Special Article

The Measurement of Pain, 1945–2000

Bill Noble, MBChB, MRCGP, MD, David Clark, PhD, Marcia Meldrum, PhD, Henk ten Have, MD, PhD, Jane Seymour, RGN, PhD, Michelle Winslow, PhD, and Silvia Paz, MD

Academic Palliative Medicine Unit (B.N., S.P.), University of Sheffield, Sheffield, United Kingdom; International Observatory on End-of-Life Care (D.C., M.W.), Institute for Health Research, Lancaster University, Lancaster, United Kingdom; Department of History (M.M.), University of California at Los Angeles, Los Angeles, California, USA; Department of Ethics, Philosophy and History of Medicine (H.t.H.), University of Nijmegen, Nijmegen, The Netherlands; and School of Nursing and Midwifery (J.S.), University of Sheffield, Sheffield, United Kingdom

Abstract

Three strands of activity can be identified in the history of pain measurement. The first, psychophysics, dates back to the nineteenth century and measures the effect of analgesia by quantifying the noxious stimulation required to elicit pain, as well as the maximum stimulation tolerated. The second uses standardized questionnaires for patients, developed to categorize pain according to its emotional impact, distribution, character, and other dimensions. The third asks patients to report on pain intensity using rating scales, and is used in clinical trials where analgesics are evaluated and results can be combined to influence clinical guidelines and protocols. Although all three strands have found a place in modern clinical practice or drug development, it is the reporting of pain by patients undergoing treatment using simple scales of intensity which has emerged as the crucial method by which analgesic therapies can now be evaluated and compared.

Key Words

Pain measurement, history of pain, analgesic evaluation

Introduction

…but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.

—Lord Kelvin Lecture “Electrical Units of Measurement” May 3, 1893

In the second half of the twentieth century, pain came to be understood as an experience which can be reported only by the sufferer. Although certain physiological responses or behaviors prompted by pain may be observed, modern medicine has produced no single parameter to represent our understanding of what we know to be pain. Pain has come to be viewed as a subjective phenomenon with many features, of which severity or intensity is, as Melzack comments, “the salient dimension of pain.” It is intensity which has been the subject of most methodological innovation in pain research.1 The underlying driver for a reliable, valid, and
sensitive measure of pain intensity has, of course, been the need to establish the efficacy of analgesics and other therapies for painful conditions. Clinical assessment of pain as an aid to diagnosis has also informed the development of methods which encompass other dimensions of pain. The period since World War II has seen the development in the pain field of diverse research methods and contrasting methodologies, with an ensuing debate on the utility of the information generated. Although some latitude in experimental design exists, the dominant methodology of the randomized controlled trial, with its measured clinical outcomes, now informs both clinical decisions and the provision of cancer pain services.

This article explores strands of scientific effort in the period since 1945, some of which have come to inform present day practice in pain medicine and cancer care. In Table 1, we set out the chronology of these developments. Our analysis focuses around three crucial models of pain measurement: psychophysics, multidimensional questionnaires using standardized descriptors, and scales which rate the intensity of pain. These models have not appeared in a linear fashion; rather they occupy overlapping time periods and can variously be seen as mutually reinforcing, often conflicting and occasionally complementary.

