

# Outcome Measures

January 2019



**THE BRITISH  
PAIN SOCIETY**  
EXPERTISE WHERE IT MATTERS



**FACULTY OF  
PAIN MEDICINE**  
of the Royal College of Anaesthetists

# Table of Contents

<b>1</b>	<b>Executive Summary</b>	<b>3</b>
<b>2</b>	<b>Introduction</b>	<b>4</b>
<b>2.1</b>	<b>Background</b>	<b>4</b>
<b>3</b>	<b>Methodology</b>	<b>5</b>
<b>4</b>	<b>Objectives</b>	<b>6</b>
<b>5</b>	<b>Pain Scales</b>	
5.1	<i>Numerical Pain Rating Scale (NPRS)</i>	<b>7</b>
5.2	<i>Visual Analogue Scale</i>	<b>9</b>
5.3	<i>Verbal Rating Scale (VRS)</i>	<b>11</b>
<b>6</b>	<b>Pain Interference</b>	
6.1	<i>Roland &amp; Morris Disability Index</i>	<b>13</b>
6.2	<i>Oswestry Low Back Pain Disability Questionnaire</i>	<b>15</b>
<b>7</b>	<b>Physical Functioning</b>	
7.1	<i>Brief Pain Inventory</i>	<b>17</b>
<b>8</b>	<b>Emotional Distress And Functioning</b>	
8.1	<i>Beck Depression Inventory (BDI-II)</i>	<b>19</b>
8.2	<i>Centre For Epidemiologic Studies-Depression Scale</i>	<b>21</b>
8.3	<i>Hospital Anxiety And Depression Scale (HADS)</i>	<b>23</b>
8.4	<i>Short Form MOS-36 (SF-36)</i>	<b>25</b>
8.5	<i>Profile Of Mood States</i>	<b>26</b>
8.6	<i>Pain Catastrophising Scale</i>	<b>28</b>
8.7	<i>The Patient Health Questionnaire – 2</i>	<b>30</b>
8.8	<i>The Pain Self-Efficacy Questionnaire (PSEQ)</i>	<b>31</b>
<b>9</b>	<b>Quality Of Life</b>	
9.1	<i>Euroqol 5D</i>	<b>32</b>
<b>10</b>	<b>Patient Reported Outcomes</b>	
10.1	<i>Patient Global Impression Of Change (PGIC)</i>	<b>34</b>
<b>11</b>	<b>Table</b>	<b>35</b>
<b>12</b>	<b>The Working Party</b>	<b>38</b>

# 1 Executive summary

This document is a combined project between the Faculty of Pain Medicine (FPM) and the British Pain Society (BPS). Health service outcomes can be used to benchmark services against one another and against targets, and to improve patient care; NHS England is currently considering various Patient Reported Outcome Measures (PROMs).

This document describes outcome scales appropriate to pain management. It does not cover diagnostic and screening tools, nor direct measurement of physical performance, only scales that are completed by patients themselves before and after treatment. The scales are designed for use by cognitively intact adults over 16; other scales and methods are suitable for children and for people of any age with cognitive impairment. The document is necessarily a brief practical guide to the commonly used scales in the UK, not a comprehensive account of the qualities and performance of all possible scales.

Chronic pain has multiple effects on patients, so outcome measures cover several domains:

- Pain Quantity
- Pain Interference
- Physical Functioning
- Emotional Functioning
- Quality of Life
- Patient reported global rating

The Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT: [www.immpact.org](http://www.immpact.org)) outputs are a valuable resource, and several of its recommendations of scales are included here. For each scale, a brief description is given with main advantages and disadvantages, and information on use and copyright.

## 2 Introduction

This document is a joint project of the Faculty of Pain Medicine (FPM) of the Royal College of Anaesthetists and the British Pain Society (BPS) to provide guidance on the various available outcome measures used by pain services. This project has been supported by the national clinical reference group in pain.

### 2.1 Background

The National Health Service (NHS) is moving towards outcome-based commissioning which encourages value for money and better outcomes for patients. NHS pain services are currently commissioned by different clinical commissioning groups (CCGs) based on their local infrastructure and requirements. NHS performance indicators, such as 18 weeks wait, are used to assess quality of services. Outcome measures are not normally required to show service effectiveness, but this is currently changing and some CCGs are requesting outcomes to inform commissioning.

Pain is defined as an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage. NHS pain services follow a biopsychosocial approach to manage pain. Chronic pain has a substantial effect on mood, quality of life and relationships. It can affect the ability to work, leading to financial implications for the patient and the wider economy.

This document was prepared to guide pain services across the country in selection of the most appropriate outcome measures for their needs. No single scale can meet all needs: the choice of outcome/s for a service depends on the treatments offered, the aims of treatment, and on the population treated.

### 3 Methodology

The Faculty of Pain Medicine and The British Pain Society appointed the working group. The primary aim was to provide guidance on the various available outcome measures used by pain services. By having a degree of uniformity to such measures, we will be in a better position to influence the service level agreements both at a local and a national level.

The group was tasked to review already available and free to use measures, easily used in the clinical and telephone follow-up environment. It should provide well-established and validated outcome measures that cover the domains of pain improvement, functional improvement, psychological improvement, and overall satisfaction applicable in the secondary and tertiary care setting. It is likely that a few measures will be needed for different aspects of services and it is accepted that none will be perfect. New original research, and systematic/meta-analysis reviews were outside the remit of this group.

The working group communicated via email and teleconference due to geographical spread and to promote efficiency. Most of the validated questionnaires analysed were from the “Core Outcome Measures for Chronic Pain Clinical Trials: IMMPACT Recommendations”. Additional outcome measures were added if the group agreed that they were commonly used in pain clinics, and were well validated with evidence in the use of chronic pain. It should be noted that the analysis of all available outcome measures was not done, due to the clear objectives assigned to the working group, and the time frame involved. It is envisaged that this document will continually evolve, based on feedback, experience of its use, and the availability of further evidence.

## 4 Objectives

This document provides a shortlist of scales used to assess outcomes of pain management, with reasonable psychometric properties and most without undue restrictions on use. Some are copyright protected, or require payment, particularly if for commercial (profit-making) rather than research use.

**Reliability:** The quality of being trustworthy or of performing consistently well. The degree to which the result of a measurement, calculation, or specification can be depended on to be accurate. A measure has high reliability if it produces similar results under consistent conditions. If one person takes the same personality test several times and always receives the same results, the test is reliable.

**Validity:** In its purest sense, this refers to how well a scientific test or piece of research actually measures what it sets out to, or how well it reflects the reality it claims to represent. If the results of the personality test claimed that a very shy person was in fact outgoing, the test would be invalid.

Reliability and validity are independent of each other. A measurement may be valid but not reliable, or reliable but not valid. Suppose a bathroom scale was reset to read 10kgs lighter. The weight it reads will be reliable (the same every time you step on it) but will not be valid, since it is not reading your actual weight.

**Sensitivity to change:** It must be able to detect small, but clinically significant changes in a phenomenon over time. This may be done by comparing longitudinal measures of the same outcome i.e., at baseline and discharge or a series of measures over time.

If departments are looking for one scale, EuroQol (EQ5D-5L) is currently being adopted by the NHS England as a key outcome measure in other areas such as knee and hip surgeries. We hope to have a NHS license to use this in the near future. Until then, a request to the EuroQol team obtains a temporary user agreement.

