Pain 110 (2004) 310–317

PAIN

www.elsevier.com/locate/pain

Electronic pain questionnaires: A randomized, crossover comparison with paper questionnaires for chronic pain assessment

Andrew J. Cook*, David A. Roberts, Michael D. Henderson, Lisa C. Van Winkle, Dania C. Chastain, Robin J. Hamill-Ruth

Department of Anesthesiology, Division of Pain Management, University of Virginia Health System, P.O. Box 801008, Charlottesville, VA 22908-1008, USA

Received 26 November 2003; received in revised form 20 February 2004; accepted 5 April 2004

Abstract

Electronic questionnaires for pain assessment are becoming increasingly popular. There have been no published reports to establish the equivalence or psychometric properties of common pain questionnaires administered via desktop computers. This study compared responses to paper (P) and touch screen electronic (E) versions of the Short-Form McGill Pain Questionnaire (SF-MPQ) and Pain Disability Index (PDI), while examining the role of computer anxiety and experience, and evaluating patient acceptance. In a randomized, crossover design 189 chronic pain patients completed P and E versions of the SF-MPQ and PDI, and self-ratings of anxiety, experience, relative ease and preference. Psychometric properties were highly similar for P and E questionnaires. For the SF-MPQ, 60% or more of subjects gave equivalent responses on individual descriptors and PPI scale, with 80% rating within ± 1 point for an 11-point VAS. Correlations for the SF-MPQ scales ranged from 0.68 to 0.84. For the PDI, 60% or more of subjects responded within ± 1 point on individual questions, and the total score correlation was 0.67. Comparison of mean difference scores revealed no significant differences between modes for any of the questionnaire items or scores. Anxiety and experience scores showed no significant associations through correlations and high/low comparisons. Although nearly half of subjects reported no computer training, anxiety ratings were low, and considerably more subjects rated the E questionnaires as easier and preferred. Findings are consistent with test-retest reliability data, and support the validity and acceptance of electronic versions of the SF-MPQ and PDI.

Keywords: Computers; Electronic data; Assessment; Short-form McGill Pain Questionnaire; Pain Disability Index; Chronic pain

1. Introduction

Standardized, psychometrically validated self-rating scales and questionnaires are important tools in pain assessment. The vast majority of these questionnaires have been validated for a paper-based, self-report format of administration.

Electronic technologies have offered much promise for health assessment: data management capabilities, improved adaptability, enhanced ergonomics and appeal, and more truthful responding for sensitive questions (Bock et al., 1999; Fawdry, 1989; Hufford and Shields, 2002; Kleinman et al., 2001; Lucas et al., 1977). However, questions regarding the equivalency of electronic and paper formats have appropriately been raised: Are there perceptual, cognitive, or emotionally-based differences? (Booth-Kewley et al., 1992; Wilson et al, 1985). Professional organizations such as the American Psychological Association (1986) have recommended that equivalency be empirically evaluated. A growing base of research supports the use, equivalence, and acceptance of electronic questionnaires in diverse areas of health assessment (Fawdry, 1989; Kleinman et al., 2001; Paperny et al., 1990; Pouwer et al., 1998; Skinner and Allen, 1983).

Electronic pain assessment systems have been developed and evaluated for more than a decade (Swanston et al., 1993), but have only begun to gain widespread appeal through the advent of palmtop computers (PTCs). A PTC-based version of a 5 cm visual analogue scale (VAS) has been validated through psychophysical techniques (Jamison et al., 2002). Electronic diaries for pain assessment have been shown to have good psychometric properties (Jamison et al., 2001; Peters et al., 2000). A 6-point pain intensity

* Corresponding author. Tel.: +1-434-243-5682; fax: +1-434-243-5689.
E-mail address: ajcook@virginia.edu (A.J. Cook).

0304-3959/$20.00
doi:10.1016/j.pain.2004.04.012

scale within a transcutaneous electrical nerve stimulator was found to have good concurrent validity with traditional pain scales (Scudds et al., 2003). The size, format and applications for these devices have focused their content on brief rating scales.

