Arthritis Care & Research
Vol. 71, No. 6, June 2019, pp 768–776
DOI 10.1002/acr.23711

American College of Rheumatology
Empowering Rheumatology Professionals

Daily Pain Measurements and Retrospective Pain Measurements in Hip Osteoarthritis Patients With Intermittent Pain

Carolien H. Teirlinck, Dieke S. Sonneveld, Sita M. A. Bierma-Zeinstra, and Pim A. J. Luijsterburg

Objective. To examine the value of daily pain measurements in patients with hip osteoarthritis (OA), and whether the reliability of retrospective measurements was lower in patients with intermittent pain than in patients with more constant pain.

Methods. We used data from a randomized controlled trial that investigated the effectiveness of general practitioner care plus exercise therapy in 203 patients with hip OA. During the first 6 weeks, patients scored their pain each day. These daily measurements were available for 185 patients. At 6-week follow-up, patients filled in a questionnaire rating their pain during the previous week. We examined whether the daily measurements provided results for pain other than those provided by retrospective measurements, using a linear mixed-effects model. We also explored differences between subgroups, based on the frequency and severity of intermittent pain, during the pain course and reliability between retrospective measurements and daily measurements.

Results. Daily measurements showed no different effect of exercise therapy on pain compared with retrospective measurements. We found statistical differences (by analysis of variance) during the course of pain between the subgroups based on the intensity of intermittent pain. Reliability between retrospective and daily measurements was lower in the subgroup with severe intermittent pain (Cronbach's α = 0.642) than in other subgroups (Cronbach's α >0.843).

Conclusion. In this specific trial, daily measurements did not yield more precise or additional information compared with retrospective measurements at the 6-week follow-up. However, reliability of retrospective measurements may be lower in patients with a higher intensity of intermittent pain.

INTRODUCTION

Pain is one of the most important symptoms in patients with osteoarthritis (OA) of the hip (1). Therefore, pain is one of the main outcomes in OA clinical trials (2). Besides using general pain measures such as the visual analog scale and the numerical rating scale (NRS), OA researchers also use disease-specific questionnaires, such as the pain subscales of the Western Ontario and McMaster Universities Arthritis Index and the Hip Disability and Osteoarthritis Outcome Score (HOOS).

Some of these questionnaires ask the patient to rate his or her pain on the present day, while others ask patients to rate pain as a mean of the previous several days or weeks (recall period). During follow-up, most trials will use several time points to measure pain in order to provide a course of the pain. However, it is unclear how many time points are needed to follow the course accurately and, in addition, to be sure not to miss a temporal effect between the time points. Although keeping a diary can be a burden for a patient, would daily measurements provide more precise or additional information?

In patients from 2 rheumatology practices, Broderick et al compared daily measurements with retrospective measurements in which longer recall periods were used (3). Those investigators observed that a short recall period (1–7 days) corresponded better with the mean of daily measurements compared with a longer recall period. In addition, pain scores from retrospective measurements were higher than pain scores from daily measurements. Research in patients other than rheuma-

Supported by a grant from the Dutch Arthritis Foundation and from The Netherlands Organization for Health Research and Development.

Carolien H. Teirlinck, MD, Dieke S. Sonneveld, MD, Sita M. A. Bierma-Zeinstra, PhD, MS, Pim A. J. Luijsterburg, PhD: Erasmus MC University Medical Centre, Rotterdam, The Netherlands.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Carolien H. Teirlinck, MD, Department of General Practice, Erasmus MC, University Medical Centre Rotterdam, PO Box 2040, 3000 CA, Rotterdam, The Netherlands. E-mail: c.teirlinck@erasmusmc.nl.

Submitted for publication November 17, 2017; accepted in revised form July 17, 2018.

SIGNIFICANCE & INNOVATIONS

- Daily pain measurements provided no information different from that provided by retrospective pain measurements in patients with hip osteoarthritis.
- The reliability of retrospective measurements seemed lower in the subgroup of patients with a higher intensity of intermittent pain.

tology patients revealed similar results, although patients both overestimated and underestimated the scores in retrospective measurements compared with daily measurements (4,5).

Another reason why patients may find it difficult to express in a questionnaire how they experience pain was suggested by Hawker et al (6). In that qualitative study, patients with hip or knee OA experienced 2 types of pain: dull, aching, constant pain, and more intense intermittent pain. Consequently, a new pain measurement scale was developed: the Intermittent and Constant Osteoarthritis Pain (ICOAP) instrument. The ICOAP questionnaire captures both types of pain and therefore better represents the pain experience in patients with hip OA (7).

Intense intermittent pain might be even harder to recall or summarize compared with more constant pain. Therefore, we hypothesized that daily measurements, despite the burden for patients, could be of additional value in trials in patients with hip OA, not only for measuring the course of pain but especially for patients experiencing more intermittent pain. To explore this hypothesis, we used data from patients with hip OA who participated in a previous RCT conducted by our group (8,9).

