THE CLINICAL ASSESSMENT OF ANALGESIC DRUGS

D.W. LITTLEJOHNS & D.W. VERE

Department of Pharmacology and Therapeutics,
The London Hospital Medical College,
London E1 2AD

To strike a proper balance between the efficacy and the adverse effects of an analgesic, it is obviously necessary but notoriously difficult to weigh those two factors individually. Any study of analgesic efficacy in man requires a measure of the subjective variable pain, or pain relief. The difficulty of forming even a satisfactory definition of pain (Fordyce, 1978) should warn of the conceptual problems that can lead to measures lacking the intended validity. Early studies in healthy volunteers are often needed, and require some means of inducing pain as well, if analgesic efficacy is to be demonstrated. The ethical problems involved in these, and in the essential studies of analgesic effect in clinical pain, are perhaps more dramatic than complex. Biometric problems can trap the unwary, even when relevant measures have been employed. So it is fortunate that at least the prediction, detection and evaluation of the adverse effects of analgesics, and their pharmacokinetics, are beset with only a few difficulties peculiar to this group of drugs.

What do you mean by 'pain'?

Methods of pain measurement began to be developed in a form suitable for the assessment of analgesics about 30 years ago, and the work of that decade is still a useful introduction to the subject (Hewer, Keele, Keele & Nathan, 1949; Hardy, Wolff & Goodell, 1953; Beecher, 1957; Lasagna, 1960). More recent reviews incorporate more work comparing one method with another (Huskisson, 1974b), or provide a higher level of psychometric sophistication (Wolff, 1978).

How pain is measured depends on how it is regarded. A scheme is presented in Figure 1 which employs a terminology adapted from Loeser (1980). It can be regarded as compatible with, but as being a simplified version of, the map of pain experience offered by Melzack & Dennis (1978), which was derived from somewhat speculative correlations between neurophysiological mechanisms of pain sensation and psychological observations on the dimensions of pain experience. The fine detail of neurophysiology (Nathan, 1977; Melzack & Dennis, 1978), valuable to mode of action studies (e.g. Yaksh & Rudy, 1978), cannot be resolved by the coarse methods of investigation available in man. On the other hand, further simplification would obscure differences between the information provided by different modes of measurement, whereas the specificity of the measure, for instance, is always of particular concern when dealing with subjective variables. It is of interest that the plan employed here resembles that used similarly by Hindmarch (1980) in this series to
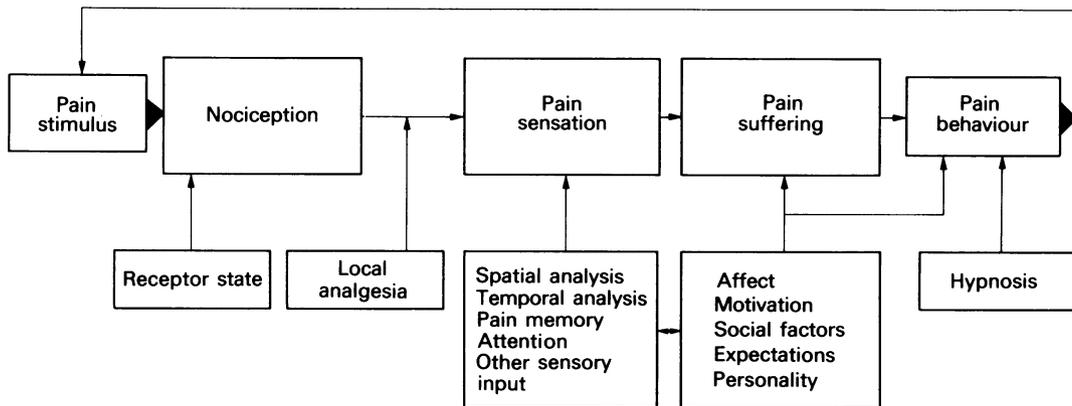

Figure 1 Relationships between the events in pain perception.

order the relationships between tests of psychomotor function, though the two were independently conceived and are not expressed alike.

Though it is seldom obvious that an exact choice has been consciously made, it is possible to choose to measure the stimulus, or nociception, or pain sensation, or pain suffering, or pain behaviour, as seems most appropriate. The use of more than one measure can help in the interpretation of results, especially if those measures have a different specificity within the above range. Those targets, gathered under the generic term 'pain', require to be defined more closely.

(i) The *pain stimulus* is accessible to measurement only when employing induced pain; then, it can often be measured with some precision. For analgesic studies it must be combined with a measure of response, and, though it is attractive to demonstrate analgesic efficacy as a blunting of the stimulus-response relationship, it is not straightforward to design such studies to be sensitive to drug effects.

(ii) *Nociception* is the sequence of processes in the periphery by which noxious stimuli are detected by receptors and converted into impulses carried by small-fibre ($A\delta$ and C) afferent nerves. These processes have so far proved largely inaccessible to measurement in man, though sensory nerve recording, while difficult, has for some time been possible (Valbo & Hagbarth, 1968).

(iii) *Pain sensation* is nociceptive input, further modulated by another series of processes, but this time in the spinal cord and brain, and emerging as a sensory experience. The quest to measure it has made it at times seem the Holy Grail of pain research. While some advocates of Signal Detection Theory have felt they were close (Chapman, 1977), others less committed have expressed fundamental doubts (Rollman, 1977). Another view is that cognitive, evaluative processes can precede and influence sensory perception (Melzack & Dennis, 1978); it would follow that attempts to separate completely the sensory and judgmental aspects of pain is more like the search for the end of the rainbow. Nevertheless, any method of measurement that diminishes the judgmental and affective ('attitudinal') elements would have obvious value in the assessment of analgesics.

(iv) *Pain suffering* is close to the pragmatic clinical concept of pain as being a subjective experience, defined by the patient rather than the doctor. It comprises not only pain sensation, but also the associated distress, the essential unpleasantness of pain, and the affective response—typically anxiety, anger and/or depression. The measures of pain most widely used reflect chiefly pain suffering. To what extent they represent pain sensation is unknowable, but probably varies among both individuals and methods. Distress from any cause other than pain is apt to intrude into all measures of pain suffering, rendering them less specific than intended.

(v) *Pain behaviour* has less frequently been measured. Like stimulus recording, it is susceptible to very precise and objective measurement but is remote from what matters to the patient or subject. Spontaneous verbal responses have been much studied, although particularly affected by social, cultural and environmental factors (Craig, 1978). It is possible to measure the purely physiological responses to pain; there is reason to suppose that they are linked to pain suffering rather than to pain sensation.

Table 1 Differences between acute and chronic pain

	Acute pain	*Chronic pain*
Conduction pathways	Rapid	Slow
Tissue injury	Clearly causal	Minor or absent
Autonomic response	Present	Absent
Biological value	High	Low
Mood	Anxiety	Depression, anxiety
Social effects	Slight	Marked
Effective treatment	Analgesics	Variable, sometimes none

The *duration* of pain influences analgesic effects. There are substantial differences between acute and chronic pain (Table 1), though the pains due to advancing malignancy may persist for many months and still have predominantly the characteristics of acute pain. Moreover, it is useful to subdivide acute pain into *first (phasic) pain* (carried by $A\delta$ fibres) and *second (tonic) pain* carried by C fibres. Nitrous oxide relieves both, but morphine chiefly the second (Melzack & Dennis, 1978), and most clinical pain is in this second category (Mumford & Bowsher, 1976).

The *intensity* of pain may also affect the results of analgesic assessment. The low pain levels most acceptable in experimentally-induced pain are one factor in making clinical pain more susceptible to analgesics than experimental pain, and the pain tolerance level more sensitive to analgesics (Smith, Lowenstein, Hubbard & Beecher, 1968; Wolff, 1977) than pain threshold. Even the placebo effects of medication are more marked at high pain intensities (Levine, Gordon, Bornstein & Fields, 1979).

If the generic term 'pain' thus requires a more specific definition, so does the generic term 'analgesic'. For instance, should an antidepressive drug that diminishes chronic pain by alleviating the depression associated with it be regarded as having analgesic properties? It is clear that relief of pain suffering makes a drug clinically useful; relief of pain sensation implies a mechanism of benefit. A drug that influenced only pain behaviour, without relieving pain suffering, might be useful, but perhaps it should not be regarded as an analgesic; there is some evidence that the relief of pain by hypnosis falls into this category (Hilgard, 1978).