## 5 Pain Scales

### 5.1 Numerical Pain Rating Scale (NPRS)

The Numerical Rating Scale (NPRS-11) is an 11-point scale for self-report of pain. It is the most commonly used unidimensional pain scale. The respondent selects a whole number (integers 0–10) that best reflects the intensity (or other quality if requested) of his/her pain. The anchors are 0 = no pain and 10 = extreme pain/worst possible pain (there are various different wordings of the upper anchor). It is often categorised into: no pain = 0, mild pain = 1-3, moderate pain = 4-6, severe pain = 7-10, but these categories do not necessarily reflect patient meanings, and are poor for any assessment of change. The categories might be used to set targets for intervention outcome. The NPRS can be administered verbally (therefore also by telephone) or graphically for self-completion.

Copyright: None

#### Validity and reliability

For construct validity, the NPRS was shown to be highly correlated with the visual analogue scale (VAS) in which pain is shown spatially as distance along a straight line, in patients with rheumatic and other chronic pain conditions (pain > 6 months): correlations range from 0.86 to 0.95.

High test-retest reliability has been observed in both literate and illiterate patients with rheumatoid arthritis ( $r = 0.96$  and  $0.95$ , respectively) before and after medical consultation.

#### Conditions used

NPRS is used to measure pain in adults and children 10 years old or older.

#### Pros and Cons

Pros:

- The NPRS takes <1 minute to complete
- The NPRS is easy to administer and score
- Minimal language translation requirements support the use of the NPRS across languages
- It is more reliable than the VAS, particularly for older patients or those with poor literacy. It has more points (and therefore is better for expressing change) than a verbal rating scale
- It does not require any further procedure for scoring
- The NPRS can be used for a single dimension of pain, commonly pain intensity, or for other dimensions such as pain distress, pain interference, etc. The question that leads into the scale changes, as do the scale anchors, but the scale itself is the same

Cons:

- The instruction with which the NPRS is given is most suitable for unidimensional constructs, so it often evaluates only one of multiple dimensions of pain

Jensen, M.P. and Karoly, P. (2011). "Self-report scales and procedures for assessing pain in adults", in Turk, D.C. and Melzack, R. (eds). Handbook of pain assessment, 3rd edition. New York, Guilford Press, 19-44.

Ferraz MB, Quesada MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. "Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis". J Rheumatol. 1990 Aug;17(8):1022-4.

Childs JD, Piva SR, Fritz JM. "Responsiveness of the Numeric Pain Rating Scale in Patients with Low Back Pain". Spine. 2005 Jun 1;30(11):1331-4.

Williamson A, Hoggart B. "Pain: a review of three commonly used pain rating scales", Journal of Clinical Nursing, Wiley/Blackwell (10.1111); 2005 Aug 1;14(7):798-804.



## 5.2 Visual Analogue Scale

A Visual Analogue Scale (VAS) is an instrument for subjective rating of pain. The pain VAS is most often used as a unidimensional measure of pain intensity.

The VAS is a straight horizontal line of fixed length, usually 100 mm. The ends are defined as the extreme limits of the parameter to be measured, hence the left end is usually labeled 'no pain', and the right end 'extreme pain' or (as with the NPRS) some other verbal anchor. Scales can also be used vertically but are no more reliable.

Using a ruler, the score is determined by measuring the distance (mm) on the 10-cm line between the "no pain" anchor and the patient's mark, providing a range of scores from 0–100. As with the NPRS, categories may be imposed on this (no pain 0–4 mm; mild pain 5–44 mm; moderate pain 45–74 mm; severe pain 75–100 mm) but this is arbitrary and does not necessarily reflect patients' meanings.

Copyright: None

### Validity and reliability

Criterion validity cannot be evaluated for VAS in the absence of a gold standard. For construct validity, in patients with a variety of rheumatic diseases, the pain VAS has been shown to be highly correlated with a 5-point verbal descriptive scale ("nil," "mild," "moderate," "severe," and "very severe") and a numeric rating scale (with response options from "no pain" to "unbearable pain"), with correlations ranging from 0.71–0.78 and 0.62–0.91, respectively). The correlation between vertical and horizontal orientations of the VAS is 0.99. Test–retest reliability has been shown to be good, but higher among literate ( $r = 0.94$ ,  $P = 0.001$ ) than illiterate patients ( $r = 0.71$ ,  $P = 0.001$ ) before and after attending a rheumatology outpatient clinic.

### Conditions used

VAS can be used for various domains including pain in adults. The same scale can be modified to measure other variables such as mood, sleep, and other functions.

### Pros and Cons

Pros:

- VAS is more sensitive to small changes than those with fewer categories
- The VAS takes < 1 minute to complete
- No training is required other than the ability to use a ruler to measure distance to determine a score
- Minimal translation requirements have led to an unknown number of cross-cultural adaptations

Cons:

- The VAS requires paper and pencil and good visual acuity
- People with little education, particularly in the older population, find this a difficult scale to use and tend to write on the line
- Caution is required when photocopying the scale as this may change the length of the 10-cm line
- The instruction with which the VAS is given is most suitable for unidimensional constructs, so it often evaluates only one of multiple dimensions of pain

D. Gould et al. "Visual Analogue Scale (VAS)", *Journal of Clinical Nursing* 2001; 10:697-706

Jensen, M.P. and Karoly, P. 2011 "Self-report scales and procedures for assessing pain in adults" in Turk, D.C. and Melzack, R. (eds) *Handbook of pain assessment*, 3rd edition. New York, Guilford Press, 19-44.

Joos E, Peretz A, Beguin S, Famaey JP. "Reliability and reproducibility of visual analogue scale and numeric rating scale for therapeutic evaluation of pain in rheumatic patients", *J Rheumatol.* 1991 Aug;18(8):1269-70.

Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. "Studies with pain rating scales", *Annals of the Rheumatic Diseases*, BMJ Publishing Group; 1978 Aug;37(4):378-81.

Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. "Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis", *J Rheumatol.* 1990 Aug;17(8):1022-4.

### 5.3 Verbal Rating Scale (VRS)

The Verbal Rating Scale consists of a list of adjectives describing different levels of pain intensity. Patients are asked to select the adjective that best represents their pain. This should reflect the extremes of this dimension; from 'no pain' to 'extremely intense pain' and sufficient intervening adjectives to capture gradations of pain intensity that may be experienced between extremes.

no pain = 0

mild pain = 1

moderate pain = 2

severe pain = 3

very severe pain = 4

VRSs are scored as above but these are ranks, not equal intervals.

Copyright: None

#### Validity and Reliability

The VRS is easy to administer on paper or verbally, and to understand. It is often preferred by patients over the NPRS and VAS. Compliance is excellent due to its simplicity. Scales are not very sensitive to change because there are few data points.

#### Conditions used

VRS is used to measure pain in adults and children more than 10 years old.

#### Pros and Cons

Pros:

- Easy to administer and comprehend
- Good compliance

Cons:

- VRS assumes equal intervals between the adjectives. This might not be accepted as the interval between no pain and mild pain may be much smaller than that between moderate pain and severe pain, yet the interval is scored as if the difference were equivalent

Other issues:

- Patients may not find a descriptor that accurately describes their perceived pain intensity
- In patients who are illiterate, they are less reliable than other pain intensity measures

- It is easier to find terms for pain intensity for a VRS than for other dimensions of pain such as pain distress or pain interference, so it is usually only an indicator of pain intensity.