Because the current trial uses retrospective measurements and daily measurements in the first 6 weeks, we can explore whether daily measurements reveal an effect that was not seen in the retrospective measurement at 6 weeks. Also, we can compare the pain course described in the diaries between patients with different levels of intermittent pain. Accordingly, we formulated 3 sub-hypotheses, as follows: 1) Daily measurements can reveal a clinically important effect overall or during a specific period in time, even if retrospective measurements showed no effect; 2) The course of pain, measured daily, will show more fluctuation in patients indicating that they experience more severe or more frequent intermittent pain; and 3) The agreement between daily measurements and retrospective measurements will be lower in patients indicating that they experience more severe or more frequent intermittent pain. The goal of our research is to provide more insight into the value of daily measurements and provide new suggestions for future research to improve pain measurement in patients with OA.

PATIENTS AND METHODS

In the current study, we used data from an earlier RCT conducted by our group (8,9). That trial investigated the effectiveness of general practitioner (GP) care plus exercise therapy in patients with hip OA. This trial was approved by the Medical Ethics committee of the Erasmus Medical Center, and all participants provided written informed consent. Patients were included if they visited the GP for a new episode of non-

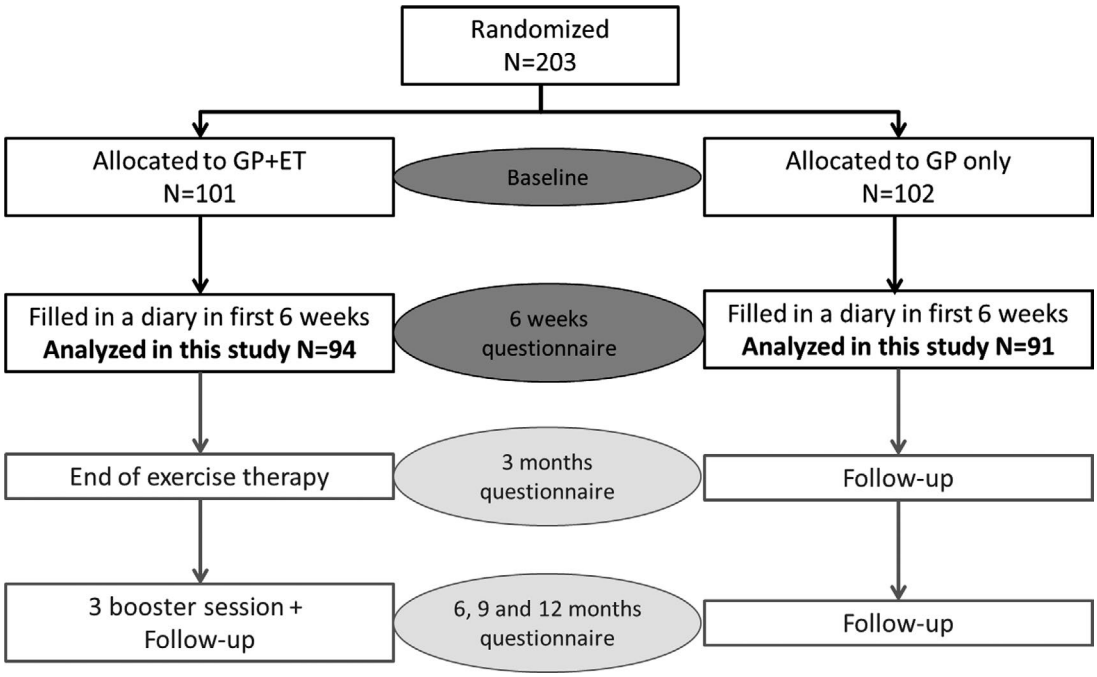

Figure 1. Flow diagram showing randomization and the time points at which measurements were taken and questionnaires were used. GP = general practitioner; ET = exercise therapy.

traumatic hip symptoms, were older than age 45 years, and complied with the clinical criteria for hip OA of the American College of Rheumatology (10). Exclusion criteria were exercise therapy in the previous 3 months, hip pain score of <2 on an 11-point NRS, a high level of physical function (score <2 on an algofunctional index), hip surgery or on a waiting list, and contraindication for exercise therapy because of co-morbidity, mentally incapable of participation, and insufficient comprehension of the Dutch language.

During the first 3 months, patients in the exercise therapy group received 12 sessions of strengthening, aerobic, and flexibility exercises followed by 3 booster sessions in the fifth, seventh, and ninth months. Exercise therapy was used in addition to standard GP care. Patients in the control group received GP care alone. Pain was measured on an NRS over the previous week and 1 day, as well as with the ICOAP and HOOS questionnaires at baseline (moment of randomization), at 6 weeks and at 3, 6, 9, and 12 months (Figure 1). At baseline, an anteroposterior radiograph of the pelvis was obtained, and 2 independent trained raters determined the Kellgren/Lawrence scores (11). Information on randomization, interventions, and other outcome measures are available in the published protocol (12) and our earlier reports (8,9).

