Prospective Validation of Clinically Important Changes in Pain Severity Measured on a Visual Analog Scale

From the Department of Emergency Medicine, Albert Einstein College of Medicine, Bronx, NY.

Author contributions are provided at the end of this article.

Received for publication January 30, 2001. Revision received June 19, 2001. Accepted for publication July 19, 2001.

Reprints not available from the authors.

Address for correspondence: E. John Gallagher, MD, Department of Emergency Medicine, Albert Einstein College of Medicine, Montefiore Medical Center, 111 East 210th Street, Bronx, NY 10467; 718-920-7459, fax 718-798-0730; E-mail jgallagh@montefiore.org.

Copyright © 2001 by the American College of Emergency Physicians.

0196-0644/2001/$35.00 + 0


Prospective Validation of Clinically Important Changes in Pain Severity Measured on a Visual Analog Scale

E. John Gallagher, MD
Michele Liebman, MD
Polly E. Bijur, PhD

Background: In a landmark hypothesis-generating study, Todd et al found that a difference of approximately 13 mm (95% confidence interval [CI] 10 to 17 mm) on a visual analog scale (VAS) represented the minimum change in acute pain that was clinically significant in a cohort of trauma patients.

Study objective: We test the hypothesis that the minimum clinically significant change in pain as measured by the VAS in an independent, more heterogeneous validation cohort is approximately 13 mm.

Methods: This was a prospective, observational cohort study of adults presenting to 2 urban emergency departments with pain. At 30-minute intervals during a 2-hour period, patients marked a VAS and were asked if their pain was “much less,” “a little less,” “about the same,” “a little more,” or “much more.” All data were obtained without reference to prior VAS scores. The minimum clinically significant change in pain was defined a priori as the difference in millimeters between the current and immediately preceding VAS scores when “a little more” or “a little less pain” was reported.

Results: Ninety-six patients enrolled in the study, providing 332 paired pain measurements. There were 141 paired measurements designated by patients as “a little less” or “a little more” pain. The mean clinically significant difference between consecutive ratings of pain in the combined “little less” or “little more” groups was 13 mm (95% CI 10 to 16 mm). The difference between this finding and that of Todd et al was 0 mm (95% CI –4 to 4 mm).

Conclusion: These data are virtually identical to previous findings indicating that a difference of 13 mm on a VAS represents, on average, the minimum change in acute pain that is clinically significant.

INTRODUCTION

Pain is the most frequent reason for emergency department visits in the United States. Accurate assessment of pain intensity, which is a necessary prerequisite to rational choice of analgesics, represents a clinically challenging proposition. This challenge is related to the complex nature of pain, which requires cortical interpretation of sensory stimuli within the context of personal and cultural norms for expression of discomfort.

One of the most frequently used pain rating scales is the visual analog scale (VAS). The VAS is a unidimensional scale, with several appealing characteristics. It is easy to use, requires no verbal or reading skills, and is sufficiently versatile to be employed in a variety of settings.

In an important, hypothesis-generating study, Todd et al concluded that a 13-mm difference (95% confidence interval [CI] 10 to 17 mm) on the VAS represented the smallest measurable change in acute pain severity that was clinically important. In this study, the minimum clinically significant difference was defined as the mean difference between VAS scores taken 20 minutes apart among patients experiencing “a little more pain” or “a little less pain.” Despite the relatively small sample size (N=48) consisting exclusively of trauma patients, the findings of this study have become the criterion standard for defining the smallest analgesic effect worth obtaining in the management of acute pain in the ED.

The present study was designed to test the hypothesis that the minimum clinically significant difference in acute pain intensity, as measured by the VAS, is approximately 13 mm in an independent validation cohort of ED patients with heterogeneous causes of pain.

MATERIALS AND METHODS

A prospective, observational cohort study was performed in 2 urban EDs. All English- and Spanish-speaking patients 18 years of age or older presenting to either ED with acute pain as a component of their chief complaint were eligible for inclusion. Acute pain was operationally defined as pain of recent onset (within 24 hours) or exacerbation (also within 24 hours) of preexisting pain of sufficient severity to cause the patient to seek emergency care. Patients with altered mental status or decreased visual acuity were excluded because of inability to score the VAS, as were patients who spoke neither English nor Spanish, and patients with unchanged chronic pain. The study was approved by the institutional review boards of both hospitals and the medical school.

