# Pain Outcomes: A Brief Review of Instruments and Techniques

**Jarred Younger, PhD**, **Rebecca McCue, BA**, and **Sean Mackey, MD, PhD**

## Abstract

Pain is a difficult outcome to measure due to its multifaceted and subjective nature. The need for selecting proper outcome measures is high because of the increasing demand for scientifically valid demonstrations of treatment efficacy. This article discusses some basic topics in the measurement of pain outcomes and addresses issues such as statistical versus clinical significance, daily home data collection, appropriate length of outcome measurement packets, and the possibility of objective pain measurements. This article also reviews some of the more commonly used tools for measuring pain and pain-related disability. By selecting the proper tools and employing them correctly, we can obtain highly reliable and valid measures of pain outcomes in research and clinical care.

## Introduction

Pain is a complex and subjective experience that poses a number of measurement challenges. However, in the current culture of evidence-based medicine, it is important that clinicians and researchers utilize sensitive and accurate pain outcome measures. Currently, there exists no valid and reliable method of objectively quantifying an individual's experience of pain. Therefore, we rely mainly on self-report measures to determine the impact of pain. Despite the challenges that pain measurement presents, a number of tools and approaches can be employed to collect useful pain estimates. This article discusses some important considerations in selecting and utilizing pain outcome measures for clinical and research purposes. We also review some of the commonly used pain measurement tools, provide a few additional recommendations in preparing a pain measurement procedure, and discuss the future of objective pain measurement. For a more comprehensive look at pain measurement, we refer readers to the excellent 2001 text by Turk and Melzack, *Handbook of Pain Assessment* [1].

## Statistical Versus Clinical Significance

Outcome measures provide a metric by which to gauge changes in the experience of pain; however, human interpretation is always needed to determine if a meaningful change has occurred. Generally, a treatment should demonstrate a statistically significant effect and clinically significant effect [2]. These two types of significance are complementary and answer two different questions. One type of significance cannot be inferred from the other. In a basic sense, statistical significance answers the question, "Is this a real effect?" and clinical significance answers the question, "Is this an important effect?"

Statistical and clinical significance are oftentimes confused when one misinterprets a *P* value as an indicator of a treatment's strength. The *P* value is an extremely common parameter that

**Corresponding author**: Jarred Younger, PhD, Department of Anesthesia, Division of Pain Management, Stanford, University School of Medicine, 780 Welch Road, Suite 208C, Palo Alto, CA 94304-1573, USA. Jarred.Younger@stanford.edu.

is used to indicate how likely it is that the observed treatment effect occurred purely by chance, and does not truly indicate the treatment's efficacy. If a statistical test showed drug A to be superior to drug B at reducing pain with $P = 0.05$, there is a 5% chance that the observed difference is not real, and would probably not be seen if the study was conducted again with new patients. At $P = 0.0001$, there is only a 0.01% chance that the results are erroneous, and we can be much more confident that the drug is having a true impact on the outcome. However, smaller $P$ values say nothing about the actual strength of the effect. Many factors have a drastic impact on $P$ values. For example, as sample size goes up, $P$ values go down, even if the strength of the effect never changes. This phenomenon can lead very large studies (1000–2000 patients) to show significant statistical results even when the changes in outcomes are very small, and meaningless to patients and clinicians.

To circumvent this problem of interpretability, readers should rely more on the effect size. For clinical trials, effect size is generally expressed in clinically relevant terms, such as the reduction of pain scores that occurs when switching from placebo to the active drug. To determine clinical significance, the clinician or researcher must first choose a metric (eg, percent reduction of pain), and then choose a cutoff that indicates a clinically meaningful change (eg, 30% reduction of pain).

## Determining Cutoffs for Clinical Significance

When deciding on a cutoff value for clinical significance, we must determine the minimal amount of change in pain that would be valuable and important to patients. Many different approaches to determining this minimal important clinical difference have been proposed. A well-researched cutoff method suggests that a 30% reduction of pain can be considered clinically significant [3]. This level corresponds with a "much improved" or "very much improved" response from patients on a global impression of change, or 2 points on a 0 to 11 pain intensity numerical rating scale (these scales are discussed in more depth later in the article). Farrar et al. [4] offer several other conversions of this standard, and Kvien et al. [5•] offer other methods for determining clinical significance. For example, the PASS (Patient Acceptable Symptom State) method has been used to determine that, for many chronic pain conditions, a reduction in daily pain of 35 mm on a 100-mm visual analogue scale indicates a satisfactory result for the patient. A recent consensus statement by the IMMPACT (Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials) group provides additional recommendations on detecting clinically important changes [6••].