### Table 1

<table>
<thead>
<tr>
<th>Year</th>
<th>Key Innovations in Pain Measurement (1939–1999)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1939</td>
<td>List of pain words categorized into five groups (Dallenbach)</td>
</tr>
<tr>
<td>1940</td>
<td>Psychophysical methods. Threshold, JND and tolerance (Hardy, Wolff, Goodell)</td>
</tr>
<tr>
<td>1941</td>
<td>Tail flick test (D’Amour and Smith)</td>
</tr>
<tr>
<td>1948</td>
<td>Simple descriptive scales for pain intensity and pain relief (Keele)</td>
</tr>
<tr>
<td>1950</td>
<td>Challenge to validity of psychophysical methodology (Beecher)</td>
</tr>
<tr>
<td>1952</td>
<td>Established principles and practices for measurement of subjective response (Beecher)</td>
</tr>
<tr>
<td>1953</td>
<td>Average Pain Index (Beecher)</td>
</tr>
<tr>
<td>1956</td>
<td>Controlled clinical trials of analgesics. Crossover RCTs (Houde and Beecher)</td>
</tr>
<tr>
<td>1964</td>
<td>10-cm line VAS (Bond)</td>
</tr>
<tr>
<td>1966</td>
<td>Submaximum effort tourniquet technique. An attempt to replicate clinical pain in laboratory (Smith)</td>
</tr>
<tr>
<td>1971</td>
<td>Language of pain expanded list in 3 categories (Melzack and Torgeson)</td>
</tr>
<tr>
<td>1975</td>
<td>McGill Pain Questionnaire (Melzack)</td>
</tr>
<tr>
<td>1979</td>
<td>Pain defined by IASP (Mersey)</td>
</tr>
<tr>
<td>1981</td>
<td>Face scale for children (Rogers)</td>
</tr>
<tr>
<td>1982</td>
<td>Pain behaviors described and UAB Pain behavior scale (Keefe and Block, Richards)</td>
</tr>
<tr>
<td>1983</td>
<td>Sensory Decision Theory (Clark and Yang)</td>
</tr>
<tr>
<td>1983</td>
<td>Ethical guidelines for the investigation of pain in conscious animals (Zimmerman)</td>
</tr>
<tr>
<td>1987</td>
<td>Metanalysis in clinical research (L’Abbé)</td>
</tr>
<tr>
<td>1996</td>
<td>RCT quality assessment for inclusion in systematic reviews of pain therapy (Jadad, McQuay, Moore)</td>
</tr>
<tr>
<td>1996</td>
<td>TOTPAR, SPID and NNT (McQuay)</td>
</tr>
<tr>
<td>1999</td>
<td>Pain—the fifth vital sign (Joel)</td>
</tr>
</tbody>
</table>

Using this approach, it became possible to calculate the pain interval or pain difference as the gap between the pain threshold and the maximum tolerance. The Dol scale emerged as a graphic representation of this. Crawford Clark at the New York State Psychiatric Institute in the late 1960s developed Sensory Decision Theory, a mathematical model for separating sensory and emotional components of pain to measure the ability of subjects to discriminate between high and low intensity stimuli. He distinguished two parameters in individuals—discriminability and response bias. Thus the effects of placebos, analgesics, or psychotropics on components of the response to pain could be described.3

Berthold Wolff, a psychologist born in Germany, who came to work in New York after World War II, categorized methods of providing the painful stimulus as threefold: cutaneous, deep somatic, and visceral.4 To the contemporary reader, the devices whereby these methods were executed sound rather fearsome and include the pressure algometer, tooth pulp electrical stimulation, the canthadrin blister, and the weighted

### Say Stop When It Hurts

The science of psychophysics was the principal paradigm in experimental psychology in the nineteenth century. Its founder, the German Gustav Fechner, described the relationship between perceived intensity of a stimulus and its physical intensity. Working within this tradition at Cornell in the 1930s, James Hardy, Helen Gleadall, and Harold Wolff refined a methodology that recognized two discrete components of pain. In this model, components of “sensation” and “reaction” were separately recorded by determining: a) the intensity of a stimulus which will be reported as painful (the pain threshold); b) the smallest change in stimulus intensity which is detected as a change in the pain (the just noticeable difference or JND); and c) the intensity of stimulus which is the point of maximum tolerance.2
Lucite knife. In the early 1950s, Harold Wolff and colleagues described eight requirements for a stimulus suitable for evoking pain:  

- A measurable aspect associated with changes causing pain  
- An identified reproducible pain threshold  
- Gradation to a degree finer than the JND  
- An association with differences in pain throughout the range of intensity  
- Minimal tissue damage  
- Capable of evoking a separate quality of pain, such as burning or pricking  
- Convenience of application  
- Clear-cut identification of pain  

Four further prerequisites were added by Beecher in 1959:  

- Possibility of repeated stimulation above pain threshold without interference with subsequent measures  
- Sensitivity to the effect of weak analgesics  
- Ability to differentiate between analgesic doses  
- Applicability to humans and animals  

