Recall of medication use, self-care activities and pain intensity: a comparison of daily diaries and self-report questionnaires among low back pain patients

Kate M. Dunn, Kelvin P. Jordan and **Peter R. Croft**

Arthritis Research Campaign National Primary Care Centre, Keele University, UK

Aim: We aimed to compare recalled information on medication use, self-care activities and pain intensity among primary care low back pain consulters with diary records of the same events. **Background:** Concerns are often expressed regarding the validity of recalled information about past experience of health events such as pain or its treatment. Comparing with information collected using daily diaries is one method of validating recalled findings. **Methods:** Patients completed diaries recording their medication use, self-care activities and pain intensity each day for two weeks. Immediately following this period, patients completed questionnaires asking for recall of their medication use, self-care activities and least, worst, usual and current pain for the previous two weeks. The recalled information obtained from the questionnaires was compared with data from the daily diaries using intraclass correlation coefficients (ICC) and κ, with one-sided 95% confidence intervals. **Findings:** All 29 participants returned 11 or more diaries. Validity of questionnaire-based recall for medication use and self-care activities was good, with everyone who reported use in the diaries also reporting this on the questionnaires (both $\kappa = 1.0$). However, some specific medications (eg, diclofenac) were over-reported in the questionnaires, and some self-care activities (eg, exercises) were under-reported. Combinations of pain intensity ratings were more accurate than single ratings; the mean of the recalled least, usual and current pain intensities was closest to the diary ratings (ICC 0.94, mean difference 0.13). The generalisability of these findings to other settings, recall periods and patient groups remains to be established.

Key words: back pain; diaries; pain intensity; questionnaires; recall; validity

Introduction

In studies of symptoms such as back pain, self-reported information on medication use, self-care and pain intensity are commonly collected and are important for the assessment of patients and the prediction of prognosis, and in the planning of symptom management. However, there are no accepted recommendations for the standardised collection of medication and self-care data in common pain syndromes such as low back pain (LBP). There are recommendations for the assessment of pain intensity, which suggest characterising pain severity over a period of recall rather than at a single point in time (Von Korff

Correspondence to: Dr Kate M. Dunn, Arthritis Research Campaign National Primary Care Centre, Primary Care Sciences, Keele University, Keele, Staffordshire, ST5 5BG, UK. Email: k.m.dunn@cphc.keele.ac.uk

et al., 2000), due to the variability of pain, but pain intensity is often simply reported without reference to a time period. In addition, the validity and meaning of self-reported recalled measurements of pain have been questioned, for example problems such as memory decay and telescoping (incorrectly including or excluding an experience from the time period), and characteristics of the pain experience itself (eg, the quality of the pain, pain intensity, course of pain, etc.) may influence the self-report of pain (Verbrugge, 1980; Raspe and Kohlmann, 1994; Von Korff, 2001).

Although there is work evaluating the recall of medication use among specific patient groups, few studies have assessed the validity of recalled self-care activities in common symptoms like pain. Guzmán *et al*. (1999) used diaries to validate interview data in occupational LBP patients; they found that health care utilisation was accurately recalled, but medication use was consistently over-reported. The generalisability of this study to self-completion data from patients recruited from health care settings is questionable, first, because interviews were used, where interviewers could clarify responses, and second, as highlighted by the authors, because of possible conflicts with patients' healthcare providers and compensation claims. Other studies have used diaries to collect data on self-care activities (Freer, 1980; Bentzen *et al*., 1989) and have reported that they are a reliable method for collecting this information, but none have compared diary and questionnaire estimates of self-care use, and no studies have focused on these issues among pain patients. Examination of the validity of self-reported data on self-care in pain patients in primary care settings would improve the basis for the investigation of these activities.

Previous studies have compared ratings of pain collected at frequent intervals and retrospective ratings. Two studies of chronic pain patients using paper ($n = 107$, Salovey *et al*., 1992) and electronic diaries ($n = 36$, Jamison *et al*., 2001), and two studies of LBP patients using paper ($n = 200$, Bolton, 1999) and electronic ($n = 21$, Jamison *et al*., 2006) diaries found close association between weekly pain recall and daily recordings. However, one other electronic diary study of chronic pain patients ($n = 68$, Stone *et al*., 2004) and two studies of pain recall ($n = 12$, Linton and Melin, 1982; $n = 15$, Linton and Götestam, 1983)

reported that recall led to an overestimation of pain, and one study of long-term recall indicated an underestimate of pain ($n = 144$, Dawson *et al*., 2002). The larger studies appear to show that recall of pain intensity is accurate over a week, although the largest study (Bolton, 1999) indicated that their results might not be generalisable as the patients included had declining pain levels over the study period. Recent research has also investigated the relationship between period of recall and accuracy (using electronic pain assessment), and concluded that recall was accurate up to 28 days (Broderick *et al*., 2008). None of these studies were carried out in primary care LBP consulters. Therefore, studying recall of pain intensity among primary care LBP patients would improve our understanding of the topic in this patient group.

