# *Evaluation of Analgesics by the Rating of Patient Behavior*

*John E. Steinhaus, M.D., Ph.D., William Bevan, Ph.D., Sam C. Webb, Ph.D., W. R. Thompson, Jr., B.S.*



THE EVALUATION of analgesics has been the subject of much investigation within recent years. The extensive work of Beecher and his co-workers, summarized in his review,[1] has emphasized the importance of the double blind study, the problem of placebo reactors, and the advantages of using pathological pain for the study of analgesic agents. Stressing the subjective nature of pain, most investigators have determined the presence and relief of pain by using the patient's verbal response.[1,2,3] In contrast, one finds that the decision to give an analgesic drug in the postoperative management of pain is often determined by the patient's actions and appearance rather than his verbal response. This suggested the possibility of evaluating analgesic effects in terms of the rating of patient behavior which might be associated with or be symptomatic of the experience of pain. If such behavior could be identified, its use in the experimental analysis of analgesic agents might result in increased information and consequently facilitate the management of pain in these patients.

The purpose of this study was to evaluate the analgesic effect of four drugs on postoperative pain using ratings of patient behavior. For comparative purposes verbal estimates of pain provided by patients were analyzed in a similar manner. The four drugs selected for evaluation were meperidine, and three low potency analgesics—aspirin, acetophenetidin, and A.P.C. (acetylsalicylic acid, acetophenetidin and caffeine tablets, N.F.). Although analgesic action has been demonstrated for aspirin,[4] others have experienced difficulty in demonstrating analgesic action for this class of drugs.[5]

Received from the Departments of Anesthesiology and Psychology, Emory University, Atlanta, Georgia; accepted for publication October 21, 1963. Dr. Bevan's present address: Department of Psychology, Kansas State University.

## Procedure

The selection of behavioral traits and construction of rating scales were made following a preliminary study of 10 patients who had clinically recognizable pain and who gave verbal response to that effect. These patients were observed for the duration of their stay in a recovery room following operation. Based on these observations a comprehensive, narrative description of each patient's response during this period was written. The narratives were then analyzed and a frequency count was made for the occurrence of various behavior traits in the sample of 10 cases. The most frequently identified traits selected were as follows: general demeanor, gross bodily movement, degree of pain, facial lines, forehead tension, mouth-jaw expression, facial expression, breathing regularity, breathing rate, and breathing depth.

In the first phase of this study it was noted that certain of these traits changed very little and they were consequently dropped. Three of the traits which revealed the most marked changes are listed in table 1. A more detailed description of these traits is as follows: (1)

TABLE 1. Representative Rating Scales for Pain, Breathing Regularity, Bodily Movement and Facial Expression

| Breathing Regularity | Bodily Movement | Facial Expression |
|---|---|---|
| 7. None | 1. None | 7. Extremely tense |
| 6. Very little | 2. Very little | 6. Very tense |
| 5. Little | 3. Little | 5. Moderately tense |
| 4. Moderate | 4. Moderate | 4. Slightly tense or relaxed |
| 3. Much | 5. Much | 3. Moderately relaxed |
| 2. Very much | 6. Very much | 2. Very relaxed |
| 1. Extreme | 7. Extreme | 1. Extremely relaxed |

TABLE 2. Description of Patients

| Ages | No. | Sex | No. | Operations | | | |
|---|---|---|---|---|---|---|---|
| 15–31 | 21 | Female | 41 | Abdominal hysterectomy | 12 | Herniorrhaphy | 4 |
| 31–40 | 24 | Male | 24 | Vaginal hysterectomy | 9 | Mastectomy | 3 |
| 41–50 | 15 | | | Tubal ligation | 8 | Caesarean section | 2 |
| 51–65 | 5 | | | Reduction of fracture | 4 | Thoracotomy | 2 |
| | | | | Skin graft | 4 | Miscellaneous | 17 |

