

Can we assess baseline pain and global health retrospectively?

P.M. ten Klooster¹, K.W. Drossaers-Bakker^{1,2}, E. Taal¹, M.A.F.J. van de Laar^{1,2}

¹*Institute for Behavioural Research, University of Twente, Enschede;*

²*Department of Rheumatology, Medisch Spectrum Twente, Enschede, The Netherlands.*

Abstract

Objective

To study the agreement between patients' actual baseline assessments of pain and global health before treatment and retrospective assessments collected 2 weeks after treatment.

Methods

Data were collected in a prospective study of 200 rheumatology outpatients treated with a local corticosteroid injection. At baseline and 2-week follow-up, localized pain and global health were assessed on 100 mm visual analogue scales. The follow-up questionnaire was extended with a retrospective assessment of pain and global health before treatment.

Results

At follow-up patients slightly overestimated the severity of pain and global health before treatment. Actual and retrospective assessments were adequately correlated (pain: $r_s = 0.73$; global health: $r_s = 0.67$). Bland-Altman analysis showed that both pain and global health were characterized by high intra-individual variation between actual and retrospective assessments, with the 95% limits of agreement (-37.3 to 32.3 mm for pain and -49.7 to 37.8 mm for global health) far exceeding proposed values for minimal clinically important differences.

Conclusion

Over a 2-week interval, patients' retrospective assessments of baseline pain and global health are fairly accurate and adequately correlated with actual baseline scores. At the group level, retrospective assessments can provide acceptable data on baseline pain and global health. The wide variability between actual and retrospective assessments, however, indicates that even over short time intervals there is poor individual agreement between the two methods.

Key words

Prospective measurement, retrospective measurement, recall, visual analogue scale, pain, global health.

Peter M. ten Klooster, MSc; K. Wiepke Drossaers-Bakker, MD, PhD; Erik Taal, PhD; Mart A.F.J. van de Laar, MD, PhD, Professor.

Please address correspondence to: Peter M. ten Klooster, Institute for Behavioural Research, Faculty of Behavioural Sciences, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. E-mail: p.m.tenklooster@utwente.nl

Received on March 24, 2006; accepted in revised form on July 27, 2006.

© Copyright CLINICAL AND EXPERIMENTAL RHEUMATOLOGY 2007.

Introduction

The assessment of changes in patient-perceived pain and global health plays a key role in both clinical trials and routine practice. In clinical practice, physicians often rely on patients' retrospective accounts of previous states or perceived changes in state to evaluate the effectiveness of treatment. In clinical trials, on the other hand, retrospective measurement is usually discouraged (1) and patients' retrospective perceptions of change or baseline states are rarely measured. However, prospective research designs are usually expensive and time consuming, and sometimes impractical or even impossible (2, 3). In these situations, retrospective assessments of baseline health states collected at follow-up could provide an attractive alternative, provided that these assessments yield reasonably accurate data.

The main concern with retrospective research designs is the extent to which patients are able to accurately recall their symptoms or overall health before treatment (1, 4-6). In patients with arthritis, pain is the most prominent symptom and is best measured with a visual analogue scale (VAS) (7). Several studies have investigated the accuracy of pain recall, but their findings vary considerably. Whereas some found that patients are quite able to recall previous pain states (8-12), others concluded that recall is inaccurate or systematically biased (13-19).

Besides patient-perceived pain, the VAS for patient global health status has become a central outcome measure in rheumatology. In contrast to pain recall, however, very little is known about patients' ability to remember previous global health states. Two studies that have examined similar constructs, indicate that recall of global health may be susceptible to error and bias (20, 21). Moreover, it would seem plausible that patients generally will have more difficulties in accurately recalling general health states than concrete symptoms such as pain (6, 20).

One important factor in recalling pre-treatment pain or global health is the time between the actual and the retrospective assessment. Most studies

on pain recall in chronic pain patients have used long time intervals between both assessments, ranging from several months to years. Since errors in pain recall generally get worse with the passage of time (22-24), retrospective assessments after a relatively short time interval may yield sufficiently reliable data.

Finally, an additional drawback of studies comparing actual and retrospective assessments is their reliance on comparison of means and correlation analysis, which are likely to overestimate the actual agreement. A more informative measure of agreement was developed by Bland and Altman (25), who suggested to plot the absolute individual differences between both methods against their mean and comparing their 95% limits of agreement with a clinically acceptable difference between the two methods.