Although the effects of distraction, stress, and drugs on psycho-physical parameters could be investigated in the laboratory, the context of clinical pain with the demands of physical activity and the stress of a life threatening illness gave rise to questions about the generalizability of psychophysics and its validity when evaluating analgesics for clinical use. Indeed, the distinction between “laboratory” and “clinical” pain had been the subject of debate since the early 1950s. In time, key failures to detect analgesic effects of widely used drugs when compared with placebo were to inflict fatal damage on psychophysical methodology as the final arbiter of clinical practice  

Since 1983, the use of animals in experiments which induce pain and where it is a requirement that the animal is conscious has been regulated by specific guidelines from the International Association for the Study of Pain. The guidelines are aimed at limiting animal studies to those that offer a likely benefit to humans and at reducing the duration and intensity of pain necessary to conduct the experiment. A commonly used method in current pain research, the von Frey hair, uses a tactile stimulus to invoke pain and is derived from some of the earliest psychophysical experiments.
Over thirty years later, Ronald Melzack at McGill University in Montreal, working with Warren Torgerson, Professor of Psychology at Johns Hopkins, used panels of students, patients, and doctors to identify three separate dimensions of pain: sensory, affective, and evaluative. From these were constructed 16 subclasses, 10 sensory and 5 affective (Table 2). The words within each subclass were ranked and then rated independently according to intensity by separate panels of students, doctors, and patients. The work showed that there was agreement on the category of the words in terms of properties of pain. In spite of the different backgrounds of panel members, they agreed on certain words which were then placed on an intensity scale. This work was to be the basis of the McGill Pain Scale.

Since the 1970s, the number of pain scales has proliferated, some with specific pain applications and others attempting to capture more thoroughly a particular theoretical dimension of pain. Gracely and Dubner in a 1981 report proposed five properties of an ideal verbal pain measure:

1) Sensitive measurement, free of biases inherent in different methods.
2) Provision of immediate information about the accuracy and reliability of the subjects’ performance in the task.
3) Separation of the sensory-descriptive aspects of the pain experience from its hedonic qualities.
4) Usefulness for clinical as well as experimental pain measurement, allowing reliable comparisons between these fundamentally different types of pain.
5) Absolute measures that increase the validity of pain comparisons between groups and within groups over time.

A widely used example is the Brief Pain Inventory (BPI), a multidimensional pain assessment tool which records present, worst, and least pain intensity as well as location on the body, character, effect on mood, function, and enjoyment. It has been translated into many languages and along with the McGill short form was recommended by the European Association for Palliative Care expert working party for studies defining pain syndromes, prevalence studies, and trajectory.

| 1 | Flickering | 6 | Tugging | 13 | Fearful |
| 2 | Jumping | 7 | Throbbing | 14 | Hot |
| 3 | Pricking | 8 | Flushing | 15 | Itch |
| 4 | Sharp | 9 | Boring | 16 | Stinging |
| 5 | Pinching | 10 | Cutting | 17 | Aching |
| 6 | Quivering | 11 | Pulsing | 18 | Pulling |
| 7 | Throbbing | 12 | Pounding | 19 | Wrenching |
| 8 | Jumping | 13 | Beating | 20 | Scalding |
| 9 | Flushing | 14 | Thrusting | 21 | Burning |
| 10 | Sharp | 15 | Stabbing | 22 | Burning |
| 11 | Pinching | 16 | Lancinating | 23 | Brief |
| 12 | Pressing | 17 | Cramping | 24 | Momentary |
| 13 | Gnawing | 18 | Crushing | 25 | Transient |
| 14 | Gagging | 19 | Squeezing | 26 | Rhythmic |
| 15 | Piercing | 20 | Percenting | 27 | Periodic |
| 16 | Cramping | 21 | Excruciating | 28 | Intermittent |
| 17 | Exhausting | 22 | Continuous | 29 | Steady |
| 18 | Squeezing | 23 | Constant | 30 | Intermittent |
studies where pain is described repeatedly over the course of a disease.\textsuperscript{20}

**Mild, Moderate, or Severe?**

Away from the laboratory, Kenneth Keele, a practicing cardiologist and medical historian with an interest in the history of pain, developed different methods to evaluate the intensity of pain arising in patients suffering because of disease or surgery. In this framework, the subjects were asked to categorize their present pain as “0” none, “1” mild, “2” moderate, or “3” severe and a categorical verbal rating scale for pain relief was devised of: “0” none, “1” slight, “2” moderate, “3” good, “4” complete.\textsuperscript{21}