Pain stimulus

The physical measurement of the pain stimulus seldom causes problems, but the stimulus is inevitably somewhat remote from the processed and modulated phenomena of pain suffering which concern the patient. Indeed, experimentally-induced pain is likely to be registered more as pain sensation, compared to the more threatening, and often more severe, pain due to disease, and it is often of the 'first pain' type. These reasons presumably underlie the difficulty of demonstrating the effects of even well-known analgesics under experimental conditions. Within these limitations, the criteria for an appropriate stimulus have been discussed for many years (Hardy et al., 1953; Beecher, 1959; Wolff, 1978). It is widely agreed that any stimulus should have these properties:

(i) it should be convenient to apply
(ii) it should be recognised as painful, whatever other sensations it produces, and the pain should be diminished by known analgesics.
(iii) the measurable content of the stimulus should be closely associated with the changes causing pain.
(iv) its intensity should be measurable physically with greater precision than sensory discrimination can achieve.
(v) its intensity should range from providing threshold to maximum tolerable pain.
(vi) it should provide a reproducible measure of pain threshold under standard conditions.
(vii) repetitions even above the threshold level should not affect subsequent determinations, which implies the avoidance of tissue damage.

Of these criteria (v) and (vii) are particularly difficult to meet, but four methods of pain induction come close enough to the ideal to be widely used.

Electrical stimulation

The sensation produced is more discomfort than pain, and the associated sensations are unlike those to any other stimulus. Healthy tooth pulp is widely used, but fingers, hand, forearm and deep stimulation to the gluteus maximus have all been used. The stimulator should be given a constant current, since contact resistance varies and the stimulus is related more to current than voltage. The subject must be electrically isolated for safety. Responses are highly repeatable (Wolff, 1978) but the validity of the sensation is open to question.

Radiant heat

The original paper describing this method is well known (Hardy, Wolff & Goodell, 1940) and most instruments have continued to use the principle of controlled heat and light focussed from a projector bulb onto any non-hairy, blackened skin surface. The flexor surface of the forearm, the forehead or the back are suitable sites. Heart intensity is measured as mcal s^{-1} cm^{-2}, and is controlled by a variable resistance. The subject can indicate the time to pain threshold and to maximum pain tolerance by pressing buttons, the button for tolerance terminating the stimulus by closing a shutter. Above about 400 mcal s^{-1} for 3 s tissue damage is likely, being shown by subsequent erythema (Hardy, Wolff & Goodell, 1947). Responses are acceptably repeatable.

Pressure algometer

A spring-loaded plunger of standard size and shape is pressed at a constant rate of increase in force onto the skin against a bony surface e.g. forehead, knuckle, medial malleolus (Keele, 1954). The stimulus measured is the force (as the distance the spring is depressed) that has to be applied for pain threshold and tolerance. The method provides acceptably repeatable results for both variables (Merskey & Spear, 1964).

Cold water pain

The skin temperature of one hand is standardised by putting it into water at a given temperature (e.g. 37°C) for a few minutes; then it is plunged into a stirred bath of ice fragments in water. The subject indicates the onset of pain and the end of tolerance; these times are the measurements. The method is, of course, also used for studies of cardiovascular responses, and especially the hypertensive response to pain. The stimulus is standard but not physically quantified. It cannot be rapidly repeated, but the pain appears to resemble that due to disease, and it is a sensitive test of analgesic activity.

Other methods

A variety of other methods has been employed, but none has been so generally successful. The slow onset of pain at four different intensity levels after standard arm exercise while the arterial supply is occluded has been used to demonstrate the effects of analgesics (Smith, Egbert, Marcowitz, Mosteller & Beecher, 1966; Smith et al., 1968) but few investigators have used it and it has not been proved to be repeatable. The same is true of inflating a sphygmomanometer cuff with internal projections (Poser, 1962). The application of chemicals to a blister produced by cantharidin (Armstrong, Dry, Keele & Monkham, 1953) is more a technique for studying peripheral mechanisms of nociception, but could be used to test analgesics. It can also be used to test local anaesthetics (Mongar, 1955).

Pain sensation

Two approaches to the measurement of pain sensation are possible. One is to measure both pain suffering itself and those factors (see Figure 1) which modify pain sensation to result in pain suffering, and subtract their influence; the other is to measure pain sensation directly. The former method would involve non-pain measurements almost as controversial as those of pain itself, with no way of predicting accurately in the individual case how powerfully these entities influence pain suffering. However, it is advantageous to make at least some measurements of relevant personality and affective factors when measuring pain suffering; marked differences in them might illuminate the pain measurements. For example, lessened anxiety can reduce pain suffering more than aspirin does (von Graffenried, Adler, Abt, Nuesch & Spiegel, 1978).

The only method that might seem to offer a measure of pain sensation is Signal Detection Theory (SDT). It originated in the statistics of hypothesis testing, and was developed by radar engineers interested in the detection of a weak signal against background noise. Lucid descriptions of an admittedly complex subject by Swets (1973) introduced the concepts to sensory psychology, and a helpful introduction to its use in pain studies has been written by Clark (1974). It is sometimes known as sensory decision theory.

The basic assumptions of SDT are that:

(i) the sensory experience to a standard stimulus is not fixed but variable, with a median value and some dispersion, and so can be represented best by a distribution curve (Figure 2).
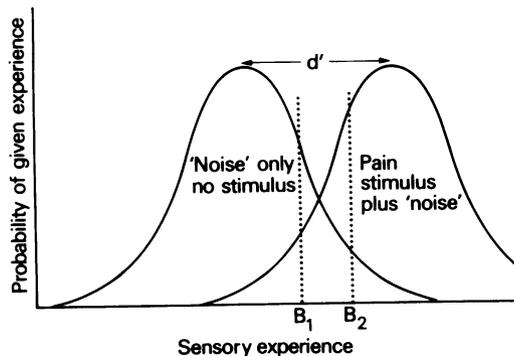

Figure 2 The variable sensory experience in the absence of a stimulus is regarded as 'noise'. The discriminability (d') between a repeated mild pain stimulus and this noise depends on how large the difference in pain sensation is. For a near-threshold stimulus, overlap of experience means that reports of 'pain' and 'no pain' do not perfectly match 'stimulus' or 'no stimulus'. Different settings of response bias (B_1 and B_2) alter the proportion of false negative and false positive reports.

(ii) increasing stimuli give distribution curves with a displaced median, but all share the same dispersion; the discriminability of two stimuli then depends on the difference in the two medians (d' in Figure 2).

(iii) there is at any time a particular value of sensory experience which corresponds to a person's report of a threshold (e.g. between 'no pain' and 'pain', or between 'tolerable pain' and 'intolerable pain'). This is referred to as response bias (B_1 and B_2 in Figure 2), or criterion, and it is influenced by cultural factors, mental attitude, expectancy of pain etc. For instance, in Figure 2 any value of B will involve some 'false positive' and/or 'false negative' reports of pain, but B_2 represents greater stoicism than B_1—fewer false positives, at the cost of more false negatives.

The method of SDT is to present pairs of stimuli repeatedly, record the reported responses, and analyse to give values for discriminability and response bias. It is attractive because it seems that all the attitudinal influences are gathered together as 'response bias'. The calculations may assume normal distributions, or use non-parametric methods, and a variety of procedures is possible (Swets, 1973; Chapman, 1978).

On the whole, it does seem that analgesics change discriminability whereas psychological modifiers and psychotropic drugs change response bias (Wolff, 1978). But nitrous oxide changes both measures (Chapman, Murphy & Butler, 1973), as do morphine and diazepam, though diazepam has more effect on response bias and morphine more on discriminability (Yang, Clark, Ngai, Berkowitz & Spector, 1979). Though these methods are obviously useful, the criticisms of the validity of additional assumptions that it is easy to make from SDT Theory are a necessary warning (Rollman, 1977). For instance, to measure discriminability is to measure something related to pain sensation, but it does not measure pain sensation itself. It would be an error to define analgesics as substances that reduced the discriminability of pain stimuli. It is easy to commit this error, because as most stimuli increase, so does the difference between just-discriminable stimuli, as measured in linear physical units; logarithmic units such as decibels are the result of the logarithmic Weber-Fechner Law. But the calculations are upset by adaptation (the sensory equivalent to tachyphylaxis) as shown by Rollman (1979), and a drug might blunt the discriminatory ability of the CNS without altering sensation.

Pain suffering

When measured as a response to a pain stimulus, two points on the intensity scale are fairly clearly fixed— the pain threshold and pain tolerance. The former contains the greater amount of information about pain sensation, but the latter is more valuable in

analgesic studies. The measurement of clinical pain is more problematical, but by no means impossible.

