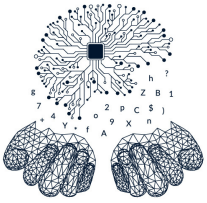

The Strawberry Hub
Empowering Patient Engagement

# Comprehensive Clinical Guide to Patient Reported Outcome Measures

*Evidence-Based Assessment Tools for Clinical Practice*

**Version 1.0 - Final**

AB Wenban – awenban@uahcs.com

February 2026

## Table of Contents

**1. Introduction and Purpose**

**2. PROM Systems**

- PROMIS (Patient-Reported Outcomes Measurement Information System)
- PROMIS-29

**3. Pain and Disability Measures**

- PEG-3 Pain Screening Tool
- Visual Analog Scale (VAS)
- Numerical Rating Scale (NRS)
- McGill Pain Questionnaire
- Oswestry Disability Index
- Roland-Morris Disability Questionnaire
- Neck Disability Index
- Bournemouth Questionnaires (Neck & Back)

**4. Health Status and Well-being**

- SF-36 Health Survey
- Global Well-Being Scale
- Patient Global Impression of Change (PGIC)
- EQ-5D

**5. Patient Satisfaction Measures**

- VSQ-9 Visit-Specific Satisfaction Questionnaire
- MYMOP2 (Measure Yourself Medical Outcome Profile)

**6. Psychosocial Screening Tools**

- STarT Back Screening Tool
- Fear-Avoidance Beliefs Questionnaire (FABQ)

**7. Mental and Behavioural Assessments**

- PHQ-9 (Patient Health Questionnaire)
- DASS-21 (Depression Anxiety Stress Scale)
- GAD-7 (Generalized Anxiety Disorder)

- HADS (Hospital Anxiety and Depression Scale)
- Beck Anxiety Inventory (BAI)

**8. Sleep Assessment**

- Pittsburgh Sleep Quality Index (PSQI)

**9. Activity Level Measures**

- International Physical Activity Questionnaire (IPAQ)

**10. Region-Specific Measures**

- Upper Extremity Functional Index (UEFI)
- Shoulder Pain and Disability Index (SPADI)
- Lower Extremity Functional Scale (LEFS)

**11. Immune Function Assessment**

- Immune Health Questionnaire
- Immune Function Questionnaire (IFQ)

**12. Gastrointestinal Assessment**

- Short Health Scale for GI Symptoms (SHS-GI)
- IBS Symptom Severity Score (IBS-SSS)
- GSRS-IBS (Gastrointestinal Symptom Rating Scale)

**13. References and Online Resources**

---

# 1. Introduction and Purpose

*Patient Reported Outcome Measures (PROMs) are standardised, validated questionnaires completed by patients to measure their perceptions of their own health status, functional outcomes, and quality of life. PROMs capture the patient's voice and perspective, which is essential for patient-centred care and evidence-based practice.*

### What are PROMs?

PROMs are instruments designed to assess health outcomes from the patient's perspective rather than relying solely on clinician observations or biomedical markers. They provide quantifiable data on:

Symptom severity and frequency

Functional limitations and disability

Quality of life and well-being

Treatment response and satisfaction

Psychosocial factors affecting health

### Purpose of This Document

This comprehensive clinical guide provides detailed information on evidence-based PROMs across multiple health domains. It serves to:

- Support clinical decision-making with validated assessment tools
- Enable standardised outcome measurement across practice settings
- Facilitate patient-centred care through structured patient input
- Track treatment progress and effectiveness over time
- Support quality improvement and research initiatives

### How to Use This Guide

Each PROM is presented with comprehensive clinical information:

- Description and intended purpose
- Number of items and domains assessed
- Psychometric properties (reliability, validity, responsiveness)
- Scoring methodology and interpretation
- Clinical applications and appropriate populations
- Administration guidelines and practical considerations

- Access information and licensing requirements

#### Selecting Appropriate PROMs

Selection of PROMs should be guided by:

Patient population and presenting condition

Assessment goals (screening, diagnosis support, outcome tracking)

Available time and resources

Psychometric quality and validation status

Clinical utility and interpretability

## 2. PROM Systems

### 2.1 PROMIS (Patient-Reported Outcomes Measurement Information System)

#### Description

*PROMIS is a comprehensive system of highly reliable, valid, flexible, and responsive assessment tools that measure patient-reported health status across multiple domains. Developed by the National Institutes of Health (NIH), PROMIS represents a paradigm shift in patient-reported outcome assessment through the use of Item Response Theory (IRT) and Computer Adaptive Testing (CAT).*

#### Key Features

Standardised item banks covering physical, mental, and social health

T-score metric (Mean = 50, SD = 10 in U.S. general population)

Precision-matched assessment with minimal respondent burden

Available in short forms, profile forms, and computer adaptive formats

Extensively validated across diverse populations and conditions

Available in 30+ languages

#### Core PROMIS Domains

Physical Health:

Physical Function

Pain Interference

Pain Intensity

Fatigue

Sleep Disturbance

Sleep-Related Impairment

Mental Health:

Depression

Anxiety

Anger

Cognitive Function

Social Health:

Ability to Participate in Social Roles and Activities

Satisfaction with Social Roles and Activities

Companionship

Emotional Support

Informational Support

Instrumental Support

## Psychometric Properties

### Reliability:

Internal consistency (Cronbach's α): 0.84-0.98 across domains

Test-retest reliability: 0.71-0.92

IRT-based scoring ensures precision across the full range of each construct

### Validity:

Convergent validity with legacy instruments: r = 0.60-0.90

Discriminant validity confirmed across known groups

Responsive to clinical change with effect sizes of 0.30-1.00

Validated in general population and 50+ clinical conditions

## Scoring

T-Score Interpretation:

- Mean: 50 (U.S. general population)

- Standard Deviation: 10 points
- For most domains: Higher scores = MORE of the concept (e.g., more pain, more function)
- Exception: For Physical Function, higher = better function

T-Score Benchmarks:

- T-score 40: One SD better than average
- T-score 50: Average
- T-score 60: One SD worse than average (clinically significant)
- T-score 70: Two SD worse than average (severe)

### Clinical Applications

Screening for health problems across multiple domains

Establishing baseline health status

Monitoring treatment response and progress

Comparing outcomes across patients, providers, and interventions

Quality improvement initiatives

Clinical research and pragmatic trials

### Administration

Time to complete: 3-5 minutes per domain (short forms)

Reading level: 5th-8th grade

Recall period: Past 7 days (typically)

Administration modes: Paper, electronic, interview

### Access and Licensing

Cost: FREE for clinical and research use

Access: www.healthmeasures.net

Registration required (free)

Attribution required in publications

Modifications not permitted without authorization

## 2.2 PROMIS-29 Profile v2.1

### Description

PROMIS-29 is a fixed-length profile measure that assesses seven health domains with 4 items each, plus a single pain intensity item (total 29 items). It provides a comprehensive snapshot of physical, mental, and social health in approximately 5-7 minutes.

### Domains Assessed

1. Physical Function (4 items)

   Ability to carry out physical activities

2. Anxiety (4 items)

   Fear, anxious misery, hyperarousal

3. Depression (4 items)

   Negative mood, views of self, social cognition, decreased positive affect

4. Fatigue (4 items)

   Range from mild tiredness to overwhelming exhaustion

5. Sleep Disturbance (4 items)

   Perceptions of sleep quality, depth, and restoration

6. Ability to Participate in Social Roles and Activities (4 items)

   Performance of usual social roles and activities

7. Pain Interference (4 items)

   Consequences of pain on relevant aspects of life

8. Pain Intensity (1 item)

   0-10 numerical rating of pain intensity

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.88\text{-}0.97$ across all domains

Test-retest reliability: ICC = 0.81-0.94

#### Validity:

Strong correlations with corresponding legacy instruments ($r = 0.65\text{-}0.87$)

Discriminates between healthy individuals and clinical populations

Sensitive to change in clinical trials and longitudinal studies

Validated in 100+ clinical populations

### Scoring

Each domain (except Pain Intensity) produces a T-score

Pain Intensity scored 0-10 (raw score used)

Scoring available via Health Measures Scoring Service or manual lookup tables

Higher scores = More of the concept measured

### Clinical Interpretation

Clinically Significant T-Score Thresholds:

Physical Function: < 40 (significantly impaired)

Anxiety: ≥ 60 (clinically significant)

Depression: ≥ 60 (clinically significant)

Fatigue: ≥ 60 (severe fatigue)

Sleep Disturbance: ≥ 60 (severe disturbance)

Social Participation: < 40 (significantly impaired)

Pain Interference: ≥ 60 (severe interference)

Pain Intensity: ≥ 7 (severe pain)

### Clinical Applications

Comprehensive health assessment in primary care

Pre-treatment evaluation across multiple domains

Monitoring response to treatment

Population health screening

Clinical research requiring brief multidimensional assessment

### Advantages

Brief yet comprehensive (7-8 minutes)

Covers key domains of health-related quality of life

Standardised scoring allows cross-condition comparisons

Well-validated across diverse populations

Available in 30+ languages

#### Administration

29 items total

Time: 5-8 minutes

Recall period: Past 7 days

Reading level: 6th grade

Self-administered or interview format

#### Access and Licensing

FREE for clinical and research use

Download from www.healthmeasures.net

Attribution required

Available in paper and electronic formats

## 3. Pain and Disability Measures

### 3.1 PEG-3 Pain Screening Tool

#### Description

The PEG is an ultra-brief, 3-item pain assessment tool derived from the Brief Pain Inventory (BPI). It evaluates pain intensity (P), interference with Enjoyment of life (E), and interference with General activity (G). The PEG provides a rapid assessment of pain and its impact in busy clinical settings.