## Measuring Daily Outcome Variables

If a treatment is tested on a small number of patients, then obtaining statistical significance may be impossible, even if a large clinical effect is observed. One method for dealing with very small sample sizes involves running a crossover design, in which each patient receives all treatment conditions. The use of crossover designs over parallel designs (in which each patient receives one treatment condition) can reduce the required sample size by up to 90% [7]. These required sample sizes may be further reduced by collecting the outcome measure multiple times at baseline, and within each treatment condition [8].

When daily measurements are used in patients with chronic pain, these assessments usually have to be self-reported in a natural environment. There are a number of challenges in collecting these real-world measures, such as the lack of control over measure completion and the poor control over the assessment environment and circumstances. Data collection methods, such as paper diaries or daily phone calls, are widely used but each has significant drawbacks. Paper diaries are prone to backfilling (leading to inaccurate retrospective guesses), and phone calls are work-intensive to study staff and may be disruptive to the patient. A newer approach to collecting home-based outcome measures involves the use of handheld computers. New units

can be bought for under $100, and free software, such as the Experiential Sampling Program, is available for administering the measures [9]. However, these units may be hard to operate by elderly or impaired patients. Another drawback of using handheld computers is the possibility that the units will be lost, stolen, or damaged. To avoid potential data loss, it is recommended that data be frequently downloaded to the study site's computers, which may be done in person or via Internet-based transfer services. Wright [10] further discusses the important data considerations in biomedical research.

## Available Tools

Because resources and time are always limited, we are forced to make decisions on which outcomes to include in our measurements. In some cases, a simple measure of pain intensity may be the most logical primary outcome variable. In other cases, a general indicator of work or social functioning may be more clinically relevant. Pain clinicians will recognize cases in which an individual is profoundly disabled by seemingly low pain intensity, and cases in which an individual maintains a productive and fulfilling lifestyle despite reporting a high degree of pain. Some interventions may have little impact on pain intensity scores, but may benefit mood, motivation, and functioning. Therefore, one of the most important decisions to make in testing a new treatment is determining what outcomes are most clinically relevant. We now review a few of the available pain outcome measures, which range from simple and narrowly defined, to large and multidimensional. Each has its proper place in measuring pain outcomes. We also refer readers to the IMMPACT recommendation on a core set of outcome measures [11].

### Unidimensional scales

In the busy clinical setting, pain measures must be simple, quick to administer, and easily understood by the patient. Unidimensional scales fill this role by providing fast (often one-item) measures of pain that can be administered multiple times with minimal administrative effort. One commonly used unidimensional tool is the numerical rating scale (NRS). Although variations exist, the instrument typically consists of scores from 0 to 10 (or 0–100), with the far left being described as "no pain" and the far right described as "worst pain imaginable." The NRS has the advantage of being administered verbally, thus not requiring patient mobility.

As an alternative to the NRS, a visual analogue scale (VAS) may be used. The patient marks anywhere along a 10-cm line to indicate their current pain intensity, which can be measured in millimeters to yield a 101-point scale. To assist in the scoring process, slide-rule-like devices have been developed. The VAS provides a high degree of resolution and is probably the most sensitive single-item measure for clinical pain research. Another alternative is the verbal rating scale (VRS), which is sometimes used for individuals who have trouble translating their pain experience into a number value. The anchors are instead replaced by descriptors, such as no pain, mild pain, moderate pain, and severe pain. This type of measure has several statistical drawbacks and is usually used only when patient characteristics require it.

Attempts at capturing pain improvement more broadly can also be achieved with single-item measures. For example, the PGIC (Patient Global Impression of Change) scale provides a single, general estimate of improvement. The typical PGIC asks patients to rate their current status as: 1) very much improved, 2) much improved, 3) minimally improved, 4) no change, 5) minimally worse, 6) much worse, or 7) very much worse. This measure has the advantage of being applicable to a wide variety of conditions and treatments, but lacks sensitivity required for some statistical analyses.