In validity studies, recalled information would ideally be compared with an objectively measured reference standard. However, factors such as pain intensity are inherently subjective, and no objective reference standard is available. One alternative, where no objective reference standard exists, is to compare the information in question with data that has higher face validity (Abramson, 1984). Information collected at frequent short-time intervals is perceived to be more valid than data collected retrospectively through recall, but is impractical to collect in many studies due to problems of compliance and response, and may alternatively be used as a reference standard against which to compare recalled information. Data recorded at frequent intervals (by methods such as daily diaries) are assumed to reflect current experience and be free of the error that can affect questionnaires or interviews based on longer-term recall of past events or states (Verbrugge, 1980; Carp and Carp, 1981; Stone *et al*., 1991; Cruise *et al*., 1996; Lousberg *et al*., 1997). Diaries have been recommended as a reference standard for self-report questionnaires (Stewart *et al*., 2000), and have previously been used by pain researchers to assess the validity of reported pain intensity (Jensen and McFarland, 1993; Bolton, 1999).

The objective of this study was to assess the validity of questions about recalled medication use, self-care activities and pain intensity from self-completion questionnaires by comparing the responses with daily diary information on the

same topics previously gathered during the period of recall.

Methods

This work is part of a longitudinal study looking at the pain, disability and health care use of a cohort of back pain consulters from primary care, using epidemiological and qualitative methods – the Back-pain Research in North Staffordshire (BaRNS) Study. The local Research Ethics Committee approved this study.

Participants

The participants in the main BaRNS study were recruited from five computerised general practices who are members of the North Staffordshire General Practice Research Network. All patients aged 30–59 years consulting their general practitioner with LBP during the 12 months from October 2001 were invited to take part; 935 (65%) returned baseline questionnaires and 776 consented to be followed up as part of the study. Further information about the main study is available elsewhere (Dunn and Croft, 2005), but in brief follow-up consisted of monthly postal questionnaires enquiring about pain and pain-related disability in the time period since the previous questionnaire.

Patients were eligible for the study reported here if they were taking part in the main study, had already returned more than two of the monthly questionnaires, and reported still having back pain on their latest questionnaire. Including patients at an earlier stage of the study could have meant that results were subject to regression to the mean (Dunn & Croft, 2006). Each week, groups of patients fulfilling these inclusion criteria were invited to take part in an additional diary-based study; this recruitment strategy was continued until the required number of patients providing the minimum number of daily diary returns was obtained.

Sample size was calculated on the basis of the formula by Donner and Eliasziw (1987) for the intraclass correlation coefficient (ICC). Based on a power of 90% and a 5% significance level, a sample size of 29 is appropriate for assessing whether the value of the population ICC is above 0.85 with an assumed real value of 0.95. This also allows detection of a minimal clinically important difference of 1 point on a pain numerical rating scale (Salaffi *et al.*, 2004) with 90% power and significance level of 1%.

Procedures

Potential study participants for the diary study were sent an information sheet, a consent form, a reply-paid envelope and an explanatory letter. If the participant returned the consent form, the researcher telephoned and arranged to visit them in their home two weeks before they were due to receive their next main study questionnaire. At the visit, the researcher explained the study and how to complete the diaries, and obtained written informed consent. Patients were told that the researchers wanted to collect detailed information in order to understand how much pain people feel and what they do to ease their pain each day.

Participants were then given 14 reply-paid diaries to complete and return daily. This ensured that data were collected at frequent intervals (as participants could not complete the diaries at the end of the study period) as recommended for diary studies (Stone *et al.*, 1991; Goossens *et al.*, 2000). At the end of the two-week diary period, patients were sent their usual monthly questionnaire, with a letter thanking them for taking part and asking them to return the completed questionnaire.