#### Items

1. Pain intensity (average pain in past week, 0-10 scale)

2. Interference with enjoyment of life (0-10 scale)

3. Interference with general activity (0-10 scale)

#### Psychometric Properties

Reliability:

Internal consistency: $\alpha = 0.89$

Test-retest reliability: $r = 0.72\text{-}0.85$

Validity:

High correlation with BPI (r = 0.89)

Responsiveness to change: Effect size = 0.44-0.92

Discriminates between pain severity levels

Validated in primary care, oncology, and musculoskeletal populations

### Scoring

Each item scored 0-10

Total score = Average of 3 items (range 0-10)

0 = No pain/interference

10 = Worst imaginable pain/complete interference

### Interpretation

Score 0-3: Mild pain impact

Score 4-6: Moderate pain impact

Score 7-10: Severe pain impact

Minimum clinically important difference (MCID): 1.0 points

### Clinical Applications

Rapid pain screening in primary care

Quick assessment in busy clinical settings

Monitoring pain treatment effectiveness

Population health surveillance

Quality improvement initiatives

### Advantages

Ultra-brief (< 1 minute)

Assesses both intensity and impact

Easy to score and interpret

Well-validated across diverse populations

Responsive to clinical change

### Administration

3 items only

Time: 30-60 seconds

Recall: Past 7 days

Reading level: 5th grade

Self-administered

#### Access

Public domain - FREE

Available at www.mytopcare.org

No permission required

May be freely reproduced

## 3.2 Visual Analog Scale (VAS)

#### Description

The Visual Analog Scale is a single-item measure of pain intensity consisting of a 100mm horizontal line anchored by word descriptors at each end. The patient marks the line at the point that best represents their current pain intensity.

#### Format

100mm horizontal line

Left anchor: 'No pain' (0mm)

Right anchor: 'Worst pain imaginable' (100mm)

Patient marks position on line

Measured in millimeters from left edge

#### Psychometric Properties

##### Reliability:

Test-retest reliability: r = 0.71-0.94

High correlation with numerical rating scales: r = 0.71-0.94

Varies by pain condition and population

##### Validity:

Strong construct validity with other pain measures

Sensitive to pain changes with treatment

Widely validated across numerous pain conditions

Responsive to clinical change: Effect sizes 0.30-1.20

### Scoring

Measure distance from left edge in millimeters (0-100mm)

Score = Distance in mm

0-4mm: No pain

5-44mm: Mild pain

45-74mm: Moderate pain

75-100mm: Severe pain

### Interpretation

Minimum clinically important difference: 10-13mm (10-13 points)

Substantial clinical benefit: 20-30mm reduction

Higher scores indicate worse pain

### Clinical Applications

Assessment of acute and chronic pain

Pre- and post-treatment comparison

Research studies requiring continuous pain measurement

Clinical trials

### Advantages

Simple and quick (< 1 minute)

Provides continuous data for statistical analysis

Widely understood by patients and clinicians

Language-independent (minimal text)

Sensitive to small changes in pain

### Limitations

Requires manual measurement

Difficult to administer by phone

Some patients find it confusing

Photocopying can distort the 100mm line

Electronic versions may not be equivalent to paper

#### Administration

1 item

Time: 15-30 seconds

Timeframe: Flexible (current, past 24 hours, past week)

Self-administered

Ruler required for scoring

#### Access

Public domain - FREE

No copyright restrictions

Can be freely reproduced

### 3.3 Numerical Rating Scale (NRS)

#### Description

The Numerical Rating Scale is a simple, single-item pain intensity measure where patients rate their pain on an 11-point scale from 0 (no pain) to 10 (worst pain imaginable). It is one of the most commonly used pain assessment tools worldwide.

#### Format

11-point scale: 0 to 10

0 = No pain

10 = Worst pain imaginable

Patient selects single number

#### Psychometric Properties

##### Reliability:

Test-retest reliability: r = 0.95-0.96

High correlation with VAS: r = 0.86-0.95

Excellent inter-rater reliability in verbal format

##### Validity:

Strong construct validity with other pain measures

Discriminates between pain severity categories

Sensitive to treatment effects

Validated across all age groups (6+ years)

Validated in acute, chronic, cancer, and postoperative pain

### Scoring

Direct numerical score: 0-10

0 = No pain

1-3 = Mild pain

4-6 = Moderate pain

7-10 = Severe pain

### Interpretation

Minimum clinically important difference: 2 points (or 30% reduction)

Moderate improvement: 3-4 points reduction

Substantial improvement: ≥5 points reduction or ≥50% reduction

### Clinical Applications

Pain assessment in all clinical settings

'Fifth vital sign' in hospitals and clinics

Pre- and post-procedure pain monitoring

Chronic pain management

Acute pain assessment (emergency department, post-surgical)

Telephone or electronic assessment

### Advantages

Extremely simple and quick (< 30 seconds)

Easy to administer verbally, on paper, or electronically

Universally understood across cultures

Suitable for phone administration

No special equipment or measurement required

Preferred by many patients over VAS

Generates discrete data

### Variations

NRS-11 (0-10): Standard version

NRS-101 (0-100): Used in some research settings

Verbal NRS: Spoken administration

Graphic NRS: Numbers displayed on line or boxes

#### Administration

1 item

Time: 15-30 seconds

Timeframe: Flexible (current, worst, average, least pain)

Common timeframes: Now, past 24 hours, past 7 days

Self-administered or verbal

#### Access

Public domain - FREE

No copyright restrictions

Can be freely used and reproduced

### 3.4 McGill Pain Questionnaire (MPQ)

#### Description

The McGill Pain Questionnaire is a comprehensive multidimensional pain assessment tool that evaluates the sensory, affective, and evaluative dimensions of pain experience. It uses qualitative pain descriptors to capture the complex nature of pain beyond simple intensity.

#### Versions

1. Long Form (LF-MPQ): 78 pain descriptors

2. Short Form (SF-MPQ): 15 pain descriptors

3. Short Form-2 (SF-MPQ-2): 22 descriptors (includes neuropathic pain descriptors)

#### SF-MPQ-2 Domains

Continuous Pain (6 items)

Intermittent Pain (6 items)

Neuropathic Pain (6 items)

Affective Descriptors (4 items)

#### Psychometric Properties

Reliability:

Internal consistency: α = 0.88-0.95 (SF-MPQ-2)

Test-retest reliability: ICC = 0.71-0.87

Good stability over short periods

### Validity:

Discriminates between different pain syndromes

Sensitive to treatment effects

Correlates with other pain and QOL measures (r = 0.40-0.75)

SF-MPQ-2 validated in multiple pain conditions

Neuropathic subscale discriminates neuropathic from non-neuropathic pain

## Scoring

SF-MPQ-2 Scoring:

22 items, each rated 0-10

0 = None, 10 = Worst possible

Total score: 0-220 (sum of all items)

Subscale scores:

- Continuous: 0-60

- Intermittent: 0-60

- Neuropathic: 0-60

- Affective: 0-40

## Interpretation

Higher scores indicate more severe pain

Subscale analysis helps characterize pain quality

Neuropathic subscale score ≥12-14 suggests neuropathic pain

Useful for tracking changes in pain character over time

## Clinical Applications

Comprehensive pain assessment in specialty pain clinics

Characterizing complex pain syndromes

Distinguishing neuropathic from nociceptive pain

Monitoring changes in pain quality with treatment

Research on pain mechanisms and treatment responses

#### Advantages

Captures multidimensional nature of pain

Provides qualitative pain description

Helps identify neuropathic pain components

Widely translated and validated

SF-MPQ-2 relatively brief (5-10 minutes)

#### Limitations

Longer completion time than simple scales

Requires higher literacy level

Some descriptors may be unfamiliar to patients

May be challenging for patients with cognitive impairment

#### Administration

SF-MPQ-2: 22 items

Time: 5-10 minutes

Recall: Past week typically

Reading level: 7th-8th grade

Self-administered or interview

#### Access

SF-MPQ-2: FREE for clinical use

Available from Dr. Ronald Melzack's website

Permission required for commercial use or modification

Available at www.npcrc.org/

### 3.5 Oswestry Disability Index (ODI)

#### Description

The Oswestry Disability Index is the most widely used condition-specific outcome measure for spinal disorders. It assesses the impact of low back pain on 10 activities of daily living.