## Multidimensional scales

In many situations, a simple, one-item instrument is not sufficient to truly capture pain or quality of life. Many comprehensive measures of pain exist. These instruments typically measure several dimensions of pain, with differing combinations of (among other things): pain intensity, quality, affect, interference with functioning, and effects on general quality of life. By assessing the pain experience in a more complex way, these scales may circumvent the commonly observed lack of association between pain intensity and disability.

**The short-form McGill Pain Questionnaire—**The short-form McGill Pain Questionnaire (SF-MPQ) is a well-validated measure with extensive clinical research use [12]. Patients rate their pain in sensory terms (eg, sharp or stabbing) and affective terms (eg, sickening or fearful), with 15 total descriptors. Each item is rated on a 4-point scale that ranges from none to severe. The SF-MPQ also has a single VAS item for pain intensity and a VRS for rating the overall pain experience.

**Brief Pain Inventory short form—**The Brief Pain Inventory short form (BPI-SF) captures two broad pain domains: 1) the sensory intensity of pain, and 2) the degree to which pain interferes with different areas of life [13]. The 17-item scale also captures pain location, pain medication use, and response to treatments.

**West Haven-Yale Multidimensional Pain Inventory—**The West Haven-Yale Multidimensional Pain Inventory (WHYMPI) is a comprehensive pain outcomes measure that contains 52 items and 12 subscales [14]. Some of the subscales include perceived interference of pain in a variety of areas, response from significant others, pain severity, perceived life control, affect, and participation in various work, social, and personal activities. Items are assessed on a 7-point scale. The scale can also yield clinically useful types of pain patients, such as dysfunctional, interpersonally depressed, and adaptive copers.

**Treatment Outcomes of Pain Survey—**The Treatment Outcomes of Pain Survey (TOPS) is one of the longest and most comprehensive instruments for patients with chronic pain [15]. The survey is an extension of the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36), an instrument designed to measure quality of life and function in a variety of patient populations. The survey also contains items from the WHYMPI, BPI, the Oswestry Disability Questionnaire, and items that assess coping style, fear avoidance beliefs, substance abuse, patient satisfaction with treatment, and demographic variables. The full scale contains 120 items, and there is a 61-item follow-up version. The following scores are obtained from the TOPS: pain symptoms, functional limitations, perceived family/social disability, objective family/social disability, patient satisfaction, fear avoidance, passive coping, solicitous responses, work limitations, and life control. The questionnaire may be burdensome to some patients because of its length.

## Behavioral scales

Behavioral scales are often used in noncommunicative patients for whom direct assessment of pain self-report is not possible. These tools often measure facial or bodily movements as proxies for pain. Li et al. [16] offer a critical review of the tools available for adult patients. Hummel and van Dijk [17] offer a similar review for nonverbal infants. These scales are clinically necessary in some cases, but are generally not acceptable as outcomes for clinical trial reporting.

## Length of Outcome Packet

There is no established maximum length for an outcome packet. In general, as a questionnaire packet increases in length, there is an increasing risk that the patient will lose interest in the scale. Scales placed later in the packet may have incomplete or unreliable data. An individual's ability to persist through a questionnaire depends on a number of individual and environmental factors (eg, attention span, interest in the scale, dedication to the project, incentives, outside distracters, or item complexity). Conservatively, questionnaire packets should be able to be completed by the majority of individuals in under 25 minutes. These longer packets may also be combined with more frequently administered, single-item measures to provide a balance of depth of information and temporal resolution.

## What About an Objective Marker of Pain?

The field of pain management would benefit enormously from an objective, physiologic marker of pain [18]. Several physiologic variables have been measured for this purpose, such as skin conductance [19] and heart rate [20]. In general, however, these markers do not correlate strongly enough with pain to warrant their use as a surrogate measure of pain [21-23]. Pain can exist in the absence of changes in these measures, and these measures can fluctuate drastically with no change in pain. These peripheral measures indicate general autonomic activity, which can be influenced by many factors other than pain, such as other forms of arousal. Also, treatments may directly impact those physiologic variables, further reducing their reliability as a clean pain measure. Work still continues in this area, with tests of more sophisticated measurement approaches [24,25] or biomarkers of pain intensity [26].