During the first 6 weeks of the trial, patients also filled in a diary (Figure 1). Every evening they scored their average hip pain and function over that day on an NRS, with scores ranging from 0 to 10 (0 = no pain/maximum function, 10 = worst imaginable pain/not able to function at all). Of the 203 eligible patients, 18 (9%) did not fill in a diary and were excluded, leaving 185 patients for the current analysis.

We used different scales to measure pain. An NRS score was measured (as described above) and recorded in the diaries for both momentary pain and recalled pain during the previous week. The HOOS questionnaire consists of 5 subscales: pain, symptoms, function in daily living, function in sports and recreation, and hip-related quality of life. Each subscale contains multiple questions, asking the patient to score the previous week retrospectively. Scores for the questions from each subscale are summed and converted to a score ranging from 0 (indicating extreme problems) to 100 (indicating no problems). The ICOAP questionnaire has 5 questions on constant pain and 6 questions on intermittent pain. This questionnaire also asks patients to (retrospectively) score pain in the previous week but to distinguish between constant pain ("that is present all the time") and intermittent pain ("that comes and goes"). The questions ask about intensity, frequency (only intermittent pain), distress, and quality of life; all answers are scored on a 5-item Likert scale. Scores are calculated for constant pain (from 0 to 20) and intermittent pain (from 0 to 24) separately and also combined as a total pain score. Total scores range from 0 (no pain) to 100 (extreme pain).

To explore the potential additional value of daily measurements and the role of intermittent pain, we performed 3 different analyses, as described below.

1. We examined whether there was a difference in the pain level between the exercise therapy group and the control group when the pain level was measured daily. The rationale for this analysis was that the 6-week questionnaire asks the patients to summarize the effect over the previous week, whereas daily measurements might uncover a temporary effect at some time during these 6 weeks. We created a linear mixed-effects model with repeated measurements

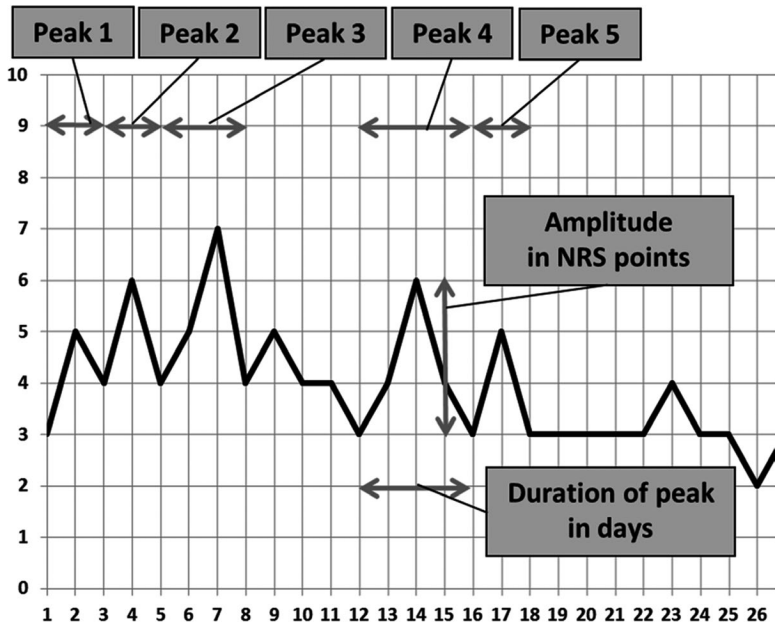

Figure 2. Graph showing the course of pain over the first 25 days in a representative patient. The numbers of peaks, the duration of a peak, and the amplitude of a peak are shown. NRS = numerical rating scale.

using the daily NRS scores for each patient. To calculate the mean scores for the groups for each separate day, the diaries had to be synchronized by date. Day 1 was set as the day on which the patient was physically examined and received the diary. A mixed-effects model incorporates both fixed effects and random effects. Fixed effects are considered to be measured without random error and are estimates of the mean differences or mean slopes, such as the treatment effect in the current study. Random effects are subject-specific and represent the natural variability between patients. The covariance structure of the model takes into account the correlation between repeated measures. Akaike's information criterion was used to determine the significance of each stage of model development, and the Toeplitz matrix was chosen as the covariance structure. This covariance structure allows different correlations for each step between repeated measurements (in this model, for example, measurements 1 day apart from each other, measurements 2 days apart from each other, measurements 3 days apart from each other, etc.). Therefore, the model will take into account that measurements made 1 day apart will correlate more strongly than measurements made 2 or 3 days apart from each other.