A touch-screen electronic assessment of chronic pain using the McGill Pain Questionnaire has recently been reported to have good acceptance and completeness across gender, age, ethnicity, education, and computer experience (Wilkie et al., 2003). However, reported equivalence of electronic and paper responses was based on descriptive analysis of a small sample of unpublished data. Watson et al. (2002) reported good validity and acceptance for a computer-assisted postoperative pain interview for children. However, there have been no published reports to establish equivalence or psychometric properties of pain questionnaires administered by desktop computers.

Receptiveness and responses for electronic assessments can be influenced by computer anxiety and experience (Igbaria and Iivari, 1995; Tseng et al., 1997), possibly mediated by computer self-efficacy (Igbaria and Iivari, 1995). Computer anxiety has been found to be more common among individuals who are female, older, lack computer experience, with lower expectations and trait or math anxieties (Gaudron and Vignoli, 2002; Heinssen et al., 1987; Igbaria and Iivari, 1995).

The goals of this study were: (a) to compare distributions and reliabilities of responses to electronic and paper versions of two widely used pain questionnaires (Short-form McGill Pain Questionnaire; Pain Disability Index), (b) to evaluate agreement in responses between the two modes, (c) to determine if differences were associated with computer anxiety or experience, and (d) to determine subjects' relative preferences and perceived ease of use.

2. Methods

2.1. Subjects

As part of their initial evaluation at a multidisciplinary, university-based pain management clinic, 200 patients with chronic pain were recruited for participation in the study. Completion of the questionnaires under study, in either paper or electronic format, was a requirement of their clinical evaluations. Eleven randomized subjects did not complete the study or provided incomplete data on paper questionnaires and were excluded from analyses. They did not significantly differ from the remaining sample in terms of age, gender, education level, pain location or duration ($P > 0.05$). The study was approved by the Human Investigation Committee of the university's Institutional Review Boards, and all subjects provided informed consent to participation. Patients were excluded from participation if they were less than 18 years of age, or were illiterate. Subjects received no compensation for their participation.

Subjects were 63% female, 84% Caucasian and 12% African American with a mean age of 47.5 years (SD = 12.8, range 18–82). Primary pain locations included low back/-sacral (21%), lower limbs (24%), head/face/neck (18%), other single sites (19.1%), or a combination of major body sites (18%), based on the classification system of the International Association for the Study of Pain. The average duration of pain at time of assessment was 68 months (SD = 94). Fifty-one percent of subjects were married, with 30% having a high school education and another 44% having some level of college training. The educational level of the sample was higher than that of the general clinic population due to the literacy requirement.

2.2. Materials

The *Short-form McGill Pain Questionnaire* (SF-MPQ; Melzack, 1987) and *Pain Disability Index* (PDI; Pollard, 1984) were selected for the equivalency comparison because of their inclusion in standard assessments in our clinical setting, widespread use in chronic pain assessment, acceptable psychometric properties, ease of comprehension, and brevity.

The SF-MPQ is a widely used measure of the sensory, affective and intensity dimensions of pain. It has been translated from English into several other languages (Melzack and Katz, 2001). It is comprised of 15 pain descriptors (11 sensory, 4 affective) from the McGill Pain Questionnaire (MPQ) (Melzack, 1975), rated on a 4-point severity scale, with ratings summed to produce the Pain Rating Index (PRI), as well as sensory and affective subscales. Pain intensity is assessed with a 10 cm visual analogue scale (VAS) and a 6-point present pain intensity (PPI) scale. Good consistency, validity, treatment sensitivity and discriminant ability have been demonstrated (Melzack, 1987; Melzack and Katz, 2001). Test-retest reliabilities have not been reported for English PRI or PPI scales, with the exception of data showing low mean change scores over 3–8 week periods in a small sample of cancer patients (Dudgeon et al., 1993). They are moderate to high for the corresponding PRI scales of the MPQ ($r = 0.76–0.83$), but vary considerably for the pain descriptor categories ($r = 0.29–0.83$) (Love et al., 1989). They are moderate to high for Greek and Swedish versions of the SF-MPQ (correlations 0.58–0.98, same day to 1 month periods) (Burkhardt and Bjelle, 1993; Georgoudis et al., 2000). The VAS has been extensively validated as an independent scale (Jensen and Karoly, 1992). Test-retest reliabilities are moderate to high for experimental pain (mean differences 0.7–1.2 on 10-unit scales, to $r = 0.90$), and for pain affect ratings for clinical pain ($r = 0.70–0.90$) (Price et al., 2001). Because the input format of our electronic questionnaire limited sensitivity on the VAS to 11 points (0–10), responses from the paper format were scored in the same manner by rounding. Equivalence of 101-point

