Pain rating by patients and physicians: evidence of systematic pain miscalibration

Laetitia Marquié\textsuperscript{a}, Eric Raufaste\textsuperscript{a}, Dominique Lauque\textsuperscript{b}, Claudette Mariné\textsuperscript{a}, Marie Ecoiffier\textsuperscript{b}, Paul Sorum\textsuperscript{c}

\textsuperscript{a}. LTC, UMR 5551 CNRS, Toulouse, France
\textsuperscript{b}. Département des Urgences, Hôpital Purpan, CHU Toulouse, France
\textsuperscript{c}. Departments of Medicine and Pediatrics, Albany Medical Center, Albany, NY

Address: Laetitia Marquié (lmarquie@univ-tlse2.fr)
LTC
Université Toulouse-II
31058 Toulouse Cedex 1
France
(+33) 551 503 542 (tel); (+33) 551 503 533 (fax);
Abstract

This study is an investigation of the existence and potential causes of systematic differences between patients and physicians in their assessments of the intensity of patients’ pain. In an emergency department in France, patients \( (N=200) \) and their physicians \( (N=48) \) rated the patients’ pain using a visual analog scale (VAS), both on arrival and at discharge. Results showed, in confirmation of previous studies, that physicians gave significantly lower ratings than did patients of the patients’ pain both on arrival (mean difference -1.33, standard error \([SE]= 0.17\), on a scale of 0-10, \( P <.001 \)) and at exit (-1.38, \( SE=0.15\), \( P<.001 \)). The extent of “miscalibration” was greater with expert than novice physicians and depended on interactions between physician gender, patient gender, and the obviousness of the cause of pain. Thus physicians’ pain ratings were influenced by non-medical factors.

Keywords: Pain assessment; Accuracy; Visual Analog Scale
1. Introduction

Pain is a personal experience of the patient. Yet the key diagnostic judgments and treatment decisions depend largely on the physician’s perception of this pain (Bartfield et al., 1997). It is the patient’s physician who must decide, for example, how serious the problem is, how much pain medication should be given, and whether the patient can leave the emergency room or the hospital.

Researchers have studied patients’ perception of pain intensity, physicians’ perception of patients’ pain, and the difference between the perceptions of patients and their physicians (Hodgkins et al., 1985; Todd et al., 1994; Thomas et al., 1999). In general, physicians give lower ratings of patients’ pain than do the patients themselves. In France, Blettery and colleagues (1996) found that patients’ pain was rated lower by physicians using behavioral cues of pain such as pain mimicry (i.e., a “hetero-evaluation scale”) than by their patients using a visual analog scale (“VAS”, Price et al., 1983); the difference in scales could, however, have been responsible for the difference in ratings (Boureau and Doubrere, 1988). Our aims were not only to confirm this “miscalibration” between French patients and physicians, but more importantly to study some of the factors associated with it.

Multiple factors influence the pain perceptions of patients and their physicians. Some factors are particular to each individual, while others involve categories of individuals. Information about the impact of categorial factors is limited. Several researchers have reported gender differences among patients (Brennum et al., 1989; Hall and Davies, 1991; Jensen et al., 1992; Riley et al., 1998; Sarlani and Greenspan, 2002; Keogh and Herdenfeldt, 2002). Yet others did not find clear differences (Jensen et al., 1986; Lautenbacher and Strain, 1991; for a review, Hoffmann and Tarzian, 2001). In addition, patients’ age (Egbert, 1991; Pickering et al., 2002) and ethnicity (Edwards and Fillingim, 1999; Edwards et al., 2001; but not Todd et al., 1994) have been shown to affect patients’ pain judgments. Male and female physicians may perceive illness differently (Borum, 2002), but to our knowledge the effect of gender or other physician characteristics on their ratings of patients’ pain has not been studied.
We focused on several potential causes of miscalibration: the impact of, and interaction between, patient gender and physician gender; the obviousness of the cause of the pain; the degree of physician expertise in treating pain; and the extent of patient-physician interaction prior to the judgments. Weisse and colleagues (2001) discovered that gender can have a complex impact on physicians’ treatment decisions; it seemed likely to cause disjunctions in pain perception as well. The obviousness of the cause of pain might affect the degree of miscalibration through its impact on the patient’s anxiety (which is known to affect pain [Jones and Zachariae, 2002]), i.e., the less apparent the cause, the greater the patient’s anxiety, and the greater the difference between the patient’s and the physician’s rating. Physician expertise also might affect miscalibration, although it is unclear in which direction. On one hand, experts might calibrate better than novices because experts have learned to use more relevant cues; they perform better in diagnosis than novices (Raufaste et al. 1998). On other hand, experts’ perceptions of patients’ pain might differ more than novices’ because they have learned to correct for patients’ affective overreactions to pain; experts’ greater miscalibration would thus play a “functional” role (Bisseret, 1981). Finally, calibration might be better at the end of a patient-physician encounter than at the beginning as the patient and physician become better acquainted and the nature of the patient’s illness becomes clearer. We could not, however, study the likely impact of patients’ and physicians’ ethnicity because in France it is neither ethically nor legally acceptable to record information about patients’ ethnicity.

