - Employee of Health Research Associates, a consulting firm, which has received funding from the PRO Consortium Depression Working Group
- **This research was sponsored by the individual member firms of the PRO Consortium Depression Working Group, which includes:**
  - **Abbott Laboratories**, Abbott Park, IL, USA; **Bristol-Myers Squibb Company**, Wallingford, CT, USA; **Eli Lilly & Company**, Indianapolis, IN, USA; **Forest Laboratories, Inc.**, Jersey City, NJ, USA; **Janssen Global Services, LLC***, Raritan, NJ, USA, **Pfizer Inc.**, New York, NY, USA; **Shire Development Inc.**, Wayne, PA, USA; **Sunovion Pharmaceuticals Inc.**, Marlborough, MA, USA
- **Critical Path Institute's PRO Consortium** is supported by grant No. U01FD003865 from the United States Food and Drug Administration and by Science Foundation Arizona under Grant No. SRG 0335-08.

2

*Janssen joined PRO Consortium Depression Working group subsequent to submission of abstract

Acknowledgements

- **The authors would like to thank and acknowledge the following individuals/organizations for their contributions to this research**
- **Critical Path Institute (C-Path Staff):**
  - Stephen Joel Coons, Theresa Flemming, Theresa Griffey, Karla Lehmann, Jason Lundy, Elizabeth Piault-Louis
- **Depression Working Group member firm representatives:**
  - **Abbott:** Nicholas Greco, Steven Haas, Christy Houle; **BMS:** Justin Doan; **Lilly:** Susan Ball, Elizabeth N. Bush, Risa Hayes; **Forest:** Steven Blum, Maju Mathews, Abhilasha Ramasamy; **Janssen:** Carol Jamieson; **Pfizer:** Lucy Abraham, Brendon Binneman, Philip Ninan (retired, non-member participant); **Shire:** Linda Deal, Bryan Dirks, Manisha Madhoo, Dylan Supina; **Sunovion:** Mariam Hassan, Kitty Rajagopalan
- **Health Research Associates:**
  - Don Bushnell, Cecilia Dedios, Mona Martin, Kelly McCarrier
- **Expert Panel Members:**
  - Linda L. Carpenter, MD - Butler Hospital/Brown University, Providence, RI
  - Michael E. Thase, MD - University of Pennsylvania, Philadelphia, PA
  - Madhukar H. Trivedi, MD - UT Southwestern Medical Center, Dallas, TX

About Critical Path Institute and the PRO Consortium

- The **Critical Path Institute (C-Path)** is an independent, non-profit institute, unaffiliated with any single entity or interest group, created in 2005 by the University of Arizona and the U.S. Food and Drug Administration (FDA).
  - It is dedicated to bringing scientists from the FDA, industry, and academia together to improve the path for innovative new drugs, diagnostic tests and devices to reach patients in need.
- C-Path, in cooperation with the FDA and the medical products industry, formed the **Patient-Reported Outcome (PRO) Consortium** in 2008 for the purpose of developing, evaluating, and qualifying PRO instruments with the FDA for use in clinical trials designed to evaluate the safety and efficacy of medical products
  - Current working groups include: Asthma, Cognition (mild cognitive impairment), **Depression**, Functional Dyspepsia, Irritable Bowel Syndrome, Non-Small Cell Lung Cancer, and Rheumatoid Arthritis
- http://www.c-path.org/PRO.cfm

Working Group Objectives

- The Depression Working Group is working toward qualification of a PRO measure intended for use as a primary or key secondary endpoint in clinical trials for major depressive disorder to assess treatment efficacy from the patient's perspective.

- Measure will have to satisfy the FDA PRO Guidance and will be qualified through the FDA's Drug Development Tools (DDT) process
  - DDT Draft guidance issued October 2010

Research Objective

- As part of the Content Validity Stage of development we conducted a systematic review to assess key characteristics of a selection of existing instruments

- This study along with other data will help to guide working group on whether to:
    - Seek qualification of an existing instrument,
    - Recommend modifications to an existing instrument; or
    - Develop a new instrument (items from existing instruments could contribute to item bank).

- The instrument review helped to inform on the revision of conceptual (disease) model, identification of preliminary key concepts to measure, and development of our qualitative interview guide

Background – PRO Guidance

- In December 2009, FDA issued the final guidance for industry: ***Patient-Reported Outcome Measures: Use in Medical Product Development to Support Label Claims***
    - The "PRO Guidance" describes how the FDA reviews and evaluates existing, modified, or newly created PRO instruments used to support claims in approved medical product labeling in the United States
- The evaluation of a PRO instrument by FDA includes the following considerations (or context of use):
    - The population enrolled in the clinical trial
    - The clinical trial objectives and design
    - The PRO instrument's conceptual framework
    - The PRO instrument's measurement properties

Background – PRO Guidance

- Because the purpose of a PRO measure is to capture the patient's experience, an instrument will not be a credible measure without evidence of its usefulness from the target population of patients.

  - Sponsors should provide documented evidence of patient input during instrument development and of the instrument's performance in the specific application in which it is used (i.e., population, condition).

- An existing instrument can support a labeling claim if it can be shown to reliably measure the claimed concept in the patient population enrolled in the clinical trial.

Source: Guidance for Industry - Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Final version, December 2009

Aims of the Study

- To identify existing self-reported measures for Major Depressive Disorder (MDD) symptoms and assess:

  - Their developmental history,

  - Concepts measured,

  - Measurement properties, and

  - Use in clinical studies

Methods

- Systematic review of existing MDD symptom measures and related published literature was conducted using PubMed, University of Oxford PRO Measurement Group and the Cochrane Library
  - The following combinations of keywords were used for the search: *"Patient Reported Outcome(s)", "Clinician Reported Outcome(s) AND "Depression", "Depressive Symptoms", "Depressed Mood", "Depression Index", "Depression Scale(s)", "Depression Instrument(s)", "Depression Measure(s)"*
- Conducted searches of the following internet sources
  - ProQolid (Patient Reported Outcome and Quality of Life Instruments Database)
  - OLGA (Online Guide to Quality of Life Assessment)
  - ISPOR Databases (International Society for Pharmacoeconomics and Outcomes Research)

Methods

Search limited to those articles and instruments in English for which information on both their development process and psychometric properties were available.

Category	Inclusion Criteria	Exclusion Criteria
Languages	English only	Other languages
Population	Humans only	Animal studies
Type of Studies	Cross-sectional or longitudinal; Used original data; included patients with depression	Letter, editorial, commentary, discussion paper, non-systematic reviews that have no original data, practice guidelines
Type of Instruments	Patient reported outcomes; Clinician Reported outcomes; Measure depressive symptoms; Had to describe development process; Had to include psychometric properties	Instruments which do not measure depressive symptoms or primarily measure health related quality of life, functional status and satisfaction

Methods

- Instruments were ranked based on number of citations from the Institute of Scientific Information Web of Science database with detailed reviews conducted for top-cited instruments.
- Areas of inquiry for candidate instrument review:
  - Name of Measure, Acronym, and Purpose
  - Concepts measured, number of items in each (sub)scale
  - Overall content coverage of targeted concepts
  - Developmental History
  - Published use in mental health studies & in clinical trials
  - Reliability results (internal consistency, test-retest reproducibility)
  - Results related to convergent and discriminant validity
  - Ability to detect change over time
  - Other useful information