Pain threshold

This is usually determined by the method of limits, though other methods are possible. Starting with very weak stimuli, stimulus strength is gradually increased in successive trials until pain is reported; then a stronger stimulus is applied, followed by a descending series, noting when pain is no longer felt. The threshold of the ascending and the descending series is averaged. The reliability of the result depends on the subject, the method of inducing pain (Lynn & Perl, 1977), the interval between measurements, and the number of pairs of trials—e.g. five pairs of trials are sufficient for work with radiant heat, and the first value is best disregarded as it shows the greatest variance about the same mean.

Pain threshold measurements impose least discomfort on the experimental subject, and thresholds have been extensively studied, with divergent results (Beecher, 1959; Woodforde & Merskey, 1972). Keele (1968) found the pain of myocardial infarction related to pain threshold, but Huskisson & Hart (1972) using the same method found it accounted for only about 5% of the variance in analgesic consumption in patients with rheumatoid arthritis, and this may have been due to pain threshold correlating with disease severity. Thus it does not seem to be a major determinant of clinical pain, and though it can be used to detect the effects of analgesics, it is not particularly sensitive to them (Smith *et al.*, 1968; Wolff, Kantor, Jarvick & Laska, 1969; Wolff, 1977).

Pain tolerance

Pain tolerance is the point at which the subject feels obliged to end the pain stimulus. It is more easily modulated by psychological factors than is pain threshold, and so it has a rather higher variance, but is more sensitive to analgesic effects and is therefore an important measure in the study of analgesics, despite the risk of tissue damage and the discomfort to the subject. Obviously, to use it is ethical only if the subject can terminate the stimulus.

Wolff (1977) has used an intermediate point on the scale, the *drug request point*—e.g. the pain level at which the subject would normally take an over-the-counter analgesic if the pain occurred spontaneously. This measure is also sensitive to analgesic effects, but the sensitivity may depend on how close it lies to pain tolerance.

Ordinal pain scales

The pain produced by radiant heat applied to the forehead or forearm can be divided into a series of 21 'just noticeable differences' (JNDs) ranging between threshold pain and the maximum pain possible with this technique (Hardy *et al.*, 1947). Since such precise measurement was possible only under ideal conditions, two JNDs were considered to be the practical pain unit, called a 'dol'. However, these observations have not been validated for other pain stimuli, and the 10-point scale of pain has not been widely used.

In measurements of clinical pain, a 5-category ordinal scale is well understood and easily used (Keele, 1948), though it is effectively a 4-point scale because the highest value is seldom used. We have found the words 'no pain/slight pain/moderate pain/severe pain/excruciating pain' are seldom misunderstood. It is not difficult to apply such scales to the assessment of analgesics, an observer asking the patient to rate the pain at regular intervals (Hewer *et al.*, 1949; Beecher, 1957; Wallenstein, Heidrich, Kaiko & Houde, 1980; Dundee, 1980). It may be useful to represent the scale as a series of lights (Nayman, 1979). Scales any more finely divided are difficult to provide with suitable words to each category, but Downie, Leatham, Rhind, Wright, Branco & Anderson (1978a) employed a 'numerical rating scale' of adjacent boxes labelled from 1 to 10 and showed advantages from using this. The novel numerical scale may have attracted more interest and explanation from the experimenters, but would be of interest if the results can be repeated by others. A similar technique is to use a numerical scale with zero as 'no pain', 10 as pain tolerance, but open to higher scores, with no top limit (Hilgard, 1978).

Visual analogue scale (VAS)

Visual analogue scales were introduced into medical use for the measurement of feelings (Aitken, 1969; Aitken & Zealley, 1970). High reliability and validity had already been shown in other contexts, and Aitken added other experiments on validity, using loaded breathing; not all have agreed about validity (Downie, Leatham, Rhind, Pickup & Wright, 1978b). Aitken pointed out the advantages of greater sensitivity of such scales compared to verbal category scales, and a more recent opinion (Nicholson, 1978) is that VAS methods are valuable, though it is not safe to rely on them when used alone. Validity for pain is impossible to estimate, but there is good agreement between ordinal pain scales and visual analogue scales both for spontaneous pain changes and those due to analgesics, but the VAS is more sensitive (Downie *et al.*, 1978a; Ohnhaus & Adler, 1975; Joyce, Zutshi, Hrubes & Mason, 1975; Woodforde & Merskey, 1972; Scott & Huskisson, 1976). A 100 mm scale with the ends marked, for example, 'No pain/Pain as bad as it could be' is now widely used, and is usually measured to the nearest millimetre (Huskisson, 1974b). Some authors have found it difficult to use, but we and many others

have had few difficulties even with patients of low intelligence. A standard introduction is useful—e.g. 'This is a way of showing me how bad your pain is, without having to use words. A mark here (indicates 'No pain' end) means you are completely free from pain. Marks along the line (moves pointer along it) would mean gradually worse pain, until at this end (indicates other end) is the most extreme pain imaginable. Please, would you put a stroke through the line at the place that shows what your pain is like now.' Subsequent use requires no instruction.

The visual analogue scale has been much studied. Words set beside the scale merely attract a high density of responses close to them (Scott & Huskisson, 1976), and a 50 mm line is less discriminatory than 100 mm (Revill, Robinson, Rosen & Hogg, 1976); either change reduces the sensitivity of the measure. There is no appreciable difference between a scale set vertically and one set horizontally (Downie et al., 1978a). It is usual to prevent the subject being able to see previous pain ratings, which users may prefer, but this precaution is unnecessary or perhaps deleterious (Joyce et al., 1975). Memory for a random mark is good even 24 hours later, but consistency in rating a previous pain is better still, and is unaffected by pethidine 150 mg intramuscularly (Revill et al., 1976).

Verbal responses

Spontaneous verbal responses are better classified as pain behaviour, but in measuring pain it is usual for structured responses to be elicited for that purpose alone. The only method widely used is the McGill Pain Questionnaire (Melzack, 1975), which takes some time to administer, and is unsuitable for frequent use in single-dose analgesic studies; it was designed for the evaluation of clinical pain, and is concerned with many factors other than pain intensity. The subject is asked to mark any of 102 adjectives which have been found to be reliable in their meaning independent of age, culture etc. in somewhat limited validation studies (Melzack & Torgerson, 1971). As many are marked as seems appropriate to the subject.

The number of words marked provides one index of response, and a pain rating index calculated from numerical values given to each word provides another. A 5-point ordinal pain scale is also included. Unfortunately, the words employed were originally chosen by the investigators rather than being selected from those used by patients, and some (e.g. 'lancinating') are not in common use. It has been suggested that the analysis of the questionnaire results could be more appropriately performed using somewhat different factors, derived experimentally rather than rationally (Crockett, Prkachin & Craig, 1977). Verbal reports contain considerable information about the affective and sensory aspects of pain; intensity is by no means as predominant a factor as might

be supposed (Bailey & Davidson, 1976). Although the McGill Questionnaire is thus open to improvement, chosen adjectives for pain do seem to be surprisingly reliable indicators of sensory intensity (Gracely, McGrath & Dubrier, 1978a), and are changed as would be expected by diazepam (Gracely, McGrath & Dubrier, 1978b). As a measure of dental pain it was effective, though the subscales contained rather too much common information (van Buren & Kleinknecht, 1979), and head pain was accurately recalled when using it on occasions up to 5 days apart (Hunter, Philips & Rachman, 1979). The McGill Questionnaire would probably be a useful assessment measure in an analgesic trial carried out over a day or longer.

Stimulus matching

It is possible for the subject to match clinical pain with a second pain due to another, variable stimulus (Kast, 1962), or alternatively to express pain as a proportion or multiple of a standard pain stimulus. The latter method avoids the danger of causing physical injury by a stimulus of an intensity as high as that of clinical pain, but it is not easy to match two pain stimuli that differ in other sensory qualities, and most methods of inducing experimental pain involve equipment too cumbersome to bring to the bedside. Pain intensity can also be expressed in terms of the magnitude of another sensory modality (e.g. brightness, loudness), but the value of this is limited (Woodforde & Merskey, 1972). Armstrong, Dry, Keele & Monkham (1953) used the force of squeezing a sphygmomanometer cuff, and adapted this to graphical recording. The visual analogue scale uses linear distance likewise, and has already been discussed.