2. Methods

2.1 Participants

This prospective study took place in an emergency department middle-sized city in France. All the participants were volunteers. The patients were recruited by the experimenter in the waiting room. They had to (i) feel a pain; (ii) be aged over 14; (iii) be lucid; and (iv) know the French language well enough to understand the instructions fully. They ($N = 200$) consisted of 121 men and 79 women, aged between 14 and 83 (mean $= 32$; $SD = 14$). The “physicians” ($N = 48$) included two groups, the experts ($N = 15$) and the novices ($N = 33$). The experts were trained and certified in
emergency medicine. Their group consisted of 7 men and 8 women, aged between 32 and 50 (mean = 39; \( SD = 4.61 \)). The novices were medical students and residents who were receiving training in emergency care. Their group consisted of 18 men and 15 women, aged between 22 and 28 (mean = 25; \( SD = 2.83 \)).

2.2 Pain evaluation

The experiment did not disturb the usual way of handling patients. The patient’s first pain rating was made on arrival in the emergency department. The triage nurse asked the patient to assess his or her pain using a visual analog scale, by placing a mark on a line 10 cm in length at an appropriate distance between two endpoints labeled "no pain" and "most intense pain imaginable." This evaluation was placed in the patient's folder and, therefore, was accessible to the physician. After finishing the initial examination of the patient, the physician evaluated the patient’s pain using the VAS. It was made clear that, despite knowing the patient's rating, each physician had to give a pain rating according to his or her own perception of the patient’s pain. The physician’s rating was made out of sight of the patient. Both the patient and the physician made second, exit evaluations when the patient left the hospital. Final diagnoses were recorded for each case.

2.3 Analyses

The data were analyzed by means of analyses of variance (ANOVAs). The intra-case factors were the moment of rating (arrival or exit) and the rater (patient or physician). One of the inter-case factors was the obviousness of the cause of the patient’s pain. Two levels of obviousness were considered. The “obvious causes” (155 cases) included patients with environment-related, external causes (such as a blow, fall, fracture, or burn). These causes were apparent to the patient. The “non-obvious causes” (45 cases) included pains for which the causes were internal and less apparent to the patient (for example, abdominal pain, headache, and lower back pain). The cases were classified by the physicians who managed the patients. Other inter-case factors were physician expertise (novice vs. expert), physician gender, and patient gender. Patient age was controlled for by introducing it as a covariant in the ANOVAs.
Two dependent variables were used in the ANOVAs: the VAS ratings and a "pain miscalibration" measure. For a given case and a given moment (arrival or exit), pain miscalibration was computed as the physician’s pain rating minus the patient’s pain rating. A zero value indicates a perfect pain calibration. Negative and positive values indicate, respectively, “undervaluation” and “overvaluation” of the patient’s pain by the physician. These terms do not convey a judgment of quality because no outside standard is available to compare physicians’ and patients’ evaluations; they simply convey the direction of the difference, with the patient’s rating as baseline. For all factors, the possibility of distribution biases was controlled for using chi-square procedures.

Significance was defined as $p<.05$, two-tailed.

3. Results

3.1 Comparison of pain ratings in patients and physicians

As shown in Table 1, the mean pain ratings, averaging arrival and exit data, were 4.39 ($SE = 0.18$) in patients and 2.97 ($SE = 0.16$) in physicians. The mean “pain miscalibration” was significant, $m = -1.42$ ($SE = 0.15$), $t(171) = 9.67$, $P < .001$. Miscalibration remained significant whichever the moment of hospitalization (arrival or exit) or the obviousness of the cause.