Data were collected on a convenience sample of patients by trained research associates during the hours of 8 AM to 8 PM on selected days including weekends, during a 9-month period. Research associates approached patients with acute pain or an acutely painful illness or injury recorded on the nursing triage note. Training consisted of orientation by one of the investigators, followed by ongoing, close supervision until each research associate demonstrated proficiency in data collection.

After providing written informed consent, patients were asked to make a vertical mark through a 100-mm horizontal VAS bounded by the descriptors “least possible pain” on the far left and “worst possible pain” on the far right. At 30-minute intervals for the next 2 hours, patients were asked by the same data collector to repeat the measurement on an unmarked VAS, without access to any previous VAS ratings. This resulted in a maximum of 5 VAS scores per patient. A minimum of 2 VAS ratings of pain made 30 minutes apart was required for inclusion in the analyses. After completing the VAS at the end of each 30-minute interval, patients were also asked to contrast current pain with the pain at the time of the previous measurement, using one of the following 5 categorical descriptors: “much less pain,” “a little less pain,” “about the same pain,” “a little more pain,” or “much more pain.”

We used data from Todd et al’s derivation cohort to estimate the number of patients needed for our validation cohort. Todd et al reported that 80 out of 248 paired pain contrasts (approximately 1 out of 3) were recorded as either “a little less pain” (N=41) or “a little more pain” (N=39). To confirm that Todd et al’s mean difference of 13 mm corresponded to either a “little less” or a “little more” pain, we estimated that we would need at least 100 pain contrasts classified as a “little less” or a “little more” pain. The decision to increase our target sample size to exceed that of Todd et al was based on the following 3 factors: (1) the need to obtain an equal or higher degree of precision (95% CI 10 to 17 mm) than Todd et al; (2) the expectation that a more heterogeneous sample, which by design included both trauma and nontrauma patients, would result in an increased variance in pain scores; and (3) the assumption that, similar to the data reported by Todd et al, we would find that approximately 1 pain contrast in 3 would correspond either to a “little more” or a “little less” pain. Thus, we estimated that we would require approxi-
mately 300 total paired pain contrasts to obtain at least 100 paired pain contrasts categorized as a “little more” or “little less” pain.

All data were entered into Epi-Info (version 6, Centers for Disease Control and Prevention, Atlanta, GA), checked for entry errors, and exported into SPSS (version 9, SPSS, Chicago, IL) statistical software for statistical analysis. The mean and SD of the difference between each consecutive pair of measurements on the VAS were calculated for each of the 5 categories of pain descriptors provided by the patients at 30-minute intervals. All VAS differences associated with the verbal descriptors contributed to the means and SD, including those that were discordant with the descriptor (eg, a VAS difference of a 5-mm increase in pain associated with a verbal descriptor of “a little less pain”). The main outcome variable was mean change in VAS scores for the categories “a little less pain” and “a little more pain,” which were combined and compared with the mean change observed by Todd et al. To combine the differences in the 2 categories of “a little less pain” and “a little more pain,” the sign of the ratings of “a little less pain” was reversed by multiplying them by –1, identical to the technique used by Todd et al. The precision of estimates and the clinical and statistical significance of differences in mean VAS scores were expressed with 95% CIs.

The data in this study and in the study by Todd et al do not come from simple random samples of pain ratings, but are clustered by individuals (ie, 2 to 5 pain ratings were made by each individual in the current study). Clustering may result in underestimation of variance if data within individuals are correlated. For example, if some patients consistently rated their pain as 10 mm less at each of the four 30-minute intervals, that would represent correlated data, and the variability of pain ratings would be smaller than if each change in VAS score came from different individuals. However, in this study, the intraclass correlation was 0.003, indicating virtually no intrasubject correlation between measures of change in pain during the 2-hour study period. SDs calculated with a mixed model analysis of variance did not differ from ordinary SDs by more than 1 mm. Therefore, the SDs were calculated as usual, treating change in VAS scores as uncorrelated measures.