Pain behaviour

Pain behaviour can conveniently be regarded as activities consequent on pain, not necessarily under voluntary control, but of a sort not especially introduced for the purposes of measurement. The objec-

Table 2   Physiological responses used as measures of pain

Physiological response	Sample reference
Cortical evoked response	Sitaram et al. (1977)
	Buchsbaum et al. (1977)
Skin conductance	Craig & Neidermayer (1974)
Plasma cortisol	Lascelles et al. (1974)
Urine catecholamines	Huskisson (1974a)
Serum non-esterified fatty acid	Knitza et al. (1979)
Serum β-lipoproteins	Keele & Stern (1973)
Acid-base balance	Evans (1972)
Vital capacity	Parkhouse & Holmes (1973)

tivity of an external observer may indeed reduce the impact of emotional factors (Houde, Wallerstein & Rodgers, 1960) or prevent some patients exaggerating their pain (Parkhouse & Holmes, 1963) but there is no valid way in which an observer can consistently conclude from pain behaviour what the pain sensation may be; the most valid measures of a subjective sensation are subjective responses. In practice, the chief effect is to increase variance, but if observation of analgesic effects by an observer is carried out on a large enough scale, analgesic effects can be detected (Jick, Slone, Shapiro, Lewis & Siskind, 1971).

Physiological responses have the advantage of precise and objective measurement, and the disadvantages of uncertain validity. Some examples are given in Table 2. Many are related to the activity of the autonomic nervous system. These are poor measures of first pain, being extinguished by repetition of the stimulus when the pain is not (Wolff, 1978). They might be more valid in clinical pain, but it is commonly observed that the most easily-measured autonomic responses are increasingly diminished as pain persists longer. Cortical evoked responses are an attractive measure, but magnitude and latency may well reflect arousal, the most likely explanation for the effect of physostigmine (Sitaram, Buchsbaum & Gillin, 1977).

Biometric problems

In any biometric work it is necessary to consider:
(i)   what is measured—its biological identity (validity)
(ii) the measures and the measurers—to minimise bias
(iii) the mathematical properties of the measures, including aspects related to (i)
(iv) sensitivity and repeatability.

Pain measurement is particularly difficult in each respect; despite this striking progress has been made in the field. Pain measurements can now yield monotonic results (i.e. measures move in a uniform direction in relation to one another, and to stimulus intensity, when that is known), the variances of measurements can be strikingly small, and the predictive power of research for clinical practice can be substantial (Wallenstein et al., 1980). Nevertheless, close attention to detail is needed to ensure useful results.
(i) In relation to *what is measured*, there must be careful consideration of:
(a) the intensity levels of the studied pain
(b) the clinical type of pain (Beaver & McMillan, 1980)
(c) the relationship of analgesic blood levels to the onset, peak and duration of drug effect, and
(d) the place of the observed data within the range of possible drug effects (Beaver & McMillan, 1980).

Any experiment in which these features are ill-matched is likely to fail. For example, a measurement of analgesic effects in one type of pain may not predict them for another, variable-dose tests crowded into one end of the range of biological response will fail to shown sensitivity to differences in pain stimuli, tests made with single doses of a cumulative drug will show a surprising lack of effect, and so on. Recurring pain behaves differently from non-recurring (Dundee, 1980) and expectations may depend upon the time of day at which measurements are made (Glynn & Lloyd, 1976). Obviously, variance of the results is minimised by measuring the effect of an analgesic on a similar pain stimulus in each subject. This is most easily achieved when pain is experimentally-induced, but the value of such methods is limited (see above and Figure 3), and so studies of clinical pain are more than just ways of generating paper for regulatory authorities. A number of clinical situations have been investigated in some detail, notably post-partum pain (Bloomfield, Barden & Mitchell, 1976; Sunshine, 1980), dental pain (Cooper & Beaver, 1976; Levine et al., 1979) and cross-over studies in cancer pain (Twycross, 1975). Instances in which pain is likely to be more variable, but which represent important uses of analgesics, are headache (von Graffenried & Nuesch, 1980), rheumatoid and osteoarthritis (Deodhar, Dick, Hodgkinson & Buchanan, 1973) and post-operative pain (Nayman, 1979; Gómez-Jiménez, Franco-Patino, Chargoy-Vera & Olivares-Sosa, 1980).
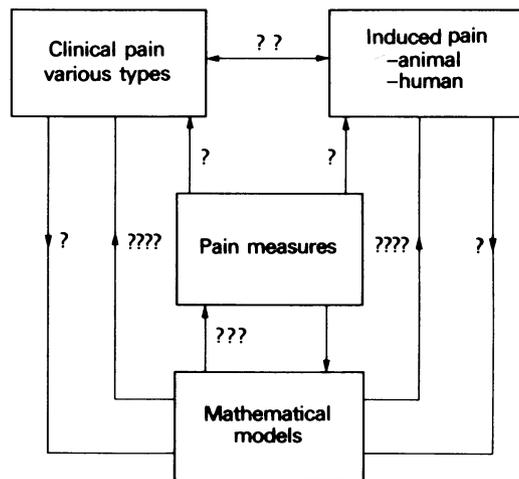

Figure 3 Relationships between components of pain measurement. The arrows represent inferences, and the more question marks associated with each, the more are they open to doubt.

(ii) The *measures* used are awkward for several reasons, not least because the same measure may at times yield data nearly normally-distributed and suitable for parametric analysis, but at other times yield data having a highly skewed distribution (Twycross, 1975). The *measurers* are involved partly because pain is an emotive matter, likely to lead to bias in reporting, but mostly for the reason that in varying degrees pain measurements are made through at least two minds, those of the person in pain and the observer, which can lead to bias compounded by a sort of folie à deux. It is often rightly stressed that double-blind conditions are essential in analgesic trials to avoid this, but the effectiveness of the blindfold should be tested. In Figure 3 the complexity of the measurement problems is stressed by separating out the components of pain measurement, and by indicating by the number of question marks on the arrows between them which inferences are most susceptible to logical errors.

(iii) The *mathematical properties* involved are the problems set by non-normal distributions, by the use of mathematical models, and by the occurrence of time-dependent variation. There is no problem when ordinal data, or VAS scales are analysed using appropriate non-parametric statistical tests, which make no assumptions about the distribution of the measurements, albeit at the sacrifice of some statistical power. For instance, the VAS using a 'No pain/Worst pain imaginable' scale is unipolar, and it is not surprising if the measurements are sometimes crowded towards the 'No pain' end. If pilot trials suggest that only one end of the scale is much used, it may be better to devise another scale on which they are more evenly spread. Certainly, if a distribution is skewed or bimodal (see Figure 4), it would be misleading to
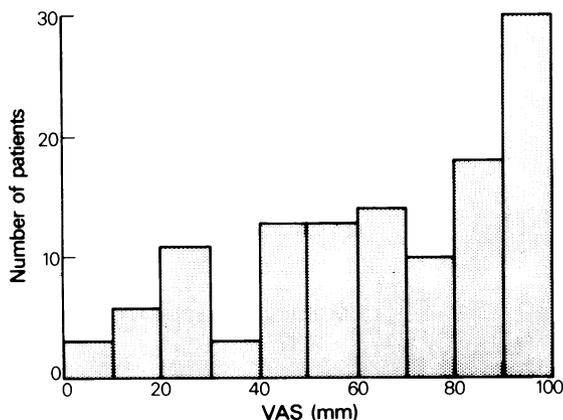
attempt any simple parametric analysis of the data. Bimodal responses may be found because there was ambiguity. Patients then include two kinds of response upon the same scale (e.g. pain and suffering). It may be possible to separate these responses later, but usually it is best to repeat the observations, having changed the scales so as to avoid ambiguity. If the data are simply skewed (there are now tests for skewness and kurtosis in the statistical packages available for many programmable calculators) a transformation (also now an easy thing to do) may be required to render the data suited to parametric analysis. Although Aitken (1969) mentioned an arc-sine transform for VAS data, the choice of transform is quite arbitrary, and the methods were discussed at length by Oldham (1968). The arc-sine transform is described and tabled in the Geigy Tables (Diem & Lentner, 1970). Many workers ignore it as conferring no discriminative advantage upon these methods. In summary, the ways around the problems of non-normality of data distributions are to look for non-normality, and, should it be present, design another scale, or use a distribution-free test statistic, or a normalising transform, as seems appropriate to the problem.