Because pain is more than just the peripheral and spinal transmission of nociceptive information, many new attempts at objectively measuring pain have focused on the brain. These neurologic markers of pain attempt to encompass the large number of emotional, situational, and attentional factors that can accentuate or attenuate the pain experience. The pursuit of a neuroimaging approach to measuring pain has intensified with better technology and increases in spatial and temporal resolution. Several brain regions show pain-related activations, and some degree of pain intensity encoding has been described [27]. Methods such as magnetoencephalography, functional MRI, and positron emission tomography are used to explore the supraspinal, neural correlates of pain. However, no neuroimaging technique has been established as a reliable method of measuring pain. The technology has not yet advanced to the point where we can accurately classify someone as being in pain based purely on a scan; we are further still from judging the amount of pain change based on changes in neural signals.

Even if a measure of acute pain is developed, it does not necessarily follow that such a measure could accurately measure chronic pain. The way an individual interprets the pain, or learns to cope with its unpredictability, can greatly moderate the way pain affects his or her life. Given the complex interplay between sensory, affective, and cognitive components of pain, it is unlikely that an objective measure will soon emerge that captures the experience of pain in a clinically useful way.

Many physical functioning and performance tests, such as range-of-motion, exist and have been used as a proxy for objective pain measurement [28]. Examples of standardized performance/functioning tests for chronic pain include the following: the loaded forward-reach test for chronic back pain [29], timed "Up & Go" test for osteoarthritis [30], and grip strength for rheumatoid arthritis [31]. In general, these performance tests only modestly predict self-reported pain, with correlations rarely exceeding 0.30 [32-36]. These results suggest that pain is just one component of physical performance, and other factors, such as fear of pain, may heavily impact performance scores [37,38]. Therefore, although clinic-based tests of

functioning can complement self-reported pain measures in chronic conditions, they are not useful as a pain-report substitute.

## Conclusions

Despite the difficulty inherent to measuring pain, there are a number of accepted tools for tracking pain-related treatment outcomes. The proper use of these tools can allow clinicians and researchers to demonstrate both statistically and clinically significant treatment effects. These instruments range from quick, one-item assessments of pain intensity, to long surveys that tap into multiple dimensions of the pain experience and overall functioning. Until more objective physiologic/neurologic measurement techniques are perfected, clinicians who study pain will rely on the careful use of established self-report pain measures.

## References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:

• Of importance

•• Of major importance

1. Turk, DC.; Melzack, R. Handbook of Pain Assessment. Guilford Press; New York: 2001.

2. Todd KH. Clinical versus statistical significance in the assessment of pain relief. Ann Emerg Med 1996;27:439–441. [PubMed: 8604855]

3. Farrar JT, Young JP, LaMoreaux L, et al. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–158. [PubMed: 11690728]

4. Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: a validation study. J Pain Symptom Manage 2003;25:406–411. [PubMed: 12727037]

5•. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/ MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis 2007;66(Suppl 3):iii40–iii41. [PubMed: 17934093] Provides guidelines for empirically setting clinically significant outcome measure cutoffs.

6••. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–121. [PubMed: 18055266] Outlines a methodological standard for conducting clinical trials in chronic pain.

7. Louis TA, Lavori PW, Bailar JC, et al. Crossover and self-controlled designs in clinical research. N Engl J Med 1984;5:24–31. [PubMed: 6689736]

8. Vickers AJ. How many repeated measures in repeated measures designs? Statistical issues for comparative trials. BMC Med Res Methodol 2003;3:22. [PubMed: 14580266]

9. Experiential Sampling Program. [Accessed August 2008]. Available at http://www.experience-sampling.org/esp.html.

10. Wright DB. Making friends with your data: improving how statistics are conducted and reported. Br J Educ Psychol 2003;73:123–136. [PubMed: 12639281]

11. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19. [PubMed: 15621359]

12. Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30:191–197. [PubMed: 3670870]

13. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994;19:129–138. [PubMed: 8080219]

14. Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). Pain 1985;23:345–356. [PubMed: 4088697]

15. Rogers WH, Wittink H, Wagner A, et al. Assessing individual outcomes during outpatient multidisciplinary chronic pain treatment by means of an augmented SF-36. Pain Med 2000;1:44–54. [PubMed: 15101963]

16. Li D, Puntillo K, Miaskowski C. A review of objective pain measures for use with critical care adult patients unable to self-report. J Pain 2008;9:2–10. [PubMed: 17981512]