Breathing regularity: this refers to the evenness of respiration without regard to rate or depth. The dimension extends from extremely uneven or irregular breathing to the even or regular breathing seen in quiet sleep. (2) Bodily movement: this refers to the movement of arms, legs and trunk. Attention is particularly directed toward the non-specific variety rather than those directed toward the performance of some act. Tossing and turning would be an example of the behavior categorized under this trait; so, also, would hand wringing or kicking the legs. The trait varies between no bodily movement to extreme bodily movement. (3) Facial expression: this characteristic is related to the general demeanor. At one extreme it is completely calm and relaxed; at the other it conveys the impression of marked tension with distortion. This trait overlaps certain other traits listed in the larger set, including expression of mouth and jaw, which refers to the musculature and set of the jaw ranging from open-mouthed relaxation on the one hand to teeth clenched, or tight lips on the other. It would also include forehead tension, which refers particularly to the furrowing of the brow. Facial lines, a related trait, includes the lines in the region of the nose and eyes and would range from smooth relaxed features to those with marked lines and creases. Another trait listed in table 4, degree of pain (rater), was the overall assessment of pain (discomfort) made by the rater, from no discomfort to extreme discomfort. Presumably this would include the above traits plus additional subtle aspects of appearance and behavior, which are commonly identified with discomfort.

A seven point rating scale[6] was constructed for each of these traits. The selection of seven points was made for convenience and because it was found in the preliminary ratings that judgments could be separated to that degree. It should be noted, however, that scales of 3 or 9 numbers could also be used efficiently. It should be pointed out in the design of rating scales that one point in a scale differs from the adjacent point in value, *i.e.*, either greater or lesser; however, the difference in degree between points may not be uniform. For example, in a seven point scale where 1 is the greatest and 7 is the least, 5 is always greater than 6; however, the difference between 5 and 6 may not be identical to that between 3 and 4. It is obvious that such data cannot be analyzed by methods of analysis in which the differences between points are uniform.

In addition, a five point rating scale ranging from no pain to intolerable pain, was employed to secure patient ratings of pain (verbal response) in four areas: area of operation, area incident to procedure (for example, site intravenous infusion), area not incident to procedure (postural discomfort) and headache. The rating scale was similar to that

TABLE 3. Analysis of Trend Schema

| Source | *DF* | *F*-Ratio |
|---|---|---|
| Between Group Means | 64 | T/SWT |
| Treatments (T) | 4 | |
| Subjects Within Treatments (SWT) | 60 | |
| Within Subjects | 260 | |
| Overall Trend (OT) | 4 | OT/BIT |
| Between Group Trends (BGT) | 16 | BGT/BIT |
| Between Individual Trends (BIT) | 240 | |
| **Trend Analysis** | | |
| Linear Trend | | |
| Overall ($O_L$) | 1 | $O_L/BI_L$ |
| Between Group ($BG_L$) | 4 | $BG_L/BI_L$ |
| Between Individual ($BI_L$) | 60 | |
| Quadratic Trend | | |
| Overall ($O_Q$) | 1 | $O_Q/BI_Q$ |
| Between Group ($BG_Q$) | 4 | $BG_Q/BI_Q$ |
| Between Individual ($BI_Q$) | 60 | |

TABLE 4. Summary of Analysis of Trends for Rated Traits

| Source | *DF* | Pain (Verbal Response) | Degree of Pain (Rater) | Breathing Regularity | Facial Expression | Gross Bodily Movement | General Demeanor | Facial Lines | Forehead Tension | Mouth-Jaw Expression |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean Sq. | Mean Sq. | Mean Sq. | Mean Sq. | Mean Sq. | Mean Sq. | Mean Sq. | Mean Sq. | Mean Sq. |
| Between subject means | 64 | | | | | | | | | |
| Treatments | 4 | 4.40 | 29.11* | 10.53† | 10.14 | 30.19* | 26.31* | 4.76 | 3.08 | 18.51* |
| Subjects within treatments | 60 | 6.49 | 5.81 | 3.72 | 4.03 | 5.99 | 4.87 | 3.81 | 5.07 | 2.97 |
| Within subjects | 260 | | | | | | | | | |
| Overall trend | 4 | 18.17* | 43.50* | 23.57* | 28.19* | 23.00* | 22.48* | 12.73* | 25.73* | 19.96* |
| Between group trends | 16 | 0.79 | 3.96* | 2.32† | 2.63* | 3.84* | 3.63* | 1.43† | 1.75 | 2.39* |
| Between individual trends | 240 | 0.79 | 1.42 | 1.36 | 0.97 | 1.43 | 1.17 | 0.81 | 1.28 | 0.93 |

* Significant at 1 per cent level of confidence.
† Significant at 5 per cent level of confidence.

of Houde and co-workers[4] except that an additional degree, intolerable pain, was added. The patient was questioned as to how he was feeling. If this indicated any degree of discomfort he was asked to categorize his pain in terms of slight ("it doesn't bother me too much"), moderate ("it hurts but I can stand it"), severe ("it hurts real bad") and intolerable ("I can't stand it"). An analysis of the initial phase of the study indicated that we could limit the inquiry to the area of operation.