2. We explored whether patients who indicated at 6 weeks that they experienced a high intensity of intermittent pain or a high frequency of intermittent pain displayed a different course of pain compared with patients with a low intensity or low frequency of intermittent pain. We divided the patients into 5 intensity subgroups, based on their answer to the first question on the intermittent pain subscale in the ICOAP questionnaire: "What was the severity of the pain that 'comes and goes' during the previous week?" Patients could report no, mild, moderate, severe, or extreme pain. The same was done for the frequency of pain using the second question of the intermittent pain subscale in the ICOAP questionnaire: "How often did you have pain that 'comes and goes' during the previous week?" The frequency subgroups were never, rarely, sometimes, often, or very often. For each patient, the pain course during the first 6 weeks of the trial was plotted and described using 5 different measures: maximum amplitude in the plot, mean of amplitudes, standard deviation of the plotted line, and frequency and duration of the peaks. A peak was defined as an increase of ≥2 points on the NRS, because studies have shown a minimum clinically important difference of 2 points on an NRS in patients with chronic pain (13,14). A peak started at the first increase after a steady line (flat line for at least 2 days) or decrease and ended at the point at which a new increase started or stayed flat for at least 3 days after the line was decreasing from the highest point of the peak (Figure 2). Analysis of variance (ANOVA) tests were used to compare the means of the 5 measurements for the pain course during the first 6 weeks between the different subgroups of intensity and frequency. A linear trend test was performed to test our hypothesis that a higher intensity or frequency of intermittent pain would show more fluctuation in the course of pain and, therefore, higher means of the 5 measurements.

Table 1. Baseline characteristics of patients*

Characteristic	Exercise group (n = 94)	Control group (n = 91)
Age, mean ± SD years	64 ± 8.4	67 ± 9.8
Sex		
Female	61 (64.9)	49 (53.8)
Male	33 (35.1)	42 (46.2)
BMI, mean ± SD kg/m^2	27 ± 4.0	28 ± 4.2
Education, higher vocational education/ university	14 (15)	13 (14)
Visited specialist in past 3 months	9 (10)	10 (11)
Kellgren/Lawrence score		
0	15 (16)	15 (16)
1	25 (27)	25 (27)
2	25 (27)	30 (33)
3	15 (16)	12 (13)
4	2 (2)	3 (3)
No radiograph available	12 (13)	6 (7)
Duration of current hip symptoms, median (IQR) days	365 (862)	385 (799)
Self-exercised in past 3 months	22 (23)	38 (42)
Used pain medication daily in past 3 months	20 (21)	27 (30)
NRS score for hip pain in previous week, mean ± SD	4.4 ± 2.0	4.6 ± 1.8
HOOS, mean ± SD (0–100 scale)		
Pain	61.9 ± 16.2	61.5 ± 15.3
Function	64.2 ± 18.0	62.0 ± 16.1
ICOAP questionnaire score, mean ± SD		
Intermittent (0–24 scale)	7.9 ± 4.0	8.2 ± 4.2
Constant (0–20 scale)	5.3 ± 3.5	5.6 ± 3.5
Total (0–100 scale)	29.9 ± 15.9	31.3 ± 16.6

* Except where indicated otherwise, values are the number (%). BMI = body mass index; IQR = interquartile range; NRS = numerical rating scale; HOOS = Hip Osteoarthritis Outcome Score; ICOAP = Intermittent and Constant Osteoarthritis Pain.

3. We calculated reliability between the daily measurements and the retrospective NRS score with a recall period of 1 week (assessed at the 6-week follow-up). We compared the retrospective NRS scores with the mean of the daily NRS pain scores for the previous 7 days and the previous 2 days. In addition, we calculated the correlation between the retrospective NRS score and the daily NRS pain scores on the day on which the retrospective NRS score was assessed. Differences in these correlations were examined between patients with different levels of intensity and

frequency of intermittent pain. Cronbach's alpha was used to calculate reliability. All analyses were conducted using SPSS for Windows, version 21.

RESULTS

Among the 185 patients, 94 were randomized to the exercise therapy group and 91 to the control group. The mean age of patients was slightly higher in the control group than in the exercise therapy group. More patients in the control group than in the exercise therapy group used daily pain medication or had already performed exercise at home in the previous 3 months (Table 1). Comparisons between the 185 patients who were analyzed and the 18 nonresponders revealed only that more nonresponders visited a specialist in the previous 3 months (39%) than did the patients in this study (10%). In all 185 diaries, 9% of the data were missing. On average, patients had their first session of exercise therapy on day 10.

Difference in daily pain between the exercise therapy group and the control group. Figure 3 shows the pain scores in the GP care plus exercise therapy group and the GP care only group over the first 6 weeks of the trial. A mixed-effects model was used to investigate differences between the 2 groups, per day and overall, during the 6 weeks. We were not able to add possible confounders to the model because of the large number of repeated measurements. Adding confounders resulted in convergence problems. We found no statistical difference in the overall estimate of pain (estimate 0.14; 95% confidence interval [95% CI] –0.47, 0.19) between the GP care plus exercise therapy group and GP care group during the 6 weeks. Although pain scores in the group receiving GP care plus exercise therapy were statistically lower on day 21 (estimate –0.52; 95% CI –0.97, –0.08), this difference did not appear to be clinically relevant.