Diary and questionnaire measures

Each diary was divided into two sections, one to be completed each morning and one in the evening to capture pain throughout the day. Both sections contained questions about pain at the time of completion and average pain through the preceding night or day. In the evening, there were additional questions on medication use and self-care activities during that day. Each diary comprised one page, and only took a few minutes to complete each morning and evening. This method fulfils the recommendations for diary studies to be brief and easy to complete (Stone *et al.*, 1991; Jensen and McFarland, 1993).

The questions about medication use and self-care were identical on the diaries and the monthly questionnaires except that the latter asked about

the previous two weeks instead of 'today'. Information was collected about the use of standard analgesics (paracetamol, co-codamol, co-proxamol, co-dydramol or dihydrocodeine), non-steroidal anti-inflammatory drugs (NSAIDS – ibuprofen, aspirin, diclofenac or naproxen) and strong analgesics (tramadol). Self-care activities for LBP were lying down, using creams or sprays, doing exercises or stretches, using heat (eg, heat packs or lamps), cold (eg, cold packs or ice), massage, having a hot bath or jacuzzi and use of lumbar supports or corsets.

The questionnaire contained four pain intensity questions: worst, least and usual pain over the previous two weeks, and pain at the time of questionnaire completion (current pain). Pain intensity was measured on the diaries and in the questionnaires using 0–10 numerical rating scales. These are reported to be sensitive instruments for pain intensity measurement (Jensen *et al.*, 1994), and have been recommended for use in clinical populations in preference to visual analogue scales or verbal rating scales (Raspe and Kohlmann, 1994; Von Korff, 2001). The anchors were 0 = 'No pain' and 10 = 'Pain as bad as could be'. We used four different pain intensity questions as combining information from more than one rating has previously been stated to give more accurate estimates of pain intensity than single questions (Jensen *et al.*, 1999; Von Korff, 2001).

The diary and the relevant questionnaire-based questions can be found in the appendix that appears online at http://journals.cambridge.org/phc.

Data analysis

Respondents were excluded if they returned fewer than four diaries in each week, or if they completed the questionnaire more than six days after the last diary. This was done to reduce variability, provide more reliable estimates, and to reduce disparity between the period covered by the diaries and questionnaires (Jensen and McFarland, 1993; Jensen *et al.*, 1996; Schwartz and Stone, 1998; Goossens *et al.*, 2000).

Kappa was used to measure the agreement between the diary and questionnaire measures of medication and self-care use, and is presented with one-sided 95% confidence intervals (CI).

The reference standard pain intensity score for each participant (which will be referred to as diary-based pain) was calculated as the arithmetic mean of all of the pain numerical rating scales completed in the diaries during the study period. The ICC (2,1) (Shrout and Fleiss, 1979) was used to compare diary-based pain and the single pain measures recorded in the questionnaire, with one-sided lower 95% confidence limits to show the likely minimum agreement (Jordan *et al.*, 2000). Mean differences between diary-based pain and questionnaire measures were calculated; with 95% CI for paired data. Also calculated were 95% limits of agreement (Bland and Altman, 1999). The limits of agreement indicate the range in which the difference between the diary and questionnaire pain scores would be expected to lie for 95% of subjects. This was repeated on combinations of the questionnaire pain ratings using the means of the included ratings to determine which item or group of items had the highest agreement and lowest difference to diary-based pain.

Analysis was carried out using SPSS for Windows 11.0. (2001).

Results

In order to obtain complete data on 29 participants, 151 people were invited. Thirty-four people (23%) agreed to take part but five were subsequently excluded (one only completed five diaries, two did not return the questionnaire and two returned the questionnaire too late). The participants' mean age was 46 years (range 32–59 years, s.d. 8.58), and 13 were male (45%). Most people (79%, $n = 23$) returned all 14 diaries, four returned 13 and two people returned 11 diaries; almost 50% of diaries were returned either on the day they were supposed to be completed or the following day, 70% responded within two days and 89% within three days (mean 1.9 days, median two days). The mean time between completion of the last diary and completion of the monthly questionnaire was two days (median one day, range 0–6 days); 55% of people completed the questionnaire within one day of the last diary, and 79% completed it within two days. There was no difference between diary study participants and the total study population in terms of age and gender, and people participating in the diary study had similar but slightly higher disability

(mean modified Roland-Morris Disability Questionnaire (Roland and Morris, 1983) score 12.4 versus 10.2) and depressive mood (Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983) depression score 8.2 versus 7.4) scores than the whole group. Similarly to the main study, most of the diary study participants had leg pain at the start of the study (86%, $n = 25$), and two-thirds reported having their back pain for more than a year (69%, $n = 20$). Fifty five percent of the diary study participants were in employment at the start of the study, 44% of those were in routine or semi-routine occupations and 48% had followed education beyond the age of 16 years; these figures are similar to the whole study population.