The aim of the present study was therefore to examine the agreement between patients' actual assessments of baseline pain and global health and retrospective assessments collected after a relatively short period of 2 weeks, using additional Bland-Altman analyses.

Materials and methods

Patients

The data for this study were collected at the outpatient rheumatology clinic. Arthritis patients older than 16 years who experienced localized musculoskeletal pain and who were treated with a local corticosteroid injection were eligible for inclusion. Informed consent was obtained from all patients.

Measurements

The study consisted of two serial assessments. The baseline assessment was completed during the patient's visit at the outpatient clinic, just before the injection procedure. The 2-week follow-up questionnaire was mailed to the patients. At baseline and follow-up, average localized pain and global health in the past week were measured on 100 mm, unmarked VASs, anchored by "no pain – unbearable pain" and "very well – very poor." At the end of the follow-up questionnaire, patients

were asked to recall their average level of pain and global health in the week before the injection on identical VASs (e.g., In general, how much pain did you experience in the affected joint in the week before the local injection?).

Statistical analysis

Normal distribution of age, disease duration, VAS scores and differences between actual and retrospective VAS scores was examined by the Kolmogorov-Smirnov (K-S) test and inspection of normality plots. Since several VAS scores were not normally distributed (K-S, $p < 0.05$), all comparisons were conducted using non-parametric tests. Differences between actual and retrospective assessments were tested using paired Wilcoxon signed ranks tests, with Hodges-Lehmann estimates for median differences and 95% confidence intervals (CI). Correlations between actual and retrospective assessments were expressed using Spearman's rank correlation coefficient (r_s). Individual agreement between the two methods of baseline assessment was assessed by plotting the difference between both assessments against their mean (25).

Results

Patient characteristics

In the period between May and December 2004, 200 consecutive patients were recruited. Six patients (3%) did not return the follow-up questionnaire and 13 patients (6.5%) did not complete the retrospective assessments. Data from these patients were excluded from further analyses. Baseline characteristics of the excluded patients did not differ from the included patients. The descriptive characteristics of the 181 included patients are shown in Table I.

Difference between actual and retrospective baseline assessments

Two weeks after treatment patients slightly overestimated the severity of their baseline pain (estimated median difference -2.5, 95% CI: -4.5 to 0) and global health (estimated median difference -5.0, 95% CI: -8.0 to -2.5). The difference between actual and

retrospective assessments was correlated with the respective actual level of pain or health before treatment (pain $r_s = 0.28$, 95% CI: 0.14 to 0.41; global health $r_s = 0.34$, 95% CI: 0.21 to 0.47) and the prospective change in pain or health between baseline and follow-up (pain $r_s = 0.27$, 95% CI: 0.13 to 0.40; global health $r_s = 0.50$, 95% CI: 0.38 to 0.60). Patients with low baseline pain or global health tended to exaggerate its severity afterwards, while patients with high baseline scores tended to underestimate baseline states. Moreover, prospectively improved patients tended to underestimate baseline severity, whereas patients whose condition deteriorated tended to overestimate baseline severity. Differences between both methods of baseline assessment were not significantly correlated with patients' baseline characteristics and present level of pain or health status at the moment of recall.

Correlation between actual and retrospective baseline assessments

The retrospective assessments of baseline pain and global health correlated adequately with the actual baseline assessments (pain $r_s = 0.73$, 95% CI: 0.65 to 0.79; global health $r_s = 0.66$, 95% CI: 0.57 to 0.74).

Agreement between actual and retrospective baseline assessments

Bland-Altman analysis of the difference between actual and retrospective baseline assessments against the mean of both methods (Figs. 1 and 2) confirmed that the systematic bias between actual and retrospective assessments was small. Both pain and global health were, however, characterized by high intra-individual variation, with the 95% limits of agreement ranging from -37.3 to 32.3 mm for pain and 49.7 to 37.8 mm for global health.

Discussion

Prospective measurement of changes in patient-reported outcomes such as pain and global health is the gold standard for clinical research. In this study we investigated whether patients' baseline pain and global health states can be reliably assessed retrospectively. The results of the study indicate that although retrospective assessments of baseline pain and global health are fairly accurate at the group level and adequately correlated with actual baseline scores, there is poor agreement within individual patients.