Also, in the late 1940s, Hewer working with Keele demonstrated that by using these categorical responses to enquiries at regular intervals or by asking patients to record the pain score on a daily pain chart, a graphic representation of pain over time could be constructed. These pain charts were used to demonstrate the effect of analgesic administration on pain in an individual patient in studies where changes were recorded at designated intervals after dosing and the differences were subjected to statistical tests of significance when compared to a placebo.\textsuperscript{21}

Michael Bond, Issy Pilowsky, and Graham Spear, psychiatrists working in Erwin Stengel’s Sheffield group in the UK in the 1960s, adapted a non-verbal measure of other psychological phenomena. Continuous data of pain intensity unrelated to verbal description were generated by a new technology that came to be known as the visual analogue scale (VAS); a 10-cm line labeled “no pain” at one end and “the pain is as much as I can bear” at the other. Here, patients were given an explanation of the line and asked to mark a point upon it which corresponded to their pain.\textsuperscript{23} A variation published in 1974\textsuperscript{24} used the labels “no pain” and “severe pain” and here, severity is expressed as the distance between the “no pain” end of the line and the mark. Recording of pain intensity is uniform over the range of stimuli in an unselected population and the method correlates well with other categorical or numerical tools. However, the approach did generate a problem concerning pain scores falling around 6.2 cm, because subjects have difficulty in accurately reproducing marks around this point, known as the ‘golden section.’

A third type of unidimensional scale developed within this framework after the verbal rating scale (VRS) and the visual analogue scale (VAS) is the numerical rating scale (NRS). Here, the subject is asked to describe the intensity of pain on a scale of 0–10.\textsuperscript{25,26} All three of these scales are usually administered using pen and paper but the NRS may also be administered verbally. The report of the recommendations of the European Association for Palliative Care (EAPC) expert working group in 2001 produced a valid version of the VRS in several languages and recommended the use of a standard 0–10 NRS and a 100-mm horizontal VAS while pointing to the poorer compliance associated with the VAS.\textsuperscript{20}

Awareness of the inadequate assessment of pain in busy clinical units, together with the availability of simple measures of pain intensity, led in the 1990s to numeric rating scales being promoted in the clinical context.\textsuperscript{27} This numeric quality has made them particularly useful clinically, in the same way that temperature, as the most measurable quality of fever, was charted by Carl Wunderlich in the nineteenth century.

In the late 1990s, pain came to be termed ‘the fifth vital sign’ in the nursing literature.\textsuperscript{28} Although pain is neither vital nor a sign, its routine recording after temperature, pulse, blood pressure, and respiration was a powerful reminder of clinicians’ duty to attend to the suffering of patients in their care. The regular use of numeric pain rating scales by nurses assessing pain in the clinical context thus became routine in many care settings. This development appears to reinforce the validity of the methods adopted by clinical researchers when evaluating analgesics. Pain, recorded as a ‘vital sign’ that according to patients’ reports varies in severity over time, puts the phenomenon on a par with other objective clinical parameters.

These simple technologies have allowed a detailed and prolonged recording of clinical pain in patients undergoing randomized controlled trials of analgesic agents. The principles of the methodology were set out by Beecher as early as 1952 and are shown in Table 3.\textsuperscript{27}

This strand of investigation was continued by Raymond Houde at the Memorial Sloan-Kettering Hospital, New York, in studies involving nurse observers such as Ada Rogers, who interviewed patients about their pain at hourly intervals. Trials of analgesics were conducted double-blind and the placebo control was allocated randomly in a cross-over design.\textsuperscript{29} In these studies, “placebo
reactors” were deemed to be patients who responded favorably to a saline injection. They formed the majority in some series, and hence methods were discussed for analyzing data on intensity and relief of pain following either analgesic drug or placebo. These included the Average Pain Index (API), which represented the intensity of pain under study; the Sum Pain Intensity Difference (SPID), a measure of change following administration of drugs; and the Total Pain Relief (TOTPAR), which is a summary of relief of pain following an intervention. This last development was essential before data from different studies could be combined and treatments compared.