RESULTS

Of 166 patients in acute pain approached by research assistants, 28 refused to participate, and 42 were unable to do so because of altered mental status, poor vision, lack of comprehension of instructions, or inability to provide informed consent. The remaining 96 patients were entered into the study, generating 332 paired pain measurements for analysis. Seventy-two individuals completed the maximum number of 5 VAS measurements. Mean age of the cohort was 37 years (range 19 to 71 years), of whom 55% were women. The racial and ethnic distribution was: 39% black, 35% white, 19% Hispanic, and 7% unknown. The cause of pain was traumatic in 37 (39%) patients, non-traumatic in 36 (38%) patients, and unknown in 23 (24%) patients. The distribution of pain was as follows: abdomen, 44%; extremities, 23%; back, 11%; head, 8%; chest, 6%; other, 7%; and missing for 1 patient.

Figure 1 shows the distribution of VAS scores at baseline. These scores varied from 0 to 100 mm. The distribution was skewed toward higher scores with a median of 71 mm.

As shown in the Table, the mean VAS scores increased linearly as pain severity categorically worsened from “much less pain” to “much more pain.” For patients whose pain was “about the same,” the median and mean change in VAS score was 0 mm (95% CI –1 to 2 mm). As the descriptors moved toward the extremes of “much less pain” or “much more pain,” their distributions became increasingly skewed, as demonstrated by the differences between the means and medians for these groups (Table).

Figure 2 shows the widely dispersed distribution of changes in VAS scores for ratings described as “a little less
pain." There were 12 (13%) ratings indicating either no change or more pain on the VAS, even though the descriptor chosen by the patient was “a little less pain.” Similarly, in Figure 3 there were 9 (19%) ratings indicating either no change or less pain on the VAS, even though the descriptor chosen by the patient was “a little more pain.”

The combination of VAS scores of ratings with “a little less” and “a little more pain” is shown in Figure 4. The mean change in pain was 13 mm, with a 95% CI of 10 to 16 mm and a median of 11 mm. Twenty-one (15%) of these ratings were either 0 or were scored in the opposite direction from their verbal descriptors of “a little less” or “a little more pain.”

When the mean change in pain of the 141 ratings indicating “a little less” or “a little more pain” was tested against the mean change in VAS of the 80 corresponding ratings from the study by Todd et al,7 the observed difference was 0 mm (95% CI –4 to 4 mm).

**DISCUSSION**

The VAS entered the realm of pain research in the 1980s, demonstrating a greater sensitivity to increments or decrements in pain than other instruments.9,10 In contrast with verbal rating scales, which require a patient to choose a categorical pain descriptor (eg, mild, moderate, severe), the VAS provides a smooth continuum of choices. If a patient’s pain decreases slightly over time, a verbal scale may lack the sensitivity to detect it because a qualitative shift in degree of pain would be required to trigger a change in choice of descriptor. In contrast, the VAS allows patients to record small changes in pain severity.

**Table.**

VAS scores and corresponding categorical descriptors of change in pain.

<table>
<thead>
<tr>
<th>Categorical Descriptor of Pain</th>
<th>No. of Comparisons</th>
<th>Mean (SD), mm</th>
<th>95% CI, mm</th>
<th>Median, mm</th>
<th>Range, mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Much less pain”</td>
<td>30</td>
<td>–24 (29)</td>
<td>–35 to –13</td>
<td>–16</td>
<td>–100 to 0</td>
</tr>
<tr>
<td>“A little less pain”</td>
<td>93</td>
<td>–15 (17)</td>
<td>–18 to –11</td>
<td>–11</td>
<td>–78 to 21</td>
</tr>
<tr>
<td>“About the same pain”</td>
<td>144</td>
<td>0 (11)</td>
<td>–1 to 2</td>
<td>0</td>
<td>–66 to 43</td>
</tr>
<tr>
<td>“A little more pain”</td>
<td>48</td>
<td>10 (17)</td>
<td>5 to 15</td>
<td>10</td>
<td>–62 to 49</td>
</tr>
<tr>
<td>“A lot more pain”</td>
<td>15</td>
<td>19 (22)</td>
<td>6 to 31</td>
<td>13</td>
<td>–1 to 87</td>
</tr>
<tr>
<td>“A little less pain” or “A little more pain”</td>
<td>141</td>
<td>13 (17)</td>
<td>10 to 16</td>
<td>11</td>
<td>–62 to 78</td>
</tr>
</tbody>
</table>

**Figure 2.**

Frequency distribution of VAS change verbal descriptions of “a little less pain” (n=93).