Search Strategy and Outcome
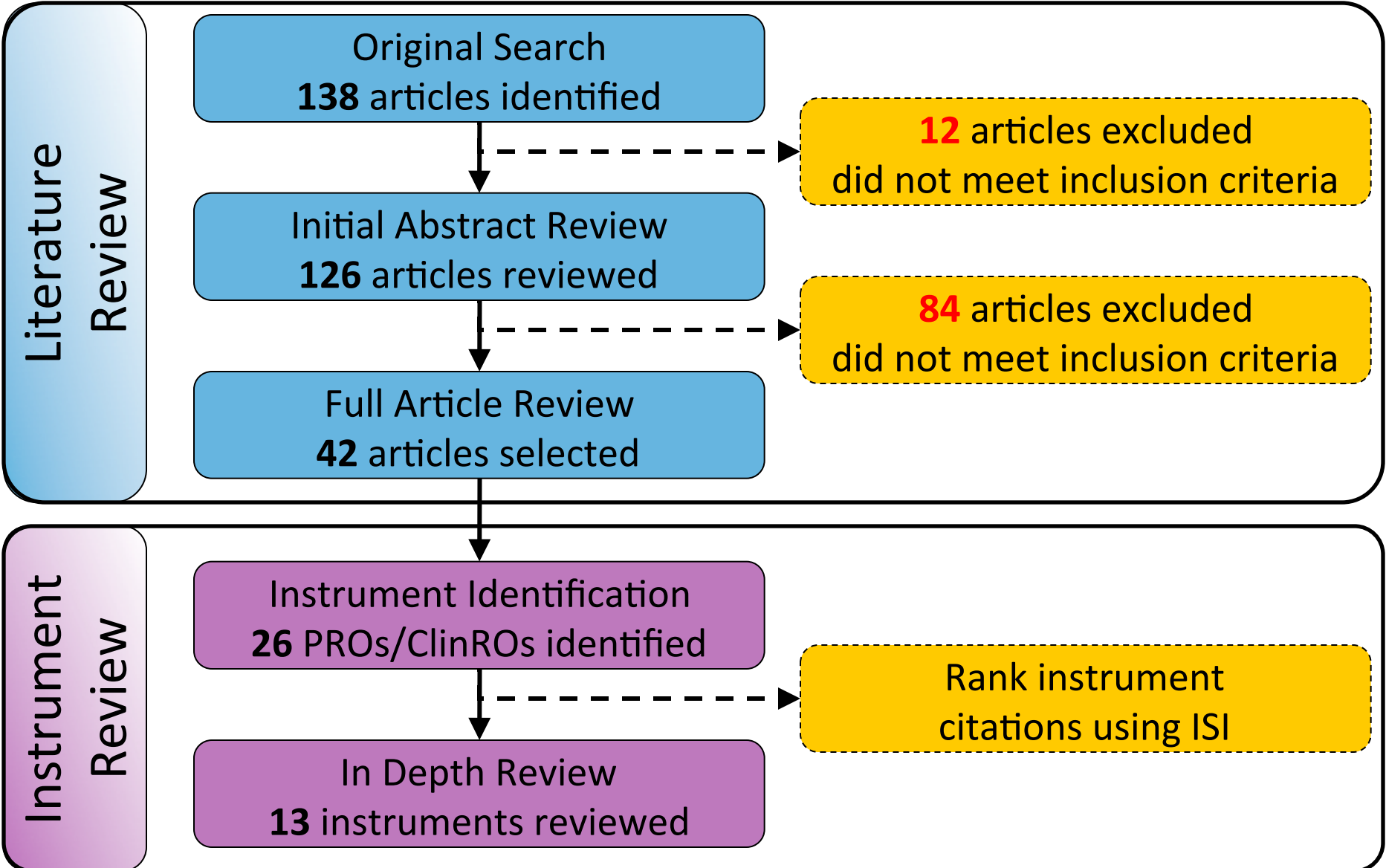
Results – Instruments Selected

Instrument	Full Name	Citations
BDI/BDI-II	Beck Depression Inventory	5,893
GHQ	General Health Questionnaire	2,759
HADS	Hospital Anxiety and Depression Scale	2,519
HAM-D	Hamilton Rating Scale for Depression	1,777
GDS	Geriatric Depression Scale	1,522
POMS/POMS-SF	Profile of Mood States	1,183
MADRS/MADRS-S	Montgomery-Åsberg Depression Rating Scale	1,089
CES-D	Center for Epidemiological Studies – Depression Scale	976
BSI	Brief Symptom Inventory	938
PHQ-9	Patient Health Questionnaire	667
ZUNG SDS	Zung Self-Rating Depression Scale	498
IDS-C/IDS-SR QIDS-C/QIDS-SR	Inventory of Depressive Symptomatology Quick Inventory of Depressive Symptomatology	133
PROMIS	Patient-Reported Outcomes Measurement Information System	48

Results – Other Instruments Identified
Not Selected for In-Depth Review

Instrument	Full Name	Citations
PRIME-MD	Primary Care Evaluation of Mental Disorders	158
CGIS	Clinical Global Impression of Severity	88
MDI	Major Depression Inventory	59
HDI	Hamilton Depression Inventory	26
EURO-D	EURO-D Scale	18
CRS	Carroll Rating Scale for Depression	15
QLDS	Quality of Life Depression Scale	8
DUKE-AD	Duke Anxiety Depression	5
SDI	Short Depression Inventory	4
DS	Depression Scale	0
RDS	RAND Depression Screener	0
SDSS	Signs of Depression Screening Scale	0

Results – Instrument Versions

- There are often a number of **Different Versions** of instruments, including versions with differing lengths
  - For the 13 instruments evaluated, over 40 different iterations or versions were identified
  - Clinician- (ClinRO) versus Patient-reported (PRO) versions
  - Versions ranged in length from 1 item (GDS-1) to 65 items (Original POMS)
- *Implication: need to ensure that instrument has adequate coverage of key concepts while at the same time not creating unnecessary patient burden*

Results – Concepts Measured

- Instruments vary in terms of **Concept Coverage**: which concepts are measured and how they are measured.

  - Coverage varies across different instruments as well as between different versions of the same instrument (due to length or additional concepts being added/ deleted in newer versions).

  - Some instruments have items which include descriptions of multiple concepts in a single item.

Results – Concepts Measured
Emotions/Mood

Domain/Concept — Emotions/Mood	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Anger						X							
Crying	X		X		X					X			
Decreased Pleasure	X	X						X		X		X	
Despair				X	X	X							
Empty					X				X				
Frustration						X							
Irritability/Hostility	X		X	X	X	X		X	X	X		X	
Less Compassion						X							
Mood Swings												X	
Rage						X							
Sadness	X		X	X	X	X	X	X		X	X	X	X

Results – Concepts Measured
Negative Affect/Sense of Self

Domain/ Concept	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Negative Affect													
Feeling Lonely											X		
Focus on Negative	X		X										
Guilt	X		X			X			X				
Hopeless/Helpless	X			X	X	X	X	X		X	X	X	X
Shame						X			X				
Worthlessness	X			X	X	X					X		X
Sense of Self													
Self-Hate	X												
Self-Efficacy	X			X		X		X		X			
Self-Esteem	X		X				X			X		X	
Self-Blame	X												

Results – Concepts Measured
Cognition

Domain/ Concept	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Cognition													
Cognitive Lethargy			X		X	X							
Distraction					X					X			
Feel Overwhelmed				X		X							
Fixation on Problem				X									
Indecisiveness	X		X	X	X	X		X				X	
Intrusive Thoughts		X	X		X								
Memory Issues					X								
Concentration	X			X	X	X	X		X				
Other Cognitive			X					X					

Results – Concepts Measured
Energy/Fatigue/Motivation

Domain/Concept	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Energy/Fatigue													
Fatigue/Exhaustion	X		X	X		X	X					X	
Lethargic						X				X		X	
No/Low Energy	X			X	X	X	X			X		X	
Daytime Sleepiness												X	
Tiredness				X		X		X					
Weak											X		
Motivation													
Level of Drive		X		X									
Level of Interest	X					X					X	X	
Interest in Self-Care		X											

Results – Concepts Measured
Physical Symptoms

Domain/ Concept / Physical Symptoms	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Breathing Problems			X								X		
Chest Pressure				X									
Dizziness											X		
GI Problems	X		X	X				X			X	X	
Heart Palpitations			X					X				X	
Bodily Pain	X		X									X	
Restlessness		X	X			X	X	X			X	X	
Stomach Discomfort	X	X											
Sweat				X									
Tingling Extremities											X		
Other Physical												X	

Results – Concepts Measured
Eating Behavior/Anxiety

Domain/Concept	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Eating Behavior													
Decreased Appetite	X		X				X		X	X		X	
Increased Appetite	X						X					X	
Weight Gain												X	
Weight Loss	X		X					X				X	
Other Eating Behav								X					
Anxiety													
Anxiety	X	X			X	X					X	X	
Fear		X	X	X		X				X	X	X	
Nervousness				X		X					X	X	
Panic Attacks		X		X		X			X		X	X	
Worry			X	X	X								

Results – Concepts Measured
Social Function/Activity Impairment

Domain/ Concept	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Social Function													
Few Social Engage	X				X								
Isolation				X		X				X			
Impact Relationship				X			X						
Sexual Activity	X		X					X				X	
Activity Impairment													
General Difficulty w/ Daily Activities			X	X				X	X				
Household Activity							X						
Personal Care				X									
Work Impacts	X		X	X			X						

Results – Concepts Measured
Sleep Impacts/Self-Harm

Domain/Concept	BDI	HADS	HAM-D	GHQ	GDS	POMS	PHQ	ZUNG	MADRS	CES-D	BSI	IDS	PROMIS
Sleep Impacts													
Early Awakening			X	X								X	
Difficulty Fall Asleep			X	X			X					X	
General Sleep Diff	X							X				X	
Oversleeping	X						X					X	
Diff Remain Asleep			X	X			X		X	X			
Self-Harm													
Better Off Dead			X	X	X			X	X				
Self-Harm	X								X				
Suicidal Ideation	X		X	X			X				X	X	
Thoughts of Death				X								X	

25

Results – Response Scale/Recall Period

- A variety of **Response Scales** are used including numerical, categorical/Likert-type and dichotomous

- *Implication: response scale may vary depending on whether measuring severity, frequency, duration of signs, symptoms and impacts. Response scale needs to be interpretable for patients*

- **Recall Periods** vary from "present time" (BDI-I, HAM-D, Zung) to "last few weeks" (GHQ)
  - 7-day recall period was the most utilized recall period (50% of instruments reviewed).
  - Recall period may vary between mode of administration: MADRS-C (no specific period), MADRS-S: 3 days

- *Implication: need to ensure that recall period is sensitive enough to detect changes in status and variability of symptom or impact being measured*

