The Visual Analog Scale for Pain

Clinical Significance in Postoperative Patients

Carol A. Bodian, Dr.P.H.,* Gordon Freedman, M.D.,† Sabera Hossain, M.S.,§ James B. Eisenkraft, M.D.,‖ Yaakov Beilin, M.D.‡

Background: The visual analog scale is widely used in research studies, but its connection with clinical experience outside the research setting and the best way to administer the VAS forms are not well established. This study defines changes in dosing of intravenous patient-controlled analgesia as a clinically relevant outcome and compares it with VAS measures of postoperative pain.

Methods: Visual analog scale measurements were obtained from 150 patients on the morning after intraabdominal surgery. On the same afternoon, 50 of the patients provided a VAS score on the same form used in the morning, 50 on a new form, and 50 were not asked for a second VAS measurement.

Results: Visual analog scale values and changes in value were similar for patients who were given a new VAS form in the afternoon and those who used the form that showed the morning value. The proportions of patients requesting additional analgesia were 4, 43, and 80%, corresponding to afternoon VAS scores of 30 or less, 31–70, and greater than 70, respectively. Change from morning VAS score had no apparent influence on patient-controlled analgesic dosing for patients with afternoon values of 30 or less or greater than 70, but changes in VAS scores of at least 10 did discriminate among patients whose afternoon values were between 31 and 70.

Conclusions: When pain is an outcome measure in research studies, grouping final VAS scores into a small number of categories provides greater clinical relevance for comparisons than using the full spectrum of measured values or changes in value. Seeing an earlier VAS form has no apparent influence on later values.

THE current investigation was motivated by discussions about planning studies to compare the effectiveness of various analgesic agents in relieving postoperative pain. The main issue was how to characterize the intensity of pain so that the study results would reflect clinically important differences between groups.

The visual analogue scale (VAS) is commonly used as the outcome measure for such studies. It is usually presented as a 100-mm horizontal line on which the patient's pain intensity is represented by a point between the extremes of "no pain at all" and "worst pain imaginable." Its simplicity, reliability, and validity, as well as its ratio scale properties, make the VAS the optimal tool for describing pain severity or intensity.[1]

There remain some outstanding questions regarding the use of the VAS in research studies. We address three of these in this report. First, an optimal connection between VAS values and clinical experience outside the research setting is not well established. Certain differences between groups in their VAS scores or changes in score may have no clinical relevance, even if they achieve statistical significance.[2–6] Moreover, a patient's perception of meaningful change in pain may depend on the initial level of pain.[7] The latter suggestion came from a study that used hypothetical assessments about a meaningful pain reduction from various starting levels. In our study, the patient's request for additional analgesia provides a concrete definition of clinical relevance to help interpret absolute values and changes in values of VAS scores.

Second, when the VAS is administered repeatedly to the same patient, the question of whether or not earlier values should be visible to the patient on succeeding forms has not been well answered.[8,9] Third, another issue to consider in measuring pain for research purposes is the possibility that the process of collecting data for research might itself influence the clinically relevant outcome. We have included a control group in the current study to shed light on that question.

We address these three issues in the setting of postoperative pain assessment in a series of patients who agreed to use intravenous patient-controlled analgesia (PCA) after intraabdominal surgery with general anesthesia. The reliability of VAS measurements as a measure of pain intensity[10] and of differences in VAS measurements as a measure of change in pain sensation for patients experiencing mild to moderate pain[11] has been demonstrated in postoperative patients.

Materials and Methods

After approval by the Institutional Review Board of the Mount Sinai School of Medicine (New York, New York), written informed consent was obtained from patients scheduled to undergo intraabdominal surgery with general anesthesia to participate in the current prospective, randomized study. All patients had agreed to the use of intravenous PCA for their postoperative pain management.

No changes from routine use of the PCA pump were made for the study. The pump was regulated to allow

---

* Associate Professor of Biomathematical Sciences, † Assistant Professor of Anesthesiology and Obstetrics, Gynecology and Reproductive Sciences, ‡ Associate Professor of Anesthesiology, § Programmer Analyst of Biomathematical Sciences, ‖ Professor of Anesthesiology.