**Figure 3.**

Frequency distribution of VAS change verbal descriptions of “a little more pain” (n=48).
Although this sensitivity to small changes in pain increases the validity of pain measurement, it can be problematic when using the VAS to compare effectiveness of different analgesic strategies. With large samples, small differences in mean VAS score can be declared "statistically significant," even though they may be of little clinical significance to the patient. Therefore, to advance studies of pain management in the ED, it is important to identify a minimum clinically significant difference in pain that can be used as a criterion for assessing differences between analgesic regimens.

In Todd et al’s original work and in the present study, the minimum clinically significant difference in pain was defined as the mean change in pain associated with a rating of “a little more pain” or “a little less pain.” The results of the current study are nearly identical to the findings of Todd et al. The mean minimum clinically significant difference of the VAS score was 13 mm in both studies, with a 95% CI of 10 to 16 mm in the current study and 10 to 17 mm in the study by Todd et al. The medians were similar: 11 mm in our study and 10 mm in the study by Todd et al. A third study, which was published while our study was underway, addressed this same question in an Australian patient population and found a slightly smaller minimum clinically significant difference in pain of 9 mm. However, the confidence interval of the Australian study (95% CI 6 to 13 mm) overlapped with those of the 2 US studies.

Our study has several limitations, the first of which is shared by all 3 VAS studies, that is, reliance on measures of central tendency (means or medians) to characterize the minimum clinically significant difference in pain severity. It is clear from the Figures that there is a great deal of variability around the mean and median. Some patients were highly sensitive to small differences in pain (eg, a decrease of 1 mm rated as a “little less pain”), whereas others required far larger differences before categorical change was noted. Included in the calculation of the measures of central tendency are the pain changes reported as either “a little less” or “a little more,” which corresponded to VAS ratings that either represented no change (0 mm) or scored in the opposite direction (eg, patients reporting their pain as “a little more,” who scored their VAS at least 1 mm lower than their previous ratings). In fact, 15% of the VAS difference scores in our study were discordant with the verbal descriptors; 18% of Todd et al’s ratings were similarly discordant.

In this study and the study by Todd et al, the mean and median VAS difference scores associated with increasing pain were lower than the mean and median VAS difference scores associated with decreasing pain. In Todd et al’s study the absolute value of the mean associated with “a little less pain” was –16 mm, whereas the mean associated with a “little more pain” was 10 mm. Similarly, in our study, the corresponding values were –15 mm and 10 mm, respectively. It may be that the threshold for detection of increased pain is lower than the threshold for decreased pain.

There may be other definitions of the minimum clinically significant difference in pain that are more clinically useful than the one used in this and similar studies. Measures of pain relief such as the proportion of patients reporting no additional need for analgesia might be more useful than the average change in millimeters on a VAS corresponding to “a little more or less pain” in a population of patients. Dichotomizing patients into those who did require additional analgesia for pain relief versus those who did not offers the further advantage of facilitating calculation of the number needed to treat associated with a given analgesic strategy.

The large variability around the means and the discordance between categorical pain descriptors and VAS scores may reflect a problem with the reproducibility or reliability of the VAS. We were unable to find any published studies of the reproducibility of the VAS in measurement of acute pain in the ED setting. One study of postoperative pain that examined reproducibility of VAS pain assessments during a 3-minute interval found that 95% of the repeat ratings were between ±20 mm. Thus, individuals who indicated a change in pain of 30 mm on the VAS may have a true difference as small as 10 mm or as large as 50 mm. If there is a similar level of imprecision in
the ED setting, then the minimum clinically significant difference in pain found in the 2 US studies7 and 1 Australian study12 may be too small for meaningful interpretation.

In summary, the evidence from this study confirms the finding by Todd et al7 that the minimum clinically significant difference in pain, as defined by the perception of a little more or little less pain, is in the range of a 13-mm change in VAS score. Attempts to apply measures of central tendency, such as means or medians, to individual changes in pain reveal some disagreement between categorical and numeric ratings, which may represent a fundamental limitation of the VAS. Research is needed on the reproducibility of this instrument to determine whether unreliability of VAS measurements in the ED might be responsible for this discordance.

Author contributions: EJG and ML designed the study. ML supervised the conduct of the trial and data collection. PEB managed the data, provided statistical consultation, and performed the data analysis. EJG, ML, and PEB drafted the initial manuscript and each contributed substantially to its revisions. EJG takes responsibility for the paper as a whole.

REFERENCES