It has been said already that the practice of putting words or marks along the course of a VAS has several measurement effects which are mostly disadvantageous. It is not always realised that these marks also change the nature of the data. Their effect is to convert interval into ordinal data (so losing information) and to change their distribution from a continuous to a stepped form (i.e. moving towards a rating scale). An impressive result of this is that parametric tests are no longer applicable; only non-parametric methods should be used; once more there is then a small sacrifice of power. This may be worthwhile if it improves comprehension. Dundee (1980) gives a figure with several alternative forms of VAS. These may be satisfactory alternative measures, but it must be realised that their results should probably be analysed in different ways.

Model fitting and model forcing

Often experimental data are well fitted by a model; the model used most frequently is a straight line, but a logarithmic or other curve is sometimes better. The logic of using models often seems ill-understood. It is essential to realise that there are two kinds of model, an arbitrary one which just happens to fit the data well (an empirical model), and one with an implied chemical, physiological or pharmacological meaning (a bio-theoretic model). An example would be a log dose-response curve, where a straight line may be found to give a good fit over the central 60% of the curve *either* because it happens to be almost straight, *or* because such a line would be expected from mass-

Figure 4 Distribution of marks made by 123 hospital inpatients with pain on a visual analogue scale for pain intensity.

action receptor kinetics, or both. It happens that the monotonic relationship between visual analogue and ordinal pain scales is curved; curved lines give a significantly better fit than straight lines, and a power function:

$$y = ax^n$$

where y is the VAS, x is the numerical pain rating score and a and n are constants, gives a better fit than a logarithmic curve:

$$y = f (\log x) \text{ (Wallenstein } et\ al., 1980)$$

However, it is important not to infer from what is known or pain in other situations—that it behaves as a power function of stimulus intensity (Tursky, 1975; Gracely *et al.*, 1978a)—to what its behaviour *should* be in numerical or VAS measures of clinical pain. The model may fit well for purely arbitrary mathematical reasons (e.g. a sigmoid log dose-response curve can be fitted equally well by a mass-action law formula, by a hyperbolic tangent, by a probit, or by a cubic curve). Indeed, there might be reasons to expect the curve which relates VAS to rating scores to be sigmoid because the scales are tied at both ends, and there is even a suggestion that this could be so in Figure 7 of Wallenstein *et al.* (1980). What is worse, a sigmoid fitted to such data would be sure to yield a better fit than the power curve; a curve with two bends can always accommodate varied data better than a curve with only one. However, this does not suggest that a sigmoid should be used as a better bio-theoretic model. It is therefore not possible to agree with Wallenstein *et al.* (1980) that 'categorical reports of pain and magnitude estimates of pain are best described in terms of power curve relationship'; it seems wise to suspend judgment on the point until much more is known. The logical point is exposed with full rigour in Gödels Theorem, but is illustrated in less formal ways by many examples of clinical 'reasoning' (Hofstadter, 1979). It is quite safe to use a model simply because it happens to fit the data, and to stop the chain of inference there.

Another major difficulty is *time-dependent variation* or self-modification of recorded pain. The features of a single episode of pain may depend on other pain which precede it and adaptation to a pain stimulus is a well-known psychological observation. Methods exist to study self-dependent variation; they are a particular form of Bayes' logic. When the probability of a given event depends only on that of the event which precedes it immediately, the sequence of events is called a Markov chain (Lindley, 1965). But if within that sequence of events each is independent, the appropriate mathematical model for counting numbers of attacks is that used to describe the likelihood of radioactive decay (Vincent, 1973). These two probabilistic approaches provide mathematical models against which to test biological event patterns to test whether or not they are indistin-

guishable from the independent or Markovian models. We are unaware of any clinical attempts to do this, but the relevance to topics such as pain is obvious. Thus, Sunshine (1980) discusses whether patients should be asked to rate their pain in abstract, or in relation to the severity of the last pain they experienced. The problem also arises in cross-over trials, if inadequate 'wash-out' periods are used, since the second course of treatment is then influenced by the first (Hills & Armitage, 1979).

Cross-over designs are indeed prevalent in pain assessment because, in such a subjective field, it is hoped that each subject is best used at his own control. Sophisticated methods, such as the twin crossover design (Finney, 1964) adapted to pain measurement (Wallenstein & Houde, 1975) can be very useful provided that the power of the analysis of variance (ANOVA) test is computed, and provided tests for sequential effects are made (Laska & Sunshine, 1973). It is not always realised that when the same subjects appear in more than one cell of the ANOVA design, the ordinary computation must be modified to cope with the loss of independence amongst the data. (The situation is a development from the simpler 'related' and 'unrelated' forms of the *t*-test familiar to clinical investigators). A description is given in any standard work on ANOVA, e.g. Dixon & Massey (1969). Moreover, a significant ANOVA test reveals merely an inhomogeneity amongst the means somewhere in that block of data. The site of the inhomogeneity must then be sought amongst all pairs of data sets using *t*-tests; it may not be found to reside amongst the single treatment contrasts at all. Indeed, this seems to have been the basis for the difference between claims of analgesic effects of codeine between Bloomfield *et al.* (1976) and Sunshine (1980). Problems readily arise if biostatistical 'packages' are applied routinely to data without careful attention to their properties, or if statistical advice is obtained without giving a full explanation of the biological problem to the statistician.

(iv) The *sensitivity and reliability* of pain measures can be tested very well provided that positive and negative controls (i.e. a known analgesic and a placebo) are both incorporated in each test, and moreover several doses are used to obtain a graded response. This 'belt and braces' approach is unnecessary to show only that a drug has analgesic efficacy, but is advisable to determine relative potencies (Sunshine, 1980). As we have remarked, it still does not suffice for a drug which cumulates appreciably (Prescott, 1980a).

Often, measures of drug effects are modified to provide derived indices. An example is the S.P.I.D. (Sum of Pain Intensity Differences) (Sunshine, 1980). Such indices may have obvious clinical value in interpreting pain alteration as it applies to the patient, but any index will have different statistical properties

from those of the measures-from which it is derived. For example, the S.P.I.D. is in fact a cusum, and might better be called that, since the term is standard and the statistics of its use well studied (Van Dobben de Bruyn, 1968). If we have:

$$I = \text{initial pain score}$$
$$H_i = \text{hourly pain scores}$$

then S.P.I.D. $= \sum_{o}^{t} |H_i-I|$

The 'variable percentage S.P.I.D.' is also used, which is:

$$\frac{\text{S.P.I.D. found}}{\text{S.P.I.D. max}} \times 100$$

Sunshine (1980) states that analysis of variance was done upon such scores. However, problems readily arise when scores from different individuals are combined (Oldham, 1968); ANOVA is applicable only when data are drawn from subjects randomly chosen from populations having roughly equal variances (Dixon & Massey, 1969); and interactions between data groups radically alter the ways in which ANOVA can be applied.

Now it is important to note the effects of taking indices upon these measures. If someone's initial score is close to the maximum, his pain has more room to improve than someone whose initial score was well below the maximum. The point is illustrated in Figure 5. Who had more relief of pain? What is the appropriate statistical test for such data?

Such problems become extremely difficult to analyse when the variance of the results depends on initial value (one form of heteroscedasticity). Often it is impossible to tell from published methods how indices have been derived, or what their statistical properties might be. It is therefore essential to publish individual and unmodified data in addition to the indices, advice which will be resented by journal editors but which is essential to assess the scientific merit of original work (Oldham, 1968).

Ethics of analgesic studies

This was recently discussed by Dundee (1980). It is essential that the subject should be fully informed. This does not prejudice double-blind technique, for it is necessary to conceal only the timing, not the fact, of placebo use. There must be a voluntary escape mechanism by which a subject can either discontinue an induced pain stimulus, or take 'rescue' medication, as soon as spontaneous pain becomes intolerable, even though this causes difficulty in analysing the results (Lasagna, 1980). Even with this provision, it is still important to design experiments to avoid
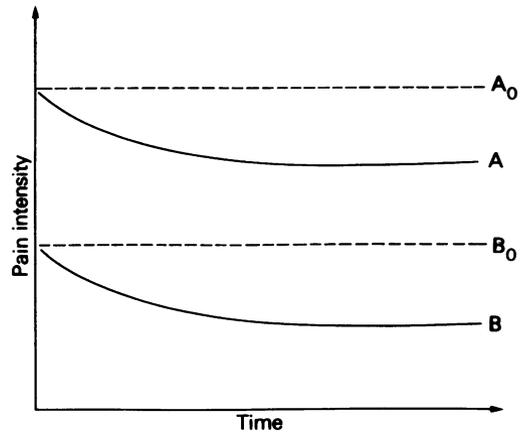
Figure 5 Imaginary pain scores in two subjects, A and B, differing in the initial intensities (A_o and B_o) of their pains. A obtains 25% pain relief, B obtains 50%, but both have the same SPID index.

injury during analgesic treatment. For example, even if a radiant heat stimulus to blackened skin can be turned off by the subject he may not do so before he has been burned if he has taken a strong analgesic.