paper and electronic VAS ratings has been demonstrated (Jamison et al., 2002).

The PDI is a brief measure of pain-related disability, assessing perceived impairment of functioning across seven areas of daily functioning. It has been found to be an internally consistent ($\alpha = 0.86$) measure of perceived disability, with good concurrent, criterion-related and discriminative validity in a variety of inpatient and outpatient pain populations (Pollard, 1984; Tait et al., 1990). Its reported stability is moderate (test-retest $r = 0.44$), though published data were for an 11–307 day interval and may reflect factors other than the measure's reliability (e.g. external treatment effects, regression to the mean). Although a two-factor structure was initially identified for the PDI, the balance of research supports a single underlying factor (Jacob and Kerns, 2001).

For computer anxiety and experience we developed a 4-item questionnaire, titled the *Computer Anxiety and Experience Assessment* (CAEA). Its content is based on a previously validated computer anxiety scale (Heinssen et al., 1987) and a study on computer self-efficacy (Igbaria and Iivari, 1995). The two anxiety items, each rated on 5-point categorical rating scales, were 'I have avoided computers because they are unfamiliar and somewhat intimidating to me', and 'I am confident that I can learn computer skills'. These were summed with the latter reverse scored to generate a computer anxiety score. The two experience questions, rated on anchored 7-point likert scales, were 'Please indicate the amount of experience you have with computers', and 'Please indicate how often you use a computer'. These were summed to generate a computer experience score. Two additional questions presented at the completion of the study assessed perceived ease of use and preference for the questionnaire formats, with three response choices: paper, computer, no preference. The CAEA and preference questions were administered solely in paper format.

2.3. Equipment

The electronic questionnaires were based on Microsoft Access 2000 (Microsoft Corporation, USA) format, replicating as closely as possible the paper formats. Four desktop computers with Pentium III, 600 MHz processors, 196 MB RAM, 4 MB video memory and Intel 82810E video cards (Intel Corporation, USA) were located in a quiet and private room within the clinic, operating on Microsoft Windows 2000 Professional (Microsoft Corporation, Seattle, WA) operating systems. Data were entered via 17-inch, 1280×1024 pixel resolution, Accutouch resistance touchscreen LCD monitors (Elo Touchsystems, USA), and standard keyboards. A 2-button mouse was also available for those who preferred. The screen was positioned according to subject preference, typically at chest height and a distance of approximately 12 in., and secured using an adjustable monitor arm. Approximately 15 questions were viewable per screen, and the forms were structured such that all items required responses before subjects could advance to the subsequent questionnaire. A gray background and black text were used for all screens.

2.4. Procedure

All subjects completed the paper (P) and electronic (E) versions of the SF-MPQ and PDI in a randomized, crossover design. After completing the anxiety and experience questions (1 through 4) of the CAEA, subjects were randomly assigned to 1 of 2 experimental conditions. Condition 1 completed the E followed by P questionnaires, while condition 2 completed the P followed by E questionnaires. Subjects were not able to consult the previously completed forms. Between the two versions of the SF-MPQ and PDI, subjects completed a series of electronic forms that included demographic, pain, and psychosocial questionnaires. These additional electronic questions provided a latency of approximately 45 min. This delay was desired to minimize rote repetition between formats while maximizing consistency of symptoms and perceptions being reported and avoiding excess fatigue. The mean completion time for the full set of questionnaires was 62 min (SD = 18).