Address reprint requests to Dr. Bodian: The Mount Sinai Medical Center, Department of Biomathematical Sciences, Box 1023, One Gustave L. Levy Place, New York, NY 10029-6574. Address electronic mail to: bodian@camelot.mssm.edu. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

Table 1. Demographic Data and Morning VAS Scores

	New Form Group	Same Form Group	Control Group
No. of patients	50	50	50
Age (yr)	49 (37–61)	49 (35–60)	53.5 (38–71)
Height (cm)	170 (167–178)	167 (163–173)	170 (163–178)
Weight (kg)	69 (60–80)	61 (55–72)	68 (57–80)
Gender (% male)	62%	42%	52%
Day 1 postoperative morning VAS score	38.5 (22–51)	34.5 (19–58)	49.5 (25–71)

Data are presented as median and interquartile range, or percent.

New form group = patients marked a new visual analog scale (VAS) form in the afternoon; Same form group = afternoon VAS markings were put on same form as morning markings; Control group = no afternoon pain scores were obtained.

the patient to self-administer morphine sulphate in amounts between 1.0 and 1.5 mg (demand dose), depending on the patient's age and weight. The demand dose was limited to between six and eight doses per hour, also depending on age and weight. None of the patients received a basal infusion of morphine from the PCA pump, and all patients were encouraged to ambulate, according to the routine of the Department of Surgery.

On the morning of postoperative day 1, a research assistant presented each patient with a VAS for pain, followed by a McGill pain questionnaire.[12] The VAS consisted of a 100-mm horizontal line anchored at one end with the words "no pain" and at the other end with the words "worst pain imaginable." The research assistant asked the patient to mark the line at the point that best represented the intensity of his or her pain. The VAS numeric value is the distance in millimeters from "no pain" to the point marked by the patient.

Patients were randomly assigned to one of three groups defined in terms of the protocol for afternoon pain assessment. A table of random numbers generated the randomization sequence, using a restricted randomization scheme to assure equal numbers in each group. Group assignments were sealed in opaque envelopes and opened sequentially by the investigators.

On the afternoon of postoperative day 1, between 8 and 12 h after the morning assessment, patients in the first group (new-form group) were asked to complete a new VAS form, and patients in the second group (same-form group) were asked to mark their current pain status on the same VAS form they had used in the morning. The VAS forms were removed, and patients in both groups completed another McGill pain questionnaire. Patients in the third group (control group) were not given any pain scales to complete in the afternoon.

Rating scales for pain, such as the VAS, are measures used in research studies, but they are not part of our standard clinical practice. In our standard clinical procedure, an anesthesiologist asks the patient how he or she is feeling and then decides whether or not to adjust the settings of the PCA pump. In the current study, this was performed in the afternoon, after all the other forms and scales had been completed and removed, and with the anesthesiologist unaware of their content. If the patient was not satisfied with the degree of pain relief and wanted additional pain medication, the pump settings were increased (clinical change). If the patient complained about side effects from the medication, such as pruritus, the medication may also have been changed, but in the current study *clinical change* refers only to changes related to pain management, not to those related to side effects.

Statistical Analysis

All data were entered into an Excel database and converted to a SAS file (SAS/STAT User's Guide, version 6; SAS Institute Inc., Cary, NC) for statistical analysis. Ordinal and categorical data were compared using the chi-square test or the chi-square test for trend. The Mantel-Haenszel test was used to compare groups stratifying on quartiles of morning VAS scores, to control for group differences in the morning scores. The Wilcoxon or Kruskal-Wallis test was used to compare continuous variables. Differences were considered significant at $P < 0.05$.

Logistic regression analysis was used to study the connection between use of additional analgesia and VAS scores, the fit being tested by chi-square goodness-of-fit criteria. Estimates of the probability of requesting additional analgesia at specified VAS values were obtained from the fitted logistic regression model as $p = \text{inv}(1 + e^{-(a+bx)})$, where a is the value of the fitted intercept, b the slope, and x the VAS measurement.

Results

We enrolled 150 patients, 50 in each group. Baseline demographic data and initial pain scores are shown in table 1. The groups were similar except that the patients in the control group started with higher VAS scores than did those in the other two study groups.

There was fairly good correlation between the VAS values and the corresponding McGill questionnaire scores, with Spearman rank correlation coefficients of 0.69 for the 150 pairs of morning measurements, 0.69 for the 100 pairs of afternoon measurements, and 0.54 for the 100 pairs of changes from morning to afternoon.