The patients were selected from those returning to the recovery room following a variety of surgical procedures. A description of these patients is given in table 2. The selection was limited to patients between the ages of 15 and 65 who exhibited significant pain and discomfort during the first hour in the recovery room. Approximately 65 per cent of this group were women. Patients who had complicated procedures or who were seriously ill were excluded. Since some of the analgesic agents used were given orally, patients suffering from nausea and vomiting were likewise eliminated. One patient developed such intense pain that additional analgesic agent was required and consequently he was eliminated from the study.

Using a double blind technique, four different drugs and a placebo were administered in random order to 65 patients with the qualification that 13 be in each group. These drugs included meperidine, 75 mg., subcutaneously, acetylsalicylic acid 0.6 g., orally, A.P.C. (acetylsalicylic acid, 0.390 g., acetophenetidin, 0.360 g., and caffeine, 0.027 g.) orally and acetophenetidin 0.6 g., orally. Each patient received two tablets and a subcutaneous injection. Either the tablet or the hypodermic injection acted as a placebo except in the placebo series in which both were placebos.

The first set of ratings was collected just prior to the administration of the drugs and additional sets were secured every 30 minutes thereafter for a period of three hours. All ratings were made by one technician. The following standard procedures were employed to insure independence of the behavioral ratings. By a random procedure some scales were numbered so that severe pain corresponded to the high end of the scale, while