Results – Patient Input

- Many existing instruments have limited documentation of patient involvement during the initial development process.
  - HADS: patients completed self-assessment scale and were interviewed
  - PROMIS Depression Item Bank included patient focus groups and cognitive interviews.
- Many instruments were developed based on expert clinical opinion or were derived from other instruments
- *Implication: need to avoid use of medical terminology and ensure language used is accessible and understandable by patients*
- Most instruments used patients during "validation period"
  - Most instruments demonstrate acceptable scores for reliability and validity
- *Implication: documentation of "content validity" required for qualification of a PRO instrument under the FDA PRO Guidance.*
  - *According to the PRO Guidance, content validity must be established prior to evaluating construct validity.*

Results – Clinical Trial Use

- Many instruments have been used in a variety of clinical trials assessing treatment benefit in different disease states
- *Implication: PRO Guidance states that an instrument must be "fit for purpose" within a specific "context of use".*
  - *Without supporting evidence of content validity within the specific population studied, it is unlikely these instruments would be considered "fit for purpose" in these trials – and would not be considered appropriate endpoints in clinical trials from a regulatory perspective*
- The obvious exception to this is the use of Clinician-Reported Outcomes (MADRS, HAM-D), which have traditionally been used as primary endpoints in MDD clinical trials.
  - Other measures which have been used in the past include the BDI, BDI-II, Zung and PHQ-9 – however this was prior to the PRO Guidance

Limitations

- Information on patient involvement in the development process for existing measures is limited.
    - Importance of "content validity" and "patient-centered" outcomes research have increased in recent years.
    - Many older published articles focus primarily on the psychometric measures for validity and do not describe the instrument development process in detail.
    - This review included reports of patient-involved to the extent that such reports were published.
- This systematic review focused on instruments predominantly measuring MDD symptoms, and not those that measure more distal concepts like quality-of-life, functional status and satisfaction.
    - FDA is less likely to accept measures which measure more distal impacts (such as social function) or more general HRQOL concepts

Limitations

- The detailed review did not include less frequently cited instruments.

    - Potentially excludes new instrument or ones that have been less frequently used

- Use in clinical trials is limited to uses identified in publications included in our review, which focused on the articles related to the instrument development process and psychometric properties.

    - It did not include articles pertaining use of the instrument in clinical trials for assessing efficacy, as this was beyond the scope of our research objectives

Conclusions

- There are a large number of existing patient-reported outcome measures for major depressive disorder
  - These instruments vary greatly in terms of concepts measured, instrument length, response options, anchoring, scoring algorithms, and recall period
  - Limited information is available in the literature on the developmental history of existing depression symptom inventories, making it unclear whether the inventory items were developed with direct patient input.

Conclusions

- In the absence of additional information about the patient's involvement during development of these instruments it is unlikely these scales will satisfy the requirements set forth in the FDA PRO Guidance
  - Qualitative research could support qualification of an existing scale - if an existing measure could be shown to have content validity.
  - An existing scale could also provide the basis for modification or development of a new scale.
- Further research is required to understand whether the concepts measured by these instruments are adequately supported by direct patient input and can provide a basis for qualification according to the PRO/DDT Guidance.
  - Alternatively, development of a new patient-reported outcome measure could be considered

Thank you

Appendix I: Instrument Profiles

Beck Depression Inventory (BDI)

Beck Depression Inventory
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** To assess measure depression symptoms and their severity in adolescents and adults population (age 13 and older).
- **Content:** Contains items that reflect the cognitive, affective, somatic and vegetative symptoms of depressions.
- **Version:** There are 4 versions: BDI, BDI-IA, BDI-II and BDI-FS.
- **Number of Items:** BDI=21 items, BDI-IA =21 items, BDI-II =21 items, BDI-FS=7 items.
- **Subscale:** None

Beck Depression Inventory
Scoring

- **Response Scale:** Each item is rated on 4-point scale: none (0), mild (1), moderate (2), severe (3). BDI-BDI-IA, BD-II each has range score of 0-63. BDI-FS has range score of 0-21.

- **Method of Scoring:** Total score is the sum of each item.

- **Patient Burden:** None reported. Time to complete=5-10 minutes.

- **Interpretation of Scores:** BDI & BDI-IA= minimal depression (0-9), mild (17-29), moderate (17-29), severe (30-63); BDI-II= minimal depression (0-13), mild (14-19), moderate (20-28), severe (29-63); BDI-FS= minimal depression (0-3), mild (4-8), moderate (9-12), severe (13-21).

Beck Depression Inventory
Development History

- **Patient Involvement:** Affirmative in original BDI; No report on the BDI-IA, BDI-II and BDI-FS.
- **Development Population:** Psychiatric and normal population.
- **Cross Cultural Input:** No cross cultural input during the development. Translated into Spanish, French for Belgium, French for Canada, Korean, Chinese, Romanian, Japanese, Arabic, German, Xhosa, Icelandic, Greek, and Dutch.
- **Known Use in Clinical Studies:** Studies on various settings: patients with Parkinson's disease, HIV+, cancer, arthritis, multiple sclerosis, insomnia, pain, and renal disease.

Beck Depression Inventory
Measurement Properties

- **Reliability:** Internal consistency: BDI (α=0.81-0.86), BDI-IA (α=0.79-0.90), BDI-II (α=0.83-0.93), BDI-FS (α=0.83-0.89). Test-retest reliability: BDI-II (r=0.89-0.94).

- **Validity:** Construct validity: BDI-II & SCL-90-R (depression subscale r=0.89; anxiety subscale r=0.47-0.71). Convergent validity: BDI-II & Revised HPRSD (r=0.71); BDI-II & HSRD (r=0.66); BDI-II & CDRS-R (r=0.72); BDI-II & RADS (r=0.84); meta-analysis (r=0.27-0.89). Concurrent validity: BDI-II & CES-D (r=0.68-0.74); BDI-II & CATI-Depression (r=0.58); BDI-II & CDS (r=0.69); BDI-II & MADRS. Discriminant validity: BDI-II & SPS (r=-0.42); BDI-II & DPD (r=0.59).

- **Ability to Detect Change:** Effect size: BDI-II (0.282) & BDI-FS (0.231) in chronic pain sample at 6-month follow up. BDI-FS (1.5) & BDI (1.1) in treated-untreated patients with multiple sclerosis. Smallest real difference: BDI-II (3.3) to assess severity of depression in patients with Parkinson's disease.

Hospital Anxiety and Depression Scale (HADS)

Hospital Anxiety and Depression Scale
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** To assess anxiety and depressive symptoms and severity in patients attending a general medical clinic.
- **Content:** Contains 7 items measuring cognitive and emotional aspects of depression, and 7 items measuring cognitive and emotional aspects of anxiety.
- **Version:** Only 1 original version.
- **Number of Items:** There are 14 items in the scale.
- **Subscale:** There are two subscales: the HADS-A (Anxiety subscale, 7 items) and HADS-D (Depression subscale, 7 items)

Hospital Anxiety and Depression Scale
Scoring

- **Response Scale:** The anxiety and depression items were both rated on four-point (0-3) rating scale. The score range is 0-42 for HADS scale, 0-21 for HADS-Anxiety subscale, and 0-21 for HADS-Depression subscale.

- **Method of Scoring:** Total score of HADS is the sum of the 14 items responses. Subscale score of HADS-A is the sum of the 7 anxiety items responses; subscale score of HADS-D is the sum of the 7 depression items responses.

- **Patient Burden:** None reported. Reading level= Grade 7-8. Time to complete=5-10 minutes.

- **Interpretation of Scores:** HADS-A and HADS-D: no-case (0-7), doubtful case (8-10), and case (11-21).

Hospital Anxiety and Depression Scale
Development History

- **Patient Involvement:** During item generations, patients were asked to complete a self-assessment scale questionnaire.

- **Development Population:** Patients in general medical outpatients clinics, age 16-65, who suffered from a variety of complaints and illness.

- **Cross Cultural Input:** No cross cultural input during the development. Translated into Arabic, German, Hebrew, Swedish, Italian, Spanish, Danish, Dutch, Finish, French, Norwegian, Portuguese, Greek, Chinese, Japanese, Thai, and Urdu.

- **Known Use in Clinical Studies:** Studies on various settings: stroke, cancer, osteoarthritis, cardiac disease, non-cardiac chest pain, brain injury, COPD, Parkinson's disease, ESRD, chronic fatigue syndrome, cardiomyopathy, and spinal cord injury.

Hospital Anxiety and Depression Scale
Measurement Properties

- **Reliability:** Internal consistency: HADS-A (r=0.68-0.93) and HADS-D (0.67-0.90); Test-retest reliability: HADS-A (0.70-0.899) and HADS-D (0.70-0.944).