At arrival, no factor had a significant effect on pain miscalibration (all $Ps > .05$). At exit, expertise had a significant effect: novices undervalued patient pain intensity less than experts, 1.26 ($SE = 0.20$) vs. 2.39 ($SE = 0.43$), $F(1,171) = 5.79$, $P = .017$. A 3-way interaction (Fig. 2) among physician expertise, physician gender, and patient gender approached significance ($F(1,171) = 3.8$, $P = .053$).
To investigate further the origin of pain miscalibration, we analyzed separately the factors of patients' and physicians' ratings.

3.2 Analyses of patients’ ratings

At arrival, only obviousness of cause influenced significantly patients' pain ratings. As expected, patients coming with obvious causes of their pain gave lower ratings than patients with non-obvious causes, 4.52 ($SE = 0.22$) vs. 6.37 ($SE = 0.32$), $F(1, 198) = 20.30, P < .001$. At exit, both obviousness of cause and physician expertise influenced patients’ ratings. Patients coming with obvious causes again provided lower ratings than patients with non-obvious causes, 3.70 ($SE = 0.20$) vs. 4.74 ($SE = 0.42$), $F(1,175) = 9.06, P = 0.03$. In addition, patients handled by novice physicians provided lower ratings than patients handled by experts, 3.82 ($SE = 0.20$) vs. 4.56 ($SE = 0.45$), $F(1,175) = 4.75, P = 0.031$. This result is most likely not an effect of expertise but a simple artifact of the normal functioning of the service: experts are called for difficult cases, in which pain may also be less reduced by analgesics.

3.3 Analyses of physicians’ ratings

At arrival, several variables contributed to physicians’ ratings. There was a slight but significant positive correlation between physicians' ratings and patients’ age, $r(200) = .153, P = 0.031$; this variable was controlled for in subsequent analyses. Physicians graded the pain of patients coming for obvious causes lower than the pain of patients coming for non-obvious causes, 3.34 ($SE = 0.19$) vs. 4.63 ($SE = 0.35$), $F (1,183) = 4.80, P = .03$. As the upper part of Figure 3 shows, there was also a significant 3-way interaction among patient gender, physician gender, and obviousness of cause ($F(1,183) = 4.44, P = .036$). More precisely, when the cause of pain was obvious, male physicians rated the pain of female patients lower than the pain of male patients, while female physicians’ ratings did not differ between patients' genders. When the cause was not
obvious, male physicians rated male patients' pain lower than female patients' pain, whereas female physicians rated female patients' pain lower than male patients' pain.

Insert Figure 3

At exit, patients’ age no longer had a significant effect, $F(155) = 3.31$, $P>0.05$. Otherwise physicians' ratings exhibited the same pattern of results as at arrival. They rated pain in patients with obvious causes lower than in those with no obvious causes, $2.32$ ($SE = 0.17$) vs. $3.29$ ($SE = 0.41$), $F(1,155) = 4.19$, $P=.042$, and, as shown in the lower part of Fig. 3, there was approximately the same interaction among patient gender, physician gender and obviousness of cause ($F(1, 155) = 4.92$, $P=.028$).

3.4 Other results

Table 2 shows the patients’ ratings according to the type of disease causing their pain. ANOVAs with pain ratings or miscalibration as dependent variables and type of disease as the independent variable found no effect (all $P$s > .05).

Between arrival and exit, there was a significant difference in pain ratings. Overall ($F(163) = 15.48; p<.001$) and when considering separately the ratings of patients and physicians.

Patients provided higher pain ratings at arrival than at exit, $4.89$ ($SE = 0.20$) vs. $3.95$ ($SE = 0.19$), $t(175) = 7.31$, $P<.001$. Among the physicians the pattern was similar: $3.38$ ($SE = 0.17$) at arrival vs. $2.56$ ($SE = 0.17$) at exit, $t(171) = 6.76$, $P<.001$. These numbers are not quite the same as those in Table 1 because the paired comparisons do not involve the same number of cases.

Participants’ pain assessments were, when averaging arrival and exit data, significantly higher for non-obvious than for obvious causes of pain, $3.28$ ($SE = 0.42$) vs. $2.32$ ($SE = 0.17$), $F(1, 163) = 10.68$, $P=.001$). The results were almost the same when considering patients and
physicians separately (Fig. 4). For patients there was a significant difference between obvious and non-obvious diseases, \( t(174) = -3.46, P = .001 \), and for physicians the difference was also significant, \( t(170) = -3.16, P = .002 \).