The results showed that, as a group, patients tended to overestimate both the severity of baseline pain and global

Table I. Patient baseline characteristics and actual baseline, follow-up and retrospective baseline VAS scores from the included patients.

Age (years), median (IQR)	60 (51–71)
Female, n (%)	128 (71)
Primary diagnosis	
Rheumatoid arthritis, n (%)	67 (37.0)
Osteoarthritis, n (%)	33 (18.2)
Psoriatic arthritis, n (%)	16 (8.8)
Tendinitis / bursitis, n (%)	15 (8.3)
Other, n (%) [*]	50 (27.6)
Disease duration (years), median (IQR)	4 (0–11)
Baseline pain (VAS, 0–100 mm), median (IQR)	61.0 (46.0–78.0)
Follow-up pain (VAS, 0–100 mm), median (IQR)	25.5 (10.0–47.0)
Retrospective baseline pain (VAS, 0–100 mm), median (IQR)	67.0 (45.5–79.0) [†]
Baseline global health (VAS, 0–100 mm), median (IQR)	38.0 (10.5–59.0)
Follow-up global health (VAS, 0–100 mm), median (IQR)	31.0 (9.0–48.0)
Retrospective baseline global health (VAS, 0–100 mm), median (IQR)	46.0 (20.5–63.5) [‡]

IQR: interquartile range, VAS: visual analogue scale.

^{*}Includes several diagnoses such as polymyalgia rheumatica, shoulder complaints, and gout.

[†]Significantly different from actual baseline pain, Wilcoxon (2-tailed), $Z = -2.02$, Hodges-Lehmann estimated median difference -2.5 mm, 95% CI: -4.5 to 0, $p = 0.044$.

[‡]Significantly different from actual baseline global health, Wilcoxon (2-tailed), $Z = -3.60$, Hodges-Lehmann estimated median difference -5.0 mm, 95% CI: -8.0 to -2.5, $p < 0.001$.

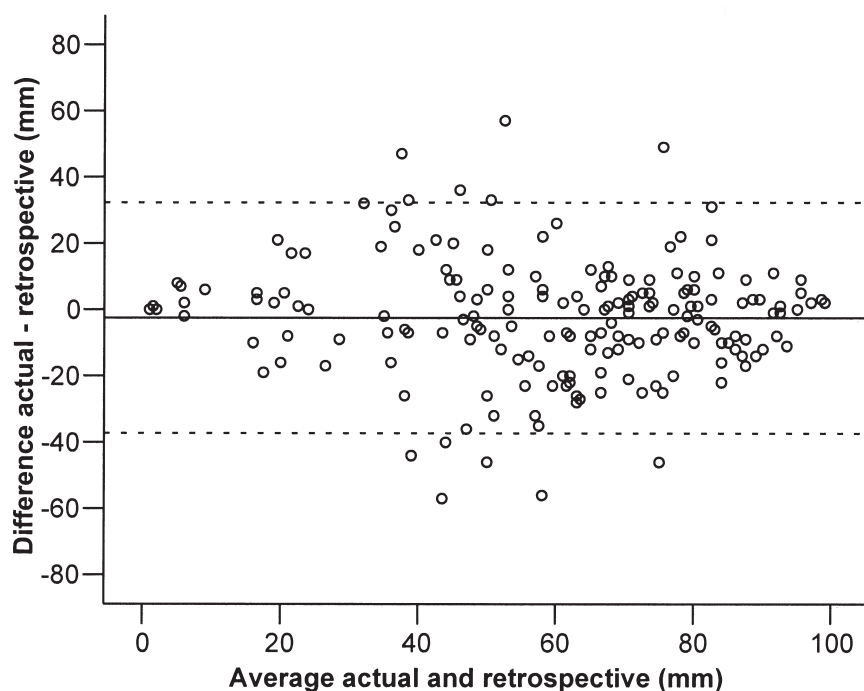


Fig. 1. Individual agreement between actual and 2-week retrospective assessments of baseline pain on the VAS. The horizontal solid line represents the mean difference (bias) between both change scores (-2.5 mm). The dashed lines represent the 95% limits of agreement (mean difference \pm 1.96 standard deviation of the difference), ranging from -37.3 to 32.3 mm.