Prior to starting the E questionnaires, subjects were given a brief information session to familiarize them with the relevant computer components and response methods. A trained facilitator was available to provide limited procedural assistance to subjects as needed, while providing for adequate privacy. Subjects were encouraged to take brief rest or stretching breaks as needed. For all subjects, after both P and E questionnaires were completed, they were asked to respond to the final two CAEA (preference and ease) questions in paper format.

2.5. Data analysis

Data screening procedures revealed negatively skewed distributions for the majority of the SF-MPQ and PDI items and the total PDI scores. The distributions were improved by statistical transformations (square root of reflected scores). Relevant analyses were compared for observed and transformed versions of these variables. Given the absence of significant impact on the observed results, reported results are based on the non-transformed variables (Tabachnick and Fidell, 1989). All within-subjects difference scores (P–E) were normally distributed. To avoid bias due to observed violations of bivariate normality, all correlational analyses were conducted with Spearman's nonparametric rank-order correlation coefficient (ρ).

Comparison of P and E responses to the two questionnaires included examination of score distributions, calculation of Cronbach's coefficient α for internal consistencies of scale scores, review of frequency distributions for within-subjects difference scores, correlations of

test items and scale scores, and calculation of mean difference scores (P–E) with 95% confidence limits (to define the likely range of true values in the population from which we drew our sample). Correlation coefficients for individual questions were given low weight in the interpretation of results as they are highly influenced by variability of responses between subjects, and as such are not good indicators of agreement between measurement methods when variability is low (Bland and Altman, 1990). Given the intent to accept the null hypothesis (absence of differences), minimization of the type II error rate (β) was desired. Therefore, the per comparison rate for type I (α) error was set at $P = 0.05$ and overall type I error for the multiple confidence intervals was not controlled.

Setting the type II error rate at 0.10 and type I error at 0.05, we calculated minimum detectable differences (paper vs. electronic) for the main test scales. Variances for the P and E forms were assumed to be equal, and were based on prior sample values from our clinical population. Correlations between P and E tests were estimated from available test-retest reliability data (SF-MPQ scales $r = 0.80$, PDI score $r = 0.50$). With sample size 200, we estimated 90% power to detect mean differences between the two modes of 0.87 points Sensory, 0.46 Affective, 1.24 Total PRI, 0.22 VAS, 0.14 PPI, and 2.84 total PDI. Post-hoc analyses with observed difference scores and $n = 189$ were consistent with these estimates, indicating 90% power to detect minimum mean differences of 0.88 Sensory, 0.50 Affective, 1.17 Total PRI, 0.30 VAS, 0.19 PPI, and 2.71 total PDI.

The CAEA anxiety and experience scores were positively skewed, and not improved by attempted transformations. Good internal consistency was found for the anxiety (Cronbach's $\alpha = 0.57$) and experience ($\alpha = 0.81$) questions. Divergent validity was supported by a predicted strong negative correlation between anxiety and experience scores ($\rho = -0.65$). Using education level as a measure of external validity, an expected negative association was found for anxiety scores ($\rho = -0.48$), and a positive correlation with experience scores ($\rho = 0.57$). Anxiety scores were found to have a predicted modest correlation with age ($\rho = 0.21$). Because of the highly skewed distributions, the CAEA scales were dichotomized by median splits to form low/high categorizations of computer anxiety and experience (Tabachnick and Fidell, 1989). The relationships of anxiety and experience to within-subjects SF-MPQ and PDI difference scores were evaluated through correlations and grouped (high/low) comparisons, with Bonferonni adjustment of type I error rates to maintain total $P = 0.05$.