- **Validity:** Concurrent validity: HADS & BDI (r=0.62-0.749); HADS-A & BDI (r=0.61-0.83); HADS-D & BDI (r=0.62-0.73); HADS-A & GHQ-28 (r=0.50-0.68); HADS-D & GHQ-28 (r=0.50-0.66); HADS-A & CAS (r=0.69-0.75); HADS & STAI (r=0.758); HADS-A & STAI (r=0.52-0.774); HADS-D & STAI (r=0.485-0.71); HADS-A & SCL-90 (r=0.49-0.73); HADS-D & SCL-90 (r=0.6); HADS & MADRS (r=0.51-0.77); HADS-D & MADRS (r=0.62-0.81); HADS-A & HARS (r=0.34-0.44); HADS-A & BAI (r=0.6-0.7; HADS-A & DASS (r=0.88); HADS-D & DASS (r=0.93).

- **Ability to Detect Change:** Anchor-based method using FRQ & FT, the MID=1.41-1.68 for HADS. Effect size approach, the MID=1.4 for HADS-D, 1.32 for HADS-A, and 1.17 for HADS.

Hamilton Rating Scale for Depression (HAM-D)

Hamilton Rating Scale for Depression
General Description

- **Type:** Clinician Reported Outcome
- **Purpose:** To assess severity of depressive symptoms in patients who already been diagnosed with depressive disorder.
- **Content:** Contains 17 items measuring depressed mood, vegetative and cognitive symptoms of depression and comorbid anxiety symptoms and 4 items measuring diurnal variation, depersonalization/ derealization, paranoid symptoms, and obsessional symptom.
- **Version:** There is 1 original and more than 20 modified version.
- **Number of Items:** The original HAM-D has 21 items, but only the first 17 items are scored.
- **Subscale:** Several factor scores have been proposed to identify HAM-D subscales.

Hamilton Rating Scale for Depression Scoring

- **Response Scale:** Of the 17 scoreable items, 8 items are scored on 5-point scale (0-4) and 9 items are on 3-point scale (0-2), and half-points maybe used by skilled rater. The total score ranges from 0 to 52.

- **Method of Scoring:** Total score of HAM-D is the sum of the 17 items responses.

- **Patient Burden:** HAM-D is a semi-structured interview by clinician. Time to complete=15-20 minutes.

- **Interpretation of Scores:** APA recommended the following cutting points: 0-7 (normal), 8-13 (mild depression), 14-18 (moderate depression), 19-22 (severe depression), 23 or higher (very severe depression).

Hamilton Rating Scale for Depression
Development History

- **Patient Involvement:** No information available.
- **Development Population:** No information available.
- **Cross Cultural Input:** No cross-cultural input during the development. Translated into Arabic, Afrikaans, Finnish, German, Norwegian, Cantonese, , French, German for Austria, Slovak, Dutch, French for Canada, Italian, Swedish, Turkish, Spanish, Castilian, and Thai.
- **Known Use in Clinical Studies:** Studies on various settings: stroke, multiple sclerosis, epilepsy, Parkinson's diseases, Wilson's disease, obsessive compulsive disorder (OCD) , and ESRD.

Hamilton Rating Scale for Depression
Measurement Properties

- **Reliability:** Internal consistency: 0.46-0.89. Intraclass reliability: 0.46-0.99. Inter-rater reliability: 0.27-0.87. Test-retest reliability: 0.81-0.90

- **Validity:** Convergent validity: HAM-D & BDI (r= 0.27-0.85), HAM-D & BPRS (r=0.56-0.89), HAM-D & CES-D (r=0.65), 0.56 and HAM-D & CGIS (r=0.77), HAM-D & CRSD (r=0.41-0.92), HAM-D & GAS (r=-0.47 and -0.86), HAM-D & HADS (r=0.38 and 0.67), HAM-D & MADRS (r=0.68-0.85), HAM-D & MDI (r=0.86), HAM-D & MMPI (r=0.20-0.67), HAM-D & RDS (r=0.65 & 0.81), and HAM-D & VAS (r=-0.65). Only two instruments displayed correlations below acceptable threshold: HAM-D & SCID (r=0.37), and HAM-D & Zung SDS (r=0.49).

- **Ability to Detect Change:** Effect size of HAM-D (0.37) in treatment of fluoxetine, in treatment of tricyclic antidepressants, HAM-D (0.25).

General Health Questionnaire (GHQ)

General Health Questionnaire
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** The GHQ is a self-report screening instrument designed to indicate psychological well-being and to detect current diagnosable psychiatric disorders in populations of primary medical care patients or general medical outpatients.
- **Content:** To examine functioning in the condition of an inability to carry out one's normal healthy functions, and the appearance of new phenomena of a distressing nature. It covers 4 aspects of distress: depression, anxiety, social impairment, and hypochondriasis.
- **Version:** The GHQ-60 is the original version of the instrument. There are four other versions which are shorter versions of the original GHQ, namely GHQ-30, GHQ-20, GHQ-12 and the "Scaled GHQ" or GHQ-28.
- **Number of Items:** The original GHQ contains 60 items. The GHQ-30, GHQ-20, GHQ-12 and GHQ-28 contain 30 items, 20 items, 12 items, and 28 items, respectively.
- **Subscale:** The GHQ-28 is the only GHQ version that provides subscales measuring 4 domains, i.e. Somatic Symptoms, Anxiety and Insomnia, Social Dysfunction, and Severe Depression.

- **Response Scale:** Several scoring methods:
  - Rating scale: 0-1-2-3
  - Modified rating scale: 0-0-1-2
  - GHQ Scoring: 0-0-1-1
  - CGHQ: 0-1-1-1 (for the negative phrase), and 0-0-1-1 (for the positive phrase).
- **Method of Scoring:** The total score of GHQ is the sum of the item responses.
- **Patient Burden:** The GHQ is easy to administer and score. Time to complete= 6-8 mins (GHQ-60), 3-4 mins (GHQ-30/ GHQ-28), 2 mins (GHQ-12).
- **Interpretation of Scores:** Threshold may vary from place to place. Mean or median of GHQ score is recommended as a rough guide for the best threshold.

General Health Questionnaire
Development History

- **Patient Involvement:** The original 60 items of GHQ was derived from factor analysis of a pool of 140 items. There is no information available regarding patient involvement in the item generating process of GHQ.

- **Development Population:** There is no information available on the characteristics of the population used in the development of the instrument.

- **Cross Cultural Input:** GHQ has been used in diverse cultural groups. Translated into 38 languages.

- **Known Use in Clinical Studies:** Studies on general populations and various disease settings including: stroke, ischemic heart disease, tuberculosis, COPD, neurological diseases, myocardial infarction, menorrhagia, Parkinson's disease, musculoskeletal disease, chronic fatigue syndrome, cancer, food-hypersensitivity, endocrinological illness, multiple sclerosis, pancreatic neuroendocrine tumor, skin disease, lower urinary tract symptoms, and leprosy.

General Health Questionnaire
Measurement Properties

- **Reliability:** Internal consistency: 0.70-0.95. Test-retest reliability: 0.85-0.90

- **Validity:** Construct validity: GHQ and standardized psychiatric assessments (r=0.59-0.76), GHQ-28 & SCL-R-90 (r≥0.83), GHQ & CIS (r=0.76-0.81), GHQ & PSE (r=0.88), GHQ & HAM-D (r=0.88), GHQ & Zung SDS (r=0.86). Discriminant validity: Discriminate between mild (GHQ median=27) and severe (GHQ median=44) depression in patients with stroke at baseline and also at 6 months follow up with GHQ median=28 (mild depression) and GHQ median=48 (severe depression). Sensitivity/Specificity: GHQ-12 sensitivity/Specificity = (84%-89%) / (76%-80%); GHQ-28 sensitivity/ specificity= (79.7%-84%) / (79.2%-82%); GHQ-30 sensitivity/specificity= 74% / 82%; GHQ-60 sensitivity/specificity= 78% / 85%.

- **Ability to Detect Change:** Effect Size: In patients with multiple sclerosis, the GHQ-12 displayed effect size of 0.60 and 0.87 for rehabilitation and steroid sample.

Geriatric Depression Scale (GDS)

Geriatric Depression Scale
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** The GDS is a self-rating scale designed specifically as a simple screening instrument for depression in elderly people.
- **Content:** The GDS items covered wide range of depression topic in elderly including cognitive symptoms, motivation, future/ past orientation, self-image, losses, agitation, obsessive traits and mood.
- **Version:** The original GDS-30 contains 30 items. The short version GDS-20, GDS-15, GDS-10, GDS-4 and GDS-1 has respectively 20, 15, 10, 4 and 1 item.
- **Subscale:** There are no subscales of the GDS.

Geriatric Depression Scale
Scoring

- **Response Scale:** dichotomous yes/no.
- **Method of Scoring:** A score of 1 is applied for each answer indicating depression and 0 otherwise. The total score is the sum of the depression responses. The total score ranges from 0 to 30 for the original GDS and from 0 to 15 for the GDS-15.
- **Patient Burden:** The original 30 item GDS may take 8 to 10 minutes to complete. The 15-item version may take 5 to 7 minutes to complete.
- **Interpretation of Scores:** Higher score indicates greater depression. For the original GDS, scores of 0 to 10 are generally considered as normal, 11 to 20 indicate mild depression, and 21 to 30 indicate moderate to severe depression. For the GDS-15, scores of 0-4 are considered normal, 5-8 indicate mild depression, 9-11 indicate moderate depression, and 12-15 indicate severe depression.