Intermittent pain and the course of pain. Subgroups of frequency of intermittent pain were as follows: never (n = 9), rarely (n = 19), sometimes (n = 74), often (n = 68), and very often (n = 7). Subgroups of intensity consisted of no intermittent pain (n = 11), mild intermittent pain (n = 47), moderate intermittent pain (n = 90), severe intermittent pain (n = 28), and extreme intermittent pain (n = 1). The single patient with extreme pain was excluded from the ANOVA.

No statistical differences were found for the defined measures of the course of pain between the subgroups based on the frequency of intermittent pain. In contrast, statistical differences were found for several defined measures of the course of pain between the subgroups of intensity of intermittent pain. Patients with a higher intensity of intermittent pain had a higher frequency of peaks, higher mean and maximal amplitude of peaks, and a higher standard deviation (significance level $\alpha = 0.005$ [Bonferroni-adjusted]). The average duration of peaks did not differ statistically between the subgroups of intensity. Table 2 shows the results of the ANOVA and the trend test.

Correlation between scores on the ICOAP for intermittent and constant pain at the 6-week follow-up was high (Pearson's r = 0.836). Therefore, we repeated our analysis after dividing the

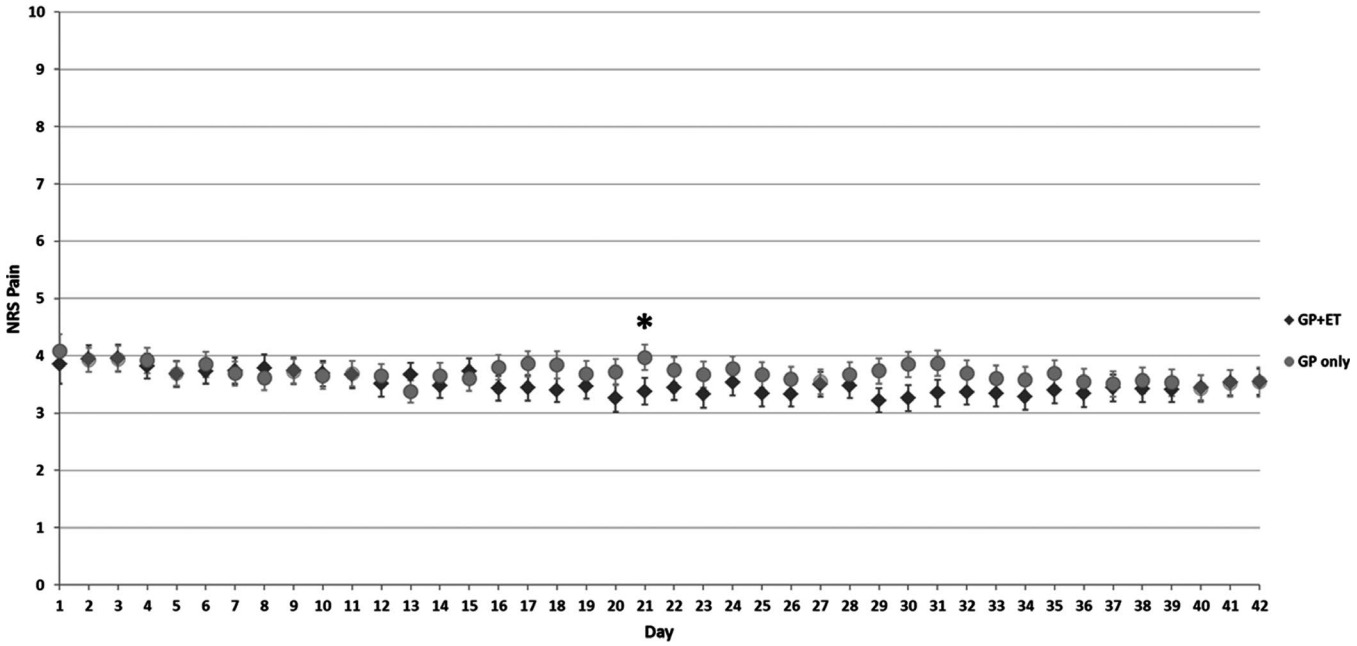

Figure 3. Pain scores in the intervention (general practitioner care [GP] plus exercise therapy [ET]) group and the control group (GP care only) over the first 6 weeks of the trial. Values are the crude mean ± SD pain scores per day. NRS = numerical rating scale; * = $P < 0.05$ (in the mixed-effects model analysis).