3. Results

Distributions of item responses and calculated scale scores for the SF-MPQ and PDI were very similar for paper (P) and electronic (E) questionnaires. Descriptive statistics for the main scale scores for the two questionnaires are presented in Table 1. Cronbach's αs were also found to be highly similar between modes of administration, as shown in Table 1. Within-mode (P or E) correlation matrices for the SF-MPQ and PDI scales were compared and were also highly similar for the P and E response sets.

Table 1
Descriptive statistics for scales from paper (P) and electronic (E) questionnaire administrations

Scale	Min	Max	Median	Mean	SD	α
Short-form McGill Pain Questionnaire						
P_Sensory	2	33	17	17.7	7.2	0.81
E_Sensory	2	33	18	17.4	7.1	0.81
P_Affective	0	12	5	5.5	3.5	0.78
E_Affective	0	12	6	5.7	3.5	0.81
P_Total PRI	2	45	22	23.2	9.8	0.86
E_Total PRI	2	42	23	23.2	9.8	0.87
P_VAS	2	10	8	7.5	1.8	
E_VAS	1	10	8	7.3	1.8	
P_PPI	0	5	3	2.8	1.2	
E_PPI	0	5	3	2.9	1.3	
Pain Disability Index						
P_Total	0	65	42	40.0	16.0	0.85
E_Total	0	66	42	41.1	15.5	0.85

Table 2 presents correlations *between* P and E responses and scores for the SF-MPQ, as well as the percentages of exact matches, and percentages within ±1 point for individual items or ±10% scale range for summed scores. The correlations were generally high, particularly for

Table 2
Correlations, percentage of subjects with zero differences, and percentage within ±1 point or 10% scale range for paper vs. electronic formats of the Short-form McGill Pain Questionnaire

	Spearman's ρ	% zero difference	% within ±1 point
Throbbing	0.64	61.4	92.7
Shooting	0.72	64.6	93.7
Stabbing	0.69	60.8	91.4
Sharp	0.62	66.1	90.4
Cramping	0.80	69.8	94.1
Gnawing	0.73	67.2	91.5
Hot-burning	0.75	64.6	92.1
Aching	0.57	69.3	97.9
Heavy	0.70	62.4	90.5
Tender	0.65	63.0	90.5
Splitting	0.71	64.0	89.9
Tiring-exhausting	0.57	60.8	89.9
Sickening	0.70	62.4	91.5
Fearful	0.69	64.0	91.0
Punishing-cruel	0.70	61.4	90.0
Sensory PRI	0.83	16.4	70.3[a]
Affective PRI	0.77	29.6	60.3
Total PRI	0.84	11.1	77.6[b]
Visual analogue scale	0.68	38.4	80.0
Present pain intensity	0.74	64.0	93.1

[a] Percentage within ±3 points.
[b] percentage within ±5 points.

Table 3
Correlations, percentage of subjects with zero differences, and percentage within ±1 point or 10% total scale range for paper vs. electronic formats of the Pain Disability Index

	Spearman's ρ	% zero difference	% within ±1 point
Family/home	0.62	47.6	66.6
Recreation	0.56	38.1	60.3
Social activity	0.57	35.4	64.5
Occupation	0.65	46.0	65.0
Sexual behavior	0.74	57.7	76.8
Self care	0.64	37.6	63.0
Life-support activity	0.61	34.4	59.8
Total PDI	0.67	9.5	69.1[a]

[a] percentage within ±7 points.

the summed scores, which had greater variability. The majority of Pain Rating Index (PRI) descriptors and the Present Pain Intensity (PPI) scale had close to two-thirds exact matches between formats, and over 90% within ±1 point. The summed PRI scores had 60–77% of subjects scoring within ±10% scale range. Slightly more than a third of subjects had exact matches on the VAS, but 80% fell within ±1 point. The effect of scale ranges and variability of scores on these comparisons can be seen in the PRI 'aching' item (4 point range), which had one of the lowest correlations but highest percentage of exact matches, and the PRI total score (46 point range) with the highest correlation but smallest percentage of exact matches.