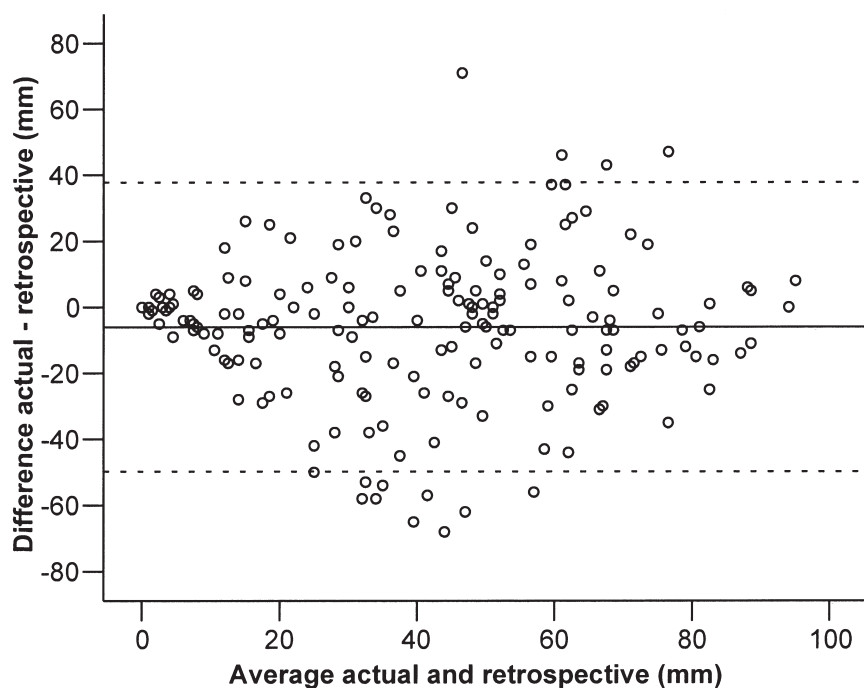


Fig. 2. Individual agreement between actual and 2-week retrospective assessments of baseline global health on the VAS. The horizontal solid line represents the mean difference (bias) between both change scores (-6.0 mm). The dashed lines represent the 95% limits of agreement (mean difference \pm 1.96 SD of the difference), ranging from -49.7 to 37.8 mm.

health retrospectively. This tendency of patients to overestimate the severity of their pre-treatment situation has been reported in previous studies (13, 14,

17, 19-22). Two possible theoretical explanations have been proposed for this systematic bias in recall. The first explanation is motivational bias

(e.g., cognitive dissonance or social desirability), where patients who have undergone a treatment will be motivated to exaggerate the benefits of that treatment (4). The second explanation is response shift bias, which refers to a change in the meaning of one's self-evaluation of their health status as a result of a change in their internal standards, values or conceptualization of the measured construct (26). However, since the patients in this study were asked to recall their baseline status, as opposed to give a renewed judgment with the insights they have now (a so-called then-test), true response shift could not be assessed.

In accordance with other pain studies (16, 22), the differences between actual and retrospective assessments in this study were related to the actual baseline level of pain or global health and its prospective change. However, the accuracy of recall was not influenced by the present level of pain or global health at the moment of recall, as previously suggested (11, 22, 27, 28).

Although the group differences between actual and retrospective assessments in this study were statistically significant, their small magnitude suggests that they are not likely to be of clinical significance. Several studies have demonstrated that patient-perceived pain and global health on the VAS have poor test-retest reliability and high random measurement error compared to multi-item measures (29, 30). The observed differences on the VAS can therefore not be reliably distinguished from random error.

The small average differences between retrospective and actual assessments and the adequate correlation between them, would suggest that retrospective assessments after a 2-week period can capture quite reliable data on baseline pain and global health at the group level. However, within individual patients, the difference between actual and retrospective assessments proved to be highly variable and subject to error. Although there are no established rules for clinically acceptable differences between the two methods, using retrospective assessments should at the least not lead to different conclusions about

the efficacy of treatment. In this study, however, the 95% limits of agreement of the Bland-Altman plots far exceeded proposed values of approximately 15-20 mm for minimal clinically important improvements in pain and global health (31, 32). Using patients' retrospective instead of actual baseline assessments to measure change over treatment, could thus result in a high number of patients being incorrectly classified as having significantly improved or deteriorated.