Individuals may volunteer for pathological reasons; it is difficult to protect someone who volunteers for masochistic reasons or from false heroism. It helps to know personally the group from whom volunteers are chosen and to elicit consent singly and in private, perhaps through a neutral person. People vary greatly in the 'gains' which give them satisfaction and provide their inner drives. These comments have special relevance to tests of narcotic analgesics. Particularly dangerous methods include routine approaches, where everyone who has undergone a certain operation is asked to volunteer, without regard to their personal differences. Substantial financial rewards can elicit a conditioned acceptance response from many people, but pharmaceutical companies commonly find themselves unable to test a promising new analgesic unless they offer high rewards. It seems equally unwise to ignore the motives that might attract experimenters into this field. So much depends on how the experiment is done, and not just on what is done, that Ethics Committees would be wise to expect particularly detailed protocols before giving such work their approval, and to consider who will make the actual tests as well as the person who submitted the protocol.

Adverse effects

Such adverse effects as are predictable by the investigator of a new analgesic will, of course, be indicated

by its preclinical pharmacology and by the pattern of adverse effects of any familiar compounds related to it in chemical structure or mode of action. Some can be investigated in early human pharmacology, but others must await larger-scale investigation. It is easy to measure the histamine-releasing potency of opioids in man by intradermal injection of them in solution, measuring the size of the weal and/or flare, and using classical bioassay methods of calculation (Paton, 1957; Fowle, Hughes & Knight, 1971). The measurement of respiratory depression is not easy (Jennett, 1968; Saunders, 1980), but the use of respiratory minute volume or blood gas estimations is an important safeguard in initial tolerance studies. Useful observations of haemostasis, including platelet aggregation, can also be made on small numbers of subjects (e.g. Mielke, Kahn, Muschek, Tighe, Ng & Minn, 1980). Nausea and vomiting may also be noted early on, and can be measured with ordinal scale and VAS methods in normal volunteers, similarly and at the same time as pain; a central emetic effort may be missed if the subjects are not free to move about, as movement exaggerates this effect of opioids. In later trials, the number of subjects experiencing nausea and/or vomiting may be merely counted, probably a more valid measure, since individual susceptibility varies greatly, and a small sample of subjects may easily be misleading.

Compounds that inhibit cyclo-oxygenase may reasonably be expected to have some gastro-intestinal toxicity and to induce asthma in susceptible subjects. It is helpful and possible to study both before marketing. The controversy concerning gastro-intestinal toxicity has recently been succinctly reviewed (Duggan, 1980), and comparison of occult gastrointestinal blood loss due to the new drug with that from standard compounds possessing and lacking an effect on platelets may represent the best index of such an adverse effect that can be obtained until the results of post-marketing surveillance are available (Johnson, 1980). Asthma induced by mild analgesics is not rare nor impossible to test in suitable subjects (Szczeklik, 1980).

Animal pharmacology now goes some way to predicting the abuse liability of opioid analgesics, and studies in previously opioid-dependent subjects can provide further information (Jasinski, 1979). Nephrotoxicity of mild analgesics is due to a less predictable form of dependence, and in some countries is a fairly common disorder (Nanra, 1980), chiefly due to formulations of analgesics in combination. Animal models can help to predict the likelihood of analgesics being nephrotoxic (Axelsen, 1980), but little information can be gained from studies in normal subjects or patients. The hepatotoxicity of paracetamol sets a similar problem of a relevant animal model but no human data can be obtained for such a threshold phenomenon, save by drug abuse (Prescott, 1980b); with the type of hepatoxicity due to salicylates we lack even the animal model.

After marketing, there is an important difference between the numbers needed to detect an unwanted effect and the far larger number required to assess its incidence, though the two are closely related (Wardell, 1977). Detection requires a significant signal-to-noise ratio; it depends critically on the variance, the 'noise' level (false adverse reaction reports, perhaps due to biological ambiguities) and any bias in reporting (Finney, 1971). Neither data from circumscribed clinical experiments, nor clinical trials of the size usual in Britain, furnish sufficient data to detect or measure drug adverse reactions occuring less often than about one in a hundred users, but such measurements can be made very successfully in major trials of the size which national research councils can produce. Those interested in methods of detecting adverse reactions in larger populations will find a discussion of appropriate statistical techniques in papers by Mantel & Haenzsel (1959), Pike & Morrow (1970), Miettinen (1974) & Wardell (1977).

Pharmacokinetics

As with other classes of drugs, pharmacokinetic investigation can greatly illuminate efficacy studies, though it is particularly important that any body fluid samples are obtained painlessly, if pain is also being measured. Studies with salicylates are reviewed by Levy (1980), and with paracetamol by Prescott (1980a). Opioid analgesics have been less extensively studied, but some data are available (Berkowitz, 1976), interpretation of the results being hampered by the limited relevance of plasma drug levels for drugs that act within, and may be sequestered by, the central nervous system.

References

AITKEN, R.C.B. (1969). Measurement of feelings using visual analogue scales. *Proc. Roy. Soc. Med.*, **62**, 989–993.

AITKEN, R.C.B. & ZEALLEY, A.K. (1970). Measurement of moods. *Br. J. hosp. Med.*, **4**, 215–224.

ARMSTRONG, D., DRY, R.M.D., KEELE, C.A. & MONKHAM, J.W. (1953). Observations on chemical excitants of cutaneous pain in man. *J. Physiol.*, **120**, 326–351.

AXELSEN, R. (1980). Nephrotoxicity of mild analgesics in the Gunn strain of rat. *Br. J. clin. Pharmac.*, **10**, 309S–312S.

BAILEY, C.A. & DAVIDSON, P.O. (1976). The language of pain: Intensity. *Pain*, **2**, 319–324.

BEAVER, W.T. & McMILLAN, D. (1980). Methodological considerations in the evaluation of analgesic combinations: Acetoaminophen (paracetamol) and hydrocodone in postpartum pain. *Br. J. clin. Pharmac.*, **10**, 215S–223S.

BEECHER, H.K. (1957). The measurement of pain. Prototype for the quantitative study of subjective responses. *Pharmac. Rev.*, **9**, 59–209.

BEECHER, H.K. (1959). *Measurement of Subjective Responses.* New York: Oxford University Press.

BERKOWITZ, B.A. (1976). The relationship of pharmacokinetics to pharmacological activity: Morphine, methadone and naloxone. *Clin. Pharmacokin.*, **1**, 219–230.

BLOOMFIELD, S., BARDEN, T. & MITCHELL, J. (1976). Aspirin and codeine in two post-partum pain models. *Clin. Pharmac. Ther.*, **20**, 499–503.

BUCHSBAUM, M.S., DAVIS, G.C. & BUNNEY, W.E. (1977). Naloxone alters pain perception and somatosensory evoked potentials in normal subjects. *Nature*, **270**, 620–622.

CHAPMAN, C.R. (1977). Sensory decision theory methods in pain research: a reply to Rollman. *Pain*, **3**, 295–305.

CHAPMAN, C.R. (1978). Pain: the perception of noxious events. In *The Psychology of Pain*, ed. Sternbach, R.A., pp. 169–202. New York: Raven Press.

CHAPMAN, C.R., MURPHY, T.M. & BUTLER, S.H. (1973). Analgesic strength of 33 per cent nitrous oxide: A signal detection theory evaluation. *Science*, **179**, 1246–1248.

CLARK, W.C. (1974). Pain sensitivity and the report of pain: An introduction to sensory decision theory. *Anesthesiology*, **40**, 272–287.

COOPER, S.A. & BEAVER, W.T. (1976). A model to evaluate mild analgesics in oral surgery outpatients. *Clin. Pharmac. Ther.*, **20**, 241–250.

CRAIG, K.D. (1978). Social modeling influences on pain. In *The Psychology of Pain*, ed Sternbach, R.A., pp. 73–109. New York: Raven Press.

CRAIG, K.D. & NEIDERMAYER, H. (1974). Autonomic correlation of pain thresholds influenced by social modelling. *J. Pers. Soc. Psychol.*, **29**, 246–252.

CROCKETT, D.J., PRKACHIN, K.M. & CRAIG, K.D. (1977). Factors of the language of pain in patient and volunteer groups. *Pain*, **4**, 175–182.