Table 2. Mean measures of the course of pain according to subgroups of intermittent pain intensity*

Subgroup	No. of patients	Mean (95% CI)	*P*, ANOVA	*P*, linear trend test
Duration of the peak			0.027	0.007
No pain	11	1.83 (0.38, 3.28)		
Mild	47	2.62 (2.04, 3.21)		
Moderate	90	3.30 (2.85, 3.74)		
Severe	28	3.57 (2.97, 4.17)		
Total	176	3.07 (2.77, 3.38)		
Frequency			0.004	0.003
No pain	11	1.45 (0.24, 2.67)		
Mild	47	2.06 (1.39, 2.74)		
Moderate	90	3.14 (2.59, 3.70)		
Severe	28	3.89 (2.80, 4.98)		
Total	176	2.87 (2.48, 3.26)		
Maximal amplitude			>0.001	0.003
No pain	11	3.00 (1.38, 4.62)		
Mild	47	2.40 (1.95, 2.86)		
Moderate	90	3.41 (3.12, 3.71)		
Severe	28	4.39 (3.64, 5.14)		
Total	176	3.27 (3.02, 3.53)		
Mean amplitude			>0.001	0.002
No pain	11	1.55 (0.91, 2.21)		
Mild	47	1.53 (1.31, 1.77)		
Moderate	90	1.96 (1.80, 2.12)		
Severe	28	2.33 (1.97, 2.70)		
Total	176	1.88 (1.76, 2.01)		
Standard deviation			0.001	<0.001
No pain	11	0.55 (0.23, 0.87)		
Mild	47	0.77 (0.64, 0.89)		
Moderate	90	0.96 (0.88, 1.04)		
Severe	28	1.04 (0.89, 1.19)		
Total	176	0.89 (0.83, 0.96)		

* *P* values were considered significant if $\alpha < 0.005$ (Bonferroni-adjusted). 95% CI = 95% confidence interval; ANOVA= analysis of variance.

patients based on the intensity of constant pain and observed results similar to those for measures of the course of pain (data not shown).

Intermittent pain and reliability between daily measurements and retrospective measures of pain. Table 3 shows the reliability of retrospective measurements at the 6-weeks follow-up. For all patients combined, reliability between the mean of the previous week and the retrospective NRS score was high (Cronbach's $\alpha = 0.919$) and was not lower than reliability between the mean of the previous 2 days or the same day. This was also seen in all subgroups based on the frequency of intermittent pain. In the subgroups with moderate and severe intermittent pain, reliability was lower, particularly reliability between the retrospective NRS score and the mean of the previous week in the subgroup with severe intensity of intermittent pain (Cronbach's $\alpha = 0.642$).

DISCUSSION

To evaluate the value of daily pain measurements in patients with hip OA, we analyzed data from the diaries that were collected in an RCT in patients with hip OA. In the current study, we examined the effect of GP care plus exercise therapy. In this specific trial, we found that the diaries did not reveal a group effect in the course of pain during the first 6 weeks. We also observed that patients with a high intensity of intermittent pain showed more fluctuation in the course of pain during the first 6

Table 3. Reliability of retrospective measurement at 6-weeks follow-up*

	Same day	Previous 2 days	Previous week
Total diaries (n = 185)	0.890 (0.839, 0.924)	0.911 (0.873, 0.938)	0.919 (0.888, 0.941)
Complete diaries (n = 104)	0.883 (0.827, 0.921)	0.900 (0.853, 0.932)	0.914 (0.874, 0.942)
Intensity of intermittent pain			
No pain (n = 6)	0.994 (0.954, 0.999)	0.993 (0.952, 0.999)	0.995 (0.966, 0.999)
Mild (n = 21)	0.923 (0.809, 0.969)	0.930 (0.828, 0.972)	0.928 (0.823, 0.971)
Moderate (n = 60)	0.759 (0.597, 0.856)	0.789 (0.647, 0.874)	0.880 (0.799, 0.928)
Severe (n = 16)	0.769 (0.339, 0.919)	0.817 (0.476, 0.936)	0.642 (−0.023, 0.875)
Extreme (n = 1)	–	–	–
Mild + moderate (n = 81)	0.842 (0.755, 0.899)	0.865 (0.790, 0.913)	0.918 (0.872, 0.947)
Severe + extreme (n = 17)	0.780 (0.393, 0.920)	0.825 (0.517, 0.937)	0.663 (0.069, 0.878)
Frequency of intermittent pain			
Never (n = 5)	0.984 (0.848, 0.998)	0.992 (0.925, 0.999)	0.994 (0.946, 0.999)
Rarely (n = 9)	0.912 (0.608, 0.980)	0.928 (0.681, 0.984)	0.891 (0.518, 0.975)
Sometimes (n = 45)	0.719 (0.488, 0.845)	0.771 (0.583, 0.874)	0.843 (0.714, 0.914)
Often (n = 41)	0.885 (0.785, 0.939)	0.893 (0.799, 0.943)	0.875 (0.766, 0.933)
Very often (n = 4)	0.905 (−0.470, 0.994	0.916 (−0.291, 0.995)	0.870 (−1.009, 0.992)
Rarely + sometimes (n = 54)	0.779 (0.620, 0.872)	0.822 (0.694, 0.897)	0.874 (0.783, 0.927)
Often + very often (n = 45)	0.889 (0.797, 0.939)	0.896 (0.811, 0.943)	0.878 (0.777, 0.933)