### Domains Assessed (10 sections)

Pain intensity

Personal care (washing, dressing)

Lifting

Walking

Sitting

Standing

Sleeping

Sex life

Social life

Traveling

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.71\text{-}0.87$

Test-retest reliability: ICC = 0.83-0.99

Highly reliable in low back pain populations

#### Validity:

Strong correlation with VAS pain (r = 0.50-0.80)

Moderate correlation with SF-36 physical scales (r = 0.40-0.80)

Discriminates between different levels of disability

Validated in numerous back pain populations

Responsive to change: Effect size = 0.78

### Scoring

10 sections, each scored 0-5

Total score = (Sum of sections / Number of sections completed) × 100

Score reported as percentage: 0-100%

0% = No disability

100% = Maximum disability

#### Interpretation

0-20%: Minimal disability

21-40%: Moderate disability

41-60%: Severe disability

61-80%: Crippled

81-100%: Bed-bound or exaggerating

Minimum clinically important difference: 10-12 points

#### Clinical Applications

Assessment of low back pain disability

Pre- and post-treatment outcome measurement

Monitoring progress in physical therapy

Surgical outcome evaluation

Disability rating for compensation claims

#### Administration

10 sections, 6 items each

Time: 5 minutes

Self-administered

Paper or electronic

#### Access

Originally public domain

Version 2.0 available through Mapi Research Trust

Permission required for commercial use

### 3.6 Roland-Morris Disability Questionnaire (RMDQ)

#### Description

The Roland-Morris Disability Questionnaire is a condition-specific measure of disability due to low back pain. It consists of 24 items describing daily activities that patients find difficult due to back pain.

#### Psychometric Properties

Reliability:

Internal consistency: α = 0.84-0.93

Test-retest reliability: ICC = 0.83-0.91

### Validity:

Correlates highly with ODI (r = 0.70-0.80)

Correlates moderately with pain intensity (r = 0.40-0.55)

Sensitive to clinical change

Validated across diverse back pain populations

## Scoring

24 items, yes/no format

Each 'yes' = 1 point

Total score: 0-24

Higher scores = greater disability

## Interpretation

0-4: Minimal disability

5-14: Moderate disability

15-24: Severe disability

MCID: 2-3 points (or 30% improvement)

## Clinical Applications

Primary care back pain assessment

Short-term outcome measurement

Research in non-surgical populations

More sensitive to mild-moderate disability than ODI

## Advantages

Quick to complete (2-5 minutes)

Simple yes/no format

Easy to score

Good for less severe disability

Available in 30+ languages

#### Administration

24 items

Time: 3-5 minutes

Yes/no format

Self-administered

#### Access

Public domain - FREE

Available at www.rmdq.org

Freely reproducible

### 3.7 Neck Disability Index (NDI)

#### Description

The Neck Disability Index is the most widely used outcome measure for neck pain, adapted from the Oswestry Disability Index. It assesses the impact of neck pain on 10 activities of daily living.

#### Domains (10 sections)

Pain intensity

Personal care

Lifting

Reading

Headaches

Concentration

Work

Driving

Sleeping

Recreation

#### Psychometric Properties

Reliability:

Internal consistency: $\alpha = 0.80\text{-}0.89$

Test-retest reliability: ICC = 0.89-0.98

### Validity:

Correlates with neck pain VAS (r = 0.60-0.70)

Correlates with SF-36 physical scales (r = 0.47-0.69)

Responsive to change: Effect sizes 0.60-1.40

Validated in multiple neck pain populations

## Scoring

10 sections, each scored 0-5

Total score = (Sum / 50) × 100

Score as percentage: 0-100%

0% = No disability

100% = Complete disability

## Interpretation

0-8%: No disability

10-28%: Mild disability

30-48%: Moderate disability

50-68%: Severe disability

70-100%: Complete disability

MCID: 7.5-10 points or 20% change

## Clinical Applications

Assessment of neck pain disability

Pre/post treatment outcomes

Conservative and surgical treatment evaluation

Whiplash assessment

Cervical spondylosis outcomes

## Administration

10 sections

Time: 5-10 minutes

Self-administered

Available in 30+ languages

#### Access

Copyrighted by Dr. Vernon Howard

FREE for clinical use

Available at www.mapi-trust.org

Permission required for commercial use

## 3.8 Bournemouth Questionnaire (Neck & Back Pain)

#### Description

The Bournemouth Questionnaire assesses the multidimensional aspects of pain, disability, and psychosocial factors in neck or back pain. Separate versions exist for neck pain (BQ-N) and back pain (BQ-B).

#### Domains (7 items for each version)

Pain intensity

Physical disability

Social disability

Fear-avoidance beliefs

Anxiety

Depression

Locus of control

#### Psychometric Properties

Reliability:

Internal consistency: $\alpha = 0.79\text{-}0.92$

Test-retest reliability: ICC = 0.88-0.93

Validity:

Correlates with ODI/NDI (r = 0.60-0.75)

Correlates with SF-36 (r = 0.45-0.65)

Responsive to clinical change

Validated in primary care and chiropractic settings

#### Scoring

7 items, each scored 0-10

Total score: 0-70

Higher scores = worse status

No subscale scores (use total only)

### Interpretation

0-23: Mild

24-46: Moderate

47-70: Severe

MCID: 10-15 points or 30% change

### Clinical Applications

Biopsychosocial assessment of spinal pain

Primary care and chiropractic settings

Screening for psychosocial factors

Brief multidimensional outcome measure

### Advantages

Very brief (2-3 minutes)

Covers multiple domains

Identifies psychosocial risk factors

Widely used in chiropractic practice

### Administration

7 items

Time: 2-3 minutes

0-10 numerical scales

Self-administered

### Access

FREE for clinical use

www.bournemouth.ac.uk/research/projects/bournemouth-questionnaires

Permission required for commercial or research use

# 4. Health Status and Well-being Measures

## 4.1 SF-36 Health Survey

### Description

The 36-Item Short Form Health Survey (SF-36) is one of the most widely used generic health-related quality of life questionnaires. It measures eight domains of health across physical and mental components.

### Eight Health Domains

Physical Functioning (PF) - 10 items

Role-Physical (RP) - 4 items

Bodily Pain (BP) - 2 items

General Health (GH) - 5 items

Vitality (VT) - 4 items

Social Functioning (SF) - 2 items

Role-Emotional (RE) - 3 items

Mental Health (MH) - 5 items

Plus 1 item on health change

### Component Scores

Physical Component Summary (PCS): PF + RP + BP + GH

Mental Component Summary (MCS): VT + SF + RE + MH

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.78\text{-}0.93$ across scales

Test-retest: ICC = 0.75-0.90

Extensively validated in general and clinical populations

#### Validity:

Discriminates between groups with different conditions

Sensitive to clinical change

Strong correlation with other health status measures

Normative data available for general population

### Scoring

Each scale scored 0-100

0 = Worst health, 100 = Best health

Component summaries: T-scores (Mean=50, SD=10)

Scoring requires licensed software or manual calculation

### Interpretation

50 = U.S. population average

Each 10 points = 1 standard deviation

Score < 40: More than 1 SD below average

MCID varies by scale: 3-5 points typically

### Clinical Applications

General health assessment across conditions

Comparing outcomes across different treatments

Quality of life research

Health economics and cost-effectiveness studies

Population health surveys

### Administration

36 items

Time: 5-10 minutes

Recall: Past 4 weeks (standard), 1 week (acute)

Self-administered or interview

Available in 170+ languages

### Access

Licensed through QualityMetric/RAND

FREE for unfunded academic research (RAND version)

License required for commercial and funded research

www.qualitymetric.com or

www.rand.org/health-care/surveys_tools/mos/36-item-short-form.html

## 4.2 Global Well-Being Schedule (GWBS)

### Description

The Global Well-Being Schedule is an 18-item measure of psychological well-being and distress developed to assess mental health in general populations. It covers six domains: anxiety, depression, positive well-being, self-control, vitality, and general health. The GWBS provides a comprehensive assessment of subjective psychological and physical health status.

### Items

Anxiety (3 items)

Depression (3 items)

Positive well-being (3 items)

Self-control (3 items)

Vitality (3 items)

General health perception (3 items)

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.91\text{-}0.94$

Test-retest reliability: $r = 0.85$ (2 weeks)

High inter-scale correlations

#### Validity:

Strong correlation with other psychological well-being measures

Discriminates between clinical and non-clinical populations

Sensitive to treatment effects

Validated in multiple populations

### Scoring

Each item rated on 6-point scale

Total score range: 0-110

Higher scores = better well-being

Subscale scores can be calculated for each domain

### Interpretation

Score 0-60: Severe distress

Score 61-72: Moderate distress

Score 73-110: Positive well-being

MCID: 7-10 points

### Clinical Applications

Mental health screening in primary care

Monitoring psychological well-being over time

Treatment outcome assessment

Population health surveys

Research on mind-body interventions

### Advantages

Comprehensive assessment of multiple well-being domains

Brief administration time

Validated across diverse populations

Sensitive to change

Suitable for repeated administration

### Administration

18 items

Time: 5-7 minutes

Recall period: Past month

Reading level: 6th grade

Self-administered

Access

Available through researchers

Permission required for use

Contact: wellbeing@research.org

Modest licensing fee for commercial use

## 4.3 Patient Global Impression of Change (PGIC)

### Description

The PGIC is a single-item rating by patients of their overall improvement (or worsening) with treatment. It uses a 7-point scale ranging from 'very much improved' to 'very much worse.' The PGIC is widely used as an anchor for determining minimal clinically important differences and as an outcome measure in clinical trials.