DEODHAR, S.R., DICK, W.C., HODGKINSON, R. & BUCHANAN, W.W. (1973). Measurement of clinical response to anti-inflammatory drug therapy in rheumatoid arthritis. *Quart. J. Med.*, **166**, 387–401.

DIEM, K. & LENTNER, C. (Ed) (1970). *Documenta Geigy: Scientific tables.* 7th Edition, p. 69. Basle: J.R. Geigy.

DIXON, W.J. & MASSEY, F.J.J. (1969). *Introduction to Statistical Analysis*, 3rd Edition. New York: McGraw-Hill.

DOWNIE, W.N., LEATHAM, P.A., RHIND, V.M., WRIGHT, V., BRANCO, J.A. & ANDERSON, J.A. (1978a). Studies with pain rating scales. *Ann. Rheum. Dis.*, **37**, 378–381.

DOWNIE, W.W., LEATHAM, P.A., RHIND, V.M., PICKUP, M.E. & WRIGHT, V. (1978b). The visual analogue scale in the assessment of grip strength. *Ann. Rheum. Dis.*, **37**, 382–384.

DUGGAN, J.M. (1980). Gastrointestinal toxicity of minor analgesics. *Br. J. clin. Pharmac.*, **10**, 407S–410S.

DUNDEE, J.W. (1980). Clinical evaluation of mild analgesics. *Br. J. clin. Pharmac.*, **10**, 329S–334S.

EVANS, R.J. (1972). Acid-base changes in patients with intractable pain and malignancy. *Can. J. Surg.*, **15**, 37–42.

FINNEY, D.J. (1964). *Statistical Methods in Biological Assay*, 2nd Edition, pp. 266–272. New York: Hafner.

FINNEY, D.J. (1971). Statistical logic in the monitoring of reactions to therapeutic drugs. *Methods of Information in Medicine* (Stuttgart), **10**, 237–245.

FORDYCE, W.E. (1978). Learning processes in pain. In *The Psychology of Pain*, ed. Sternbach, R.A., pp. 49–72. New York: Raven Press.

FOWLE, A.S.E., HUGHES, D.T.D. & KNIGHT, G.J. (1971). The evaluation of histamine antagonists in man. *Eur. J. clin. Pharmac.*, **3**, 215–220.

GLYNN, C.J. & LLOYD, J.W. (1976). The diurnal variation in perception of pain. *Proc. Roy. Soc. Med.*, **69**, 369–372.

GOMEZ-JIMENEZ, J., FRANCO-PATINO, R., CHARGOY-VERA, J. & OLIVARES-SOSA, R. (1980). Clinical efficacy of mild analgesics in pain following gynaecological or dental surgery: Report on multicentre studies. *Br. J. clin. Pharmac.*, **10**, 355S–358S.

GRACELY, R.H., McGRATH, P. & DUBNER, R. (1978a). Ratio scales of sensory and affective verbal pain descriptors. *Pain*, **5**, 5–18.

GRACELEY, R.H., McGRATH, P. & DUBNER, R. (1978b). Validity and sensitivity of ratio scales of sensory and affective verbal pain descriptors: manipulation of affect by diazepam. *Pain*, **5**, 19–29.

HARDY, J.D., WOLFF, H.G. & GOODELL, H. (1940). Studies on pain: A new method for measuring pain threshold: Observations on spatial summation of pain. *J. clin. Invest.*, **19**, 649–657.

HARDY, J.D., WOLFF, H.G. & GOODELL, H. (1947). Studies on pain: discrimination of differences in intensity of a pain stimulus as a basis of a scale of pain intensity. *J. clin. Invest.*, **26**, 1152–1158.

HARDY, J.D., WOLFF, H.G. & GOODELL, H. (1953). *Pain Sensations and Reactions.* Baltimore: Williams and Wilkins.

HEWER, A.J.H., KEELE, C.A., KEELE, K.D. & NATHAN, P.W. (1949). A clinical method of assessing analgesics. *Lancet*, **ii**, 431–435.

HILGARD, E.R. (1978). Hypnosis and pain. In *The Psychology of Pain*, ed. Sternbach, R.A., pp. 219–240. New York: Raven Press.

HILLS, M. & ARMITAGE, P. (1979). The two-period crossover clinical trial. *Br. J. clin. Pharmac.*, **8**, 7–20.

HINDMARCH, I. (1980). Psychomotor function and psychoactive drugs. *Br. J. clin. Pharmac.*, **10**, 189–209.

HOFSTADTER, D.R. (1979). *Gödel, Escher, Bach.* Hassocks, Sussex: Harvester Press.

HOUDE, R.W., WALLENSTEIN, S.L. & RODGERS, A. (1960). Clinical pharmacology of analgesics: 1. A method of assaying analgesic effect. *Clin. Pharmac. Ther.*, 1, 163–174.

HUNTER, M., PHILIPS, C. & RACHMAN, S. (1979). Memory for pain. *Pain*, 6, 35–46.

HUSKISSON, E.C. (1974a). Catecholamine excretion and pain. *Br. J. clin. Pharmac.*, 1, 80–82.

HUSKISSON, E.C. (1974b). Measurement of pain. *Lancet*, ii, 1127–1131.

HUSKISSON, E.C. & HART, F.D. (1972). Pain threshold and arthritis. *Br. med. J.*, 4, 193–195.

JASINSKI, D.R. (1979). Human pharmacology of narcotic analgesics. *Br. J. clin. Pharmac.*, 7, 287S–290S.

JENNETT, S. (1968). Assessment of respiratory effects of analgesic drugs. *Br. J. Anaesth.*, 40, 746–756.

JICK, H., SLONE, D., SHAPIRO, S., LEWIS, G.P. & SISKIND, V. (1971). A new method for assessing the clinical effects of oral analgesic drugs. *Clin. Pharmac. Ther.*, 12, 456–463.

JOHNSON, P.C. (1980). A comparison of the effects of zomepirac and aspirin on fecal blood loss. *J. clin. Pharmac.*, 20, 401–405.

JOYCE, C.R.B., ZUTSHI, D.W., HRUBES, V. & MASON, R.M. (1975). Comparison of fixed interval and visual analogue scales for rating chronic pain. *Eur. J. clin. Pharmac.*, 8, 415–420.

KAST, E.C. (1962). The measurement of pain: A new approach to an old problem. *J. New Drugs*, 2, 344–351.

KEELE, K.D. (1948). The Pain Chart. *Lancet*, ii, 6–8.

KEELE, K.D. (1954). Pain-sensitivity tests: The pressure algometer. *Lancet*, i, 636–639.

KEELE, K.D. (1968). Pain complaint threshold in relation to pain of cardiac infarction. *Br. med. J.*, 1, 670–673.

KEELE, K.D. & STERN, P.R.S. (1973). Serum lipid changes in relation to pain. *J. Roy. Coll. Physcns.*, (Lond.), 7, 367–372.

KNITZA, R., CLASEN, R. & FISCHER, F. (1979). Pain-induced alterations in the individual non-esterified fatty acids in serum. *Pain*, 6, 91–97.

LASAGNA, L. (1960). The clinical measurement of pain. *Ann. N.Y. Acad. Sci.*, 86, 28–37.

LASAGNA, L. (1980). Analgesic methodology: A brief history and commentary. *J. clin. Pharmac.*, 20, 373–376.

LASCELLES, P.T., EVANS, P.R., MERSKEY, H. & SABUR, M.A. (1974). Plasma cortisol in psychiatric and neurological patients with pain. *Brain*, 87, 533–538.

LASKA, E. & SUNSHINE, A. (1973). Anticipation of analgesia—a placebo effect. *Headache*, 13, 1–11.

LEVINE, J.D., GORDON, N.C., BORNSTEIN, J.C. & FIELDS, H.L. (1979). Role of pain in placebo analgesia. *Proc. Nat. Acad. Sci.*, 76, 3528–3531.

LEVY, G. (1980). Clinical pharmacokinetics of salicylates: A re-assessment. *Br. J. clin. Pharmac.*, 10, 285S–290S.

LINDLEY, D.V. (1965). *Introduction to probability and statistics. Part 1. Probability*, pp. 25 & 215–235. Cambridge: Cambridge University Press.

LOESER, J.D. (1980). Perspectives on pain. In *Clinical Pharmacology & Therapeutics. Proceedings of the First World Conference*, ed. Turner, P., pp. 313–316. London: Macmillan.

