Can We Rely on Retrospective Pain Assessments?

Charlotte Brauer, Jane F. Thomsen, Inger P. Loft, and Sigurd Mikkelsen

From the Department of Occupational Medicine, Copenhagen University Hospital, DK-2600 Glostrup, Denmark.

Received for publication December 28, 2001; accepted for publication October 9, 2002.

The objective of this paper was to study whether subjects in a workplace setting are able to assess the intensity of musculoskeletal pain retrospectively for a period of 3 months. The intensity of average pain and maximum pain in eight anatomic regions was assessed on a numeric rating scale. The results of 12 consecutive weekly pain recordings were compared with a final retrospective assessment of pain intensity covering the same 3-month period (119 subjects). The degree of agreement was good or excellent. The subjects were able to distinguish between the worst complaints and average complaints, and the subjective perception of aggravations or improvements corresponded to the direction of a change in score. Current complaint status slightly influenced the memory of pain. The reproducibility of the questionnaire was also assessed and showed kappa coefficients between 0.44 and 0.91 (36 subjects). The results suggest that subjects are able to accurately recall and rate the severity of pain or discomfort for a period of 3 months. These findings are of practical importance in epidemiologic studies, because they imply that retrospective reports on pain intensity are sufficiently reliable.

Abbreviation: PRIM, Project on Research and Intervention in Monotonous Work.

Pain is one of the most important outcome variables in epidemiologic studies of work-related musculoskeletal disorders. Most epidemiologic studies rely on a single retrospective assessment of pain obtained by questionnaire. Usually, such questionnaires cover the intensity and/or duration of regional discomfort or pain during a specific period of time—for example, the past 3, 6, or 12 months (1–3). However, pain may not be recalled accurately (4–6). To validate self-reports of pain, some studies have compared retrospective assessments with diary reports (7–9). However, studies differ in terms of source of study subjects (cancer patients, chronic pain patients, and migraine sufferers), recall period (2 weeks to 3 months), and conclusion (inaccurate pain assessments, accurate pain assessments, and overestimation of pain intensity). Epidemiologic studies of work-related musculoskeletal disorders are usually carried out in a workplace setting. When researchers are sampling from the community, most of the participants will be healthy or experiencing mild recurrent pain; only a minority will experience intense and persistent pain. Hence, the population in workplace studies differs greatly from chronic pain patients and from the above-mentioned prior studies. To our knowledge, no previous studies have validated retrospective self-reports of pain in a workplace setting. The objective of this study was to evaluate whether subjects in a workplace setting are able to assess the intensity of musculoskeletal pain retrospectively for a period of 3 months.

MATERIALS AND METHODS

This study was performed in the context of a Danish cohort study of work-related musculoskeletal disorders, the Project on Research and Intervention in Monotonous Work (the PRIM Study), which included 3,123 participants from 19 different companies. All workers in production facilities at these companies were invited to participate (10). The PRIM Study used a self-administered pain questionnaire covering a period of 3 months (11). Pain assessments were made on a scale of 0–9 for each of eight anatomic regions (the neck, the lower back, and the right and left shoulders, elbows, and wrists/hands). A set of four questions was asked for each anatomic region, accompanied by a diagram of the region concerned (figure 1). This questionnaire will be referred to below as the PRIM questionnaire. The question-
The neck

Indicate the degree of your complaints (pain or discomfort) in the neck.
(Put an 'X' in the box that fits best to your degree of complaint. Please mark a box in each line.)

Your worst complaints during the last three months: 0 1 2 3 4 5 6 7 8 9

Your average complaints during the last three months: 0 1 2 3 4 5 6 7 8 9

How much has the neck pain or discomfort caused you to reduce your activity (work or leisure) during the last three months?: 0 1 2 3 4 5 6 7 8 9

Your complaints during the last 7 days: 0 1 2 3 4 5 6 7 8 9

FIGURE 1. Questionnaire used in the PRIM* Study (n = 119), Denmark, 1995. An example of questions about neck complaints is shown. Identical questions in the same format were asked about the lower back and about the left and right shoulders, elbows, and wrists/hands (eight anatomic regions in total). (*PRIM, Project on Research and Intervention in Monotonous Work.)

naire and the intensity scale were slightly modified versions of the ones used by Von Korff et al. (2) and Manniche et al. (12).