### Items

Single global rating of change since beginning of treatment

### Psychometric Properties

#### Reliability:

Test-retest: r = 0.90 (stable patients)

High agreement with clinician ratings

#### Validity:

Correlates strongly with change scores on disease-specific measures

Predicts satisfaction with treatment

Useful anchor for MCID determination

Validated across numerous conditions

### Scoring

1 = Very much improved

2 = Much improved

3 = Minimally improved

4 = No change

5 = Minimally worse

6 = Much worse

7 = Very much worse

### Interpretation

Scores 1-2 (very much/much improved): Clinically meaningful improvement

Score 3 (minimally improved): Possible meaningful change

Score 4 (no change): No treatment benefit

Scores 5-7 (worse): Treatment failure or disease progression

Often dichotomized as 'improved' (1-3) vs 'not improved' (4-7)

### Clinical Applications

Clinical trials as primary or secondary outcome

Clinical practice for treatment monitoring

Anchor for MCID calculations

Patient satisfaction assessment

Quality improvement initiatives

### Advantages

Single item - minimal burden

Easy to understand and complete

Patient-centered perspective

Applicable across all conditions

Correlates with objective measures

### Administration

1 item only

Time: <30 seconds

Recall: Since treatment started

Reading level: 4th grade

Self-administered or interviewer

### Access

Public domain - FREE

No permission required

Widely available in literature

Multiple language versions available

## 4.4 EQ-5D

### Description

The EuroQol-5 Dimension (EQ-5D) is a standardized measure of health status that provides a simple, generic measure of health for clinical and economic evaluation. It comprises a descriptive system with five dimensions and a visual analogue scale (EQ-VAS). The EQ-5D is one of the most widely used health-related quality of life instruments globally and is recommended for health technology assessments and economic evaluations.

## Items

Mobility - ability to walk about

Self-care - ability to wash and dress

Usual activities - ability to perform usual activities

Pain/discomfort - presence and severity

Anxiety/depression - presence and severity

## Psychometric Properties

### Reliability:

Test-retest: ICC = 0.70-0.90

Inter-rater agreement: $\kappa$ = 0.70-0.92

Consistent across multiple studies

### Validity:

Correlates with disease-specific measures (r = 0.60-0.80)

Discriminates between severity groups

Responsive to clinical change

Validated in 100+ countries

Extensive normative data available

## Scoring

EQ-5D-5L: Each dimension rated on 5 levels (1=no problems to 5=extreme problems)

Descriptive system generates 3,125 possible health states

Index value calculated using country-specific value sets (range -0.59 to 1.0)

1.0 = perfect health, 0 = death, negative values = states worse than death

EQ-VAS: 0-100 scale where 0=worst imaginable health, 100=best imaginable health

## Interpretation

Index values:

>0.80: Good health

0.60-0.80: Moderate health problems

0.40-0.60: Significant health impairment

<0.40: Severe health problems

MCID: 0.07-0.10 for index value; 7-10 points for VAS

Can compare to population norms

### Clinical Applications

Economic evaluation and cost-utility analysis (QALYs)

Clinical trials as generic health outcome

Population health surveys

Health technology assessment

Quality indicators for healthcare systems

Treatment monitoring in chronic conditions

### Advantages

Very brief (< 2 minutes)

Simple and easy to understand

Available in 200+ languages

Generates utility values for economic evaluation

Extensive normative data

Recommended by NICE and other HTA bodies

### Administration

5 items + VAS

Time: 1-2 minutes

Recall: Today

Reading level: 5th grade

Self-administered or interview

Available in paper, electronic, and phone versions

#### Access

FREE for non-commercial use

Registration required at www.euroqol.org

Commercial license available

Multiple language versions

User guide and value sets provided

## 5. Patient Satisfaction Measures

### 5.1 VSQ-9 (Visit-Specific Satisfaction Questionnaire)

#### Description

The VSQ-9 is a 9-item measure of patient satisfaction with a specific medical visit. It assesses satisfaction with visit duration, explanation received, technical quality, interpersonal manner, and overall satisfaction. Originally developed for primary care, it has been validated across multiple healthcare settings.

#### Domains

Time spent with provider (duration)

Explanation of problem and treatment

Technical skills of provider

Personal manner of provider

Overall satisfaction with visit

#### Psychometric Properties

##### Reliability:

Internal consistency: $\alpha = 0.92$

Test-retest: $r = 0.77$

Stable across different clinical settings

##### Validity:

Strong correlation with likelihood to recommend ($r = 0.85$)

Predicts return visit intention

Discriminates between satisfied/dissatisfied patients

Validated in primary care, specialty care, urgent care

### Scoring

Each item rated on 5-point scale (1=Poor to 5=Excellent)

Total score range: 9-45

Calculate mean score by dividing total by 9

Higher scores indicate greater satisfaction

### Interpretation

Score 4.0-5.0: High satisfaction

Score 3.0-3.9: Moderate satisfaction

Score 2.0-2.9: Low satisfaction

Score <2.0: Very low satisfaction

Individual item scores identify specific areas for improvement

## Clinical Applications

Quality improvement in clinical practice

Patient experience monitoring

Provider performance evaluation

Comparing satisfaction across providers/sites

Research on factors affecting satisfaction

## Advantages

Brief and focused on single visit

Easy to complete immediately post-visit

Actionable results for quality improvement

Validated across multiple settings

Sensitive to visit-specific factors

## Administration

9 items

Time: 2-3 minutes

Recall: Current visit

Reading level: 6th grade

Self-administered (paper or electronic)

#### Access

Public domain - FREE

Available from RAND Corporation

www.rand.org/health-care/surveys_tools/vsq.html

No permission required

### 5.2 MYMOP2 (Measure Yourself Medical Outcome Profile)

#### Description

MYMOP2 is a patient-generated outcome measure that allows patients to choose which symptoms or activities to track. It includes patient-specified problems (symptoms), an activity affected by the problem, general well-being, and (optionally) medication use. This individualized approach makes it highly relevant to each patient's unique situation.

#### Domains

Symptom 1: Patient chooses main symptom (0-6 scale)

Symptom 2: Patient chooses second symptom if applicable (0-6 scale)

Activity: Activity most affected by problem (0-6 scale)

Well-being: General sense of well-being (0-6 scale)

Medication: Optional medication tracking

#### Psychometric Properties

##### Reliability:

Test-retest: r = 0.89 for stable patients

Internal consistency not applicable (individualized items)

##### Validity:

Correlates with generic measures (r = 0.50-0.70)

Responsive to patient-perceived change

Predicts treatment satisfaction

Validated in complementary medicine, primary care, chronic disease

### Scoring

Each item scored 0-6:

0 = As good as it could be

6 = As bad as it could be

Profile score = average of symptom(s), activity, and well-being

Range: 0-6 (lower scores = better outcomes)

### Interpretation

Change of 1.0 points or more: Clinically important improvement

Change of 0.5-0.9: Small improvement

No change or worsening: Consider treatment modification

Individual items show which areas improved

### Clinical Applications

Monitoring treatment in complementary/alternative medicine

Tracking outcomes in chronic disease management

Patient-centered outcome assessment

Situations where standardized measures miss key issues

Research on individualized interventions

### Advantages

Patient-chosen outcomes ensure relevance

Captures issues important to individual patient

Simple and quick to complete

Suitable for repeated measurement

Works across all conditions and treatments

### Administration

4 core items (+ optional medication item)

Time: 2-3 minutes at first use; <1 minute for follow-up

Recall: Last week

Reading level: 5th grade

Self-administered with initial explanation

### Access

FREE to use

Available at www.bris.ac.uk/primaryhealthcare/resources/mymop

User guide provided

Multiple language versions available

## 6. Psychosocial Screening Tools

### 6.1 STarT Back Screening Tool (Keele)

#### Description

The STarT (Subgroups for Targeted Treatment) Back Screening Tool is a 9-item questionnaire that stratifies primary care low back pain patients into low, medium, or high-risk groups for persistent disabling pain. It identifies modifiable physical and psychosocial risk factors (yellow flags) to guide treatment decisions. The tool has been shown to improve outcomes when used to match patients to appropriate treatment intensity.