Geriatric Depression Scale
Development History

- **Patient Involvement:** None found.
- **Development Population:** Sample of 47 elderly (>55 years) people who were either normal elderly (no complaints of depression and no history of mental health) or elderly hospitalized for depression.
- **Cross Cultural Input:** None
- **Known Use in Clinical Studies:** The GDS has been used in the study of elderly people in general population samples and also in various disease settings including Parkinson's disease, Alzheimer's disease, hemodialysis, osteoarthritis, cancer, stroke, and heart failure.

Geriatric Depression Scale
Measurement Properties

- **Reliability:** Development: Cronbach's alpha of 0.94 and a split-half reliability coefficient of 0.94. Test-retest (1 week) 0.85.
- **Validity:** Significantly able to discriminate patients with different levels of depression (F-stat=99.48, p<0.001). The GDS-30 reported convergent validity against other depression measures including Zung SDS (r= 0.84) and HAM-D (r=0.83). The GDS-30 and the other two instruments were also correlated against three levels of depression of the RDC and the study found the GDS-30 and HAM-D were comparable (r=0.82 and r=0.83, respectively), but SDS was less effective in discriminating between normal, mild, and severe depressed patients (r=0.69).
- **Ability to Detect Change:** The GDS-15 was able to detect change in depressive symptoms over time after the loss of a partner. The Leiden-85 plus Study using 599 cohort sample of 85+ showed that the mean score of the GDS increased 1.2 point (p=0.013) between baseline and follow-up of 3.2 years. This change was significantly higher than the mean change of -0.06 points (p=0.032) found in the control group, after controlling demographic and time factors.

Profile of Mood States (POMS)

Profile of Mood States
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** to assess mood and transient affective states.
- **Content:** The measure contains adjectives and short phrase items to measure six identifiable mood affective states, i.e. tension-anxiety (9 items), depression-dejection (15 items), anger-hostility (12 items), fatigue (7 items), vigor-activity (8 items), and confusion-bewilderment (7 items)
- **Version:** The POMS Standard, the POMS short versions and the POMS-Bipolar.
- **Number of Items:** The original POMS consists of 65 items. The short-form POMS-SF covers 37 items, the POMS-Brief covers 11 items, and the POMS-A 24 covers items. In general, the number of items in the short-form versions of the POMS varied, ranging from 11-40 item versions.
- **Subscale:** six subscales, i.e. Tension-Anxiety, Anger-Hostility, Fatigue-Inertia, Depression-Dejection, Vigor-Activity, and Confusion-Bewilderment. The 37-item POMS-SF has the same six subscales of the original POMS. The 11-item POMS-Brief only provides one score for overall psychological distress.

Profile of Mood States
Scoring

- **Response Scale:** Each item of the instrument uses a 5-point response scale, ranging from 0 (not at all) to 4 (extremely).
- **Method of Scoring:** The subscale score is the sum of the item responses. The scores range from 0 to 36 for the tension-anxiety subscale, 0-60 (depression), 0-48 (anger-hostility), 0-32 (vigor-activity), 0-28 (fatigue), and 0-28 (confusion-bewilderment). The total score for POMS is the Total Mood Disturbance (TMD) score. To obtain the TMD, the vigor subscale score is subtracted from the sum of the other 5 subscales. The TMD score ranges from -32 (best possible TMD score) to 200 (worst possible TMD score).
- **Patient Burden:** approx 3 to 7 mins for healthy and up to 20 minutes for physically ill populations. 8th grade reading level. Burdensome for geriatric patients.
- **Interpretation of Scores:** The TMD score ranges from -32 to 200. Higher scores indicate rising mood disturbances. The interpretability of the scores, such as recommended cut-off scores and MID, was not found in this review.

Profile of Mood States
Development History

- **Patient Involvement:** None found.
- **Development Population:** Validation studies of the 65-item POMS have been conducted in populations of psychiatric outpatients of the Veterans Administration and college students. The normative data for the instrument have been established using sample of more than 1000 psychiatric patients, 856 college students, and 2360 adults involved in a smoking cessation program
- **Cross Cultural Input:** None. There are many translations.
- **Known Use in Clinical Studies:** healthy subjects and patients with various medical conditions [McNair 2003]. The POMS has been used in the studies of patients with cancer, post-menopause, HIV infection, acute myocardial infarction, seizure disorders, anxiety, diabetes, narcolepsy, rheumatoid arthritis, and other medical conditions. The POMS has also been extensively used in assessing mood states in sports and exercise environments.

Profile of Mood States
Measurement Properties

- **Reliability:** internal consistency coefficients of 0.70 for the depression-dejection subscale, 0.71 for the anger-hostility, 0.65 for the vigor-activity, 0.66 for the fatigue-inertia, and 0.68 for the confusion-bewilderment. POMS publications found internal consistency estimates between 0.87 and 0.92 for the POMS and between 0.75 and 0.92 for the POMS-SF. Acceptable test-retest reliability 0.65 – 0.74.

- **Validity:** In a study of general population, the TMD and subscale scores significantly correlated against Visual Analog Mood Scales, STAI-State, STAI-Trait, BSI Total, and GDS Total. Discriminant validity: correlations between POMS subscales scores and corresponding mood scales were consistently higher than the non-corresponding mood scales with mean r=0.67 and 0.50, respectively.

- **Ability to Detect Change:** Higher effect size of POMS associated with the greatest incremental benefit in proportion to pts improving.

Patient Health Questionnaire (PHQ-9)

Patient Health Questionnaire
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** designed as a multipurpose instrument for screening, diagnosing, monitoring and measuring the severity of depression.
- **Content:** incorporates Diagnostic and Statistical Manual Fourth Edition (DSM-IV) depression diagnostic criteria with other leading major depressive symptoms into a brief self-report tool rating the frequency of the symptoms
- **Version:** PHQ-9.
- **Number of Items:** 9 items.
- **Subscale:** No subscales.

Patient Health Questionnaire Scoring

- **Response Scale:** Patients indicate for each of the 9 depressive symptoms whether, during the previous 2 weeks, the symptom has bothered them "0" (not at all), "1" (several days), "2" (more than half of the days) and "3" (nearly every day).

- **Method of Scoring:** Items count towards the final depression score if marked "More than half the days". The 9th item about suicidal ideation is counted whenever it is present.

- **Patient Burden:** takes less than 10 minutes to complete. 88% of patients said they were "very" or "somewhat" comfortable answering the PHQ questions.

- **Interpretation of Scores:** As a severity measure, the PHW-9 score can range from 0 to 27 since each of the 9 items can be scored from 0 (not at all) to 3 (nearly every day). The PHQ has the following categories of increasing severity: 0-4, 5-9, 10-14, 15-19 and 20 or greater. These levels of severity were validated with the criterion gold standard expert mental health providers.

Patient Health Questionnaire
Development History

- **Patient Involvement:** The instrument items are directly derived from the DSM-IV criteria. No patient involvement in the PHQ-9 development was found.

- **Development Population:** two studies involving 3000 patients, 18 years or older, in 8 primary care clinics and 3000 patients, 18 years or older, in 7 obstetrics-gynecology clinics.

- **Cross Cultural Input:** None found. Many translations exist.

- **Known Use in Clinical Studies:** wide range of populations, including patients in clinical trials and primary care with various medical conditions. The PHQ-9 has been used in the studies of patients with cancer, diabetes, end-stage kidney disease, smoking, substance abuse, heart failure, chronic hepatitis C, traumatic brain injury, and HIV.

Patient Health Questionnaire
Measurement Properties

- **Reliability:** The internal reliability of the PHQ-9 had a Cronbach's alpha of 0.89 for the primary care sample and 0.86 for the gynecological sample. The test-retest was tested with a clinical interview 48 hours after completing the PHQ-9 and had a reliability of 0.84

- **Validity:** PHQ-9 correlated with SF20 mental health subscale (0.73). The PHQ-9 was also correlated with self-reported disability days (past 3 months) (0.39), clinic visits (past 3 months) (0.24) and the difficulty patients attribute to their symptoms (0.55). The single item assessing difficulty attributed to depressive symptoms correlated strongly with impairment as measured by the SF-20 (0.53 for mental health, 0.42 general health perceptions, 0.40 for social functioning, 0.38 for role functioning, 0.27 for bodily pain, and 0.27 for physical functioning

- **Ability to Detect Change:** The PHQ-9 has a sensitivity of 73% and a specificity of 94%.

Zung Self-Rating Depression Scale (Zung SDS)

Zung Self-Rating Depression Scale
General Description

- **Type:** Patient Reported Outcome

- **Purpose:** developed as a self-administered measure of depression severity.

- **Content:** Affective, cognitive, behavioral, and psychological aspects of depression

- **Version:** Zung SDS

- **Number of Items:** 20 items, with 10 items keyed negatively and 10 positively.

- **Subscale:** None

Zung Self-Rating Depression Scale
Scoring

- **Response Scale:** For each of the 20 items, the subject rates whether the item occurred 1 = a little of the time, 2 = some of the time, 3 = a good part of the time, or 4 = most of the time.