Comparison of report of medication and self-care use

The agreement between the diaries and the questionnaires for reports of medication use is shown in Table 1. There was perfect agreement between reports of use of any analgesic medication – all 27 people who reported analgesic use on the diaries also reported it on the questionnaires. The agreement was also high for taking any standard analgesic or any NSAID, indicating that the overall accuracy of recall on the questionnaires was good.

When medication use was broken down further, there was more disparity; for example, four out of the 11 people who said on the questionnaire that they had taken paracetamol during the previous two weeks had not recorded it on any of the diaries (false positive responses). A similar picture was true for diclofenac, with three out of the seven people reporting use on the questionnaire but having no record of it on the diaries. Conversely, there was never more than one person for any particular medication who reported using it in the diaries, but did not report it in the questionnaire (false negative responses).

Table 2 shows that all 29 people in the study used some form of self-care during the two-week period, and all reported it in the diaries and on the questionnaires. However, when individual self-care activities are considered, although agreement was generally good, some disparity between the reports occurs. For example, for six out of the eight self-care activities studied, two or more people reported using the activity in the diaries, but did not report using it on the questionnaires (false negative responses). Conversely,

Table 1 Agreement between reports of medication use on the diaries and the questionnaires ($n = 29$)

Variable	Number of people reporting use on diary	Number of people reporting use on questionnaire	False positive[a]	False negative[b]	κ	One-sided 95% CI
Standard analgesics						
Paracetamol	8	11	4	1	0.61	$\kappa > 0.36$
Co-codamol	9	10	1	0	0.92	$\kappa > 0.80$
Co-proxamol	7	7	1	1	0.81	$\kappa > 0.60$
Co-dydramol	1	1	0	0	1.00	–
Dihydrocodeine	3	3	0	0	1.00	–
Any standard analgesic	22	23	2	1	0.70	$\kappa > 0.44$
NSAIDS						
Ibuprofen	8	8	1	1	0.83	$\kappa > 0.63$
Aspirin	1	1	0	0	1.00	–
Diclofenac	4	7	3	0	0.67	$\kappa > 0.37$
Naproxen	2	2	0	0	1.00	–
Any NSAID	13	14	2	1	0.79	$\kappa > 0.61$
Strong analgesics						
Tramadol	4	4	0	0	1.00	–
Any analgesic medication	27	27	0	0	1.00	–

CI = confidence intervals; NSAID = ibuprofen, aspirin, diclofenac or naproxen.
[a] Reported on the questionnaire but not in the diaries.
[b] Reported in the diaries but not on the questionnaire.

Table 2   Agreement between reports of self-care use on the diaries and the questionnaires ($n = 29$)

Variable	Number of people reporting use on diary	Number of people reporting use on questionnaire	False positive[a]	False negative[b]	κ	One-sided 95% CI
Lying down	19	20	3	2	0.61	$\kappa > 0.35$
Creams/sprays	5	3	0	2	0.71	$\kappa > 0.39$
Exercises/stretches	24	21	0	3	0.71	$\kappa > 0.44$
Heat	9	9	1	1	0.84	$\kappa > 0.66$
Cold	5	3	0	2	0.71	$\kappa > 0.39$
Massage	13	10	0	3	0.79	$\kappa > 0.59$
Hot bath/jacuzzi	20	19	3	4	0.45	$\kappa > 0.16$
Lumbar support	3	3	0	0	1.00	–
Any self-care	29	29	0	0	1.00	–

CI = confidence intervals.
[a] Reported on the questionnaire but not in the diaries.
[b] Reported in the diaries but not on the questionnaire.

only for lying down or taking a hot bath or jacuzzi did two or more people report the activity on the questionnaire but not in the diaries (false positive responses).