#### Domains

Referred leg pain (item 1)

Comorbid pain (item 2)

Disability - dressing (item 3)

Disability - walking (item 4)

Fear (item 5) - psychosocial subscale

Catastrophizing (item 6) - psychosocial subscale

Bothersomeness (item 7) - psychosocial subscale

Depression (item 8) - psychosocial subscale

Overall impact (item 9)

#### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.79$ (overall); $\alpha = 0.74$ (psychosocial subscale)

Test-retest: $\kappa = 0.79$ for risk classification

#### Validity:

Predicts persistent disability at 6 months (AUC = 0.75)

Correlates with disability scores (r = 0.66)

Psychosocial subscale predicts poor outcome

Validated in multiple countries and settings

Implementation shown to improve patient outcomes

### Scoring

Items 1-4 and 9: Yes (1 point) or No (0 points)

Items 5-8: Agree (1 point) or Disagree (0 points)

Total score: 0-9 points

Psychosocial subscale: Sum of items 5-8 (0-4 points)

### Interpretation

Overall score 0-3 AND psychosocial ≤3: LOW RISK

Overall score 4-9 AND psychosocial ≤3: MEDIUM RISK

Overall score 4-9 AND psychosocial ≥4: HIGH RISK

Low risk: Minimal treatment, self-management advice

Medium risk: Physical therapy, structured exercise

High risk: Psychologically-informed physiotherapy or multidisciplinary care

### Clinical Applications

Stratified care approach for low back pain

Triage in primary care

Matching treatment intensity to risk

Identifying patients needing psychosocial intervention

Resource allocation decisions

### Advantages

Quick screening tool (<2 minutes)

Identifies modifiable psychosocial risk factors

Clear treatment allocation algorithm

Improves outcomes when used properly

Cost-effective stratified care approach

### Administration

9 items

Time: 1-2 minutes

Recall: Current episode

Reading level: 6th grade

Self-administered

### Access

FREE for non-commercial use

Available at www.keele.ac.uk/sbst

Registration required

User guide and implementation materials provided

Multiple language versions

## 6.2 Fear-Avoidance Beliefs Questionnaire (FABQ)

### Description

The FABQ is a 16-item questionnaire measuring fear-avoidance beliefs about physical activity and work in patients with low back pain. It has two subscales: Physical Activity (FAB-PA, 4 items) and Work (FAB-W, 7 items). Fear-avoidance beliefs are strong predictors of disability and treatment outcomes in musculoskeletal pain conditions.

### Domains

Physical Activity subscale (4 items):

Beliefs about harm from physical activity

Fear of injury/re-injury from movement

Work subscale (7 items):

Beliefs about work causing pain

Inability to return to normal work

Beliefs about harm from current job

(5 items not used in scoring)

## Psychometric Properties

### Reliability:

Internal consistency: α = 0.88 (FAB-PA); α = 0.88 (FAB-W)

Test-retest: r = 0.88 (FAB-PA); r = 0.82 (FAB-W)

### Validity:

Predicts disability and work loss

FAB-W strongly predicts return-to-work outcomes

Correlates with pain catastrophizing (r = 0.50-0.60)

Discriminates between pain severity groups

Extensively validated in LBP populations

## Scoring

Each item rated 0-6 (0=completely disagree, 6=completely agree)

FAB-PA score: Sum of items 2, 3, 4, 5 (range 0-24)

FAB-W score: Sum of items 6, 7, 9, 10, 11, 12, 15 (range 0-42)

Higher scores = greater fear-avoidance beliefs

## Interpretation

FAB-PA scores:

<15: Low fear-avoidance

≥15: High fear-avoidance

FAB-W scores:

<34: Low work-related fear

≥34: High work-related fear (poor prognosis for RTW)

MCID: 4-7 points (FAB-W)

## Clinical Applications

Identifying patients with maladaptive beliefs

Predicting return-to-work outcomes

Selecting patients for cognitive-behavioral intervention

Monitoring treatment response

Research on fear-avoidance model

#### Advantages

Identifies psychosocial risk factor

Strong predictor of disability and work loss

Guides treatment selection

Responsive to cognitive-behavioral interventions

Well-established clinical cutoffs

#### Administration

16 items (11 scored)

Time: 5 minutes

Recall: Current beliefs

Reading level: 7th grade

Self-administered

#### Access

Public domain – FREE

Widely available in literature

No permission required

Multiple language versions available

## 7. Mental and Behavioural Assessments

### 7.1 PHQ-9 (Patient Health Questionnaire-9)

#### Description

The PHQ-9 is a 9-item depression screening and severity measure based on DSM-5 criteria for major depressive disorder. It assesses the nine criterion symptoms over the past 2 weeks. The PHQ-9 is one of the most widely used depression screening tools globally, validated across diverse populations and settings.

#### Items

Little interest or pleasure in doing things

Feeling down, depressed, or hopeless

Trouble falling/staying asleep or sleeping too much

Feeling tired or having little energy

Poor appetite or overeating

Feeling bad about yourself or that you're a failure

Trouble concentrating

Moving/speaking slowly or being fidgety/restless

Thoughts of being better off dead or hurting yourself

## Psychometric Properties

### Reliability:

Internal consistency: $\alpha = 0.86\text{-}0.89$

Test-retest: $r = 0.84$

### Validity:

Sensitivity: 88% for major depression (cutoff ≥10)

Specificity: 88% for major depression

Correlates highly with other depression measures ($r = 0.70\text{-}0.90$)

Validated in primary care, specialty care, general populations

Sensitive to treatment effects

Useful across cultures and languages

## Scoring

Each item rated 0-3:

0 = Not at all

1 = Several days

2 = More than half the days

3 = Nearly every day

Total score: 0-27

## Interpretation

Score 1-4: Minimal depression

Score 5-9: Mild depression

Score 10-14: Moderate depression

Score 15-19: Moderately severe depression

Score 20-27: Severe depression

Score ≥10: Warrants clinical assessment

Score ≥15: Consider pharmacotherapy/specialist referral

Item 9 (suicidal ideation): Any positive response requires immediate assessment

MCID: 5 points

### Clinical Applications

Depression screening in primary care and specialty settings

Monitoring treatment response

Severity assessment to guide treatment intensity

Population health surveillance

Research on depression interventions

### Advantages

Brief and easy to administer

Based on DSM-5 diagnostic criteria

Dual purpose: screening and severity measurement

Validated across diverse populations

Free and widely available

Includes suicide risk item

### Administration

9 items

Time: 2-3 minutes

Recall: Past 2 weeks

Reading level: 6th grade

Self-administered or clinician-administered

### Access

Public domain - FREE

Available at www.phqscreeners.com

No permission required

Multiple language versions (100+)

Electronic versions available

## 7.2 DASS-21 (Depression Anxiety Stress Scale)

### Description

The DASS-21 is a 21-item self-report measure of depression, anxiety, and stress/ tension with 7 items per subscale. It distinguishes between the three negative emotional states and provides severity ratings for each. The DASS-21 is widely used in research and clinical practice to assess emotional distress.

### Subscales and Items

- Depression subscale (7 items): Dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest, anhedonia, inertia
- Anxiety subscale (7 items): Autonomic arousal, skeletal muscle effects, situational anxiety, subjective experience of anxious affect
- Stress subscale (7 items): Difficulty relaxing, nervous arousal, agitation, irritability, impatience

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.88$ (Depression), 0.82 (Anxiety), 0.90 (Stress)

Test-retest: $r = 0.71$-$0.81$

#### Validity:

Discriminates between depression, anxiety, and stress

Correlates with other measures ($r = 0.60$-$0.80$)

Sensitive to treatment effects

Validated in clinical and non-clinical samples

Factor structure supported across studies

### Scoring

Each item rated 0-3:

0 = Did not apply to me at all

1 = Applied to me to some degree

2 = Applied to me a considerable degree

3 = Applied to me very much

Sum each subscale (range 0-21)

Multiply by 2 to compare with DASS-42 (range 0-42)

## Interpretation

Depression (multiply scores by 2 for severity ranges):

0-9: Normal

10-13: Mild

14-20: Moderate

21-27: Severe

28+: Extremely severe

### Anxiety (multiply by 2):

0-7: Normal

8-9: Mild

10-14: Moderate

15-19: Severe

20+: Extremely severe

Stress (multiply by 2):

0-14: Normal

15-18: Mild

19-25: Moderate

26-33: Severe

34+: Extremely severe

## Clinical Applications

Screening for emotional distress in various settings

Distinguishing between depression, anxiety, and stress

Treatment outcome monitoring

Research on emotional disorders

Workplace mental health assessment

#### Advantages

Assesses three distinct constructs

Good discrimination between states

Free and widely available

Suitable for clinical and research use

Can identify comorbid conditions

#### Administration

21 items

Time: 5-7 minutes

Recall: Past week

Reading level: 7th grade

Self-administered

#### Access

Public domain - FREE

Available at www2.psy.unsw.edu.au/dass

No permission required

Multiple language versions

Manual and norms provided

### 7.3 GAD-7 (Generalized Anxiety Disorder-7)

#### Description

The GAD-7 is a brief 7-item measure of generalized anxiety disorder symptoms based on DSM-5 criteria. It assesses worry, tension, and other anxiety symptoms over the past 2 weeks. The GAD-7 is one of the most widely used anxiety screening tools and has excellent performance characteristics.

#### Items

Feeling nervous, anxious, or on edge

Not being able to stop or control worrying

Worrying too much about different things

Trouble relaxing

Being so restless that it's hard to sit still

Becoming easily annoyed or irritable

Feeling afraid as if something awful might happen

## Psychometric Properties

### Reliability:

Internal consistency: α = 0.92

Test-retest: ICC = 0.83

### Validity:

Sensitivity: 89% for GAD (cutoff ≥10)

Specificity: 82% for GAD

Also screens for panic disorder, social anxiety, PTSD

Correlates with other anxiety measures (r = 0.72-0.80)

Sensitive to change with treatment

Validated across diverse populations

## Scoring

Each item rated 0-3:

0 = Not at all

1 = Several days

2 = More than half the days

3 = Nearly every day

Total score: 0-21

## Interpretation

Score 5-9: Mild anxiety

Score 10-14: Moderate anxiety

Score ≥15: Severe anxiety

Score ≥10: Consider further assessment and treatment

Score ≥15: Active treatment warranted

MCID: 4 points

Can be used with PHQ-9 for comorbid depression screening

### Clinical Applications

Anxiety screening in primary care

Severity assessment to guide treatment

Treatment monitoring

Population-based anxiety surveillance

Research on anxiety interventions

### Advantages

Very brief (<2 minutes)

Excellent psychometric properties

Validated screening cutoffs

Screens for multiple anxiety disorders

Free and widely available

Often paired with PHQ-9

### Administration

7 items

Time: 1-2 minutes

Recall: Past 2 weeks

Reading level: 6th grade

Self-administered

### Access

Public domain - FREE

Available at www.phqscreeners.com

No permission required

Multiple language versions

Electronic versions available

## 7.4 HADS (Hospital Anxiety and Depression Scale)

### Description

The HADS is a 14-item screening tool for anxiety and depression, designed to exclude somatic symptoms that may be confounded by physical illness. It has two 7-item subscales for anxiety (HADS-A) and depression (HADS-D). Originally developed for hospital settings, it's now widely used across all healthcare environments.