Table 2. Afternoon VAS Scores and Changes in Scores

	New Form Group	Same Form Group	P Value*
Measured Values (median, interquartile range)			
Afternoon VAS scores	30.5 (11–54)	31.0 (16–52)	0.59
Change in VAS score from morning to afternoon	−1.5 (−19 to +6)	−2.0 (−18 to +5)	0.97
Grouped values (No. of patients)			
Afternoon VAS score			0.87
≤ 30	25	24	
31–70	23	23	
> 70	2	3	
Change in VAS			0.68
Decrease of more than 10	19	21	
Change of 10 or less	22	21	
Increase of more than 10	9	8	

* Wilcoxon test for measured values, chi-square test of trend for grouped data.

VAS = visual analog scale.

Visual Analog Scale Values by Method of Administration

The median afternoon score for the patients completing a new VAS form was 30.5 (interquartile range, 11–54), compared with 31.0 (interquartile range, 16–52) for patients given the same form in the afternoon as the one they had marked in the morning. The median changes between morning and afternoon scores were −1.5 (interquartile range, −19 to +6) and −2.0 (interquartile range, −18 to +5), respectively. Seeing the morning score at the time of marking the afternoon VAS did not significantly influence the values of the afternoon VAS scores or the changes in VAS score from the morning to the afternoon (table 2). The groups were similar, whether the data were analyzed using their full numeric scale or were grouped into three categories and whether or not the morning VAS scores were controlled for in the analysis.

Relation of Visual Analog Scale Values to Clinical Changes (Request for Additional Analgesia)

The purpose of the part of the data analysis on the relation of VAS values to clinical changes is to seek a way to characterize VAS scores or changes in score, or both, so that their use as an outcome measure for comparative studies of pain will convey clinical relevance comparable to that of comparing the proportions of patients who request additional analgesia. To this end, we first explored the relations between percentage of patients requesting a clinical change and the VAS scores registered at the afternoon measurement, as well as changes in VAS scores from morning to afternoon. Separate logistic regression models were fit to each of these VAS measures, and each provided a good fit ($P < 0.0001$) by standard criteria, thus verifying the construct validity of the VAS in the immediate postoperative setting. The estimates of the intercept and slope from the logistic regression analyses are −3.8746 and 0.0710, respectively, for the afternoon scores, and −0.9867 and 0.061, respectively, for the change in scores.

At first impression, the fit of the logistic regressions would appear to support the use of either of these VAS measurements as outcomes in comparative studies, but there is a problem that becomes evident when the logits in the model are transformed to the corresponding probabilities of a clinical change. That is the well-known "S-curve," which describes the shape of the probabilities under a linear logit model, wherein the slopes in the lower and upper sections of the predictor variable are considerably flatter than the slopes in the mid-section.[13] To demonstrate what this feature implies about clinical relevance of using VAS scores or changes in score as outcome measures in comparative trials, we converted the fitted logits to probabilities of clinical change at succeeding 10-unit intervals. The same 10-unit difference between VAS scores or between changes in score represents a wide range of differences in probabilities of clinical changes. For example, suppose a study is using afternoon VAS scores as its end point and is powered to find a significant difference between two groups if the average afternoon scores in the two groups differ by at least 10 points. If the 10-point difference results from average scores of 15 in one group and 25 in the other, the difference in the proportion of patients expected to use additional analgesia is only 5.2% (10.9% − 5.7%). On the other hand, if the 10-point difference between groups results from average scores of 35 in one group and 45 in the other, the clinically relevant difference in proportions is 13.7% (33.6% − 19.9%).