Comparison of pain intensity ratings

The agreement of diary-based pain and various individual and combinations of the recalled pain measures from the questionnaire is presented in Table 3. Of the single recalled pain ratings, only the questionnaire rating of current pain was not significantly different from diary-based pain, with an average overestimate of 0.38. However, recalled usual pain also overestimated diary-based pain by a mean of < 1 point (the minimal clinically important difference (Salaffi *et al.*, 2004)), and had higher agreement than the other questionnaire measures (ICC 0.92). Recalled worst pain overestimated diary-based pain by an average of 2.6 points, and recalled least pain underestimated it by 1.3 points. The best simple combination of questionnaire ratings was the mean of least, usual and current pain, which was a mean of only 0.13 points below diary-based pain, with very high reliability (ICC 0.94). The 95% limits of agreement for the mean of least, usual and current pain with diary-based pain were -1.55 to 1.81, which would indicate that the difference between the questionnaire and diary measures for individual subjects is < 2 points, and for the majority of subjects < 1 point. Other composite pain intensity ratings were little more accurate than the single pain ratings.

Discussion

This study has shown that recall of medication use, self-care activities and pain intensity over the previous two weeks reflects the regularly documented experience of a group of primary care LBP patients relatively accurately, giving evidence for the validity of recall over this period as a measure of the average experience during that period. Recall of medication use, although accurate for groups of medications, tended towards over reporting of individual medications. Self-care overall was similarly reported by all participants in the diaries and questionnaires, implying accurate recall, but individual self-care activities tended to be under-reported in the questionnaire. Combinations of recalled pain intensity ratings gave more accurate estimates of the cumulative experience of diary-based pain intensity than ratings using single questions, a finding that agrees with other work (Jensen *et al.*, 1999; Von Korff, 2001). Such recall over a short-time period has previously been recommended as a better estimate of current 'average' pain status than a rating at a single point in time (Von Korff *et al.*, 2000).

This study has provided new information on the validity of recall of self-care activities, which has been assessed in few other studies of pain sufferers. Recall of medication use over the previous two weeks was also found to be relatively accurate, which contrasts to studies using longer periods of recall (eg, 10 years), which found that pain

Table 3 Agreement between actual pain[a] and questionnaire pain ratings (*n* = 29)

Questionnaire rating	Mean difference[b] (95% CI)	95% limits of agreement	ICC (2,1) (one-sided 95% CI)
Single pain intensity ratings			
Worst pain	−2.55 (−3.09, −2.01)	−5.39, 0.29	0.54 ($\rho > -0.05$)
Least pain	1.31 (0.88, 1.75)	−0.99, 3.61	0.80 ($\rho > 0.18$)
Current pain	−0.38 (−0.90, 0.14)	−3.10, 2.34	0.85 ($\rho > 0.73$)
Usual pain	−0.55 (−0.89, −0.21)	−2.33, 1.23	0.92 ($\rho > 0.79$)
Composite pain intensity ratings (arithmetic mean)			
Least and usual pain	0.38 (0.03, 0.73)	−1.44, 2.20	0.93 ($\rho > 0.86$)
Least and worst pain	−0.62 (−0.93, 0.31)	−2.26, 1.02	0.91 ($\rho > 0.72$)
Least and current pain	0.47 (0.11, 0.82)	−1.40, 2.34	0.91 ($\rho > 0.82$)
Usual and worst pain	−1.55 (−1.91, −1.19)	−3.44, 0.34	0.77 ($\rho > -0.14$)
Usual and current pain	−0.47 (−0.82, −0.11)	−2.35, 1.41	0.91 ($\rho > 0.82$)
Current and worst pain	−1.47 (−1.92, −1.01)	−3.84, 0.90	0.75 ($\rho > -0.04$)
Least, usual and worst pain	−0.60 (−0.89, −0.30)	−2.15, 0.95	0.92 ($\rho > 0.74$)
Least, usual and current pain	0.13 (−0.19, 0.45)	−1.55, 1.81	0.94 ($\rho > 0.89$)
Least, current and worst pain	−0.54 (−0.87, −0.21)	−2.27, 1.19	0.92 ($\rho > 0.79$)
Usual, current and worst pain	−1.16 (−1.52, −0.80)	−3.06, 0.74	0.83 ($\rho > 0.06$)
Worst, least, usual and current pain	−0.54 (−0.85, −0.24)	−2.14, 1.06	0.92 ($\rho > 0.79$)

CI = confidence intervals; ICC = intraclass correlation coefficients.
[a] Arithmetic mean of all of pain scales completed in the diaries during the study period.
[b] Actual (diary) pain minus questionnaire pain measure; negative score indicates overestimate of diary pain on questionnaires, positive score indicates underestimate.