- **Method of Scoring:** To obtain a total severity score, positive items are reversed, and then all items are summed. A severity index may be calculated by dividing the total score by 80 (the total possible).

- **Patient Burden:**

- **Interpretation of Scores:** Zung SDS scores are interpreted as follows: <50, within normal range; 50–59, minimal to mild depression; 60–69, moderate to severe depression; >70, severe depression.

Zung Self-Rating Depression Scale
Development History

- **Patient Involvement:** None, items were selected on the basis of the diagnostic criteria for depression and factor analysis studies available at the time the scale was developed.

- **Development Population:** The instrument was initially administered to 100 normal controls and 56 patients admitted to the psychiatric service of a Veteran's hospital with a primary diagnosis of depressive disorder

- **Cross Cultural Input:** None found. Many translations exist.

- **Known Use in Clinical Studies:** Although developed for use in patient populations, the Zung SDS has also been used in primary care and community settings and as a screen for depression.

Zung Self-Rating Depression Scale
Measurement Properties

- **Reliability:** Split-half reliability studies in a psychiatric population found a correlation (r) of 0.73. In a community survey of 1,173 subjects, Cronbach's alpha was satisfactory at 0.79.

- **Validity:** The correlation between Zung SDS and HAM-D was reported to range between 0.68 and 0.76. Higher correlations were observed at mild or moderate severity levels, while the greatest disagreement between Zung and HAM-D was among patients with non-endogenous symptom patterns

- **Ability to Detect Change:** the Zung SDS is sensitive to differences in severity of symptoms across subgroups of patients with diagnosed unipolar depression but is less sensitive to change in symptoms over time than other measures. Difficulties in sensitivity appear more often in the upper ranges of the scale; better results are obtained at mild or moderate severity levels.

Montgomery-Åsberg Depression Rating Scale
(MADRS)

Montgomery-Åsberg Depression Rating Scale
General Description

- **Type:** Clinician Reported Outcome, Patient Reported Outcome

- **Purpose:** to measure the degree of severity of depressive symptoms in patients diagnosed with depressive illness.

- **Content:** The MADRS items cover the core symptoms and cognitive features of clinical depression.

- **Number of Items:** The MADRS consists of 10 items. The MADRS-S includes 9 of the 10 MADRS items.

- **Subscale:** There are no subscales in either MADRS or MADRS-S instruments.

Montgomery-Åsberg Depression Rating Scale
Scoring

- **Response Scale:** The clinician-rated MADRS uses severity scores from 0 to 6, with higher scores indicating more severe symptoms. The anchor points were placed at 0, 2, 4, and 6 rating scales. The self-rated MADRS-S uses a 4-point rating scale from 0 to 3 with possible half points.

- **Method of Scoring:** Total score of MADRS is the sum of the 10 items' responses, no weight is used. The MADRS total score ranges from 0 to 60. The total score for the self-rated MADRS-S ranges from 0 to 27.

- **Patient Burden:** The clinician-rated version is administered through a trained interviewer and takes approximately 20 to 60 minutes to complete.

- **Interpretation of Scores:** The MADRS total scores ranges from 0 to 60. Snaith et al. recommended the following cut-off points to interpret MADRS: 0-6 (absence of symptoms), 7-19 (mild depression), 20-34 (moderate depression), and 35-60 (severe depression).

Montgomery-Åsberg Depression Rating Scale
Development History

- **Patient Involvement:** There is no information regarding patient involvement in the development of the measure. The MADRS was derived from the 65-item Åsberg Comprehensive Psychopathological Rating Scale (CPRS), a measure for psychiatric treatment. The selected items for MADRS were obtained from the ACPRS' most frequently checked and most sensitive to change items

- **Development Population:** 106 depressed patients (33 men and 73 women) who participated in clinical trials of antidepressant drugs.

- **Cross Cultural Input:** reported by using patients from England (n-54) and Sweden (n=54) to reduce cultural bias. Many translations exist.

- **Known Use in Clinical Studies:** Several reports show the use of MADRS and MADRS-S in various clinical studies including a study with patients exhibiting early onset dementia, schizophrenia, Parkinson's disease, multiple sclerosis, breast cancer, HIV, type 2 diabetes, heart failure, and epilepsy.

Montgomery-Åsberg Depression Rating Scale
Measurement Properties

- **Reliability:** The internal consistency and inter-rater reliability for the MADRS score is generally acceptable (α of 0.62 to 0.87). Inter-rater reliability 0.90-0.97.
- **Validity:** The MADRS has shown to have good validity. The clinician-rated MADRS showed high correlations against the self-rated MADRS-S as reported by Svanborg and Åsberg (r=0.70) and Mundt, et al. (r=0.82). High correlation against the HADS-Depression (r=0.81) and low correlation against the HADS-Anxiety.
- **Ability to Detect Change:** Reported effect size of 0.49 indicating the ability of MADRS to detect the difference between placebo and antidepressant. In a clinical trial of amitifadine, MADRS showed effect size of -0.601 in detecting treatment changes against placebo.

Center for Epidemiologic Studies – Depression Scale (CES-D)

Center for Epidemiologic Studies – Depression Scale
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** to measure depressive symptomatology for epidemiological studies in general populations.
- **Content:** six major dimensions of depression, i.e. depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance.
- **Versions:** CES-D, CES-D Boston Form (CESD-10), CES-D Iowa Form (CESD-11)
- **Number of Items:** CES-D has 20 items, CESD-10 has 10, and CESD-11 has 11 items.
- **Subscale:** There are no formal subscales for the CES-D; however, a four factor solution was reported in the factor analysis of the original version which includes a 7-item depressive affect, a 4-item positive affect, a 7-item somatic and retarded activity, and a 2-item interpersonal.

Center for Epidemiologic Studies – Depression Scale
Scoring

- **Response Scale:** Each CES-D item has 4 response options from 0 to 3, i.e. 0=Rarely or none of the time (less than 1 day), 1=Some or a little of the time (1-2 days), 2=Occasionally or a moderate amount of time (3-4 days), 3=Most or all of the time (5-7 days).

- **Method of Scoring:** The total score is calculated by summing all items. Before calculating the total score, the item #4, 8, 12, and 16 are reversed. The total score ranges from 0 to 60 for the original CES-D 20 items.

- **Patient Burden:** 2nd grade reading level, takes 5 minutes to complete.

- **Interpretation of Scores:** Higher CES-D scores indicate more depressive symptoms. A cut-off point of 16 is generally used to distinguish between individuals considered depressed from those considered non-depressed.

Center for Epidemiologic Studies – Depression Scale
Development History

- **Patient Involvement:** None reported.
- **Development Population:** Random sample of the general population of two communities between 1971 and 1973 and a total of 2846 patients. The second survey was conducted in between 1973-1974 in Washington County, MD using a slightly revised instrument (n=1089), followed by re-interview survey (1973-1974) using 343 respondents from the original interview. Along with the survey program, two clinical validation studies were conducted using a sample of 35 outpatients with severe depression.
- **Cross Cultural Input:** None reported. CES-D has numerous translations.
- **Known Use in Clinical Studies:** The CES-D has been used in general populations and also in patients with various medical conditions including patients with arthritis, rheumatoid arthritis, coronary artery disease, cancer, type 2 diabetes, spinal cord injury, multiple sclerosis, hepatitis C, and various elderly samples.

- **Reliability:** CES-D showed reliability for both general and clinical populations. High internal consistency has been reported with Cronbach's alpha coefficients ranging from 0.84 to 0.90 in the development study of the measure. T-R reproducibility (2-week) 0.87. The instrument also showed reliability across gender, race, and age categories.

- **Validity:** As a screening tool, CES-D showed its validity for detecting depressive symptoms in psychiatric populations. The CES-D demonstrated modest to high correlations with other mental health measures such as Bradburn Negative Affect Scale (r=0.60), Bradburn Positive Affect Scale (r=-0.20), Langner 22-item scale (r=0.50) and with disability days (r=0.30), indicating reasonable evidence of discriminant validity of the instrument. Convergent validity against VAS-Fatigue was reported at r=0.52.

- **Ability to Detect Change:** In a study of 5 psychiatric populations and a community sample, Weissman et al. showed that CES-D is a sensitive tool for detecting changes in symptoms in psychiatric patients over time.

Brief Symptom Inventory (BSI)

Brief Symptom Inventory
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** Designed primarily as a screening measure for psychological distress in "a broad spectrum of adult medical patients 18 or older and adult individuals in the community who are not currently assigned patient status", and was secondarily designed as an outcome measure.
- **Content:** Items that cover somatization, depression and anxiety.
- **Version:** The BSI-18 along with the BSI-53, are shortened version of the original 90-item Symptom Checklist (SCL-90). All of these instruments are essentially derived from the Hopkins Symptom Checklist (HSCL).
- **Number of Items:** 18 items.
- **Subscale:** 3 symptom dimensions: somatization (SOM, 6 items), depression (DEP, 6 items) and anxiety (ANX, 6 items).

Brief Symptom Inventory
Scoring

- **Response Scale:** Each item of the instrument uses a 5-point response scale: 0=Not at all, 1= A little bit, 2= Moderately, 3= Quite a bit, 4= Extremely.