Frank, A.J., Moll, J.M., Hort, J.F., (1982) "A comparison of three ways of measuring pain", *Rheumatol Rehabil*, *Rheumatol Rehabil* 21, 211-7.

Jensen, M.P. and Karoly, P. (2011) "Self-report scales and procedures for assessing pain in adults" in Turk, D.C. and Melzack, R. (eds). *Handbook of pain assessment*, 3rd edition. New York, Guilford Press, 19-44.

Ohnhaus, E.E., Adler, R., (1975) "Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale", *Pain*, 1, 379-84.

## 6 Pain Interference

### 6.1 Roland & Morris Disability Index

The Roland & Morris Disability Questionnaire (RMDQ) is a widely-used disability scale for low back pain. The RMDQ can be used in research or clinical practice. The RMDQ contains 24 sentences that describe different movements or functions for which respondents tick those that apply to them that day. The total score is the number of ticks/24.

Copyright: None

#### Validity and reliability

The RMDQ has been translated in 36 different languages/versions. Almost all validation studies (including non-English) were conducted in low back pain of various origins. The RMDQ has good validity and reliability with a range of internal consistency between 0.83 and 0.95 (Cronbach  $\alpha$ ) (Stratford et al, 1996; Spanjer et al. 2011; Mousavi et al. 2006), and a range of intraclass correlation coefficients between 0.83 and 0.93 (test-retest reliability) depending on the intervals between test and retest (Stratford et al, 1996; Brouwer et al, 2004; Jordan et al, 2006; Davidson et al, 2002).

#### Conditions used

RMDQ is a scale for low back pain only.

#### Pros and cons

Pros:

- Free to use
- Easy to administer but requires paper and pen/screen
- Good validity and reliability

Cons:

- Test-retest reliability poorer over longer intervals
- Only for low back pain.

- Roland MO, Morris RW. "A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low back pain", *Spine* 1983; 8: 141-144.
- Stratford, P. W., Binkley, J., et al. "Defining the minimum level of detectable change for the Roland-Morris questionnaire." *Phys Ther* 1996, vol.76(4): 359-365; discussion 366-358.
- Mousavi, S. J., Parnianpour, M., et al. "The Oswestry Disability Index, the Roland-Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: translation and validation studies of the Iranian versions." *Spine (Phila Pa 1976)* 2006, vol.31(14): E454-459.
- Davidson, M. and Keating, J. L. "A comparison of five low back disability questionnaires: reliability and responsiveness." *Phys Ther* 2002, vol.82(1): 8-24.
- Jordan, K., Dunn, K. M., et al. "A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain." *J Clin Epidemiol* 2006, vol.59(1): 45-52.
- Spanjer J, Groothoff JW, Brouwer S. "Instruments used to assess functional limitations in workers applying for disability benefit: a systematic review", *Disabil Rehabil.* 2011; 33(23-24):2143-50.
- Brouwer, S., Kuijer, W., Dijkstra, P. U., Goeken, L. N., Groothoff, J. W. and Geertzen, J. H. "Reliability and stability of the Roland Morris Disability Questionnaire: intraclass correlation and limits of agreement", *Disabil Rehabil* 2004; 26(3): 162-5.

## 6.2 Oswestry Low Back Pain Disability Questionnaire

The Oswestry Low Back Pain Disability Questionnaire is a self-report scale for low back pain functional disability. It contains 10 topics: intensity of pain, lifting, self-care, ability to walk, ability to sit, sexual function, ability to stand, social life, sleep quality, and ability to travel. Each topic is followed by 6 statements of different levels of difficulty with the area described, and the patient checks the statement that most closely resembles his or her situation. Each question is scored on a scale of 0–5, where no difficulty = 0 and maximum difficulty = 5. The total of these is multiplied by 2 to give a score from 0 to 100.

Categories have been imposed on these scores:

- 0 to 20: Minimal disability
- 21–40: Moderate Disability
- 41–60: Severe Disability
- 61–80: Crippling back pain
- 81–100: These patients are either bed-bound or have an exaggeration of their symptoms

Copyright: The Oswestry Low Back Pain Disability Questionnaire is copyrighted. It has been translated and modified into various languages and versions. For non-commercial use, there is no fee applicable. Original 2.1.a ODI must be used.

<https://eprovide.mapi-trust.org/instruments/oswestry-disability-index>

### Validity and reliability

The Oswestry Low Back Pain Disability Questionnaire has been validated for use in a wide range of languages. Internal consistency has been shown to be of an acceptable level by different authors. Cronbach  $\alpha$  ranges from 0.71 to 0.90 (Fairbank et al. 2000; Roland et al. 2000; Grotle et al. 2012; Miekisiak et al. 2013). Test-retest reliability has been shown to be high. Intraclass correlation coefficient (ICC) values range = 0.83 to 0.99 and vary according to the time interval between measurements (Davidson et al. 2002; Fairbank et al. 2000; Roland et al. 2000; Grotle et al. 2012; Miekisiak et al. 2013). The longer the interval between assessments, the lower the score.

### Conditions used

It is a condition-specific outcome measure for patients with low back pain, and was developed for use in secondary care settings.

### Pros and cons

Pros:

- Free for non-commercial use
- Considered as 'gold standard' for low back pain
- Well-validated with good validity and reliability

## Cons:

- Time-consuming to administer (10 topics/sections) and to score
- Not applicable beyond low back pain
- 10 areas covered may not be those of most importance to some patients

Fairbank JCT & Pynsent, PB (2000) "The Oswestry Disability Index", *Spine*, 25(22):2940-2953.

Davidson M & Keating J (2001) "A comparison of five low back disability questionnaires: reliability and responsiveness", *Physical Therapy* 2002;82:8-24.

Roland M., Fairbank J. "The Roland-Morris Disability Questionnaire, and the Oswestry Disability Questionnaire", *Spine*, 2000;25(24):3115–3124.

Fritz J.M., Irrgang J.J. "A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale", *Phys Ther*. 2001;81(2):776–788.

Grotle, M., Garratt, A. M., et al. (2012). "Reliability and construct validity of self-report questionnaires for patients with pelvic girdle pain." *Phys Ther* 92(1): 111-123.

Miekisiak, G., Kollataj, M., et al. (2013). "Validation and cross-cultural adaptation of the Polish version of the Oswestry Disability Index." *Spine* (Phila Pa 1976) 38(4): E237-243.



## 7 Physical Functioning

### 7.1 Brief Pain Inventory

The purpose of the Brief Pain Inventory (BPI) is to assess the severity of pain and the impact of pain on daily functions. It can be used in patients with pain related to a chronic disease condition (such as cancer or low back pain) or pain arising from acute conditions such as postoperative pain. The BPI measures pain severity, the impact of pain on daily function, the location of pain, pain medications and amount of pain relief in the past 24 hours of past week.

There is no scoring algorithm for the Brief Pain Inventory, but “worst pain” or the arithmetic mean of the four severity items can be used as measures of pain severity; the arithmetic mean of the seven interference items can be used as a measure of pain interference.