Finally, a control was made for the distribution of cases in the cells of the ANOVA tables. Results were constantly non-significant (all \( P_s > .05 \)). The interactions cannot, therefore, be explained by distribution biases.

4. Discussion

Attending physicians, residents, and medical student clerks in a French emergency room gave systematically lower ratings to patients’ pain than did the patients themselves. The mean difference between the ratings—the “miscalibration”—was 1.42 (SE = 0.15) on a 0-10 visual analog scale. This confirms the findings in France of Blettery and colleagues (1996), but without the confounding use of different scales (Boureau and Doubrere, 1988); it also confirms the findings in other countries (Hodgkins et al, 1985; Todd et al., 1994; Thomas et al, 1999). This degree of miscalibration is greater than or equal to the minimum clinically significant differences in levels of acute pain, also on a VAS scale of 0-10, of 0.9 (95% confidence interval, 0.6 to 1.3) found by Kelly (1998), of 1.3 (95% confidence interval, 1.0 to 1.7) found by Todd and colleagues (1996), and of 1.3 (95% confidence interval, 1.0 to 1.6) found by Gallagher and colleagues (2001).

We investigated four of the possible causes of this miscalibration. We found that the time of assessment had no effect on miscalibration: the mean difference at arrival (-1.33) was almost identical to that at exit (-1.38). The amount of interaction between patient and physician was limited, however, as would be expected in an emergency room. Our findings do not, therefore, disprove the potential importance of a lack of familiarity between patient and physician. It remains
possible that calibration might improve if the patient and physician became—or were already—more familiar with each other.

The obviousness of the cause of the pain had, as expected, a marked impact on pain ratings. When the cause was obvious, both patients and physicians rated pain lower at both arrival and exit than when the cause was not obvious. Whether the anxiety of the patient played a role in magnifying less obvious pain, as we speculated, cannot be determined from our data. Yet, in spite of its influence on pain ratings, obviousness of cause did not significantly affect miscalibration.

Gender played a complex role in pain ratings. Male physicians rated the pain of female patients lower than that of male patients when the cause was obvious but higher when the cause was not obvious. In contrast, female physician rated the pain of female patients the same as that of males when the cause was obvious and lower when the cause was not obvious. These interactions were robust since they occurred both at arrival in the emergency department and at exit from it. Such patterns are also seen in decisions about treating pain (Weisse et al, 2001), but are difficult to explain. Neither patient nor physician gender had a significant effect, however, on calibration.

The expertise of the physician did, however, influence calibration, though only at exit. The experts “undervalued” the patients’ pain more—were more miscalibrated—than the novices at the time of discharge from the emergency room. When rating pain, patients are likely to have in mind the worst pain they have felt, physicians the worst pain they have seen. This difference in reference points—greater for the more experienced physicians who have treated more cases of severe pain—is likely to induce a considerable miscalibration since, in making judgments, people are biased in the direction of their reference point (Kahneman, 1992; Estrada et al, 1997). Furthermore, experts have learned to identify which cues are relevant and to emphasize them while neglecting the irrelevant cues (Bisseret, 1981); novices are less able to do this. Novices are likely to give more weight to what experts learn is an unreliable cue, the patients’ assessment of their pain. In this sense, systematic “undervaluing” of patients’ pain is functional to the experts. Yet this explanation of the difference between experts and novices would predict that the difference in their
miscalibration would be greater at arrival than at discharge. The impact of physicians’ training and experience needs further study.

Our study has, of course, several limitations. First, the sample was composed of patients and physicians in a French emergency department. The findings are consistent with those in other countries but should, nonetheless, be generalized with care. Second, in accordance with routine practice in the emergency department, the physician was not blinded to the patients’ VAS pain rating at arrival. Such knowledge would tend to reduce miscalibration. We do not know if novices were more likely to look at patients’ ratings than were experts.

In spite of these limitations, our study has confirmed that physicians in emergency departments give systematically lower ratings of patients’ pain than do the patients themselves. More importantly, it has demonstrated that at least a part of this miscalibration is not functional. The reasons for miscalibration need further investigation.
References


Price DD, McGrath PA, Rafji A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983;17: 45-46.