At baseline in the PRIM Study, all participants filled in the PRIM questionnaire and underwent a physical examination. Participants for the present study were recruited consecutively when they attended this initial physical examination at one of the companies. They underwent the examination in random order, irrespective of the presence or absence of pain. A total of 146 subjects were asked to participate, and 128 (88 percent) accepted. After the initial physical examination, the participants completed 12 consecutive weekly recordings of complaints (pain or discomfort) during the past 7 days. The questions asked at weekly intervals were the same as the regional 7-day questions on the PRIM questionnaire, and responses were registered on the same scale of 0–9. In addition, the participants stated for each anatomic region whether their musculoskeletal complaints within the last week had improved, remained unchanged, or become worse in comparison with the preceding week. At week 13, the participants filled in the PRIM questionnaire, rating their worst and average complaints during the past 3 months and their current pain.

All participants were women who were employed full-time at a Danish bank. The rate of participation in the PRIM Study for this company was 74.2 percent. One half of the participants had repetitive data entry work, and the other half had varied office work. The mean age was 40 years (range, 23–63 years). The questionnaires were sent each week to the participants at the workplace. The participants completed and returned the questionnaire on the same day. They were allowed to fill in the questionnaire at work.

The reproducibility of the PRIM questionnaire was evaluated in a separate study among 36 subjects from the same bank using the test-retest method, with the questionnaire being administered three times (day 0, day 1, and day 7). The Scientific Ethical Committee of Copenhagen County accepted the study protocol (registration number KA 95087), and participants provided signed informed consent.

Statistical analysis

For each subject and anatomic region, the maximum of the 12 weekly scores was compared with the subject’s rating of her worst complaints during the previous 3 months, as registered in the final PRIM questionnaire. Likewise, the median of the 12 weekly scores was compared with self-rated average complaints during the previous 3 months. The median was chosen because of a skewed distribution. Agreement between the weekly questionnaires and the PRIM questionnaire was assessed by means of the percentage of agreement and the weighted kappa coefficient. Kappa values greater than 0.75 were considered to represent excellent agreement beyond chance; values between 0.40 and 0.75 were considered to represent fair to good agreement; and values less than 0.40 were considered to represent poor agreement (13).

The possible influence of the subject’s current complaint status on her retrospective assessments of worst and average pain was examined by Mantel-Haenszel trend analyses.
controlling for maximum or median pain as determined from the weekly questionnaires. The rejection of a zero correlation in these analyses \( p < 0.05 \) indicates that the retrospective assessments may be biased by present pain status.

Reproducibility was estimated by the percentage of agreement and the weighted kappa coefficient. Statistical analyses were conducted with SAS software (14).

RESULTS

A total of 119 persons (93 percent) filled in the final PRIM questionnaire and consequently completed the study. A total of 1,469 weekly questionnaires were completed (96 percent). The mean complaint score during the 12 weeks ranged between 0.42 and 1.45 on the scale from 0 to 9, depending on the anatomic region (table 1). The median was 1 for the neck and 0 for all other regions, corresponding to a positively skewed distribution in all regions. The total range of scores was represented. The complaint scores were rather stable week by week. Altogether, 79 percent of the scores did not change from one week to the next during the 12-week period. If subjects with no complaints at all in a given anatomic region were excluded, the corresponding figure was 59 percent. Eleven subjects had no complaints at all during the 12-week period. Depending on the region, 42–101 subjects reported scores above zero over the 12 weeks (table 2). Forty-nine participants had a change in score in one or several regions more than six times during the 12 weeks. The proportion of responses indicating aggravated pain was equal to the proportion of responses indicating improvement during the 12 weeks. The mean complaint scores were unchanged from the beginning of the study to the end. On the PRIM questionnaire, only four responses (1 percent of the responses with a score greater than zero) had a higher score for average complaints than for the worst complaints.