Copyright: The University of Texas M.D. Anderson Cancer Center holds the copyright and permission to use the tool can be sought by filling in an online form -

<https://www4.mdanderson.org/symptomresearch/index.cfm>

Permission is routinely granted at no cost but no amendments are permitted. The website also gives contact details to use if the form is unsuitable.

#### **Validity and reliability**

The Brief Pain Inventory was originally developed to assess cancer pain but has been shown to have utility in chronic pain (Tan, Jensen, Thornby, and Shanti 2004), with acceptable internal consistency (Cronbach  $\alpha$  0.85 intensity and 0.88 for interference) a stable 2-factor structure, and sensitivity to change with treatment. Similar qualities have been demonstrated in patients with arthritis or low back pain (Keller et al. 2004), in low back pain (Song et al. 2016); and musculoskeletal pain (Celik et al., 2016). It also functions well in older adults (mean age 77) with chronic musculoskeletal pain, discriminating those at risk of falls (Stubbs, Eggermont, Patchay, & Schofield, 2015). Although there are reports of the interference factor having subcomponents (Cleeland & Ryan, 1994), this is not sufficiently well established to recommend a change in scoring (Dworkin et al., 2005).

#### **Conditions Used**

The Brief Pain Inventory can be used in both cancer-related pain and chronic pain.

## Pros and Cons

### Pros:

- Can be used in both cancer-related pain and chronic pain
- Psychometrically and linguistically validated:
  - Arabic, Cebuano, Chinese (Simplified), Chinese (Traditional), Croatian, Czech, English, Filipino, French, German, Greek, Hebrew, Hindi, Italian, Japanese, Korean, Malay, Norwegian, Russian, Slovak, Slovenian, Spanish, Spanish (Spain) and Thai.
- Linguistically Validated:
  - Afrikaans, Bengali, Bulgarian, Danish, Dutch, Estonian, Finnish, Georgian, Gujarati, Hungarian, Kannada, Latvian, Lithuanian, Malayalam, Marathi, Polish, Portuguese (Brazil), Portuguese (Portugal), Romanian, Serbian, Swedish, Tamil, Telugu, Turkish, Ukrainian, Urdu, Vietnamese, Xhosa and Zulu.
- Can identify populations at risk of falls

### Cons:

- Possible problem with the factor structure

Celik, E. C., Yalcinkaya, E. Y., Atamaz, F., Karatas, M., Ones, K., Sezer, T., et al. (2016). "Validity and reliability of a Turkish brief pain inventory short form when used to evaluate musculoskeletal pain" *J Back Musculoskelet Rehabil.*

Cleeland, C. S., & Ryan, K. M. (1994) "Pain assessment: global use of the Brief Pain Inventory", *Ann Acad Med Singapore*, 23(2), 129-138.

Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., & Katz, N. P. (2005). "Core outcome measures for chronic pain clinical trials: IMMPACT recommendations", *Pain*, 113.

Keller, S., Bann, C. M., Dodd, S. L., Schein, J., Mendoza, T. R., & Cleeland, C. S. (2004). "Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain", *Clin J Pain*, 20(5), 309-318.

Song, C. Y., Lin, S. F., Huang, C. Y., Wu, H. C., Chen, C. H., & Hsieh, C. L. (2016) "Validation of the Brief Pain Inventory in Patients With Low Back Pain", *Spine (Phila Pa 1976)*, 41(15), E937-942.

Stubbs, B., Eggermont, L., Patchay, S., & Schofield, P. (2015) "Older adults with chronic musculoskeletal pain are at increased risk of recurrent falls and the brief pain inventory could help identify those most at risk" *Geriatrics & Gerontology International*, 15(7), 881-888.

Tan, G., Jensen, M. P., Thornby, J. I., & Shanti, B. F. (2004) "Validation of the brief pain inventory for chronic non-malignant pain", *The Journal of Pain*, 5(2), 133-137.

## 8 Emotional Distress and Functioning

### 8.1 Beck Depression Inventory (BDI-II)

BDI is an assessment tool for assessing the severity of depression. It has been widely used in different medical conditions over the past 35 years. The latest version of BDI is BDI-II. BDI-II consisted of 2 domains (cognitive-affective and somatic [physical] domains) with a total of 21 items. Each item has a scoring system of 0 to 3, with the cut-offs for the total scores as below:

- Minimal depression: 0 to 13
- Mild depression: 14 to 19
- Moderate depression: 20 to 28
- Severe depression: 29 to 63

BDI-II can be self-administered and it usually applies to how participants feel in ‘the past 2 weeks, including today’. There are various ‘modified’ versions of the BDI, including those adjusted for confounded somatic symptoms. However, the international IMMPACT recommendations (Dworkin et al. 2005) stated that in chronic pain research, the original full version should be used without adjustment for presumed confounding by somatic symptoms.

Copyright: The BDI is copyrighted, and the BDI website stated a fee must be paid for each copy used and photocopying it is a violation of copyright.

<https://www.pearsonclinical.com/psychology/products/100000159/beck-depression-inventoryii-bdi-ii.html>

#### **Validity and reliability**

The BDI has been extensively studied and validated in mental health settings. The BDI-II is positively correlated with the Hamilton Depression Rating Scale with a Pearson  $r$  of 0.71 (agreement), and it was also shown to have a high one-week test-retest reliability (Pearson  $r = 0.93$ ), suggesting that it was not overly sensitive to daily variations in mood (Beck et al. 1996). Other study also suggested the test has high internal consistency (Cronbach  $\alpha = 0.91$ ) (Beck et al. 1996).

The original BDI has an unstable structure when used in patients with physical health problems, including pain, and the scores cannot be straightforwardly interpreted. A more recent systematic review on the psychometric properties of the BDI-II as a self-report measure of depression in a variety of settings and populations (Wang et al. 2013) has suggested an internal consistency of around (Cronbach  $\alpha$ ) 0.90 and the re-test reliability ranged from 0.73 to 0.96 (Pearson  $r$ ).

## Conditions used

BDI and BDI-II have been widely used in psychiatry, psychology and many chronic conditions including chronic pain conditions. A more recent systematic review on the utility of BDI-II (Wang et al. 2013) has found various validation studies of BDI-II in both primary care and hospital settings, on cardiology, neurology, obstetrics, brain injury, nephrology, chronic pain, chronic fatigue, oncology, and infectious diseases.

## Pros and cons

### Pros:

- Well-validated and widely accepted scale
- Used widely in many chronic conditions including chronic pain
- Easy to use
- Could fit onto a single page of A4

### Cons:

- Total score in pain patients inflated by somatic items
- Difficult to distinguish non-somatic and somatic aspects of emotional functioning

Beck AT, Steer RA and Brown GK (1996) *Manual for the Beck Depression Inventory-II*, San Antonio, TX: Psychological Corporation.

Beck AT, Steer RA, Ball R, Ranieri W (1996) "Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients", *Journal of Personality Assessment*. 67(3): 588-97.

Wang YP and Gorenstein C (2013) "Psychometric properties of the Beck Depression Inventory-II: a comprehensive review", *Rev. Bras. Psiquiatr.* vol.35 no.4 São Paulo Oct-Dec. 2013. <http://dx.doi.org/10.1590/1516-4446-2012-1048>

Wang YP and Gorenstein C (2013) "Assessment of depression in medical patients: A systematic review of the utility of the Beck Depression Inventory-II" *Clinics (Sao Paulo)*. 2013 Sep; 68(9): 1274–1287. doi: 10.6061/clinics/2013(09)15. PMID: PMC3782729

Williams ACdeC, Richardson PH. "What does the BDI measure in chronic pain?", *Pain*, 1993;55:259-66.