Similar disparities result when comparing change in VAS scores between groups. For example, if the average change represents a 10-point difference between groups in the *decrease* in VAS scores (*e.g.*, a decrease of 15 in one group and 25 in another), the corresponding difference in proportions of patients expected to use additional analgesia is 5.4% (12.9% − 7.5%). The same amounts of change in the opposite direction (*i.e.*, *increases* of 15 in one group and 25 in another, correspond to a 15.0% difference in proportions (63.3% − 48.3%).
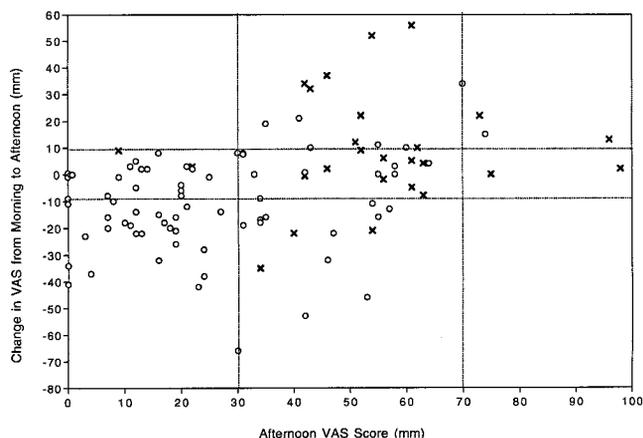
Fig. 1. Change in visual analog scale (VAS) score by afternoon visual analog scale score. X = patients who requested a change in their patient-controlled analgesia pump setting; open circle = patients with no clinical change.

An alternative way to use the VAS measurements is to group them into broader intervals by combining areas with similar probabilities of clinical change. The scatterplot of change in VAS scores by afternoon VAS score (fig. 1), on which patients who had a clinical change are denoted with "X" and those who did not with "O," suggests grouping criteria. Few clinical changes are evident among patients whose afternoon VAS was 30 or less, regardless of the amount of change in VAS since the morning. On the other hand, most of the patients with an afternoon VAS greater than 70 did request a clinical change, even if their afternoon VAS score was close to the morning score. The influence of a change in VAS on clinical change was limited to patients whose afternoon scores were in the range between 31 and 70; the more the afternoon VAS exceeded the morning VAS, the greater the need for a clinical change. These observations are summarized in table 3, where it is seen that the proportion of patients requesting a clinical change increased from 4% of patients whose afternoon VAS was 30 or less to 80% of patients whose afternoon VAS was greater than 70. For patients whose afternoon VAS was between 31 and 70, there was an increase between those extremes according to the extent of change in VAS from morning to afternoon.

These observations suggest different possible groupings. If the comparative study is aimed at maintaining a patient's comfort level to the extent that no further analgesic would be required, a simple binary outcome indicating whether or not the patient achieved a VAS score of 30 or lower would be appropriate. If the goal is to produce a downward shift across the spectrum of pain intensities, grouping the achieved VAS measurements (in our case, the afternoon values) into three broad categories (≤ 30, 31–70, and > 70) would provide a scale that is approximately linear in the corresponding probabilities of a clinical change (4, 43, and 80%, respectively, in our sample). Greater sensitivity could be achieved by subdividing the group with scores of 31–70 according to whether the achieved value is the result of a change from the initial VAS score of more than a 10-point decrease, a change of 10 points or less, or an increase of more than 10 points, as displayed in table 3.

Relation of Clinical Change in Study Groups to That in Control Group

Nine patients in the control group requested a change in their PCA settings, compared with 10 in the new-form group and 16 in the same-form group. Controlling for differences between groups in the value of their morning VAS, the difference between the same-form group and the control group was statistically significant ($P = 0.03$).

Discussion

The appropriate outcome measure for comparative studies of pain or analgesia depends on the context and purpose of the research. In some instances, the factor most relevant for patient care is the use or amount of analgesic needed, but it is not always possible to measure that directly in a uniform way in all subjects. In other instances, analgesic use itself may not be relevant but it is desirable to compare the effect of an intervention on perceived pain in a way that will focus on differences that are in some sense clinically meaningful. For example, a study comparing the effectiveness of two analgesics may produce a statistically significant difference in average VAS scores, but if the average scores in both groups are low (< 30), the clinical relevance of this finding could be suspect. Our data suggest that the proposed three- or five-point scale or the proportion of patients achieving a VAS score less than 30 would be a

Table 3. VAS Grade by Clinically Indicated Change in PCA Pump Settings

Grade	Afternoon VAS Score (mm)	Change in VAS	n	Clinical Change
I	≤ 30	Any	49	2 (4%)
II-a	31–70	Decrease of > 10	14	3 (21%)
II-b	31–70	Change of 10 or less	18	9 (50%)
II-c	31–70	Increase of > 10	14	8 (57%)
II (total)	31–70	Any	46	20 (43%)
III	> 70	Any	5	4 (80%)

VAS = visual analog scale; PCA = patient-controlled analgesia.

better outcome measure in such contexts. Further study would be needed to determine whether grouping VAS scores provides a relevant measure of drug effectiveness in other contexts.