Although several previous studies have examined patients' recall of pain, this study is one of the first to examine patients' ability to recall previous global health states. The findings support the assumption that patients' memory of global health status is even more problematic than their recall of pain. Recall bias was larger in global health assessments, and patients' actual and retrospective assessments of global health were less strongly correlated. Moreover, Bland-Altman analyses indicated that actual and retrospective assessments of global health were more susceptible to intra-individual variability. This suggests that patients have more trouble remembering previous global health states than previous pain states.

Some reservations should be made regarding the generalizability of the present findings. Firstly, the study population included patients with heterogeneous diagnoses. Since pain and global health are known to vary across different rheumatic diseases, the findings may not be applicable to specific rheumatic conditions. Moreover, since most patients experienced a major improvement in pain at the 2-week follow-up, the findings may not apply to stable pain recall. A further limitation of this study is that it is not clear whether patients at follow-up truly recalled their baseline pain and global health status, or tried to recall the physical position of their mark on the baseline VAS. Moreover, patients completed the baseline questionnaire in the clinic and in the presence of an investigator, whereas the follow-up questionnaire was mailed the patient's home. The contexts in which the data

were collected may have affected patients' reporting (33). Finally, the study design did not incorporate the influence of personality characteristics or psychosocial factors, which can contribute to the variability in the memory of previous pain or health states (14, 34-38).

In conclusion, retrospective assessments can provide fairly reliable data on aggregate baseline pain and global health and can be used for descriptive and exploratory purposes. However, at the individual level there is poor agreement between actual and retrospective assessments of baseline health states. The unacceptably high variability in the magnitude and direction of the differences confirms that even over relatively short time intervals, retrospective assessments should not be used as substitutes for individual baseline status or to measure individual changes over treatment in clinical trials.

Acknowledgements

The authors thank all the respondents for their participation in the study and the rheumatologists of Medisch Spectrum Twente for their help in the inclusion of the patients.

References

1. FELSON DT, ANDERSON JJ, BOERS M *et al.*: The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum* 1993; 36: 729-40.
2. EMBERTON M, CHALLANDS A, STYLES RA, WIGHTMAN JA, BLACK N: Recollected versus contemporary patient reports of pre-operative symptoms in men undergoing transurethral prostatic resection for benign disease. *J Clin Epidemiol* 1995; 48: 749-56.
3. KREULEN GJ, STOMMEL M, GUTEK BA, BURNS LR, BRADEN CJ: Utility of retrospective pretest ratings of patient satisfaction with health status. *Res Nurs Health* 2002; 25: 233-41.
4. ASELTINE RH, JR, CARLSON KJ, FOWLER FJ, JR, BARRY MJ: Comparing prospective and retrospective measures of treatment outcomes. *Med Care* 1995; 33: AS67-76.
5. FISCHER D, STEWART AL, BLOCH DA, LORIG K, LAURENT D, HOLMAN H: Capturing the patient's view of change as a clinical outcome measure. *JAMA* 1999; 282: 1157-62.
6. HERRMANN D: Reporting current, past, and changed health status. What we know about distortion. *Med Care* 1995; 33: AS89-94.
7. SOKKA T: Assessment of pain in rheumatic

diseases. *Clin Exp Rheumatol* 2005; 23: S77-84.