Table 1

Pain intensity assessment (on a 0-10 scale) by patients and physicians according to moment of assessment and obviousness of cause

<table>
<thead>
<tr>
<th>Obvious causes</th>
<th>Patient</th>
<th></th>
<th></th>
<th>Physician</th>
<th></th>
<th></th>
<th>Difference*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>Arrival</td>
<td>145</td>
<td>4.52</td>
<td>0.22</td>
<td>3.34</td>
<td>0.19</td>
<td>-1.18</td>
<td>0.19</td>
</tr>
<tr>
<td>Exit</td>
<td>130</td>
<td>3.67</td>
<td>0.21</td>
<td>2.32</td>
<td>0.17</td>
<td>-1.34</td>
<td>0.017</td>
</tr>
<tr>
<td>Total</td>
<td>130</td>
<td>4.02</td>
<td>0.20</td>
<td>2.69</td>
<td>0.17</td>
<td>-1.32</td>
<td>0.16</td>
</tr>
<tr>
<td>Non-obvious causes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrival</td>
<td>55</td>
<td>6.37</td>
<td>0.32</td>
<td>4.63</td>
<td>0.35</td>
<td>-1.74</td>
<td>0.34</td>
</tr>
<tr>
<td>Exit</td>
<td>42</td>
<td>4.79</td>
<td>0.43</td>
<td>3.29</td>
<td>0.42</td>
<td>-1.5</td>
<td>0.34</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>5.54</td>
<td>0.36</td>
<td>3.83</td>
<td>0.37</td>
<td>-1.71</td>
<td>0.33</td>
</tr>
<tr>
<td>Total arrival</td>
<td>200</td>
<td>5.02</td>
<td>0.19</td>
<td>3.70</td>
<td>0.17</td>
<td>-1.33</td>
<td>0.17</td>
</tr>
<tr>
<td>Total exit</td>
<td>172</td>
<td>3.94</td>
<td>0.19</td>
<td>2.56</td>
<td>0.17</td>
<td>-1.38</td>
<td>0.15</td>
</tr>
<tr>
<td>Overall</td>
<td>172</td>
<td>4.39</td>
<td>0.18</td>
<td>2.97</td>
<td>0.16</td>
<td>-1.42</td>
<td>0.15</td>
</tr>
</tbody>
</table>

*all P’s<.001
Table 2
Pain intensity assessment by patients according to type of disease

<table>
<thead>
<tr>
<th>Type of disease</th>
<th>Number of patients</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Obvious causes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contusion</td>
<td>78</td>
<td>4.27</td>
<td>2.50</td>
</tr>
<tr>
<td>Dermatological disease&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14</td>
<td>4.83</td>
<td>3.30</td>
</tr>
<tr>
<td>Cut</td>
<td>25</td>
<td>4.32</td>
<td>1.75</td>
</tr>
<tr>
<td>Fracture, sprain</td>
<td>28</td>
<td>4.28</td>
<td>2.22</td>
</tr>
<tr>
<td><strong>Non-obvious causes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td>8</td>
<td>3.59</td>
<td>2.65</td>
</tr>
<tr>
<td>Abdominal pain&lt;sup&gt;b&lt;/sup&gt;</td>
<td>20</td>
<td>4.97</td>
<td>2.54</td>
</tr>
<tr>
<td>Lower back pain</td>
<td>13</td>
<td>4.64</td>
<td>1.87</td>
</tr>
<tr>
<td>Miscellaneous disorders&lt;sup&gt;c&lt;/sup&gt;</td>
<td>14</td>
<td>5.09</td>
<td>2.91</td>
</tr>
</tbody>
</table>

<sup>a</sup> Includes: urticaria, allergy, burn, insect bite or sting.

<sup>b</sup> Includes: appendicitis, constipation

<sup>c</sup> Includes: sore throat, respiratory illness, food poisoning, gastroenteritis
Fig. 1. Comparison of pain ratings by patients and physicians depending on moment of assessment
Fig. 2. Interaction of physician expertise, physician gender, and patient gender on pain miscalibration (measured at exit)
Fig. 3. Effects of obvious and non-obvious causes, patient gender, and physician gender on physicians' ratings (estimated means), at arrival (upper graphs) and at exit (lower graphs). The horizontal dotted lines represent mean pain intensity given by patients.
Fig. 4. Mean pain intensities for non-obvious and obvious causes of pain for patients and physicians (averaging arrival and exit measures).