### Subscales and Items

Anxiety subscale (HADS-A, 7 items): Worry, tension, panic, fear, restlessness

Depression subscale (HADS-D, 7 items): Anhedonia, low mood, lack of enjoyment (excludes somatic symptoms)

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.80\text{-}0.84$ (HADS-A); $\alpha = 0.82\text{-}0.90$ (HADS-D)

Test-retest: $r = 0.70\text{-}0.90$

#### Validity:

HADS-A sensitivity/specificity: 80%/81% (cutoff ≥8)

HADS-D sensitivity/specificity: 79%/83% (cutoff ≥8)

Minimal contamination by somatic symptoms

Validated in over 700 studies

Discriminates between anxiety and depression

Extensively validated in medical populations

### Scoring

Each subscale scored 0-21:

Items rated 0-3 on various response formats

Sum items for each subscale separately

Total HADS score (0-42) sometimes used

### Interpretation

Each subscale (HADS-A and HADS-D):

0-7: Normal/no case

8-10: Mild/possible case

11-14: Moderate/probable case

15-21: Severe/definite case

Score ≥8 on either subscale: Consider further assessment

Score ≥11: Active treatment recommended

MCID: 1.5-1.7 points per subscale

### Clinical Applications

Screening in medical/hospital settings

Assessment when somatic symptoms confound diagnosis

Monitoring emotional distress in chronic illness

Research in medical populations

Post-surgical psychological screening

### Advantages

Excludes somatic symptoms (ideal for medical patients)

Brief and easy to complete

Separate anxiety and depression assessment

Extensive validation data

Widely translated (100+ languages)

### Administration

14 items

Time: 3-5 minutes

Recall: Past week

Reading level: 6th grade

Self-administered

### Access

Requires purchase/license

Available through GL Assessment

www.gl-assessment.co.uk

Cost varies by usage

Multiple language versions available

## 7.5 Beck Anxiety Inventory (BAI)

### Description

The BAI is a 21-item self-report measure of anxiety symptom severity, developed to distinguish anxiety from depression. It emphasizes physiological and cognitive symptoms of anxiety rather than symptoms shared with depression. The BAI is particularly useful for assessing panic and somatic anxiety symptoms.

### Items

21 anxiety symptoms including:

Physiological: Numbness/tingling, feeling hot, wobbliness in legs, heart pounding

Cognitive: Fear of worst happening, unable to relax, terrified

Autonomic: Sweating, flushing, trembling, dizzy

Panic symptoms: Fear of losing control, difficulty breathing, choking feeling

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.92$

Test-retest: $r = 0.75$ (1 week)

#### Validity:

Correlates moderately with Hamilton Anxiety Scale ($r = 0.51$)

Low correlation with depression measures ($r = 0.25$)

Discriminates anxiety from depression

Sensitive to treatment effects

Validated in clinical and non-clinical samples

### Scoring

Each item rated 0-3:

0 = Not at all

1 = Mildly (did not bother me much)

2 = Moderately (unpleasant but manageable)

3 = Severely (could barely stand it)

Total score: 0-63

### Interpretation

Score 0-7: Minimal anxiety

Score 8-15: Mild anxiety

Score 16-25: Moderate anxiety

Score 26-63: Severe anxiety

Score ≥16: Clinically significant anxiety

Score ≥26: Consider specialist referral

Particularly useful for panic disorder assessment

### Clinical Applications

Assessment of anxiety disorder severity

Distinguishing anxiety from depression

Monitoring treatment response

Research on anxiety disorders

Panic disorder evaluation

### Advantages

Focuses on somatic symptoms of anxiety

Distinguishes anxiety from depression

Validated across anxiety disorders

Sensitive to change

Part of Beck assessment suite (with BDI)

### Administration

21 items

Time: 5-7 minutes

Recall: Past week including today

Reading level: 5th-6th grade

Self-administered or interview

### Access

Requires purchase

Available through Pearson Assessments

www.pearsonassessments.com

License required for use

Training materials available

# 8. Sleep Assessment

## 8.1 Pittsburgh Sleep Quality Index (PSQI)

### Description

The PSQI is a 19-item self-rated questionnaire assessing sleep quality and disturbances over a 1-month time interval. It measures seven clinically-derived domains of sleep difficulty. The PSQI is the gold standard for subjective sleep quality assessment and is widely used in both clinical practice and research.

### Components

Component 1: Subjective sleep quality

Component 2: Sleep latency (time to fall asleep)

Component 3: Sleep duration (hours per night)

Component 4: Sleep efficiency (% of time in bed spent asleep)

Component 5: Sleep disturbances (frequency of problems)

Component 6: Use of sleeping medication

Component 7: Daytime dysfunction due to sleepiness

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.83$

Test-retest: $r = 0.85$ (1 month)

Component correlations support overall construct

#### Validity:

Discriminates good vs. poor sleepers (diagnostic efficiency 89.6%)

Correlates with objective sleep measures (polysomnography)

Sensitive to clinical changes in sleep

Validated across age groups and clinical populations

Predicts health outcomes related to sleep problems

### Scoring

19 self-rated items generate 7 component scores

Each component scored 0-3 (0=no difficulty, 3=severe difficulty)

Global PSQI score = sum of 7 components (range 0-21)

Lower scores indicate better sleep quality

### Interpretation

Global score >5: Poor sleep quality (sensitivity 89.6%, specificity 86.5%)

Score 0-5: Good sleeper

Score 6-10: Moderate sleep problems

Score 11-21: Severe sleep problems

### Component analysis identifies specific problem areas:

High Component 2 → Sleep onset problems

High Component 3 → Insufficient sleep duration

High Component 4 → Poor sleep efficiency

High Component 5 → Frequent disturbances

MCID: 3 points

### Clinical Applications

Sleep disorder screening and diagnosis

Monitoring treatment response for insomnia

Assessing sleep in chronic disease (pain, mental health, cardiac)

Research on sleep interventions

Identifying sleep problems in primary care

### Advantages

Comprehensive assessment of sleep domains

Widely validated and translated (60+ languages)

Gold standard for subjective sleep assessment

Identifies specific sleep problems

Good sensitivity and specificity

#### Administration

19 items (9 self-rated, 10 additional questions)

Time: 5-10 minutes

Recall: Past month

Reading level: 6th-7th grade

Self-administered with scoring by clinician

#### Access

Public domain - FREE

Available at www.sleep.pitt.edu

No permission required for clinical use

Registration requested for research

Scoring guidelines provided

## 9. Activity Level Measures

### 9.1 International Physical Activity Questionnaire (IPAQ)

#### Description

The IPAQ assesses physical activity undertaken across a comprehensive set of domains including leisure time, domestic and gardening activities, work-related and transport-related activity. Available in short (7 items) and long (27 items) forms, the IPAQ provides internationally comparable data on physical activity. It's used extensively in population surveillance and intervention research.

#### Items

Short Form (7 items):

Vigorous-intensity activity

Moderate-intensity activity

Walking

Sitting

Long Form (27 items) assesses same activities across:

Work-related physical activity

Transport-related physical activity

Housework/gardening

Leisure-time physical activity

Time spent sitting

## Psychometric Properties

### Reliability:

Short form test-retest: Spearman ρ = 0.76-0.81

Long form test-retest: ρ = 0.66-0.88

Acceptable to good repeatability across countries

### Validity:

Correlates with objective measures (accelerometry): r = 0.30-0.67

Concurrent validity with other PA questionnaires: ρ = 0.40-0.70

Criterion validity varies by domain and country

Best for moderate-vigorous activity assessment

Validated in 12+ countries

## Scoring

Calculate MET-minutes/week for each activity type:

Walking = 3.3 METs × minutes × days

Moderate activity = 4.0 METs × minutes × days

Vigorous activity = 8.0 METs × minutes × days

Total MET-minutes/week = sum of all activities

Sitting time reported separately (hours/day)

## Interpretation

Physical Activity Categories:

Low: <600 MET-min/week OR <3 days of any intensity

Moderate: ≥600 MET-min/week OR ≥3 days mod-vig activity

High: ≥3000 MET-min/week OR ≥7 days of any activity

Meeting WHO Guidelines:

≥600 MET-min/week = meets recommendations

≥150 min/week moderate OR ≥75 min/week vigorous

Sitting time:

>8 hours/day associated with health risks

### Clinical Applications

Population-level physical activity surveillance

Public health research and monitoring

Evaluating intervention effectiveness

Comparing activity across countries/regions

Identifying inactive populations for targeting

### Advantages

Comprehensive activity assessment across domains
Internationally standardized and comparable

Free and widely available

Multiple language versions (20+)

Used in 100+ countries

### Administration

Short form: 7 items, 5-7 minutes

Long form: 27 items, 10-15 minutes

Recall: Last 7 days

Reading level: 6th-7th grade

Self-administered (paper, phone, online)

### Access

Public domain - FREE

Available at www.ipaq.ki.se

Scoring protocol provided

No permission required

Multiple formats and languages

# 10. Region-Specific Functional Measures

## 10.1 Upper Extremity Functional Index (UEFI)

### Description

The UEFI is a 20-item questionnaire measuring upper extremity functional status and difficulty with activities of daily living. It assesses function regardless of which arm/hand is affected and includes items spanning personal care, household tasks, work activities, and recreation. The UEFI is applicable to any upper extremity condition.