8. SINGER AJ, KOWALSKA A, THODE HC, JR: Ability of patients to accurately recall the severity of acute painful events. *Acad Emerg Med* 2001; 8: 292-5.
9. BABUL N, DARKE AC, JOHNSON DH, CHARRON-VINCENT K: Using memory for pain in analgesic research. *Ann Pharmacother* 1993; 27: 9-12.
10. BEESE A, MORLEY S: Memory for acute pain experience is specifically inaccurate but generally reliable. *Pain* 1993; 53: 183-9.
11. SALOVEY P, SMITH AF, TURK DC, JOBE JB, WILLIS GB: The accuracy of memory for pain: not so bad most of the time. *Am Pain Soc J* 1993; 2: 184-91.
12. BRAUER C, THOMSEN JF, LOFT IP, MIKKELSEN S: Can we rely on retrospective pain assessments? *Am J Epidemiol* 2003; 157: 552-7.
13. LINTON SJ, MELIN L: The accuracy of remembering chronic pain. *Pain* 1982; 13: 281-5.
14. JAMISON RN, SBROCCO T, PARRIS WC: The influence of physical and psychosocial factors on accuracy of memory for pain in chronic pain patients. *Pain* 1989; 37: 289-94.
15. LIU WH, AITKENHEAD AR: Comparison of contemporaneous and retrospective assessment of postoperative pain using the visual analogue scale. *Br J Anaesth* 1991; 67: 768-71.
16. BRYANT RA: Memory for pain and affect in chronic pain patients. *Pain* 1993; 54: 347-51.
17. LINGARD EA, WRIGHT EA, SLEDGE CB: Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. *J Bone Joint Surg Am* 2001; 83-A: 1149-56.
18. DAWSON EG, KANIM LE, SRA P *et al.*: Low back pain recollection versus concurrent accounts: outcomes analysis. *Spine* 2002; 27: 984-93.
19. PELLISE F, VIDAL X, HERNANDEZ A, CEDRASCHI C, BAGO J, VILLANUEVA C: Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine* 2005; 30: 365-8.
20. MANCUSO CA, CHARLSON ME: Does recollection error threaten the validity of cross-sectional studies of effectiveness? *Med Care* 1995; 33: AS77-88.
21. BERNHARD J, LOWYA, MAIBACH R, HURNY C: Response shift in the perception of health for utility evaluation: an explorative investigation. *Eur J Cancer* 2001; 37: 1729-35.
22. FEINE JS, LAVIGNE GJ, DAO TT, MORIN C, LUND JP: Memories of chronic pain and perceptions of relief. *Pain* 1998; 77: 137-41.
23. MCGORRY RW, WEBSTER BS, SNOOK SH, HSIANG SM: Accuracy of pain recall in chronic and recurrent low back pain. *J Occup Rehabil* 1999; 9: 169-78.
24. JENSEN MP, CHEN C, BRUGGER AM: Postsurgical pain outcome assessment. *Pain* 2002; 99: 101-9.
25. BLAND JM, ALTMAN DG: Statistical methods for assessing agreement between two meth-

- ods of clinical measurement. *Lancet* 1986; 1: 307-10.
26. SPRANGERS MA, SCHWARTZ CE: Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med* 1999; 48: 1507-15.
 27. EICH E, REEVES JL, JAEGER B, GRAFF-RADFORD SB: Memory for pain: relation between past and present pain intensity. *Pain* 1985; 23: 375-80.
 28. SMITH WB, SAFER MA: Effects of present pain level on recall of chronic pain and medication use. *Pain* 1993; 55: 355-61.
 29. LASSERE MN, VAN DER HEIJDE D, JOHNSON KR, BOERS M, EDMONDS J: Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials. *J Rheumatol* 2001; 28: 892-903.
 30. RUSSELL AS, CONNER-SPADY B, MINTZ A, MAKSYMOWYCH WP: The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *J Rheumatol* 2003; 30: 941-7.
 31. TUBACH F, RAVAUD P, BARON G *et al.*: Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005; 64: 29-33.
 32. FARRAR JT, YOUNG JP, LAMOREAUX L, WERTH JL, POOLE RM: Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001; 94: 149-58.
 33. CAMPBELL R, QUILTY B, DIEPPE P: Discrepancies between patients' assessments of outcome: qualitative study nested within a randomised controlled trial. *BMJ* 2003; 326: 252-3.
 34. LEFEBVRE JC, KEEFE FJ: Memory for pain: the relationship of pain catastrophizing to the recall of daily rheumatoid arthritis pain. *Clin J Pain* 2002; 18: 56-63.
 35. GEDNEY JJ, LOGAN H: Memory for stress-associated acute pain. *J Pain* 2004; 5: 83-91.
 36. TASMUTH T, ESTLANDERB AM, KALSO E: Effect of present pain and mood on the memory of past postoperative pain in women treated surgically for breast cancer. *Pain* 1996; 68: 343-7.
 37. PORZELIUS J: Memory for pain after nerve-block injections. *Clin J Pain* 1995; 11: 112-20.
 38. LINTON SJ: Memory for chronic pain intensity: correlates of accuracy. *Percept Mot Skills* 1991; 72: 1091-5.