How to arrive at a clinically relevant interpretation of pain scales has been addressed in a variety of settings, but we are not aware of any such studies concerning acute pain in the postoperative patient. Carlsson[2] considered patients under treatment for chronic pain and noted that effectiveness of treatment is better reflected by absolute levels of posttreatment pain intensity than by a calculated reduction in pain. Our study also suggests that absolute values of VAS measurements are more clinically relevant than change in VAS scores, but that the values should be grouped into three categories (≤ 30, 31–70, and > 70) and that greater sensitivity can be achieved by including subdivisions for change of pain within the middle category. We do not recommend replacing the VAS with a three-category scale, because that might sacrifice the documented reliability and validity of the VAS and will not allow an assessment of the amount of change.

Grouping measured VAS values into broad categories is a way to accommodate the findings of DeLoach *et al.*,[14] who studied the precision of VAS scores obtained in the immediate postoperative period. They concluded that any single postoperative VAS score should be considered to have an imprecision of ± 20. The grouping into three clinically relevant categories that we observed closely mirrors a similar finding by Serlin *et al.*,[15] who sought a connection between numeric ratings of chronic pain in cancer patients and *clinical relevance*, which was defined as interference with functions such as activity, mood, and sleep. The researchers found that absolute values of pain on a 0–10 scale naturally grouped into three categories: 1–4 (mild pain), 5 or 6 (moderate pain) and 7–10 (severe pain).

Our finding that request for additional analgesic was confined almost exclusively to patients with VAS scores greater than 30 also is consistent with the finding of Collins *et al.*[4] that baseline VAS values in excess of 30 in patients enrolled in various randomized trials of analgesics corresponded to a verbal report of at least moderate pain. Campbell and Patterson[7] found that dental patients and surgical nurses both estimated that VAS scores would have to be reduced by 10–30 points, depending on initial intensity, to achieve clinically meaningful pain relief. Studies of acute pain in the emergency medicine setting have suggested that change in consecutive VAS scores should be at least 9 (Kelly[5]) or 13 (Todd *et al.*[3]) to be clinically relevant. We found approximately the same magnitude of change to be important, but not across the board. The influence of a change in VAS score varied according to the absolute level of pain at the later measurement. Change in VAS score was not relevant for patients with little pain at the time of the later measure-

ment (VAS score ≤ 30), only 4% of whom requested an increase in analgesic dose, or for those with considerable pain (VAS score > 70), with 80% requesting an increase. Change in VAS score was associated with requests for additional analgesia in patients experiencing middle levels of pain (VAS score, 31–70). The ones for whom that level represented an improvement of at least 10 points were less likely to request additional analgesia (21%), compared with those with little change (50%) or those with an increase of more than 10 points (57%).

A recent study[6] used the need for rescue medication in cancer patients with chronic pain to define clinically relevant cut-off points on a variety of pain scales. Although none of their scales is directly comparable to ours, their results demonstrate the value of establishing clinically important cut-off points for primary outcome measures in pain therapy clinical trials.

We are aware of only two previous studies designed to address the question of whether a new VAS form should be used for sequential ratings, one concluding that it should[8] and one that it should not.[9] Both were assessing chronic pain evaluated over longer periods in an outpatient setting. As in our study, Joyce *et al.*[8] found little difference in the values of the VAS scores between patients who used a new form and those who kept using the same one. The researchers also compared the sensitivity of the two methods with regard capturing response to different doses of analgesic. Although the results seem somewhat ambiguous, the authors concluded that using a new form was more effective. On the other hand, Scott and Huskisson[9] reported that VAS values on the new posttreatment forms tended to be higher than on the ones that showed the pretreatment marking, with increasing divergence the longer the time between ratings. The authors concluded that patients tend to overestimate their level of pain when previous scores are not available and recommended the use of the same form for sequential VAS measurements. However, at their shortest interval between evaluations (2 weeks), seven patients showed no difference between the two posttreatment scores, eight showed a positive difference, and seven showed a negative difference. This is consistent with our finding of equivalent score changes between the patients who did and those who did not see their baseline values.