The establishment of a consensus on pain measurement tools recording intensity and relief enabled data to be combined from many trials using techniques of meta-analysis. However, for data to be truly comparable, the trial design must be standardized and only included in a meta-analysis or systematic review if it passes muster. Since the mid 1990s, this technology has set a benchmark for the quality of randomized controlled trials and established an orthodoxy for pain measurement, trial protocols, and data analysis that has come to dominate the field.

Once a method for deriving dichotomous data from trials of analgesic drugs had been established, then the results from many trials could be combined and compared in order to describe the numbers needed to treat (NNT) and the numbers needed to harm (NNH) for various therapies. The information generated by such means answers the physicians’ dilemma—which of the available treatments is most likely to offer relief, which is least likely to inflict harm, and which will harm. The NNT has in turn enabled comparison between treatments to be made in the absence of RCTs directly comparing one with the other. As RCTs become more expensive and difficult to fund, the NNT can be used as a tool to inform policy makers and researchers of worthwhile trials that might establish new drugs in clinical practice.

### Table 3

**Beecher’s Principles of Clinical Trials of Analgesics**

- a) Subjective responses are the resultant of the action of the original stimulus and the psychic modification of that stimulus.
- b) Man is the essential experimental subject for a definitive answer to questions in this field, and men are easier to work with than women, for with men the controls are simpler.
- c) The investigating staff is constant during any given series of experiments.
- d) The “unknowns” technique is employed throughout. The agents tested and the time they are tested are unknown not only to the subjects but to the observers as well. This requires the use of placebos, also as unknowns.
- e) When a new agent is to be compared with the agents of past experience, and this is nearly always the case, a standard of reference is required (morphine in standardized dosage is used as the standard for analgesics, etc.).
- f) Randomization of new agent, placebo, and a standard of reference is essential.
- g) Significant comparisons of side actions of agents can be made only on the basis of doses of equal strength in terms of their primary therapeutic effect.
- h) Mathematical validation of supposed difference in effectiveness of the two agents is necessary
- i) The subjective effects of drugs can be qualified accurately and rapidly only when placebo reactors are screened out.

### Conclusion

In the period since World War II, the search for valid methods for the evaluation of analgesic therapies has fostered at least three strands of scientific effort. The first, known as psychophysics, measured the effect of analgesia by quantifying the noxious stimulation which makes the subject aware of pain, as well as the maximum stimulation which the subject was able to tolerate. This fundamental principle is used in the testing of potential analgesic compounds on animals, as well as in animal experiments researching pain mechanisms. New analgesic compounds may be identified at an early stage in drug development and new drug targets may be identified using animal models of pain.

In the second approach, the description of various aspects or qualities of pain using standardized words on validated questionnaires was developed for use by clinicians seeking to categorize pain. This is a clinical method for identifying mechanisms by which pain is generated in individual patients. The methodology has little value in assessing analgesics where pain intensity is the crucial dimension but gives some insight into the relative usefulness of different classes of analgesics in different pain states.

The practice of asking patients to report on pain intensity using verbal rating scales and visual analogue scales was considered valid in the report of the expert working group of the European
Association for Palliative Care. This third methodology allows continuous measurements while analgesic or placebo are administered and produces data that allow analgesics to be compared with each other across separate trials. Evidence of efficacy based on metanalysis of RCTs now forms the basis of national guidelines and clinical protocols for use by clinicians involved in cancer care.

The simple technology of the numeric rating scale is congruent with clinical practice to the extent that it has been adapted by health professionals as a guide to the treatment of individual patients. Although pain is a complex, subjective phenomenon, researchers and clinicians are, at present, in agreement on how to measure the response to analgesic therapy. In the same way that temperature, a measurable component of fever is charted, the subjective report of pain by a patient is to be found recorded like other objective parameters, at the foot of the bed. Although all three strands of pain measurement have found a place in modern clinical practice or drug development, it is the reporting of pain by patients undergoing treatment using simple scales of intensity that provides the crucial information by which analgesic therapies can be evaluated and compared.

Acknowledgments

Research for this article was funded by a grant to study “Innovations in Cancer Pain Relief: Technologies, Ethics and Practices,” awarded in the United Kingdom as part of the Economic and Social Research Council/Medical Research Council Program on Innovative Health Technologies (grant L218252055).

References