medications were under-reported (Dawson *et al.*, 2002). Another study (Salovey *et al.*, 1992; 1993) showed similar results, reporting that chronic pain patients accurately recalled activities such as taking aspirin or using heating pads. When specific medications and self-care activities were considered, our results showed a tendency for medications to be over-reported and for self-care activities to be under-reported in the questionnaires. A previous study in occupational back pain patients also found that medications were over-reported (Guzmán *et al.*, 1999). This may relate to the frequency of these activities, as our data show that medication use was more likely to be a daily activity than self-care, and therefore activities carried out less frequently might be expected to be under-reported in questionnaires compared with those carried out more frequently. The differences in reporting may also relate to differences in perceived importance – patients may believe it is more important to report medication use than self-care activities – or may reflect 'telescoping' bias in which patients are actually recalling experience over a longer or shorter period than the one they are requested to consider (Verbrugge, 1980). We have shown good levels of validity for our questions on medication use and self-care activities, and other researchers could use these questions where similar information is required, for example when investigating patterns between groups of people or changes over time in large epidemiological studies. However, researchers should be aware of the potential errors in reporting; further research may establish whether such errors form a type of bias (eg, differential reporting relating to characteristics of the pain experience), or are simply randomly distributed. More detailed investigation might also expound theoretical reasons for differences in recall, including the fundamental philosophy of what people are actually expressing when they recall an emotion-laden experience such as pain (Broderick *et al.*, 2006). Some research has investigated the relationship between recall of pain intensity and factors such as catastrophising (Lefebvre and Keefe, 2002), neuroticism (Raselli and Broderick, 2007) and other aspects of the pain experience (Morley, 2007), but few studies have investigated influences on recall of other parameters such as self-care use.

Averaging pain measurements to produce a single rating is recommended to reduce contextual and methodological influences and

increase reliability (Von Korff *et al.*, 2000). The most accurate combination of recalled pain ratings reported here was the mean of least, usual and current pain intensity, which is different to the combinations reported in other studies (Salovey *et al.*, 1993; Bolton, 1999). However, the next most accurate combination (least and usual pain) was the same as that reported by Jensen *et al.* (1996). This indicates that researchers should use combinations of pain intensity ratings because they give consistently more accurate estimates than single ratings alone. These findings may also be applicable to clinical practice, as clinicians are often interested in reducing patients' average levels of pain, and a single measure at one time point may not be an appropriate method of measuring this (Jensen and McFarland, 1993). Clinicians should therefore consider using more than one question to obtain information about a patient's pain experience.

There are strengths and limitations to this study. There was no evidence of fatigue in diary completion, as 97% returned a complete set of diaries, or of sensitisation or learning effects, as there were no differences in mean daily pain ratings between the start and end of the diary period. However, it must be accepted that these findings might be an optimistic reflection of recall as it is plausible that daily diary completion, and the knowledge that a questionnaire would follow the diaries, could improve recall in the questionnaires. The daily mailing of diaries ensured that completion was at frequent intervals, avoiding (as much as possible) retrospective completion of 'daily' diaries that has been reported in other studies (McGorry *et al.*, 1999). Although the majority of diaries were returned promptly after the day of completion, some diaries were returned later, and there is no guarantee that these were completed on the requested day. This may have introduced some recall bias if the diaries were completed retrospectively. Another problem with the study was that, although overlap was reasonably good between the period covered by the diaries and the recall period, it was not perfect; this difference may account for some of the discrepancies between recalled information and the daily diaries. Insufficient co-operation is a common problem with diaries, and in this study only 23% of the people invited to take part actually consented to do so. As the focus of the main study is the 12-month follow-up, the invitation to the diary study emphasised its voluntary and time-consuming nature, and the low-participation rate was not surprising. The diary study participants reflected the main study participants in terms of age and gender, and included people with a wide range of pain intensity levels. However, participants did report slightly better functional status and depressive mood on average than the total sample, and may represent a group who more closely monitored their pain intensity. Although recall may be poorer among LBP sufferers in general than among our sample, comparison with other studies in pain patients adds credence to the data reported here. Some studies have used electronic or mechanical devices for monitoring medication use; one recent study compared diary-based and electronic data among asthma patients, and reported that there was high concordance between the two sources, although electronic data was slightly more precise (Butz *et al.*, 2005). As the difference between the two sources was only slight, this would indicates that our results would change little if electronic monitoring devices had been used, but further research is needed to confirm this. The sample included in this study was small, and so the power to show generalisability across a wide sample of back pain patients was limited; however, the study sample had similar characteristics to the original target population of back pain consulters, and furthermore the core internal comparison addressing the main objectives is unlikely to have been substantially biased by any selectivity in the sample studied.