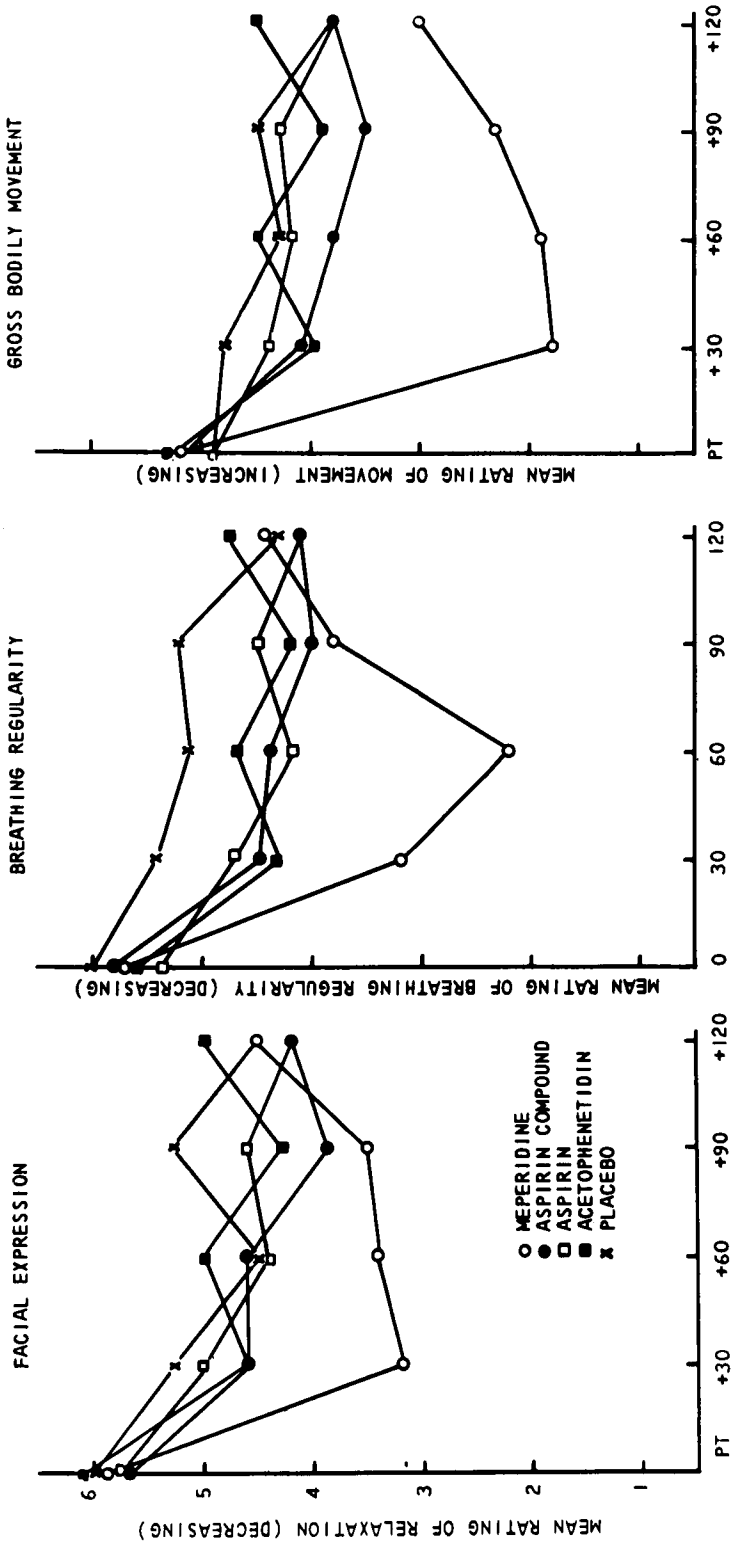


FIG. 1. Mean changes in ratings of three representative behavioral traits following analgesic drug.

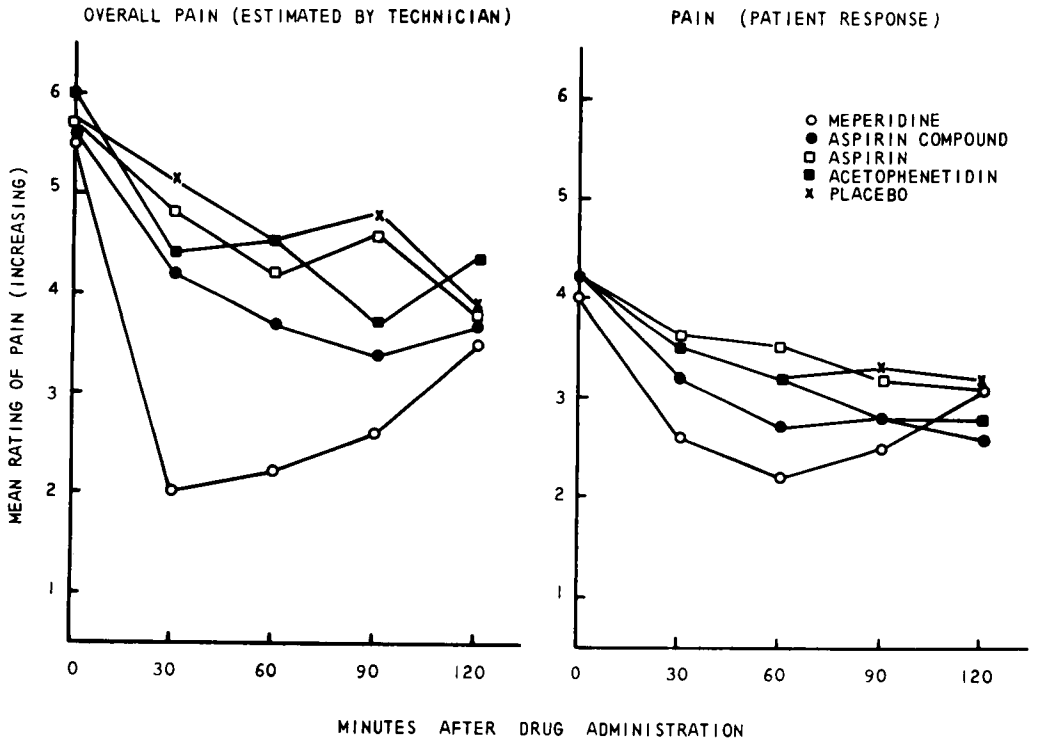


FIG. 2. A comparison of ratings by observer with verbal response of patient.

others were constructed to have severe pain represented at the low end. The order in which traits were noted was randomized within each set; all behavioral ratings were made before the patient was asked for his subjective estimate of pain. The rater was unaware of the specific drug or placebo administered.