Dworkin et al. (2005) "Core outcome measures for chronic pain clinical trials: IMMPACT recommendations", *Pain* 113 (2005) 9–19.

## 8.2 Centre for Epidemiologic Studies-Depression Scale

The CESD is based on the DSM version of depression and the respondent is asked how often in the last week s/he has felt in the ways described by the 20 items, with the response options 'rarely or none of the time (less than 1 day)', 'some or a little of the time (1-2 days)', 'occasionally or a moderate amount of time (3-4 days)', 'most or all of the time (5-7 days)', scored 0,1,2,3 respectively, and scores reversed for positive items. The revised version, consistent with DSM-5, also has 20 items. It is completed by the patient and is designed as a screening tool, not a diagnostic scale. 16/60 is considered the cutoff for risk of depression but Turk & Okifuji (1994) recommend using 19/60 for people with chronic pain.

It is designed for and standardised on people without serious medical problems, so contains somatic items on the assumption that they add to cognitive and affective symptoms to an overall depression score. This is probably a false assumption, as with other depression scales similarly constructed. It is not one of the scales recommended by IMMPACT (Dworkin et al. 2005).

Copyright: The CESD is available at [www.chcr.brown.edu/pcoc/cesdscale.pdf](http://www.chcr.brown.edu/pcoc/cesdscale.pdf) and a revised version (CESD-R) is available from <http://cesd-r.com/> Both are in the public domain. There is a 10-item version but it seems not to be used much.

### Validity and reliability

Internal consistency estimates from a small number of studies, in the community and psychiatric populations, ranging from 0.88-0.91, and test-retest is 0.45-0.87.

Validation is difficult, as with any subjective scale, but it correlated 0.25 with the pain subscale in the SF-36, and 0.75 with the mental health component score. There are mixed findings on whether in pain patients, somatic items inflate the overall score (Blalock et al 1989; Geisser et al 1997), but removing the somatic items is not recommended.

### Conditions used

Community and psychiatric populations, and populations with disabling medical problems.

### Pros and Cons

Pros:

- good reliability
- consistency with recognised (DSM) formulation of depression

Cons:

- risk of inflation of scores by high ratings on somatic items
- designed for screening, not outcome

Blalock SJ, DeVellis RF, Brown GK, Wallston KA. "Validity of the Center for Epidemiological Studies Depression Scale in arthritis populations", *Arthr Rheum* 1989;32(8):991-7.

Dworkin RH, Turk DC, Farrar JT, et al. "Core outcome measures for chronic pain clinical trials: IMMPACT recommendations", *Pain* 2005;113:9-19.

Geisser ME, Roth RS, Robinson ME. "Assessing depression among persons with chronic pain using the Center for Epidemiological Studies Depression scale and the Beck Depression Inventory: A comparative analysis", *Clin J Pain* 1997;13:163-170.

Turk DC, Okifuji A. "Detecting depression in chronic pain patients: adequacy of self-reports", *Behav Res Ther* 1994;32(1):9-16.

<https://scireproject.com/outcome-measures/outcome-measure-tool/center-for-epidemiological-studies-depression-scale-ces-d-and-ces-d-10/#1467983894080-2c29ca8d-88af>

<http://www.bmedreport.com/archives/7139>

### 8.3 Hospital Anxiety and Depression Scale (HADS)

The Hospital Anxiety and Depression Scale (HADS: Zigmond and Snaith, 1983) was designed and standardised on a population with medical problems. The HADS is properly a scale of anhedonia rather than of anxiety or depression, and recent examination of its psychometric properties recommended on strong grounds that it be treated as a single scale of distress (Cosco et al., 2012).

HADS is a 14 item scale that generates ordinal data. Seven of the items relate to anxiety and the other seven to depression. This was developed as a tool for detection of anxiety and depression in people with physical health problems. Each item on the questionnaire is scored from 0-3 and this means that a person can score between 0 and 21 for either anxiety or depression.

normal (0–7)

mild (8–10)

moderate (11–14)

severe (15–21)

Bjelland et al (2002) through a systematic review of a large number of studies identified a cut-off point of 8/21 for anxiety or depression. For anxiety (HADS-A) this gave a specificity of 0.78 and a sensitivity of 0.9. For depression (HADS-D) this gave a specificity of 0.79 and a sensitivity of 0.83.

Copyright:

The use of the questionnaire is licensed by GL Assessment ([www.gl-assessment.co.uk](http://www.gl-assessment.co.uk))

#### Validity and Reliability

Studies show good test-retest reliability on the HADS at 0-2 weeks ( $r=0.84$ ), >2-6 weeks ( $r=0.73$ ), and >6 weeks ( $r=0.70$ ) for the anxiety subscale. Results from the depression subscale were 0-2 weeks ( $r=0.73$ ), >2-6 weeks ( $r=0.73$ ), and >6 weeks ( $r=0.73$ ), indicating the HADS was stable enough to withstand situation influences for varying etiologies. Pearson product moment correlation was found to be 0.92 and 0.90 between the HADS total score and the HADS anxiety score and the HADS depression score. Kappa scores indicated there was no significant difference between the General Health Questionnaire-28 and HADS (total score) (kappa statistic =0.074, SE=0.089,  $p=0.04$ ). Good internal consistency was found (0.80 for the anxiety subscale scale and 0.81 for the depression subscale) during initial testing (Zigmond & Snaith 1983).

The correlation between the HADS depression subscale and the Beck Depression Inventory Primary Care has been found to be 0.62,  $p<0.001$ . Several studies have found that the HADS total score shows a higher correlation with depression and anxiety criterion measures.

## Conditions Used

Facilitates the early identification of both anxiety and depression simultaneously, whilst giving a separate score for each – helping aid the referral to appropriate services.

## Pros and Cons

### Pros:

- Easy to score and simple to interpret
- It is ideal for use as a screening measure, and index of clinical change, an outcome measure and for research purposes
- Highly cost effective – as a single measure of anxiety and depression it removes the need for separate measures and is suitable for all ages from 17+
- HADS is available in 115 languages and therefore suitable for researchers internationally

### Cons:

- Measures symptoms in past week only
- Not intended for severe mood disorders

Zigmond, A.S. and Snaith, R.P. (1983) "The Hospital Anxiety and Depression Scale", *Acta Psychiatrica Scandinavica*, 67, 361-370.

Cosco, T.D., Doyle, F., Ward, M., and McGee, H. (2012). "Latent structure of the Hospital Anxiety and Depression Scale: a 10-year systematic review", *Journal of Psychosomatic Research*, 72, 180-4.