Conclusions

This study has shown that LBP patients are reasonably accurate at recalling their medication use, self-care activities and pain intensity over a two-week period, and questionnaires using recall to estimate these variables are likely to be a valid reflection of experience during such a period. Recall of medication use and self-care activities showed highest validity when summary measures were used. The most accurate assessment of recalled pain intensity was obtained through combinations of ratings. We believe that these findings are likely to be generalisable to other

primary care LBP populations, but further testing is required before more general recommendations can be made.

Competing interests

None.

Authors' contributions

KD conceived and designed the study, collected and analysed the data, interpreted the findings and drafted the manuscript. KJ and PC participated in the design of the study, analysis and interpretation of the findings and preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

References

Abramson, J.H. 1984: *Survey methods in community medicine*, third edition. Edinburgh, UK: Churchill Livingstone.

Bentzen, N., Christiansen, T. and **Pedersen, K.M.** 1989: Self-care within a model for demand for medical care. *Social Science and Medicine* 29, 185–93.

Bland, J.M. and **Altman, D.G.** 1999: Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8, 135–60.

Bolton, J.E. 1999: Accuracy of recall of usual pain intensity in back pain patients. *Pain* 83, 533–39.

Broderick, J.E., Schwartz, J.E., Vikingstad, G., Pribbernow, M., Grossman, S. and **Stone, A.A.** 2008: The accuracy of pain and fatigue items across different reporting periods. *Pain* 139, 146–57.

Broderick, J.E., Stone, A.A., Calvanese, P., Schwartz, J.E. and **Turk, D.C.** 2006: Recalled pain ratings: a complex and poorly defined task. *Journal of Pain* 7, 142–49.

Butz, A.M., Donithan, M., Bollinger, M.E., Rand, C. and **Thompson, R.E.** 2005: Monitoring nebulizer use in children: comparison of electronic and asthma diary data. *Annals of Allergy, Asthma and Immunology* 94, 360–65.

Carp, F.M. and **Carp, A.** 1981: The validity, reliability and generalizability of diary data. *Experimental Aging Research* 7, 281–96.

Cruise, C.E., Broderick, J., Porter, L., Kaell, A. and **Stone, A.A.** 1996: Reactive effects of diary self-assessment in chronic pain patients. *Pain* 67, 253–58.

Dawson, E.G., Kanim, L.E., Sra, P., Dorey, F.J., Goldstein, T.B., Delamarter, R.B. and **Sandhu, H.S.** 2002: Low back pain recollection versus concurrent accounts: outcomes analysis. *Spine* 27, 984–93.

Donner, A. and **Eliasziw, M.** 1987: Sample size requirements for reliability studies. *Statistics in Medicine* 6, 441–48.

Dunn, K.M. and **Croft, P.R.** 2005: Classification of low back pain in primary care: using 'bothersomeness' to identify the most severe cases. *Spine* 30, 1887–92.

Dunn, K.M. and **Croft, P.R.** 2006: Repeat assessment improves the prediction of prognosis in patients with low back pain in primary care. *Pain* 126, 10–15.

Freer, C.B. 1980: Self-care: a health diary study. *Medical Care* 18, 853–61.

Goossens, M.E., Rutten-van Molken, M.P., Vlaeyen, J.W. and **van der Linden, S.M.** 2000: The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *Journal of Clinical Epidemiology* 53, 688–95.

Guzmán, J., Peloso, P. and **Bombardier, C.** 1999: Capturing health care utilization after occupational low-back pain: development of an interviewer-administered questionnaire. *Journal of Clinical Epidemiology* 52, 419–27.

Jamison, R.N., Raymond, S.A., Levine, J.G., Slawsby, E.A., Nedeljkovic, S.S. and **Katz, N.P.** 2001: Electronic diaries for monitoring chronic pain: 1-year validation study. *Pain* 91, 277–85.

Jamison, R.N., Raymond, S.A., Slawsby, E.A., McHugo, G.J. and **Baird, J.C.** 2006: Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. *Journal of Pain* 7, 192–99.

Jensen, M.P. and **McFarland, C.A.** 1993: Increasing the reliability and validity of pain intensity measurement in chronic pain patients. *Pain* 55, 195–203.