LYNN, B. & PERL, E.R. (1977). A comparison of four tests for assessing the pain sensitivity of different subjects and test areas. *Pain*, 3, 353–365.

MANTEL, N. & HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies on disease. *J. National Cancer Institute*, 22, 719–727.

MELZACK, R. (1975). The McGill Pain Questionnaire: Major properties and scoring methods. *Pain*, 1, 277–299.

MELZACK, R. & DENNIS, S.G. (1978). Neurophysiological foundations of pain. In *The Psychology of Pain*, ed. Sternbach, R.A., pp. 49–72. New York: Raven Press.

MELZACK, R. & TORGERSON, W.S. (1971). On the language of pain. *Anesthesiology*, 34, 50–59.

MERSKEY, H. & SPEAR, F.G. (1964). The reliability of the pressure algometer. *Br. J. Soc. clin. Psychol.*, 3, 130–136.

MIELKE, C.H., KAHN, S.B., MUSCHEK, L.D., TIGHE, J.J., NG, K.T. & MINN, F.L. (1980). Effects of zomepirac on hemostasis in healthy adults and on platelet function *in vitro*. *J. clin. Pharmac.*, 20, 409–417.

MIETTINEN, O.S. (1974). Simple interval-estimation of risk ratio. *Am. J. Epidemiol.*, 100, 515–516.

MONGAR, J.L. (1955). A study of two methods for testing local anaesthetics in man. *Br. J. Pharmac.*, 10, 240–246.

MUMFORD, J.M. & BOWSHER, D. (1976). Pain and protopathic sensibility. A review with particular reference to the teeth. *Pain*, 2, 223–243.

NANRA, R.S. (1980). Clinical and pathological aspects of analgesic nephropathy. *Br. J. clin. Pharmac.*, 10, 349S–368S.

• NATHAN, P.W. (1977). Pain. *Br. med. Bull.*, 33, 149–156.

NAYMAN, J. (1979). Measurement and control of postoperative pain. *Ann. Roy. Coll. Surg. Engl.*, 61, 419–426.

NICHOLSON, A.N. (1978). Visual analogue scales and drug effects in man. *Br. J. clin. Pharmac.*, 6, 3–4.

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, 379–384.

OLDHAM, J.D. (1968). *Measurement in Medicine*. London: English University Press.

PARKHOUSE, J. & HOLMES, C.M. (1963). Assessing postoperative pain relief. *Proc. Roy. Soc. Med.*, 56, 579–583.

PATON, W.D.M. (1957). Histamine release by compounds of simple chemical structure. *Pharmac. Rev.*, 9, 269–328.

PIKE, M.C. & MORROW, R.H. (1970). Statistical analysis of patient-control studies in epidemiology: factor under investigation an all-or-none variable. *Br. J. Prev. Soc. Med.*, 24, 42–44.

POSER, E.G. (1962). A simple and reliable apparatus for the measurement of pain. *Am. J. Psychol.*, 75, 304–305.

PRESCOTT, L.F. (1980a). Kinetics and metabolism of paracetamol and phenacetin. *Br. J. clin. Pharmac.*, 10, 291S–298S.

PRESCOTT, L.F. (1980b). Hepatotoxicity of mild analgesics. *Br. J. clin. Pharmac.*, 10, 373S–379S.

REVILL, S.I., ROBINSON, J.O., ROSEN, M. & HOGG, M.I.J. (1976). The reliability of a linear analogue for evaluating pain. *Anaesthesia*, 31, 1191–1198.

ROLLMAN, G.B. (1977). Signal detection theory measurement of pain: A review and critique. *Pain*, 3, 187–211.

ROLLMAN, G.B. (1979). Signal detection theory pain measures: empirical validation studies and adaptation-

level effects. *Pain*, **6**, 9–21.

SAUNDERS, K.B. (1980). Methods in the assessment of the control of breathing. *Br. J. clin. Pharmac.*, **9**, 3–9.

SCOTT, J. & HUSKISSON, E.C. (1976). Graphic representation of pain. *Pain*, **2**, 175–184.

SITARAM, N., BUCHSBAUM, M.S. & GILLIN, C. (1977). Physostigmine analgesia and somatosensory evoked responses in man. *Eur. J. Pharmac.*, **42**, 285–290.

SMITH, G.M., EGBERT, L.D., MARCOWITZ, R.A., MOSTELLER, F. & BEECHER, H.K. (1966). An experimental pain method sensitive to morphine in man: The submaximal effort tourniquet technique. *J. Pharmac. exp. Ther.*, **154**, 324–332.

SMITH, G.M., LOWENSTEIN, E., HUBBARD, J.H. & BEECHER, H.K. (1968). Experimental pain produced by the submaximal tourniquet technique: Further evidence of validity. *J. Pharmac. exp. Ther.*, **163**, 468–474.

SUNSHINE, A. (1980). Clinical evaluation of mild analgesics in post partum pain. *Br. J. clin. Pharmac.*, **10**, 335S–337S.

SWETS, J.A. (1973). The relative operating characteristic in psychology. *Science*, **182**, 990–1000.

SZCZEKLIK, A. (1980). Analgesics, allergy and asthma. *Br. J. clin. Pharmac.*, **10**, 401S–405S.

TURSKY, B. (1975). The development of a pain perception profile: A psychophysical approach. In *Pain: New Perspectives in Therapy and Research*, eds Weisenberg, M. & Tursky, B., pp. 171–194. New York: Plenum Press.

TWYCROSS, R.G. (1975). *The evaluation of narcotic analgesics in patients with terminal cancer.* Paul Martini Award Essay.

VALBO, A.B. & HAGBARTH, K.-E. (1968). Activity from skin mechanoreceptors in awake human subjects. *J. exp. Neurol.*, **21**, 270–289.

VAN BUREN, J. & KLEINKNECHT, R.A. (1979). An evaluation of the McGill Pain Questionnaire for use in dental pain assessment. *Pain*, **6**, 23–33.

VAN DOBBEN de BRUYN, C.S. (1968). *Cumulative Sum Tests: Theory and Practice.* London: Griffin.

VINCENT, C.H. (1973). *Random Pulse Trains, their measurement and statistical properties* I.E.E. Monograph Series, No 13. London: Peter Peregrinus.

VON GRAFFENRIED, B., ADLER, R., ABT, K., NUESCH, E. & SPIEGEL, R. (1978). The influence of anxiety and pain sensitivity on experimental pain in man. *Pain*, **4**, 253–263.

VON GRAFFENRIED, B. & NUESCH, (1980). Non-migrainous headache for the evaluation of oral analgesics. *Br. J. clin. Pharmac.*, **10**, 225S–231S.

WALLENSTEIN, S.L., HEIDRICH, G., KAIKO, R. & HOUDE, R.W. (1980). Clinical evaluation of mild analgesics: The measurement of clinical pain. *Br. J. clin. Pharmac.*, **10**, 319S–327S.

WALLENSTEIN, S.L. & HOUDE, R.W. (1975). The clinical evaluation of analgesic effectiveness. In *Methods in Narcotics Research*, eds. Ehrenpreis, S. & Neidle, A., pp. 127–145. New York: Marcel Dekker.

WARDELL, W.A. (1977). Post-marketing surveillance: case-study approach. In *Drug Monitoring*, ed. Gross, F.H. & Inman, W.H.W. London: Academic Press.

WOLFF, B.B. (1977). Analgesic drug and dose discriminations by experimental human pain response parameters. *Clin. Pharmac. Ther.*, **21**, 123–132.

WOLFF, B.B. (1978). Behavioural measurement of human pain. In *The Psychology of Pain*, ed. Sternbach, R.A., pp. 129–168. New York: Raven Press.

WOLFF, B.B., KANTOR, T.G., JARVIK, M.E. & LASKA, E. (1969). Response of experimental pain to analgesic drugs: III. Codeine, aspirin, secobarbital, and placebo. *Clin. Pharmac. Ther.*, **10**, 217–228.

WOODFORDE, J.M. & MERSKEY, H. (1972). Some relationships between subjective measures of pain. *J. psychosom. Res.*, **16**, 173–178.

YAKSH, T.L. & RUDY, T.A. (1978). Narcotic analgesics: CNS sites and mechanisms of action as revealed by intracerebral injection techniques. *Pain*, **4**, 299–359.

YANG, J.C., CLARK, W.C., NGAI, S.H., BERKOWITZ, B.A. & SPECTOR, S. (1979). Analgesic action and pharmacokinetics of morphine and diazepam in man. *Anesthesiology*, **51**, 495–502.