Bjelland, I; et al. (2002). "The validity of the Hospital Anxiety and Depression Scale. An updated literature review". *Journal of Psychosomatic Research*. 52 (2): 69–77. doi:10.1016/s0022-3999(01)00296-3. PMID 11832252

Snaith RP. *The Hospital Anxiety And Depression Scale. Health and Quality of Life Outcomes*. 2003 Aug, 1:29 (Full text article)

White D, Leach C, Sims R, Atkinson M, Cottrell D. "Validation of the Hospital Anxiety and Depression Scale for use with adolescents", *Br J Psychiatry*. 1999 Nov;175:452-4

Herrmann C. "International experiences with the Hospital Anxiety and Depression Scale- a review of validation data and clinical results" *Journal of Psychosomatic Research* 1997;42(1):17-41



## 8.4 Short Form MOS-36 (SF-36)

The short form Medical Outcomes Scale has 36 questions in 8 subscales that are combined in two-component scores. The Mental Component Summary consists of the subscales Vitality (energy/fatigue), Social functioning, Role-emotional (role limitations due to emotional problems), and Mental health. (The Physical Component Summary consists of Physical functioning, Role-physical, Bodily pain, and General health). The overall scale is recommended by IMMPACT for the quality of life, and because it is so widely used across community and clinical populations, age-sex norms are available for healthy populations and comparison scores from other populations. The scale is self-completion.

Copyright: The SF-36 is available free from [https://www.rand.org/health/surveys\\_tools/mos/36-item-short-form.html](https://www.rand.org/health/surveys_tools/mos/36-item-short-form.html) but scoring the subscales requires information (available on the website above, and elsewhere) about which items contribute to which subscale and scoring in components require payment for software, as the calculation is not straightforward.

### Validity and reliability

Internal consistency for individual subscales falls around 0.8 or above, though sometimes lower for the Social functioning scale. Test-retest reliability is a little lower on average, between 0.71 and 0.89. The physical and mental components are quite closely correlated at higher scores, and some suggest they should be used with caution (Turk & Melzack 2011). Concurrent validity suggests that the mental component score may be no more than adequate (Rehabilitation Measures Database).

### Conditions used

It has been used in a wide variety of medical conditions involving pain and not involving pain, and in many pain studies.

### Pros and Cons

Pros:

- widely used; norms and comparator scores available
- takes up to 10 minutes to complete

Cons:

- scoring is time-consuming and error-prone without expensive software
- Mental Component Score may not be valid for use
- Scales have a fairly small range of response options, and scores are multiplied up to 0-100

Dworkin RH, Turk DC, Farrar JT, et al. "Core outcome measures for chronic pain clinical trials: IMMPACT recommendations", *Pain* 2005;113:9-19.

Rehabilitation Measures Database <http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=930> accessed 22.09.17

Turk SC, Melzack R. *Handbook of pain assessment*, 3rd edition. New York; Guilford Press, 2011.

## 8.5 Profile of Mood States

The Profile of Mood States (POMS) is a rating scale for assessing general psychological/mood states, not specific to a clinical population. The POMS measures 6 different dimensions of mood swings over a period of time: 'Tension or Anxiety', 'Anger or Hostility', 'Vigor or Activity', 'Fatigue or Inertia', 'Depression or Dejection', and 'Confusion or Bewilderment'. A 5-point scale ranging from "not at all" to "extremely" is administered by researchers to participants to assess their mood states. It could also be self-administered. The scale is usually applied by asking: "How you have been feeling during the PAST WEEK, INCLUDING TODAY" or "How you feel RIGHT NOW".

There are 2 versions of the POMS: the full version consists of 65 items while the POM-SF consists of 37 items. The international IMMPACT recommendations (Dworkin et al. 2005) stated that in chronic pain research, the original full version should be used without any adjustment. There are no established cut-offs or thresholds for interpreting the total scores (domain-specific or overall) or change scores from baseline. There is some literature that suggested POMS is age-specific, and that change scores need to be examined with the normal distribution of the specific population (as a baseline).

Copyright:

POMS is copyrighted

(<https://www.mhs.com/MHS-Assessment?prodname=poms2>). However, there are some online interactive tools that are freely accessible (e.g. <https://www.brianmac.co.uk/poms.htm>).

### Validity and reliability

The validity and reliability of POMS are less well studied in medical conditions compared to BDI. The majority of validation studies were conducted in sports research. A validation study (Curran et al. 1995) has suggested that POMS has a range of internal consistency of 0.63 to 0.96 (Cronbach  $\alpha$ ). For the brief version (POMS-SF), the internal consistency rating was 0.76 to 0.95 (Cronbach  $\alpha$ ). The correlation between the subscales and the total score in POMS and POMS-SF was calculated as  $r = 0.84$ . However, the test-retest reliability is less frequently examined.

### Conditions used

There is some literature on the use of POMS in cancer research and women's health (e.g. menopause, PMS, etc.), but various modified versions of the POM-SF were used in such studies instead of the full version. There is some evidence of use in chronic pain (i.e. post-herpetic neuralgia) (Rowbotham et al. 1998).

## Pros and cons

### Pros:

- Measuring general mood states including anxiety and anger, not just depression

### Cons:

- Very long (64 items) and difficult to fit onto a 2-page A4
- Variations in scoring:
  - most studies use the sum of item responses
  - some use an average of item responses
  - some studies use T scores - and don't report the norms used to calculate T scores
- No established minimal difference (MD) or thresholds for meaningful interpretation of change
- No anchor-based (patient or clinician ratings of importance) studies found change scores need to be examined with the normal distribution of the specific population (as a baseline)
- Few norms for clinical populations

Curran SL, Andrykowski MA JL, & Sudts JL (1995) "Short Form of the Profile of Mood States (POMS-SF): Psychometric Information", *Psychological Assessment*, 1995, 7:80-83.

Rowbotham M, Harden N, Stacey B, Bernstein P, Magnus-Miller L (1998) "Gabapentin for the treatment of postherpetic neuralgia: a randomized controlled trial", *JAMA*, 280: 1837-42.

## 8.6 Pain Catastrophising Scale

The term catastrophizing arises from cognitive models of depression (Beck 1976) and refers to a negative bias about the individual's situation and particularly about the future. It can also be considered to be an emotional response to pain, a communication to potential helpers (Sullivan et al. 2001), or a poor coping strategy (Rosenstiel & Keefe 1983). It consists of 13 statements, such as 'It's awful and I feel that it overwhelms me', with the response options indicating the degree to which the respondent has the thoughts and feelings described when in pain of 'not at all', 'to a slight degree', 'to a moderate degree', 'to a great degree', and 'all the time', scored 0,1,2,3,4 respectively. The maximum score, indicating the worst possible catastrophizing, is therefore 52. The questionnaire is self-completed. There are 3 subscales, but since they are highly correlated with one another, the total score is generally used.

Copyright: Available from

<https://eprovide.mapi-trust.org/instruments/pain-catastrophizing-scale> and is copyrighted to Michael Sullivan. It is free for use in academic research, but payment is required for use in commercial research.

### Validity and reliability

There are many studies estimating internal consistency, with a range from 0.82 to 0.98; far fewer estimate test-retest reliability, with a range of estimates from 0.73 to 0.97.

Validity is more problematic, given the lack of clarity about the concept of catastrophizing (Turner & Aaron 2001), and the subjectivity of the construct. The best that can be done is to assess whether catastrophizing correlates so highly with other pain variables that it is effectively redundant, or so low that it may not be measuring an important psychological component of pain. Catastrophizing does load on a general factor of pain distress, along with anxiety, depression, pain interference, pain intensity, etc. (Mounce et al 2010, Campbell et al 2013) and clearly shares variance with them, and with mood in pain-free respondents, but proponents argue that it does not load highly enough to be subsumed by them, and that its unique variance does predict future depression and disability, in acute and chronic pain.