### Items

Personal care items: Dressing, hair care, eating

Household activities: Cooking, cleaning, laundry

Work activities: Writing, computer use, lifting

Recreation: Sports, hobbies, playing musical instrument

Other functional tasks: Opening jars, carrying objects, reaching overhead

### Psychometric Properties

#### Reliability:

Internal consistency: $\alpha = 0.94\text{-}0.96$

Test-retest: ICC = 0.94-0.95 (2-5 days)

#### Validity:

Correlates with QuickDASH ($r = -0.87$)

Discriminates between patients and controls

Responsive to change (Effect size = 1.06)

Validated across upper extremity conditions

No floor or ceiling effects

### Scoring

Each item rated 0-4:

0 = Extreme difficulty or unable to perform

1 = Quite a bit of difficulty

2 = Moderate difficulty

3 = A little bit of difficulty

4 = No difficulty

Total score: 0-80

Can be converted to percentage (score/80 × 100)

### Interpretation

Score 0-40 (0-50%): Severe functional limitation

Score 41-60 (51-75%): Moderate functional limitation

Score 61-80 (76-100%): Mild functional limitation or near-normal function

MCID: 8-14 points

Higher scores indicate better function

### Clinical Applications

Assessing upper extremity function in any condition

Monitoring treatment progress

Documenting functional improvements

Research on upper extremity interventions

Outcome measurement for shoulder, elbow, wrist, hand conditions

### Advantages

Applicable to any upper extremity condition

Excellent psychometric properties

No floor/ceiling effects

Quick to administer

Free and readily available

### Administration

20 items

Time: 3-5 minutes

Recall: Today/current status

Reading level: 6th grade

Self-administered

### Access

Public domain - FREE

Available at www.FOTO.com and in published literature

No permission required

Scoring straightforward

## 10.2 Shoulder Pain and Disability Index (SPADI)

### Description

The SPADI is a 13-item self-administered questionnaire with two subscales: Pain (5 items) and Disability (8 items). It measures the impact of shoulder pathology on pain and function. The SPADI is one of the most commonly used shoulder-specific outcome measures with extensive validation.

### Items

Pain subscale (5 items):

Pain at its worst

Pain when lying on affected side

Reaching for something on high shelf

Touching back of neck

Pushing with affected arm

Disability subscale (8 items):

Washing back

Managing toileting

Combing hair

Reaching high shelf

Carrying 10 lbs

Removing something from back pocket

Washing hair

Usual work

Usual sport/recreation

### Psychometric Properties

### Reliability:

Internal consistency: α = 0.92-0.95 (total); 0.89-0.92 (pain); 0.90-0.93 (disability)

Test-retest: ICC = 0.89-0.96

### Validity:

Correlates with other shoulder measures (r = 0.60-0.87)

Discriminates between shoulder conditions

Responsive to change (SRM = 0.80-1.43)

Validated across shoulder pathologies

Predicts surgical outcomes

## Scoring

Each item rated 0-10:

0 = No pain/No difficulty

10 = Worst pain imaginable/So difficult required help

Pain subscale: Average of 5 items (0-10)

Disability subscale: Average of 8 items (0-10)

Total SPADI: (Pain subscale + Disability subscale) / 2

Can express as percentage: (Total/10) × 100

Score range: 0-100%

Higher scores = worse pain/disability

## Interpretation

Score 0-30%: Mild impairment

Score 31-50%: Moderate impairment

Score 51-70%: Severe impairment

Score >70%: Very severe/complete disability

MCID: 10-13 points (13%)

Subscales can be interpreted separately

## Clinical Applications

Assessment of shoulder pain and disability

Monitoring treatment response

Pre/post-surgical evaluation

Clinical trials for shoulder conditions

Applicable to any shoulder pathology

### Advantages

Shoulder-specific assessment

Excellent psychometric properties

Quick and easy to complete

Validated in multiple shoulder conditions

Free and widely available

### Administration

13 items

Time: 3-5 minutes

Recall: Past week

Reading level: 6th grade

Self-administered

### Access

Public domain - FREE

Available at www.outcometracker.org

No permission required

Multiple language versions

## 10.3 Lower Extremity Functional Scale (LEFS)

### Description

The LEFS is a 20-item questionnaire measuring patients' initial function, ongoing progress, and outcome for a wide range of lower extremity conditions. It assesses difficulty with functional activities ranging from basic mobility to advanced sports activities. The LEFS is one of the most widely used and validated lower extremity outcome measures.

### Items

Usual activities: Walking, running

Mobility tasks: Stairs, squatting, standing

Activities of daily living: Getting into/out of bath, car, bed

Work/recreational activities: Light/heavy activities, hopping

Personal care: Putting on shoes/socks, rolling in bed

## Psychometric Properties

### Reliability:

Internal consistency: α = 0.94-0.96

Test-retest: ICC = 0.94-0.98

### Validity:

Correlates with SF-36 Physical Function (r = 0.80)

Discriminates between functional levels

Highly responsive to change (SRM = 1.13-2.18)

Validated across lower extremity conditions

No floor or ceiling effects in most populations

## Scoring

Each item rated 0-4:

0 = Extreme difficulty or unable

1 = Quite a bit of difficulty

2 = Moderate difficulty

3 = A little bit of difficulty

4 = No difficulty

Total score: 0-80 (sum of all items)

Maximum function = 80

Can be expressed as percentage

## Interpretation

Score 0-40 (0-50%): Severe functional limitation

Score 41-60 (51-75%): Moderate functional limitation

Score 61-80 (76-100%): Mild limitation to near-normal

MCID: 9 points

Substantial improvement: ≥15 points

Higher scores = better function

#### Clinical Applications

Assessment of any lower extremity condition

Monitoring rehabilitation progress

Pre/post-surgical evaluation

Clinical trials for LE interventions

Applicable to hip, knee, ankle, foot conditions

#### Advantages

Applicable to any lower extremity problem

Excellent psychometric properties

Highly responsive to change

Quick and simple

Free and widely available

No floor/ceiling effects

#### Administration

20 items

Time: 3-5 minutes

Recall: Today/current status

Reading level: 6th grade

Self-administered

#### Access

Public domain - FREE

Available at www.FOTO.com

No permission required

Widely available in literature

## 11. Immune Function Assessment

### 11.1 Immune Health Questionnaires

#### Description

Immune health questionnaires assess subjective immune function through patient-reported frequency and severity of infections, allergies, and autoimmune symptoms. Various formats exist including the Immune Status Questionnaire (ISQ) and infection frequency logs. These tools capture immune system performance from the patient's perspective, complementing objective biomarkers.

## Items

Upper respiratory tract infections (frequency and severity)

Lower respiratory infections

Gastrointestinal infections

Skin infections

Allergic symptoms (seasonal allergies, food allergies, eczema)

Wound healing capacity

Vaccination response (subjective)

Chronic inflammatory symptoms

Autoimmune symptom monitoring (condition-specific)

## Psychometric Properties

### Reliability:

Test-retest reliability varies by questionnaire ($r = 0.65$-$0.85$)

Internal consistency: $\alpha = 0.70$-$0.88$

Infection logs show good agreement with medical records

### Validity:

Correlates with biomarkers (moderate associations, $r = 0.30$-$0.50$)

Predicts healthcare utilization

Sensitive to immune-modulating interventions

Face validity for infection tracking

Validated in complementary medicine research

## Scoring

Varies by instrument:

Infection frequency: Count of infections per period (monthly/quarterly)

Infection severity: Rated 1-5 (mild to severe)

Symptom duration: Days per episode

Healthcare utilization: Doctor visits, antibiotics, hospitalizations

Allergic symptom scores: Frequency × severity

Composite immune health score possible for some instruments

### Interpretation

Infection frequency:

0-1 infections/6 months: Normal immune function

2-3 infections/6 months: Borderline

≥4 infections/6 months: Possible immune dysfunction

Symptom severity:

1-2: Mild (minimal impact)

3: Moderate (some functional impact)

4-5: Severe (significant functional impairment)

Compare to age-matched norms when available

### Clinical Applications

Monitoring immune function in chronic illness

Assessing effects of immune-modulating treatments

Complementary medicine research

Lifestyle intervention studies (exercise, nutrition, stress)

Epidemiological surveys of immune health

### Advantages

Non-invasive assessment of immune function

Patient-centered perspective

Longitudinal tracking possible

Complements laboratory measures

Captures functional impact of immune system

### Administration

Variable (typically 10-30 items)

Time: 5-10 minutes

Recall: Past 3-6 months

Reading level: 6th-7th grade

Self-administered

#### Access

Various instruments available

Immune Status Questionnaire (ISQ) - research use

Infection logs - can be created for practice

Some questionnaires proprietary

Simple infection tracking sheets can be free/custom

## 12. Gastrointestinal Assessment

### 12.1 Short Health Scale for GI Symptoms (SHS-GI)

#### Description

The SHS-GI is a 4-item visual analogue scale measuring common GI symptoms and their impact on quality of life. It assesses stomach pain, diarrhea, constipation, and overall well-being. The SHS is brief, simple, and validated for use in inflammatory bowel disease and functional GI disorders.