Comparisons between the maximum score on the weekly questionnaires during the 12 weeks and the self-rated worst pain in the final PRIM questionnaire showed that the extent of agreement depended on the anatomic region. The poorest agreement was found for regions where complaints were reported frequently (the neck and lower back). The percentage of identical answers varied between 45 percent (neck) and 74 percent (right elbow), and the kappa coefficients ranged from 0.63 to 0.80. Since some of this agreement could be due to subjects’ not having complaints, the same analyses were conducted for each region while leaving out subjects who had no complaints at all in that specific region during the 12 weeks. When subjects without complaints were excluded, the percentage of identical answers was 5–37 percent (table 2), and 57–74 percent of the responses agreed within a range of one score unit. The kappa coefficients ranged from 0.39 to 0.60 (table 2). On the 0–9 scale, the scores on the PRIM questionnaire differed from those on the weekly questionnaires by –0.24 to –0.59 score units, depending on the region, the weekly scores being higher than the final score (underestimating).

Better agreement was found when the median of the weekly scores was compared with the self-rated average score in the final PRIM questionnaire. Among all subjects, complete agreement was found here for 60–85 percent of the responses, and the kappa coefficients ranged from 0.66 to 0.79. When subjects without complaints were excluded, complete agreement was found for 46–61 percent of the responses, and 84–90 percent of all responses agreed within a range of one score unit (table 2). The kappa coefficients ranged from 0.52 to 0.69 (table 2). The mean difference on the scale was 0.08–0.25 score units, the final scores being higher than the weekly scores (overestimating).

The Mantel-Haenszel analyses of the relation between present pain status and retrospective assessments of worst and average pain, controlling for maximum and medium pain, respectively, showed significant positive correlations for all anatomic regions in both sets of analyses. Depending on the region, the mean score differences between worst and maximum pain ranged from \(-0.19\) to \(-0.72\) (mean = \(-0.42\)), the scores on the PRIM questionnaire differed from those on the weekly questionnaires by \(-0.24\) to \(-0.59\) score units, depending on the region, the weekly scores being higher than the final score (underestimating).
TABLE 2. Agreement between reports of pain on the PRIM* questionnaire and weekly pain reports covering the same period (n = 119), The PRIM Study, Denmark, 1995

<table>
<thead>
<tr>
<th>Anatomic region</th>
<th>Worst complaints vs. maximum complaints†</th>
<th>Average complaints vs. median complaints‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete agreement (%)</td>
<td>Agreement within a range of one score unit (%)</td>
</tr>
<tr>
<td>All subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neck</td>
<td>45.4</td>
<td>78.2</td>
</tr>
<tr>
<td>Lower back</td>
<td>51.3</td>
<td>71.4</td>
</tr>
<tr>
<td>Right shoulder</td>
<td>52.9</td>
<td>73.9</td>
</tr>
<tr>
<td>Left shoulder</td>
<td>57.1</td>
<td>80.7</td>
</tr>
<tr>
<td>Right elbow</td>
<td>73.9</td>
<td>83.2</td>
</tr>
<tr>
<td>Left elbow</td>
<td>68.9</td>
<td>87.4</td>
</tr>
<tr>
<td>Right wrist/hand</td>
<td>69.7</td>
<td>88.2</td>
</tr>
<tr>
<td>Left wrist/hand</td>
<td>71.4</td>
<td>88.2</td>
</tr>
<tr>
<td>Subjects with complaints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neck (n = 101)</td>
<td>28.9</td>
<td>68.0</td>
</tr>
<tr>
<td>Lower back (n = 93)</td>
<td>35.6</td>
<td>59.8</td>
</tr>
<tr>
<td>Right shoulder (n = 82)</td>
<td>22.8</td>
<td>58.2</td>
</tr>
<tr>
<td>Left shoulder (n = 70)</td>
<td>20.9</td>
<td>65.7</td>
</tr>
<tr>
<td>Right elbow (n = 52)</td>
<td>36.2</td>
<td>57.4</td>
</tr>
<tr>
<td>Left elbow (n = 42)</td>
<td>5.1</td>
<td>61.5</td>
</tr>
<tr>
<td>Right wrist/hand (n = 62)</td>
<td>36.8</td>
<td>73.7</td>
</tr>
<tr>
<td>Left wrist/hand (n = 46)</td>
<td>25.6</td>
<td>69.8</td>
</tr>
</tbody>
</table>

* PRIM, Project on Research and Intervention in Monotonous Work.
† Score in the final PRIM questionnaire about the worst complaints minus the maximum score in the weekly questionnaires.
‡ Score in the final PRIM questionnaire about average complaints minus the median of the scores in the weekly questionnaires.
§ CI, confidence interval.