### Conditions used

The PCS has been used across different pain conditions and for experimental and acute clinical (e.g. postoperative) pain.

### Pros and Cons

Pros:

- Good internal reliability; moderate to good test-retest reliability
- Widely used so easy to find comparison groups

- Many translated versions (but not comparable with English version or one another)
- Cons:
- Scale probably not linear
  - Interpretation can be difficult because of shared variance with other negative affect

Beck, A.T. (1976), *Cognitive therapy and the emotional disorders*, London: Penguin Books.

Campbell F, Bishop A, Dunn KM, Main CJ, Thomas E, Foster NE. "Conceptual overlap of psychological constructs in low back pain", *Pain* 2013;154:1783-91.

Mounce C, Keogh E, Eccleston C. "A principal components analysis of negative affect-related constructs relevant to pain: evidence for a three-component structure", *J Pain* 2010;11(8):710-8.

Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. "The Pain Catastrophizing Scale: further psychometric evaluation with adult samples", *J Beh Med* 2000;23(4):351-65.

Rosenstiel, A.K., & Keefe, F.J. (1983). "The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment" *Pain*, 17, 33-44.

Sullivan MJL. "The communal coping model of pain catastrophizing: clinical and research implications", *Canadian Psychology* 2012;53:32-41.

Sullivan MJ, Bishop SR, Pivik J. "The Pain Catastrophizing Scale: development and validation", *Psychol Assess* 1995;7(4):524-

Turner JA, Aaron LA. "Pain-related catastrophizing: what is it?", *Clin J Pain* 2001;17:65-71.

## 8.7 The Patient Health Questionnaire – 2

The PHQ-2 has two questions, taken directly from the PHQ-9, to answer which the respondent is asked to consider “Over the last 2 weeks, how often have you been bothered by any of the following problems”. The two items concern lack of pleasure or interest, and feeling down or depressed, with the response options of ‘not at all’, ‘several days’, ‘more than half the days’, ‘nearly every day’, scored 0,1,2,3 respectively. Scores, therefore, fall from 0 to 6, with 3 as the recommended cut-off point for screening purposes, with a sensitivity of 62%, specificity of 95%, and positive predictive value of 75%, all for any depressive disorder. The authors emphasise that it is not a diagnostic scale or a measure of severity of depression, but a screening scale, indicating patients who require further investigation.

### Copyright

It is available from [www.cqaimh.org/pdf/tool\\_phq2.pdf](http://www.cqaimh.org/pdf/tool_phq2.pdf). Like the PHQ-9, it is free to use.

### Validity and reliability

No internal reliability estimation found, but it is unlikely to be high for a 2-item scale, nor test-retest reliability. On validity, almost all psychometric data concerns accuracy for detecting a major depressive disorder, which is not what it would be used for, and for doing so in non-pain populations.

### Conditions used

No papers found concerning use in pain or related diseases or disabilities.

### Pros and Cons

#### Pros:

- Does not contain any somatic items, unlike the parent PHQ-9 and almost all depression scales
- Short and easy to administer and score
- Multiple languages

#### Cons:

- No track record in pain or related health problems
- Not designed or tested as a severity scale, i.e. not as an outcome measure: it is a screening tool. Not yet clear whether it will be useful as a severity scale.

Kroenke K, Spitzer RL, Williams JB. “The Patient Health Questionnaire-2: validity of a 2-item depression screener” *Medical Care* 2003;41:1284-94.

## 8.8 The Pain Self-Efficacy Questionnaire (PSEQ)

The Pain Self-Efficacy Questionnaire (PSEQ) is a 10-item questionnaire, developed in the 1980s by Michael Nicholas (1). It assesses the confidence of people with any type of chronic pain in activity despite pain. It covers enjoying activities, household daily activities, social life, coping in general, work, leisure activities, coping with pain without medication, accomplishing goals, living a normal lifestyle, and becoming more active, all 'despite pain'. Each is rated on a 7 point scale from 0 = not at all confident to 6 = completely confident. The total score, ranging from 0 to 60, is calculated by adding the scores for each item. Higher scores reflect stronger self-efficacy beliefs.

The PSEQ is easy to complete and has a high completion rate. It can be used in assessment, treatment planning, and outcome evaluation (Nicholas 2007).

- Low scores (< 20) - Patient is more focused on the pain (seeking pain relief first). This needs to be addressed first before embarking on an exercise program. This is also a predictor of long term disability and depression
- High scores (> 40) - Patient is likely to respond well to an exercise program (Frost et al 1993). They also are more likely to maintain this on a longer term.

Copyright: None, the PSEQ is free to use.

### Validity and reliability

Internal consistency is excellent (0.92 Cronbach's  $\alpha$ ) and test-retest reliability is high over a 3-month period (Asghari and Nicholas 2001). Validity is reflected in high correlations with measures of pain related disability, different coping strategies, and another more activity-specific measure of self-efficacy. The evidence of the PSEQ's sensitivity to change provides support for its construct validity.

### Conditions used

All pain conditions

### Pros:

- Easy to use
- Takes less than 2 minutes
- Normative data has been established for pain clinic population
- Predicts longer term disability outcome

Nicholas M.K. *Self-efficacy and chronic pain*. Paper presented at the annual conference of the British Psychological Society. St. Andrews, 1989

Nicholas M.K. "The Pain self-efficacy questionnaire: Taking Pain into account," *European Journal of Pain*, Vol 11, issue 2, Feb 2007, Pages 153-163

Asghari A, Nicholas MK. "Pain self-efficacy beliefs and pain behaviour: a prospective study", *Pain* 2001;94:85-100.

## 9 Quality of Life

### 9.1 EuroQol 5D

The EQ-5D<sup>®</sup> is a patient-reported outcome measure (PROM) that captures five dimensions of health-related quality of life: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Its primary use is in the calculation of Quality Adjusted Life Years. The measure is recommended by the NHS Patient Reported Outcome Measures (PROMs) Group and can be used across a broad range of health conditions. Normative data is available via <https://euroqol.org/>

The EQ5D exists in 3 versions: ED5D-3L, ED5D-Y, and EQ5D-5L EQ5D-3L

Each dimension uses three levels of severity corresponding to no problems, some problems, and extreme problems.

This version is currently recommended for use by The National Institute for Health and Care Excellence.

EQ5D-Y

A youth version for self-completion by adolescents ages 7-12 year of age.

Eq5D-5L

Developed to improve the sensitivity of the EQ5D-3L and to reduce ceiling effects this version maintains the same dimensions, but reliability and sensitivity were increased by using five levels of severity corresponding to no problems, slight problems, moderate problems, severe problems and extreme problems. The measures consist of two pages. It is completed by the patient by placing a tick in a box against the most appropriate level in each dimension. The patient also rates their overall health using a hash-marked visual analogue scale (VAS) measuring from 0-100. The first section acts as a qualitative assessment of patient health and the VAS acts as a quantitative measure.

Copyright

The measure is copyright and available via <https://euroqol.org/>. Registration is required to determine the cost for use.

#### **Validity and Reliability**

The EQ5D has been tested for reliability and validity in a variety of patient groups and in a broad range of languages.