Jensen, M.P., Turner, J.A. and **Romano, J.M.** 1994: What is the maximum number of levels needed in pain intensity measurement? *Pain* 58, 387–92.

Jensen, M.P., Turner, J.A., Romano, J.M. and **Fisher, L.D.** 1999: Comparative reliability and validity of chronic pain intensity measures. *Pain* 83, 157–62.

Jensen, M.P., Turner, L.R., Turner, J.A. and **Romano, J.M.** 1996: The use of multiple-item scales for pain intensity measurement in chronic pain patients. *Pain* 67, 35–40.

Jordan, K., Dziedzic, K., Jones, P.W., Ong, B.N. and **Dawes, P.T.** 2000: The reliability of the three-dimensional FASTRAK measurement system in measuring cervical

spine and shoulder range of motion in healthy subjects. *Rheumatology* 39, 382–88.

Lefebvre, J.C. and **Keefe, F.J.** 2002: Memory for pain: the relationship of pain catastrophizing to the recall of daily rheumatoid arthritis pain. *Clinical Journal of Pain* 18, 56–63.

Linton, S.J. and **Götestam, K.G.** 1983: A clinical comparison of two pain scales: correlation, remembering chronic pain, and a measure of compliance. *Pain* 17, 57–65.

Linton, S.J. and **Melin, L.** 1982: The accuracy of remembering chronic pain. *Pain* 13, 281–85.

Lousberg, R., Schmidt, A.J., Groenman, N.H., Vendrig, L. and **Dijkman-Caes, C.I.** 1997: Validating the MPI-DLV using experience sampling data. *Journal of Behavioral Medicine* 20, 195–206.

McGorry, R.W., Webster, B.S., Snook, S.H. and **Hsiang, S.M.** 1999: Accuracy of pain recall in chronic and recurrent low back pain. *Journal of Occupational Rehabilitation* 9, 169–78.

Morley, S. 2007: Thanks for the memory. *Pain* 132, 231–32.

Raselli, C. and **Broderick, J.E.** 2007: The association of depression and neuroticism with pain reports: a comparison of momentary and recalled pain assessment. *Journal of Psychosomatic Research* 62, 313–20.

Raspe, H. and **Kohlmann, T.** 1994: Disorders characterised by pain: a methodological review of population surveys. *Journal of Epidemiology and Community Health* 48, 531–37.

Roland, M. and **Morris, R.** 1983: A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8, 141–44.

Salaffi, F, Stancati, A., Alberto Silvestri, C., Ciapetti, A. and **Grassi, W.** 2004: Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *European Journal of Pain* 8, 283–91.

Salovey, P., Jobe, J.B. and **Willis, G.B.** 1992: Reporting chronic pain episodes on health surveys. *National Center for Health Statistics. Vital Health Statistics. Vital Health Statistics Series* 6, 1–71.

Salovey, P., Smith, A.F., Turk, D.C., Jobe, J.B. and **Willis, G.B.** 1993: The accuracy of memory for pain: not so bad most of the time. *American Pain Society Journal* 2, 184–91.

Schwartz, J.E. and **Stone, A.A.** 1998: Strategies for analyzing ecological momentary assessment data. *Health Psychology* 17, 6–16.

Shrout, P.E. and **Fleiss, J.L.** 1979: Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 86, 420–28.

Stewart, W.F., Lipton, R.B., Kolodner, K.B., Sawyer, J., Lee, C. and **Liberman, J.N.** 2000: Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain* 88, 41–52.

Stone, A.A., Broderick, J.E., Shiffman, S.S. and **Schwartz, J.E.** 2004: Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain* 107, 61–69.

Stone, A.A., Kessler, R.C. and **Haythornthwaite, J.A.** 1991: Measuring daily events and experiences: decisions for the researcher. *Journal of Personality* 59, 575–607.

Verbrugge, L.M. 1980: Health diaries. *Medical Care* 18, 73–95.

Von Korff, M. 2001: Epidemiological and survey methods – assessment of chronic pain. In Turk, D.C. and Melzack, R., editors, *Handbook of Pain Assessment*. New York, USA: Guildford Press, 603–18.

Von Korff, M., Jensen, M.P. and **Karoly, P.** 2000: Assessing global pain severity by self-report in clinical and health services research. *Spine* 25, 3140–151.

Zigmond, A.S. and **Snaith, R.P.** 1983: The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* 67, 361–70.