The reproducibility study (test-retest study) contained 1,128 responses out of a total of 1,152 obtainable responses (36 subjects with four questions on each of the eight regions). The assessment of reproducibility showed that 711 responses (63 percent) were identical on the three occasions and that 1,044 responses (93 percent) were identical on at least two of the days. The weighted kappa coefficients showed good agreement for all questions and all regions (mean $\kappa = 0.61–0.75$). For subjects with complaints at day 0, the kappa coefficients for the questions about the worst complaints and average complaints ranged from 0.44 to 0.91.

DISCUSSION

This study evaluated the accuracy of retrospective reports of pain intensity in a workplace setting. Good or excellent agreement was found regarding average pain as well as worst pain. Average pain intensity seemed somewhat easier to remember than worst pain, but it was slightly influenced by current pain status, which is in accordance with previous research (7, 8). Subjects tended to underestimate their worst pain and to overestimate their average pain when rating it retrospectively at the end of the study. The subjects were capable of distinguishing between their worst pain and average pain. Furthermore, the direction of a change in score corresponded very well to the subjective perception of aggravations or improvements from one week to the next. The responses were satisfactorily reproducible. These results suggest that the subjects were able to accurately recall the intensity of their average pain as well as their worst pain for a 3-month period and to express these aspects retrospectively on a scale in a questionnaire.

A study design using daily records of pain would most likely have given us a more precise pain assessment than weekly recordings. We considered administering a diary prospectively during the 12-week period, but it was not practicable, because it would have required many more resources.
from the participating firm. Furthermore, earlier diary studies have shown poor compliance with rapidly decreasing numbers of respondents in workplace settings (15, 16). An essential strength of our study was the high percentage of subjects who completed the study. This high compliance in our study was probably due to the fact that the weekly questionnaire was short, it took only a few minutes to complete, and the workers were allowed to fill it in at work.

A limitation of our study could be that the study population consisted of mainly healthy persons. However, most subjects reported symptoms of some degree during the study period, and the scores changed more than six times for almost half of the subjects. Nevertheless, subjects without complaints could have inflated the kappa values for agreement between worst and maximum complaints and average and median complaints. When subjects with no complaints were left out of the analysis, the kappa coefficients did in fact decrease, but they still showed good agreement.

The design of this study could have introduced reporting bias because the subjects focused on symptoms when they were administered a questionnaire every week. The introduction of reporting bias could have resulted in a shift towards more complaints during the study period, but the mean of the complaint scores did not differ from the beginning of the study to the end. When they are recording symptoms every week, subjects soon learn how to fill in the questionnaire, and this could have resulted in artificially high agreement estimates in the present study. However, the questions concerned eight anatomic regions to be rated on a 10-point scale. For pain levels above zero, we believe that learned responses are not very likely. Furthermore, Salovey et al. (8) examined the influence of filling in a diary on accuracy of recall using a control group of subjects who did not keep diaries. They concluded that keeping a diary does not seem to affect subsequent recollections of pain intensity (8).

Hence, there is no reason to believe that this issue should be of significant concern in the present study.

The reproducibility of the questionnaire in our study was satisfactory and was at the same level as the reproducibility of other questionnaires used in epidemiologic studies of musculoskeletal disorders (1, 17–19). In the present study, the high degree of reproducibility could be due to subjects’ reporting a score of zero, but the kappa coefficients still showed good agreement when subjects with no complaints were excluded. The short interval between administrations of the questionnaire could have resulted in subjects’ recalling previous responses. However, no difference was found with regard to the 3-month questions when day 0 was compared with day 1 instead of day 7.

In conclusion, the present study suggests that subjects are able to recall and rate the severity of their pain for a period of 3 months. These findings are of practical importance in epidemiologic studies, because they imply that retrospective reports on pain intensity are sufficiently reliable.

**REFERENCES**

6. Roche PA, Gijsbers K. A comparison of memory for induced