#### **Validity**

The EQ5D-3L demonstrated broad ordinal agreement between the EQ5D and the SF36. A large ceiling effect was found with over 95% at the ceiling for the functional dimensions. The EQ5D was less sensitive which is attributable to the smaller number of items.



## Reliability

Test-retest reliability has been assessed and found to be reliable (Intraclass correlation coefficient 0.70) over a one-week time point. Higher reliability was associated with better self-rated health status.

## Conditions Used

The EQ5D has been shown to be an effective measure of meaningful change in the quality of life across a wide range of medical conditions.

## Pros and Cons

### Pros:

- Widely recognised quality of life measure
- The quality of life construct measured by the EQ5D appears to differ from the SF6D and the SF6D has better ability to detect change beyond a 95% confidence interval in low back pain populations
- Although presented on 2 pages the measure has only 4 or 6 items (depending on version used)
- Demonstrated to detect clinically meaningful change in pain conditions

### Cons:

- The SF12 appears more sensitive to change than the EQ5D
- The wording of some questions has been suggested to be unclear e.g. “walk about” raises questions of how far and where. Also, the mid-range definitions can be challenging for patients to discriminate

Janssen, M., Pickard, A., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn, P. and Busschbach, J. (2012). “Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study”, *Quality of Life Research*, 22(7), pp.1717-1727.

Fransen, M. and Edmonds, J. (1999). “Reliability and validity of the EuroQol in patients with osteoarthritis on the knee”. *Rheumatology*, 33, pp.807-813.

Brooks, R. and The EuroQol Group (1996) “EuroQol: the current state of play”, *Health Policy*, 37, pp.53-72.

Payakachat N, Ali MM, Tilford JM, Rock L, Rock L. *HHS Public Access*, 2016;33(11):1137-54.

Obradovic, M., Lal, A. and Liedgens, H. (2013). “Validity and responsiveness of EuroQol-5 versus Short Form-6-dimension (SF-6D) questionnaire in chronic pain”, *Health and Quality of Life Outcomes*, 11(110-119).

Johnson, J. and Coons, S. (2017). “Comparison of the EQ-5D and SF-12 in an adult”, *Quality of Life Research*, 7(155-166).

## 10 Patient Reported Outcomes

### 10.1 Patient Global Impression of Change (PGIC)

Global Rating of Change Scales, also known as Patient Global Impression of Change Scales, ask the patient to rate how much his/her condition has improved or deteriorated since a pre-defined time point. The scale measures patients' self-defined construct of change rather than a specific dimension of change determined by the clinician or researcher. Scales take many forms with no set standard. They focus on the participant rating their impression of change between two fixed time points and between 7 and 15 points on a numerical rating scale or using a visual analogue scale.

Copyright: None

#### Validity and Reliability

Kamper et al provide a review of the properties of Global Rating of Change Scales. Global Rating of Change Scales have shown significant correlation with change in Roland Morris Disability Questionnaire, Oswestry Disability Index, EQ-5D and Pain Rating Scale. Test-retest reliability has been shown as ICC 0.9 on an 11-point scale and a standardised response mean of 0.2-1.7 has been shown for 7 and 15 point scales. Minimally clinically important change is 2 points on an 11-point scale

#### Conditions used

Global Rating of Change Scales has been highlighted as an important aspect of measuring participant satisfaction with treatment and their views on the clinical importance of change and have been used across a broad range of medical conditions.

#### Pros and Cons

Pros:

- As a single question, it is brief and easy to administer and the results are easy to interpret
- Strong correlations with patient satisfaction measures have been reported
- test-retest reliability over a period of 24 hours is excellent

Cons:

- The influence of current status impacts on ratings (patients struggle to take their baseline into account) and therefore it may not be reliable over a period of months

Kamper, S., Maher, C. and Mackay, G. (2009) "Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design", *The Journal of Manual and Manipulative Therapy*, 17(3), pp.163-170.

Kamper, S., Ostelo, R., Knol, D., Maher, C., de Vet, H. and Hancock, M. (2010) "Global Perceived Effect scales provide reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status", *Journal of Clinical Epidemiology*, 63, pp.760-766.

## 11 Summary Table

	Purpose	Population	Method	Copyright
NPRS	To assess pain severity	All adult pain conditions and children over 10 years old	Self-completion on paper, screen, or can be used verbally including by telephone	No
VAS	To assess pain severity	All patients. Used in various domains including pain	Self-completion on paper	No
VRS	To assess pain severity	All adult pain conditions and children over 10 years old	Self-completion on paper, screen, or can be used verbally including by telephone	No
RMDQ	Disability secondary to low back pain	Low back pain	Self-completion on paper or screen	No
ODI	Disability secondary to low back pain	Low back pain in secondary care setting	Self-completion on paper or screen	Yes
BPI	To assess pain severity and pain interference with everyday life	Acute, chronic or cancer pain	Self-completion on paper or screen	Yes

	<b>Purpose</b>	<b>Population</b>	<b>Method</b>	<b>Copyright</b>
BDI-II	To assess severity of depression	Community and psychiatric populations; not standardised in physical health	Self-completion on paper or screen	Yes
CES-D	To assess severity of depression	Community and psychiatric populations; not standardised in physical health	Self-completion on paper or screen	No
HADS	To assess distress	Adults with physical health problems	Self-completion on paper or screen	No
SF-36	To assess quality of life	Adults with any physical or mental health problem	Self-completion on paper or screen	No
POMS	To assess transient, distinct mood states	Normal adults; used in a few with health problems	Self-completion on paper or screen	Yes
PCS	To assess the emotional response to pain	Different pain conditions including acute and post-operative pain	Self-completion on paper or screen	Yes
PHQ-2	To screen for depression	Screening in community populations; no use yet in physical health conditions	Self-completion on paper or screen; or can be used verbally including by telephone	No

	<b>Purpose</b>	<b>Population</b>	<b>Method</b>	<b>Copyright</b>
PSEQ	To assess confidence in activity despite pain	Chronic pain of any type	Self-completion on paper or screen	No
EQ5D	To assess quality of life	Adults with any physical or mental health problem	Self-completion on paper or screen; can be administered over telephone	Yes
PGIC	To assess overall impression of change	Adults with any medical conditions	Self-completion on paper or screen; can be administered over telephone	No

## 12 The working party

The following people contributed to the document:

Chair: Dr G Baranidharan, Consultant Pain Medicine and Anaesthesia, Leeds Teaching Hospitals NHS Trust, Leeds

Dr Amanda Williams, Reader in Clinical Health Psychology, Research Department of Clinical, Educational & Health Psychology, University College London, London

Miss Sarah Wilson, Clinical Specialist Physiotherapist, Royal National Hospital for Rheumatic Diseases, Bath

Dr Paul Cameron, National Coordinator for Chronic Pain, Scotland & Pain Service Lead, NHS Fife Pain Management Service

Mr Toni Tan, Senior Technical Advisor, National Institute for Health and Care Excellence, UK

Board Members, Faculty of Pain Medicine

Communications Committee, British Pain Society

**The Faculty of Pain Medicine**  
of the Royal College of Anaesthetists

Churchill House  
35 Red Lion Square  
London WC1R 4SG

tel: 020 7092 1682 or 020 7092 1746  
email: [contact@fpm.ac.uk](mailto:contact@fpm.ac.uk)

[www.fpm.ac.uk](http://www.fpm.ac.uk)