* Retrospective numerical rating scale (NRS) scores (recall of pain in the previous week) were compared with the NRS score on the same day, the mean of the previous 2 days, and the mean of the previous week, as calculated from the daily measurements. Values are Cronbach's alpha (95% confidence interval).

weeks compared with patients with a low intensity of intermittent pain. Reliability between daily measurements and retrospective measurements over the previous 7 days was lower in patients with a high intensity of intermittent pain than in patients with a low intensity of intermittent pain. Surprisingly, no differences were observed in the course of pain and reliability between retrospective and daily measurements among patients with different frequencies of intermittent pain.

Daily measurements in all patients did not yield any additional value in this trial. Neither the data in the diaries of the first 6 weeks nor the retrospective measurements at 6 weeks showed a difference in pain between the intervention group (GP care plus exercise therapy) and the control group (GP care only). Thus, from our perspective, daily measurements could be of additional value only if a possible temporary treatment effect between follow-up measurements is expected and if this temporary effect is clinically relevant for patients or caregivers. In a chronic disease such as OA, an effect lasting for several days or even several weeks would not be considered relevant. However, because patients might experience a temporary negative effect of exercise therapy, it could be worthwhile to record daily measurements during the intervention to capture the magnitude and duration of this effect.

When we examined the daily measurements focusing on intermittent pain, we observed some interesting results. We found that the intensity but not frequency of intermittent pain was related to 4 of the defined measurements of fluctuation in the course of pain in the diaries. An explanation of why frequency was not related to the measurements of fluctuation may be that patients with frequent intermittent pain might experience several peaks of pain during a single day but not over multiple days. Therefore, because we measured only once a day, no peaks would be visible in the course of pain during that day. A second possibility is that even in patients with frequent intermittent pain fluctuations in intensity were low and too subtle to accurately measure a significant difference in peaks (defined as an increase of ≥2 points) and standard deviations. Therefore, only patients with a high intensity of pain were likely to show more prominent fluctuations.

The first measurement of fluctuation in the course of pain—the duration of peaks—did not reach significance. It is possible that this measurement does not represent fluctuation as well as the other 4 measurements. Also, the power decreased because of the correction for multiple testing.

One important question remains: how important is it for study investigators to differentiate between intermittent pain and constant pain when considering adding daily measurements? The correlation between the constant pain and intermittent pain scales of the ICOAP questionnaire was high in our population. We observed that patients with more constant pain also showed more fluctuations in the course of pain, similar to the results in patients with more intermittent pain. Thus, patients with more intermittent pain are more likely to also experience more constant pain and vice versa. This could imply that patients with more intermittent and/or constant pain are in a more advanced stage, and that

these patients have more difficulty summarizing their pain experience over a longer time period in retrospective measurements.

The concept of advancing pain stages was supported in the qualitative study by Hawker et al (6). Patients reported that they initially experience a dull aching pain during certain activities, that this pain evolves into a more constant pain, and that when, over time, this constant pain increased, patients moved into a more advanced stage of pain in which they also experience (unpredictable) intermittent pain. Hence, patients in a more advanced stage experience more pain in general, and intermittent pain plays a more important role, which, in turn, can lead to more recall bias in retrospective measurements.

We would like to stress that our study was explorative in design and therefore has limitations. First, we only looked at the value of daily measurement in this specific trial in patients with hip OA. The analysis of the diaries did not show results different from those of the original analysis of the retrospective measurement at 6 weeks of follow-up; therefore, no temporary effect was observed. It is possible that in other OA trials investigating participants from a different population, or patients with a different joint affected by OA, or an intervention different from the one used in our trial, daily measurements could produce additional information. In addition, we had no daily measurements from 6 weeks to 3 months of follow-up. At 3-months follow-up, we observed the largest effect; thus, it would have been interesting to compare the retrospective measurements at 3 months with daily measurements up to this follow-up time point in order to explore whether these results would differ from each other. Second, the subgroups included in the analyses of intermittent pain were small, but we believe that a trend in the subgroups and different parameters to objectify fluctuation in the course of pain was visible and suggest a plausible relationship. Third, the parameters to objectify the fluctuation in the course of pain were defined by consensus in our research group and not by validated measures. Last, we used paper diaries. Research has shown that compliance using paper diaries is lower than that using electronic diaries, and that paper diaries could also introduce a recall bias because of backfilling (15).

To our knowledge, only 1 other study has compared daily and retrospective measurements in patients with OA (16). The results of that study showed a strong correlation between "recalled pain intensities" and the "mean of daily measurements" over 7 days ($r = 0.78$), but pain intensity on the day of recall influenced this correlation. Pain variability, measured with standard deviations, did not influence the correlation. No studies were found that examined the additional value of diaries in OA patients with different levels of intermittent pain and constant pain. Therefore, this study is the first to compare these measurements and might help researchers when designing new trials in OA patients.