17. Hummel P, van Dijk M. Pain assessment: current status and challenges. Semin Fetal Neonatal Med 2006;11:237–245. [PubMed: 16647310]

18. Chizh BA, Hobson AR. Using objective markers and imaging in the development of novel treatments of chronic pain. Expert Rev Neurother 2007;7:443–447. [PubMed: 17492894]

19. Storm H. Skin conductance and the stress response from heel stick in preterm infants. Arch Dis Child Fetal Neonatal Ed 2000;83:F143–F147. [PubMed: 10952711]

20. Tousignant-Laflamme Y, Rainville P, Marchand S. Establishing a link between heart rate and pain in healthy subjects: a gender effect. J Pain 2005;6:341–347. [PubMed: 15943955]

21. Bossart P, Fosnocht D, Swanson E. Changes in heart rate do not correlate with changes in pain intensity in emergency department patients. J Emerg Med 2007;32:19–22. [PubMed: 17239728]

22. Hellerud BC, Storm H. Skin conductance and behaviour during sensory stimulation of preterm and term infants. Early Hum Develop 2002;70:35–46.

23. Harrison D, Boyce S, Loughnan P, et al. Skin conductance as a measure of pain and stress in hospitalised infants. Early Hum Develop 2006;82:603–608.

24. Ledowski T, Bromilow J, Peach MJ, et al. Monitoring of skin conductance to assess postoperative pain intensity. Br J Anaesth 2006;97:862–865. [PubMed: 17060329]

25. Schasfoort FC, Formanoy MAG, Bussmann JBJ, et al. Objective and continuous measurement of peripheral motor indicators of pain in hospitalized infants: a feasibility study. Pain 2008;137:323–331. [PubMed: 17964721]

26. Okuse K. Pain signaling pathways: from cytokines to ion channels. Int J Biochem 2007;39:490–496.

27. Coghill RC, Sang CN, Maisog JA, et al. Pain intensity processing within the human brain: a bilateral, distributed mechanism. J Neurophysiol 1999;82:1934–1943. [PubMed: 10515983]

28. Harding VR, Williams AC, Richardson PH, et al. The development of a battery of measures for assessing physical functioning of chronic pain patients. Pain 1994;58:367–375. [PubMed: 7838586]

29. Smeets RJ, Hijdra HJ, Kester AD, et al. The usability of six physical performance tasks in a rehabilitation population with chronic low back pain. Clin Rehabil 2006;20:989–998. [PubMed: 17065542]

30. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. Phys Ther 2006;86:1489–1496. [PubMed: 17079748]

31. Goodson A, McGregor AH, Douglas J, et al. Direct, quantitative clinical assessment of hand function: usefulness and reproducibility. Man Ther 2007;12:144–152. [PubMed: 16899386]

32. Simmonds MJ, Olson SL, Jones S, et al. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. Spine 1998;23:2412–2421. [PubMed: 9836355]

33. Ballinger DA, Rintala DH, Hart KA. The relation of shoulder pain and range-of-motion problems to functional limitations, disability, and perceived health of men with spinal cord injury: a multifaceted longitudinal study. Arch Phys Med Rehabil 2000;81:1575–1581. [PubMed: 11128892]

34. Salisbury SK, Choy NL, Nitz J. Shoulder pain, range of motion, and functional motor skills after acute tetraplegia. Arch Phys Med Rehabil 2003;84:1480–1485. [PubMed: 14586915]

35. Weiner DK, Rudy TE, Morrow L, et al. The relationship between pain, neuropsychological performance, and physical function in community-dwelling older adults with chronic low back pain. Pain Med 2006;7:60–70. [PubMed: 16533199]

36. Novy DM, Simmonds MJ, Lee CE. Physical performance tasks: what are the underlying constructs? Arch Phys Med Rehabil 2002;83:44–47. [PubMed: 11782832]

37. Moseley GL. Evidence for a direct relationship between cognitive and physical change during an education intervention in people with chronic low back pain. Eur J Pain 2004;8:39–45. [PubMed: 14690673]

38. Thomas JS, France CR. The relationship between pain-related fear and lumbar flexion during natural recovery from low back pain. Eur Spine J 2008;17:97–103. [PubMed: 17972113]