A preliminary inspection of the data of 10 subjects in three groups indicated that the "analgesic effect" produced by the drugs could not readily be distinguished from the placebo two hours after their administration. Secondly, the ratings for two behavioral traits, breathing rate and breathing depth, and for three subjective estimates of pain did not change significantly following drug administration. Consequently, these five categories were dropped from further evaluation; and ratings on the other variables were discontinued after the end of two hours.

## Results

The schema of analyses[7] employed in this study is given in table 3. The summary of this analysis for the changes observed in various behavioral traits in table 4 shows significant differences in five behavioral traits. The mean square value for Facial Expression was just below the level of significance, however, verbal response and two other behavioral traits did not approach the level of significant change. It should be noted that this analysis does not indicate which drugs are responsible for the significant differences observed.

The second portion of table 4 shows the analysis of trends for the changes in the rated traits. Overall trend in this analysis indicates that the ratings of all groups change with respect to time. The further analysis listed under Group Trends corroborates the results mentioned above in that the behavioral traits, degree of pain, facial expression, bodily movement, general demeanor, and mouth jaw expression showed significant differences between drug groups at the 1 per cent level. The value for facial lines and breathing regularity were significant at the level of 5 per cent, whereas

verbal response did not show a significant difference.

A further understanding of the types of relation statistically analyzed can be obtained from figure 1. This figure charts the changes in ratings produced by the five drugs for three representative traits; facial exprtession, breathing regularity and bodily movement. Each point plotted on the graph represents the average rating for 13 subjects at the indicated time period. It will be noted that the graphs are similar in contour and, in general, present the same patterns. A similar relationship can be seen between the patient's verbal response and the rater's overall estimate of pain in figure 2. Meperidine consistently produced the most marked effects as might be expected from its recognized potent analgesic effect. The changes in ratings after administration of aspirin compound were consistently greater than those following placebo, whereas the ratings for aspirin and acetophenetidin fell between.

An analysis of the data from figures 1 and 2, employing the Dunnett *T* test is shown in table 5. As would be expected, the differences produced by meperidine are highly significant. Although the low potency analgesics did not show differences as great as the 5 per cent level, the order of ranking shows aspirin compound consistently second, suggesting a greater action than aspirin or acetophenetidin.

## Discussion

Since there were significant differences between group means for five of the rated behaviors (table 4) the results strongly suggest that the action of analgesic drugs can be evaluated by the use of ratings of patient behavior. While the well-established practice of using patient's statements of experienced pain has been valuable in assessing analgesia, there is an indication that these statements might be supplemented with a rating of behavior in order to provide additional data for evaluating analgesic drugs.

Just how the aspects of patient behavior evaluated in this study are related to the verbal expression of pain obtained from the patient must be determined by additional studies. The general profiles or shapes of the curves for the rated behaviors are similar to those for the verbal reports (fig. 2) which suggests that the two types of measure may be evaluating concomitant and related expressions of the effects of some underlying variable, if not actual measurement of the variable itself. It should be further pointed out that an individual learns the meaning of the words "hurt" and "pain" in infancy as his behavior during

TABLE 5. Summary of Analysis of Means by Dunnett *T* (with Ranking)

| | | Placebo | Aspirin CMP | Phenacetin | Aspirin | Meperidine |
|---|---|---|---|---|---|---|
| Degree of pain | Mean | 4.80 | 4.11 | 4.60 | 4.63 | 3.15 |
| | Dun.*T* | | 1.67 | 0.48 | 0.41 | 3.99 |
| | (Rank) | | (2) | (3) | (4) | (1) |
| Breathing regularity | Mean | 5.18 | 4.57 | 4.72 | 4.55 | 4.06 |
| | Dun.*T* | | 1.87 | 1.37 | 1.82 | 3.33 |
| | (Rank) | | (2) | (4) | (3) | (1) |
| Gross bodily movement | Mean | 4.48 | 4.06 | 4.43 | 4.34 | 2.85 |
| | Dun.*T* | | 0.97 | 0.11 | 0.32 | 3.80 |
| | (Rank) | | (2) | (4) | (3) | (1) |
| Facial expression | Mean | 5.11 | 4.60 | 5.00 | 4.80 | 4.02 |
| | Dun.*T* | | 1.48 | 0.31 | 0.89 | 2.92 |
| | (Rank) | | (2) | (4) | (3) | (1) |