- **Method of Scoring:** The subscale score is the sum of the item responses. Since each dimension consists of 6 items, the dimension scores range from 0 to 24, i.e. SOM=0-24, DEP=0-24, ANX=0-24. The total score or the Global Severity Index (GSI) is the sum of the three dimensions, therefore the GSI ranges from 0 to 72.

- **Patient Burden:** 6th grade reading level, takes 3-4 minutes.

- **Interpretation of Scores:** The total GSI score ranges from 0 to 72 with higher scores indicating higher levels of psychological distress.

Brief Symptom Inventory
Development History

- **Patient Involvement:** None found.
- **Development Population:** Norms for the BSI-18 were derived from different populations that served as the normative base for the SCL-90-R and the BSI. The community normative data for the BSI-18 have been established using a sample of 1,136 employees at all levels of a large national corporation. The oncology norms were derived from a sample of 1,543 patients from the urban cancer center with a broad range of diagnoses.
- **Cross Cultural Input:** There was no indication of cross-cultural input in the development of the instrument. Currently, the BSI-18 has only been translated in a few languages, including Spanish, German, Icelandic, Dutch and Korean.
- **Known Use in Clinical Studies:** The BSI-18 has been used in the studies of patients with traumatic brain injury, breast cancer, survivors of childhood cancer, psychogenic movement disorders, and Parkinson's disease.

Brief Symptom Inventory
Measurement Properties

- **Reliability:** The BSI-18 demonstrated acceptable internal consistency. Using 1,134 subjects in the community normative sample, Derogatis et al. reported internal consistency coefficients of 0.89 for the total score (GSI). For the dimensions, the alpha coefficients were 0.74 for the somatization, 0.84 for the depression, and 0.79 for the anxiety.

- **Validity:** The BSI-18 demonstrated convergent and discriminant validity with other measures such as BDI, BAI and MMPI-2 in outpatients with psychiatric disorders. The GSI showed correlations of 0.82, 0.75, 0.61 with the BAI, and MMPI-Depression, respectively

- **Ability to Detect Change:** This review did not find any study on the ability to detect change of the BSI-18.

[Quick] Inventory of Depressive Symptomatology (IDS/QIDS)

[Quick] Inventory of Depressive Symptomatology
General Description

- **Type:** Clinician Reported Outcome, Patient Reported Outcome
- **Purpose:** The 30 item IDS and the 16 item QIDS are designed to assess the severity of depressive symptoms. These assessments can be used to screen for depression, although they have been used predominantly as measures of symptom severity.
- **Content:** The QIDS-C30 and QIDS-SR16 cover only the nine diagnostic symptom domains used to characterize a major depressive episode, without items to assess atypical, melancholic or their commonly associated symptoms. All 16 items on the QIDS are included within the IDS.
- **Version:** IDS-SR / QIDS-SR
- **Number of Item:** 30 items in the IDS and 16 in the QIDS.
- **Subscale:** Nine item symptom domains used to define a major depressive episode (ICD-10): depressed mood, loss of interest or pleasure, concentration/decision making, self-outlook, suicidal ideation, energy/fatigability, sleep, weight/appetite change, and psychomotor changes.

[Quick] Inventory of Depressive Symptomatology Scoring

- **Response Scale:** Each item is interval scaled from 0 to 3; 0 indicates absence of the symptom during the last 7 days.

- **Method of Scoring:** The IDS-SR30 is scored by summing responses to 28 of the 30 items to obtain a total score ranging from 0 to 84. The QIDS-SR16 total scores range from 0 to 27, the total score is obtained by adding the scores for each of the nine symptom domains of the DSM-IV MDD criteria.

- **Patient Burden:** Both the IDS and QIDS are easy to administer in either the clinician-rated (IDS-C30 and QIDS-C16) or patient self-report (IDS-SR30 and QIDS-SR16) versions; they require minimal training.

- **Interpretation of Scores:** QIDS-SR16: 0-5=No depression; 6-10=Mild; 11-15=Moderate; 16-20=Severe; and 21-27=Very severe.

[Quick] Inventory of Depressive Symptomatology Development History

- **Patient Involvement:** None found.
- **Development Population:** IDS: The self-report was completed by 289 patients (average age 38.2 years) with unipolar major depression, bipolar disorder, euthymic depression, and other psychiatric disorders; QIDS: 596 adult outpatients treated for chronic nonpsychotic, major depressive disorder.
- **Cross Cultural Input:** There was no indication of cross-cultural input in the development of the instrument, but many translations exist.
- **Known Use in Clinical Studies:** Both versions of the IDS have been used in nonpsychotic and psychotic MDD, postpartum depression, dysthymic disorder, minor depression, bipolar disorder, as well as in patients with depression comorbid with cancer and asthma.

[Quick] Inventory of Depressive Symptomatology Measurement Properties

- **Reliability:** Evidence of acceptable psychometric properties of the IDS-C30 and IDS-SR30. Cronbach's alpha was 0.94 for both the IDS-C30 and IDS-SR30 for the complete sample (n=456), and 0.67 and 0.77 for the IDS-C30 and IDS-SR30, respectively, for the sample of MDD patients (n=338). Corruble et al. found a Cronbach's alpha coefficient of 0.75 and 0.79 for the IDS-C30 and the IDS-SR30 respectively.

- **Validity:** IDS and QIDS total scores were comparable to those obtained by the HRSD17 and BDI, with Pearson product moment correlations of 0.88 between the IDS-SR30 and the HRSD17. The correlation coefficient between the IDS-C30 and IDS-SR30 was 0.91. The 20 item depression factor from the SCL-90 was correlated with the IDS-SR30 (c=0.84). The IDS and QIDS have been used to distinguish response from remission.

- **Ability to Detect Change:** Both versions are sensitive to change, with medications, psychotherapy, or somatic treatments, making them useful for both research and clinical purposes.

Patient-Reported Outcomes Measurement Information System – Depression Item Bank (PROMIS)

PROMIS – Depression Item Bank
General Description

- **Type:** Patient Reported Outcome
- **Purpose:** To measure of the impact of depression on oneself.
- **Content:** Focuses on the assessment of one's negative mood (e.g. sadness, guilt), decrease in positive affect (e.g. loss of interest), information processing deficits (e.g. problems in decision making), negative view of the self (e.g. self-criticism, worthlessness), and negative social cognition (e.g. loneliness, interpersonal alienation).
- **Version:** Short forms and a calibrated item bank that can be used for computer adaptive testing and construction of other short forms.
- **Number of Items:** 28 calibrated items.
- **Subscale:** No subscales.

PROMIS – Depression Item Bank
Scoring

- **Response Scale:** Each question in the PROMIS-Bank depression has 5 response options, e.g., 1=Never, 2=Rarely 3=Sometimes, 4=Often, 5=Always.

- **Method of Scoring:** The total raw score is calculated by summing the values of the response to each question. The total raw scores range from 8 to 40 for the 8-item short form; 6 to 30 for the 6-item short form, and 4 to 20 for the 4-item short form. The final score of the instrument is presented in a T-score format based on the US general population mean and standard deviation.

- **Patient Burden:** 1st grade reading level.

- **Interpretation of Scores:** Higher T-scores represent higher depression levels. For the negatively worded instruments, a T-score of 60.0 is one SD worse (more depressed) than the average US general population, while a T-score of 40.0 is one SD better (less depressed) than the average population.

PROMIS – Depression Item Bank
Development History

- **Patient Involvement:** Qualitative item review phase with focus groups and cognitive interviews.
- **Development Population:** U.S. general population and multiple disease populations. Two data collection designs were applied, i.e. the 'full bank' design with N=7,005 and the 'block administration' design with N=14,128 (6,245 general population, 7,883 clinical samples).
- **Cross Cultural Input:** None found. There are several translations.
- **Known Use in Clinical Studies:** wide range of populations, including healthy subjects and patients with various medical conditions such as patients with cirrhosis and their caregivers, spinal cord injury, post-polio syndrome, multiple sclerosis, muscular dystrophy, advanced-stage cancer, systemic lupus erythematosus (SLE), colorectal cancer, HIV, and various psychiatric and medical conditions

PROMIS – Depression Item Bank
Measurement Properties

- **Reliability:** The mean adjusted item to total correlation was 0.83 and the alpha coefficient for the short form was 0.95. Correlation of the item response theory (IRT) theta scores between short forms and corresponding full item banks was very high at 0.96. Similar correlation using the raw scores was also very high at 0.98.

- **Validity:** The validation study on the 28 calibrated items reported convergent validity against CES-D (r=0.83) and divergent validity against general distress (anxiety) from the Mood and Anxiety Symptom Questionnaire (MASQ) (r=0.72). The correlations of the IRT theta between depression and anxiety were high (r=0.81) raising questions whether these constructs should be modeled as a single construct

- **Ability to Detect Change:** MIDs for the PROMIS-depression between 3.0 and 4.5 with a corresponding effect size between 0.36 and 0.54 (using T-Score calculation), and MIDs between 3.0 and 4.0 with a corresponding effect size between 0.43 and 0.57 (using raw score calculation). All of the anchors passed the criteria of having Spearman correlations ≥0.3 against PROMIS-Cancer scales.