Although seeing the initial VAS score had no apparent influence on the values of the later VAS score in our study, perceived need for additional pain relief was greater in the same-form group than in the new-form group and greater than in the control group, the latter difference achieving statistical significance ($P = 0.03$). This observation could have implications for future research if confirmed in other studies. Meanwhile, it would seem prudent to use a new form each time a patient is asked to mark a VAS.

A limitation of the current study is the relative paucity in any of the groups of patients who were experiencing unsatisfactory pain control postoperatively. Thus, the associations with clinical change at the high end of the scales are less precise than at the lower end, and possible distinctions between the two VAS groups could not be assessed at these levels. Also, the usefulness of the VAS groupings that were suggested by our data will have to be subjected to statistical confirmation on an independent set of data that includes several groups of patients expected to experience different levels of pain. In addition, our gold standard applies to analgesia administered by PCA in the setting of postoperative pain assessment, and it is not known how this process compares with other forms of administration or other settings. However, the fact that we found similar minimum levels of VAS scores and changes in VAS scores as have been found to be clinically relevant in various nonsurgical settings is reassuring.

In conclusion, we found that grouping the measured values of the VAS scale into categories (≤ 30, 31–70, and > 70) produced a measure of pain that was more appropriately linked to our measure of clinical relevance than was the full spectrum of measured VAS values or changes in value. Additional sensitivity was achieved by a five-point scale that subdivided the group with scores of 31–70 according to the corresponding change in VAS (decrease of more than 10 points, change within the range of ± 10 points, and increase of more than 10 points). Although there was no indication that seeing the morning VAS pain score influenced the subsequent afternoon score, in our data the perceived need for additional analgesia differed, and we recommend routine use of a new form. We are not aware of published data about this aspect of measuring pain relief for research purposes and hope our preliminary finding will serve to motivate further study.

References

1. Katz J, Melzack R: Measurement of pain. Surg Clin North Am 1999; 79: 231–52
2. Carlsson AM: Assessment of chronic pain: I. Aspects of the reliability and validity of the visual analogue scale. Pain 1983; 16:87–101
3. Todd KH, Funk KG, Funk JP, Bonacci R: Clinical significance of changes in pain severity. Ann Emerg Med 1996; 27:485–9
4. Collins SL, Moore RA, McQuay HJ: The visual analogue pain intensity scale: What is moderate pain in millimetres? Pain 1997; 72:95–7
5. Kelly AM: Does the clinically significant difference in visual analog scale pain scores vary with gender, age, or cause of pain? Acad Emerg Med 1998; 5:1086–90
6. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL: Defining the clinically important difference in pain outcome measures. Pain 2000; 88:287–94
7. Campbell WI and Patterson CC: Quantifying meaningful changes in pain. Anesthesia 1998: 53:121–5
8. Joyce RB, Zutshi DW, Hrubes V, Mason RM: Comparison of fixed interval and visual analogue scales for rating chronic pain. Eur J Clin Pharmacol 1975; 8:415–20
9. Scott J, Huskisson EC: Accuracy of subjective measurements made with or without previous scores: An important source of error in serial measurement of subjective states. Ann Rheum Diseases 1979; 38:558–9
10. Banos JE, Bosch F, Canellas M, Bassols A, Ortega F, Bigorra J: Acceptability of visual analogue scales in the clinical setting: A comparison with verbal rating scales in postoperative pain. Methods Find Exp Clin Pharmacol 1989; 11:123–7
11. Myles PS, Troedel S, Boquest M, Reeves M: The pain visual analog scale: Is it linear or nonlinear? Anesth Analg 1999; 89:1517–20
12. Melzack R: The short-form McGill pain questionnaire. Pain 1987; 30:191–7
13. Agresti A: Categorical Data Analysis. New York, John Wiley & Sons, 1990, p 86
14. DeLoach LJ, Higgins MS, Caplan, AB, Stiff JL: The visual analog scale in the immediate postoperative period: Intrasubject variability and correlation with a numeric scale. Anesth Analg 1998; 86:102–6
15. Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS: When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. Pain 1995; 61:277–84