We would suggest that researchers in OA trials consider daily measurements during the intervention period only if a temporary effect or early effect is expected and is of clinical relevance. Retrospective measurements with a 7-day recall correlated well with daily measurements in the total group of OA patients and therefore might be used in follow-up after the intervention. Although researchers should be aware that the pain experience in patients in a more advanced stage of pain is probably more complex, and that retrospective measurements with a longer recall period probably will not represent the actual mean of pain scores during the recall period in these patients. They should consider other types of measurements to increase the reliability of the measurement, such as using retrospective measurements with a short recall period or the mean of multiple daily measurements during a short period at multiple time points of follow-up (17,18). In addition, new methods, such as in the Ecological Momentary Assessment, were developed in which momentary pain is measured electronically multiple times per day (19,20).

In conclusion, in this specific trial in patients with hip OA, daily pain measurements during the first 6 weeks did not provide additional or more precise information compared with the retrospective measurements at 6-week follow-up. However, the reliability of retrospective measurements may be lower in patients with a higher intensity of intermittent pain. To validate our results, more research on pain measurement in patients with OA is necessary and should focus on comparing the reliability of retrospective, daily, and multiple momentary measurements. Such research should also take into account subgroups of OA patients with different stages of pain and different pain experiences.

ACKNOWLEDGMENT

The authors thank A. R. Rademaker for helping with the scoring of the courses of pain.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Teirlinck had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Teirlinck, Bierma-Zeinstra, Luijsterburg.
Acquisition of data. Teirlinck, Sonneveld.
Analysis and interpretation of data. Teirlinck, Sonneveld, Bierma-Zeinstra, Luijsterburg.

REFERENCES

1. Perrot S. Osteoarthritis pain. Best Pract Res Clin Rheumatol 2015;29:90–7.
2. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
3. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. Pain 2008;139:146–57.

4. Larsson B, Fichtel A. Headache prevalence and characteristics among adolescents in the general population: a comparison between retrospect questionnaire and prospective paper diary data. J Headache Pain 2014;15:80.
5. Okupa AY, Sorkness CA, Mauger DT, Jackson DJ, Lemanske RF Jr. Daily diaries vs retrospective questionnaires to assess asthma control and therapeutic responses in asthma clinical trials: is participant burden worth the effort? Chest 2013;143:993–9.
6. Hawker GA, Stewart L, French MR, Cibere J, Jordan JM, March L, et al. Understanding the pain experience in hip and knee osteoarthritis: an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:415–22.
7. Hawker GA, Davis AM, French MR, Cibere J, Jordan JM, March L, et al. Development and preliminary psychometric testing of a new OA pain measure: an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:409–14.
8. Teirlinck CH, Luijsterburg PA, Dekker J, Bohnen AM, Verhaar JA, Koopmanschap MA, et al. Effectiveness of exercise therapy added to general practitioner care in patients with hip osteoarthritis: a pragmatic randomized controlled trial. Osteoarthritis Cartilage 2016;24:82–90.
9. Tan SS, Teirlinck CH, Dekker J, Goossens LM, Bohnen AM, Verhaar JA, et al. Cost-utility of exercise therapy in patients with hip osteoarthritis in primary care. Osteoarthritis Cartilage 2016;24:581–8.
10. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–14.
11. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–502.
12. Van Es PP, Luijsterburg PA, Dekker J, Koopmanschap MA, Bohnen AM, Verhaar JA, et al. Cost-effectiveness of exercise therapy versus general practitioner care for osteoarthritis of the hip: design of a randomised clinical trial. BMC Musculoskelet Disord 2011;12:232.
13. Farrar JT, Young JP Jr, 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.
14. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004;8:283–91.
15. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003;24:182–99.
16. Perrot S, Rozenberg S, Moyse D, Legout V, Marty M; cercle d'etude de douleur en rhumatologie (CEDR). Comparison of daily, weekly or monthly pain assessments in hip and knee osteoarthritis. A 29-day prospective study. Joint Bone Spine 2011;78:510–5.
17. Broderick JE, Schwartz JE, Schneider S, Stone AA. Can end-of-day reports replace momentary assessment of pain and fatigue? J Pain 2009;10:274–81.
18. Heapy A, Dziura J, Buta E, Goulet J, Kulas JF, Kerns RD. Using multiple daily pain ratings to improve reliability and assay sensitivity: how many is enough? J Pain 2014;15:1360–5.
19. Stone AA, Broderick JE, Porter LS, Kaell AT. The experience of rheumatoid arthritis pain and fatigue: examining momentary reports and correlates over one week. Arthritis Care Res 1997;10:185–93.
20. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. Pain 2004;107:61–9.