$df = 60 \quad k = 5$
$t\,0.95 = 2.21$
$t\,0.99 = 2.87$

discomfort leads his mother to inquire (and consequently to teach) "does it hurt?" The employment of actions to show pain is reinforced as a child grows when he needs additional evidence to substantiate his statement of pain. Even if these behaviors were considered to be independent of the subjectively felt pain emanating from the area of operation, it is possible that the pain which is experienced is, to some degree, intensified by a function of proprioceptive sensory feedback from the observed tension behaviors.

If both types of measures are assumed to be a valid expression of the same effect, several advantages can be listed for the behavioral type measures. It is commonly recognized that many patients overstate their feelings when questioned about pain, whereas other patients tend to deny such feelings. Furthermore, the accuracy with which patients who are ill and under the influence of drugs can assess their feelings may be questioned. With language difficulties and lack of mental acuity in patients, most observers have found five degrees of subjective response to be the maximum that can be used satisfactorily.[1] In contrast, the rater in this experiment found seven intervals meaningful on most scales, indicating that more discriminating estimations can be made with behavioral ratings. Errors of measurement attributable to variability in the frame of reference can be reduced with behavior ratings since all ratings are made within the frame of reference of one, or at most a few, experienced raters. When patient estimates are made, there are as many frames of reference as there are patients.

In the usual determination of verbal response, it is quite possible that unconscious observation of the patient by the experimenter has contributed to data obtained in previous studies which have been attributed to verbal response. It would be necessary to blindfold the person (rater) collecting the patient's statement if one were to determine the information to be gained solely from verbal response. Rating of patient behavior has been useful in assessing the effects of other drugs such as the drugs used for premedication.[8] Verbal responses are frequently unreliable in this situation. Because of the marked consistency of the rankings seen in table 5, subsequent studies will be designed to analyze ranks; a procedure that cannot fairly be done in retrospect.

## Summary

Analgesic drugs were evaluated in postoperative patients by rating different categories of patient behavior. These ratings appeared to be more sensitive than the patient's verbal response which was determined concomitantly. A difference in the effectiveness of the analgesics tested was indicated. Although the analysis shown in table 5 does not show significant differences between the low potency analgesics, the ranking suggest that they could be differentiated with larger numbers of patients. Rating scales for patient behavior can be adapted to measure other responses to drugs which do not have readily available parameters for measurement.

This study was presented at the Annual Meeting of The Association of University Anesthetists, January 16, 1960. The investigation was supported by a grant from the Woodward Research Corporation.

## References

1. Beecher, H. K.: The measurement of pain, Pharmacol. Rev. **9**: 59, 1957.
2. Houde, R. W., Wallenstein, S. L., and Robers, A.: Clinical pharmacology of analgesics, Clin. Pharmacol. Ther. **1**: 163, 1960.
3. Lasagna, L.: The clinical measurement of pain, Ann. N. Y. Acad. Sci., **86**: 28, 1960.
4. Houde, R. W., and Wallenstein, S. L.: Studies on analgetics at Memorial Hospital. The evaluation of analgetics in incurable cancer patients, Drug Addiction and Narcotics Bull., Appendix C: 417–433, 1953a.
5. Gruber, C. M., Jr., Doss, J., Baptisti, A., and Chernish, S. M.: The use of post partum patients in evaluating analgesic drugs, Clin. Pharmacol. Ther. **2**: 429, 1961.
6. Guilford, J. P.: Psychometric Methods, ed. 2. New York, McGraw-Hill Book Co., 1954.
7. Winer, B. J.: Statistical Principles in Experimental Design. New York, McGraw-Hill Book Co., 1962.
8. Carpenter, F. A., Steinhaus, J. E., and Thompson, W. R.: Evaluation of preanesthetic medication utilizing rating scales and untrained raters, Southern Med. J. **56**: 881, 1963.