Appendix II: References

Articles Reviewed

1. Abernethy AP, Ahmad A, Zafar SY, Wheeler JL, Reese JB, Lyerly HK. Electronic patient-reported data capture as a foundation of rapid learning cancer care. Med Care 2010 Jun;48(6 Suppl):S32-8.

2. Barrett AM . Rose-colored answers: neuropsychological deficits and patient-reported outcomes after stroke. Behav Neurol 2010;22(1-2):17-23.

3. Becker J, Fliege H, Kocalevent RD, Bjorner JB, Rose M, Walter OB, et al. Functioning and validity of a Computerized Adaptive Test to measure anxiety (A-CAT). Depress Anxiety 2008;25(12):E182-94.

4. Benecke H, Perez-Garrigues H, Bin Sidek D, Uloziene I, D K, Sondag E, et al. Effects of Betahistine on Patient-Reported Outcomes in Routine Practice in Patients with Vestibular Vertigo and Appraisal of Tolerability: Experience in the OSVaLD Study. Int Tinnitus J 2010;16(1):14-24.

5. Benzer W, Platter M, Oldridge NB, Schwann H, Machreich K, Kullich W, et al. Short-term patient-reported outcomes after different exercise-based cardiac rehabilitation programmes. Eur J Cardiovasc Prev Rehabil 2007 Jun;14(3):441-7.

6. Bremander AB, Holmstrom G, Bergman S. Depression and age as predictors of patient-reported outcome in a multidisciplinary rehabilitation programme for chronic musculoskeletal pain. Musculoskeletal Care 2011 Mar;9(1):41-8.

7. Cano SJ, Warner TT, Thompson AJ, Bhatia KP, Fitzpatrick R, Hobart JC. The cervical dystonia impact profile (CDIP-58): can a Rasch developed patient reported outcome measure satisfy traditional psychometric criteria? Health Qual Life Outcomes 2008;6:58.

8. Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. Qual Life Res 2010 Feb;19(1):125-36.

9. Cleeland CS, Sloan JA. Assessing the Symptoms of Cancer Using Patient-Reported Outcomes (ASCPRO): searching for standards. J Pain Symptom Manage 2010 Jun;39(6):1077-85.

10. Cook KF, Molton IR, Jensen MP. Fatigue and aging with a disability. Arch Phys Med Rehabil 2011 Jul;92(7):1126-33.

11. Crane HM, Lober W, Webster E, Harrington RD, Crane PK, Davis TE, et al. Routine collection of patient-reported outcomes in an HIV clinic setting: the first 100 patients. Curr HIV Res 2007 Jan;5(1):109-18.

12. Daly BJ, Douglas S, Lipson A, Foley H. Needs of older caregivers of patients with advanced cancer. J Am Geriatr Soc 2009 Nov;57 Suppl 2:S293-5.

Articles Reviewed

13. Fantino B, Moore N. The self-reported Montgomery-Asberg Depression Rating Scale is a useful evaluative tool in Major Depressive Disorder. BMC Psychiatry 2009;9:26.

14. Fernandez JL, Montgomery S, Francois C. Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder. Pharmacoeconomics 2005;23(2):155-67.

15. Fortner BV, Houts AC, Schwartzberg LS. A prospective investigation of chemotherapy-induced neutropenia and quality of life. J Support Oncol 2006 Oct;4(9):472-8.

16. Gelfand JM, Kimball AB, Mostow EN, Chiou CF, Patel V, Xia HA, et al. Patient-reported outcomes and health-care resource utilization in patients with psoriasis treated with etanercept: continuous versus interrupted treatment. Value Health 2008 May-2008 Jun;11(3):400-7.

17. Gibbons LE, Feldman BJ, Crane HM, Mugavero M, Willig JH, Patrick D, et al. Migrating from a legacy fixed-format measure to CAT administration: calibrating the PHQ-9 to the PROMIS depression measures. Qual Life Res 2011 Mar.

18. Girgis A, Breen S, Stacey F, Lecathelinais C. Impact of two supportive care interventions on anxiety, depression, quality of life, and unmet needs in patients with nonlocalized breast and colorectal cancers. J Clin Oncol 2009 Dec;27(36):6180-90.

19. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009 Sep;18(7):873-80.

20. Husson O, Mols F, van de Poll-Franse LV. The relation between information provision and health-related quality of life, anxiety and depression among cancer survivors: a systematic review. Ann Oncol 2011 Apr; 22 (4):761-72.

21. Irwin DE, Stucky B, Langer MM, Thissen D, Dewitt EM, Lai JS, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Qual Life Res 2010 May;19(4):595-607.

22. Kekow J, Moots R, Khandker R, Melin J, Freundlich B, Singh A. Improvements in patient-reported outcomes, symptoms of depression and anxiety, and their association with clinical remission among patients with moderate-to-severe active early rheumatoid arthritis. Rheumatology (Oxford) 2011 Feb;50 (2):401-9.

Articles Reviewed

23. Kelly MA, Morse JQ, Stover A, Hofkens T, Huisman E, Shulman S, et al. Describing depression: Congruence between patient experiences and clinical assessments. Br J Clin Psychol 2010 Apr.

24. Kelly MA, Morse JQ, Stover A, Hofkens T, Huisman E, Shulman S, et al. Describing depression: congruence between patient experiences and clinical assessments. Br J Clin Psychol 2011 Mar;50(1):46-66.

25. Lasch K, Joish VN, Zhu Y, Rosa K, Qiu C, Crawford B. Validation of the sleep impact scale in patients with major depressive disorder and insomnia. Curr Med Res Opin 2009 Jul; 25(7):1699-710.

26. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. J Clin Epidemiol 2010 Nov;63(11):1169-78.

27. Luckett T, Butow PN, King MT, Oguchi M, Heading G, Hackl NA, et al. A review and recommendations for optimal outcome measures of anxiety, depression and general distress in studies evaluating psychosocial interventions for English-speaking adults with heterogeneous cancer diagnoses. Support Care Cancer 2010 Oct;18(10):1241-62.

28. Nakamura Y, Lipschitz DL, Landward R, Kuhn R, West G. Two sessions of sleep-focused mind-body bridging improve self-reported symptoms of sleep and PTSD in veterans: A pilot randomized controlled trial. J Psychosom Res 2011 Apr;70(4):335-45.

29. Nelson CJ, Cho C, Berk AR, Holland J, Roth AJ. Are gold standard depression measures appropriate for use in geriatric cancer patients? A systematic evaluation of self-report depression instruments used with geriatric, cancer, and geriatric cancer samples. J Clin Oncol 2010 Jan;28(2):348-56.

30. Ow YL, Thumboo J, Cella D, Cheung YB, Yong Fong K, Wee HL. Domains of health-related quality of life important and relevant to multiethnic English-speaking Asian systemic lupus erythematosus patients: a focus group study. Arthritis Care Res (Hoboken) 2011 Jun;63(6):899-908.

31. Pezzilli R, Campana D, Morselli-Labate AM, Fabbri MC, Brocchi E, Tomassetti P. Patient-reported outcomes in subjects with neuroendocrine tumors of the pancreas. World J Gastroenterol 2009 Oct;15 (40):5067-73.

32. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item Banks for Measuring Emotional Distress From the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): Depression, Anxiety, and Anger. Assessment 2011 Jun.

Articles Reviewed

33. Rejas J, Pardo A, Ruiz MA. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. J Clin Epidemiol 2008 Apr;61(4):350-6.

34. Revicki DA, Kawata AK, Harnam N, Chen WH, Hays RD, Cella D. Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. Qual Life Res 2009 Aug;18(6):783-91.

35. Sapin C, Fantino B, Nowicki ML, Kind P. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. Health Qual Life Outcomes 2004 May;2:20.

36. Scoggins JF, Patrick DL. The use of patient-reported outcomes instruments in registered clinical trials: evidence from ClinicalTrials.gov. Contemp Clin Trials 2009 Jul;30(4):289-92.

37. Steel JL, Dunlavy AC, Stillman J, Pape HC. Measuring depression and PTSD after trauma: common scales and checklists. Injury 2011 Mar;42(3):288-300.

38. Stepanski EJ, Walker MS, Schwartzberg LS, Blakely LJ, Ong JC, Houts AC. The relation of trouble sleeping, depressed mood, pain, and fatigue in patients with cancer. J Clin Sleep Med 2009 Apr;5(2):132-6.

39. Teresi JA, Ocepek-Welikson K, Kleinman M, Eimicke JP, Crane PK, Jones RN, et al. Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. Psychol Sci Q 2009;51(2):148-80.

40. Teresi JA , Ramirez M, Lai JS, Silver S. Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. Psychol Sci Q 2008;50(4):538.

41. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol 2011 May;64(5):507-16.

42. Zullig LL, Jackson GL, Provenzale D, Griffin JM, Phelan S, van Ryn M. Transportation - A Vehicle or Roadblock to Cancer Care for VA Patients with Colorectal Cancer? Clin Colorectal Cancer 2011 Jul.