#### Items

Stomach pain severity

Diarrhea severity

Constipation severity

General well-being (overall health impact)

#### Psychometric Properties

##### Reliability:

Test-retest: r = 0.70-0.80

Internal consistency: α = 0.75

Good stability in clinically stable patients

##### Validity:

Correlates with disease-specific measures (r = 0.60-0.75)

Discriminates between active and remission states

Responsive to clinical change

Validated in IBD and IBS populations

Predicts quality of life

## Scoring

Each item rated on 100mm visual analogue scale

Scored 0-100 for each item

0 = No problem/Best well-being

100 = Worst imaginable/Poorest well-being

Total score = sum of 4 items (0-400)

Or report individual item scores

## Interpretation

**Individual items:**

0-20: Minimal symptoms

21-40: Mild symptoms

41-60: Moderate symptoms

61-100: Severe symptoms

**Total score:**

0-80: Good overall GI health

81-160: Mild GI problems

161-240: Moderate GI problems

>240: Severe GI dysfunction

MCID: 10-15mm per item

## Clinical Applications

Monitoring GI symptoms in IBD and IBS

Treatment response assessment

Clinical trials for GI interventions

Screening for significant GI problems

Longitudinal symptom tracking

### Advantages

Very brief (4 items)

Simple visual analogue format

Covers key GI symptoms

Quick to complete and score

Validated in multiple GI conditions

### Administration

4 items

Time: 1-2 minutes

Recall: Past 2 weeks

Reading level: 5th grade

Self-administered

### Access

Available in published literature

Permission may be required

Contact authors for official version

Multiple language versions exist

## 12.2 IBS Symptom Severity Score (IBS-SSS)

### Description

The IBS-SSS is a 5-item questionnaire measuring the severity of irritable bowel syndrome symptoms. It assesses pain severity/frequency, abdominal distension, bowel habit dissatisfaction, and life interference. The IBS-SSS is the most widely used IBS-specific severity measure and is recommended for clinical trials.

### Items

Abdominal pain severity

Abdominal pain frequency (number of days in pain)

Abdominal distension severity

Dissatisfaction with bowel habits

Interference with life in general

## Psychometric Properties

### Reliability:

Internal consistency: Not reported (single construct)

Test-retest: Stable patients show consistency

Good repeatability demonstrated

### Validity:

Discriminates IBS from healthy controls

Correlates with quality of life measures (r = -0.60 to -0.70)

Responsive to treatment (Effect size = 0.80-1.20)

Predicts treatment satisfaction

Validated in multiple IBS studies

Recommended by regulatory agencies for trials

## Scoring

Each of 5 items scored 0-100 on visual analogue scale

Item 1 (pain severity): 0=no pain, 100=very severe pain

Item 2 (pain frequency): 0=never, 100=all the time (converted from days/10)

Items 3-5: Similar 0-100 scales

Total score = sum of 5 items (range 0-500)

## Interpretation

Score <75: Normal/remission

Score 75-175: Mild IBS

Score 175-300: Moderate IBS

Score >300: Severe IBS

Clinical response: ≥50 point reduction

Remission: Score <75

MCID: 50 points

Individual items identify specific problem areas

### Clinical Applications

Assessing IBS severity in clinical practice

Primary outcome in IBS clinical trials

Monitoring treatment response

Categorizing patients by severity

Research on IBS interventions

### Advantages

IBS-specific assessment

Recommended outcome measure for trials

Clear severity categories

Well-validated responsiveness

Brief and simple

### Administration

5 items

Time: 2-3 minutes

Recall: Past 10 days

Reading level: 6th grade

Self-administered

### Access

Available in published literature

Original publication: Francis et al. 1997

Widely used - generally free for research/clinical use

Contact authors for definitive version

Multiple language versions

## 12.3 GSRS-IBS (Gastrointestinal Symptom Rating Scale - IBS)

### Description

The GSRS-IBS is a 13-item disease-specific instrument for assessing symptoms in IBS patients. It's adapted from the original 15-item GSRS and includes five symptom clusters: pain, bloating, constipation, diarrhea, and satiety. The GSRS-IBS is extensively validated and widely used in both clinical trials and practice.

## Items

Abdominal pain (2 items)

Bloating (2 items)

Constipation (3 items)

Diarrhea (4 items)

Satiety/early fullness (2 items)

## Psychometric Properties

### Reliability:

Internal consistency: α = 0.74-0.94 across subscales

Test-retest: ICC = 0.65-0.86

Good stability over time

**Validity:**

Discriminates IBS from healthy controls

Correlates with IBS severity (r = 0.60-0.75)

Responsive to treatment (Effect size = 0.50-1.10)

Validated in multiple countries and languages

Factor structure confirmed

## Scoring

Each item rated on 7-point Likert scale:

1 = No discomfort

7 = Very severe discomfort

Scoring:

Calculate mean for each subscale (sum of items / number of items)

Total score = mean of all 13 items

Score range: 1-7 (higher = more severe symptoms)

Can report subscale scores individually

## Interpretation

Subscale interpretation (mean score):

1-2: None to minimal symptoms

2-4: Mild symptoms

4-5: Moderate symptoms

5-7: Severe symptoms

Total score:

<3.0: Good symptom control

3.0-4.0: Moderate symptom burden

>4.0: Significant symptom burden

MCID: 0.5-1.0 points

Subscales identify predominant symptom pattern (C-IBS vs D-IBS)

### Clinical Applications

Comprehensive IBS symptom assessment

Clinical trials for IBS treatments

Monitoring symptom patterns over time

Identifying IBS subtype (constipation vs diarrhea predominant)

Treatment response evaluation

### Advantages

Comprehensive symptom coverage

Well-validated across populations

Identifies symptom patterns (useful for subtyping)

Responsive to change

Available in many languages

### Administration

13 items

Time: 3-5 minutes

Recall: Past week

Reading level: 6th-7th grade

Self-administered

### Access

Available from copyright holders

Permission required for use

Contact Mapi Research Trust

www.mapi-trust.org

Licensing fees may apply

## 13. References and Online Resources

### References


1. Bowling A. Measuring Health: A review of scales of well-being. 4th ed. London: Open University Press; 2017.

2. Fayers PM, Machin D. Quality of Life: The assessment, analysis and reporting of Patient-Reported Outcomes. 3rd ed. Chichester: Wiley-Blackwell; 2016.

3. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A practical guide to their development and use. 6th ed. Oxford: Oxford University Press; 2024.

4. Cella D, Choi SW, Condon DM, et al. PROMIS® Adult Health Profiles: Efficient short-form measures of seven health domains. Value Health. 2019;22(5):537-544. doi:10.1016/j.jval.2019.02.004

5. U.S. Department of Health and Human Services FDA. Guidance for industry: Patient-Reported Outcome Measures: Use in medical product development to support labeling claims. Health Qual Life Outcomes. 2006;4:79. doi:10.1186/1477-7525-4-79

6. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-745. doi:10.1016/j.jclinepi.2010.02.006

7. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3

8. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of Patient-Reported Outcomes in clinical trial protocols: The SPIRIT-PRO extension. JAMA. 2018;319(5):483-494. doi:10.1001/jama.2017.21903

9. Lavallee DC, Chenok KE, Love RM, et al. Incorporating Patient-Reported Outcomes into health care to engage patients and enhance care. Health Aff (Millwood). 2016;35(4):575-582. doi:10.1377/hlthaff.2015.1362

10. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res. 2012;21(8):1305-1314. doi:10.1007/s11136-011-0054-x

11. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD; Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health. 2007;10 Suppl 2:S94-S105. doi:10.1111/j.1524-4733.2007.00272.x

12. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167. doi:10.1136/bmj.f167

## ONLINE RESOURCES

- PROMIS Health Organization: www.healthmeasures.net

- COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments): www.cosmin.nl

- Patient-Reported Outcomes Measurement Information System: www.promishealth.org

- PROQOLID (Patient-Reported Outcome and Quality of Life Instruments Database): www.proqolid.org

- ePROVIDE (Searchable database of patient-reported outcome measures):

  eprovide.mapi-trust.org

- ICHOM (International Consortium for Health Outcomes Measurement): www.ichom.org

- ISOQOL (International Society for Quality of Life Research): www.isoqol.org