The correlations and percentages of exact matches and ±10% difference scores for the PDI are presented in Table 3. Item and total score correlations were moderate to high. The percentage of exact matches between P and E item responses was modest (31–55%). The sexual behavior item had the highest agreement, but also the highest variance in responses for both P and E formats, suggesting higher specific memory for responses to this item. Total score responses varied somewhat between modes of administration, with two-thirds of subjects having difference scores within ±7 points for the 71-point scale.

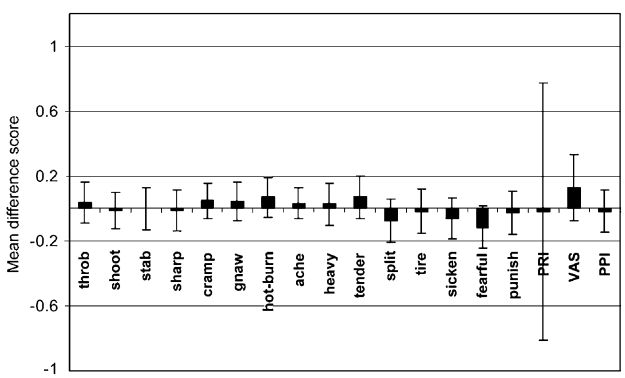

Fig. 1. Short-form McGill Pain Questionnaire (SF-MPQ) difference scores. Mean difference scores and 95% confidence intervals for within-subjects comparisons of paper minus electronic SF-MPQ responses and total PRI score.

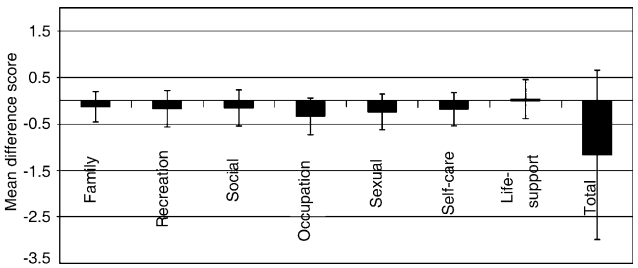

Fig. 2. Pain Disability Index (PDI) difference scores. Mean difference scores and 95% confidence intervals for within-subjects comparisons of paper minus electronic PDI responses and total score.

Mean difference scores with 95% confidence intervals for the questionnaire items and scores are shown in Fig. 1 for the SF-MPQ and Fig. 2 for the PDI. With 95% confidence, all of these ranges of likely true mean differences included zero.

Potential order effects were evaluated through within-subjects, P vs. E item comparisons for the two orderings of the questionnaires. For each ordering, a difference was found for one or two questionnaire items. For subjects completing the P forms first, P responses were higher on SF-MPQ item 7 ('hot-burning', mean 1.35 vs. 1.12, $P = 0.04$). For subjects completing the E forms first, E responses were higher on SF-MPQ item 14 ('fearful', mean 1.22 vs. 0.94, $P = 0.006$), and PDI item 4 ('occupation', mean 7.3 vs. 6.7, $P = 0.05$). Although the type I error rate was not adjusted for the previously stated reasons, it is recognized that with a probability level $P = 0.05$, one to two significant differences would be expected as a result of random error for each set of multiple comparisons.

The potential for a time or activity related effect in difference scores was evaluated by comparing responses to first (P or E) vs. second completed questionnaires. Paired *t*-tests for Sensory, Affective and Total PRI, PPI, VAS and total PDI scores revealed no significant differences across time (first to second testing) for any of the pain or disability scores ($P = 0.08–0.80$). Additionally, no significant correlations of pain severity at the time of testing (mean of P and E PPI ratings from SF-MPQ) with any of the difference scores were found ($\rho < 0.12$ for all comparisons).

The CAEA response distributions indicated generally low amounts of computer experience, but also low rated computer anxiety levels. Forty-eight percent of subjects reported no formal computer training, while 37% reported using a computer less than once per month. CAEA anxiety and experience scores had weak correlations with all SF-MPQ and PDI item and scale difference scores ($\rho \leq 0.15$ anxiety, $\rho \leq 0.13$ experience). The only outlier was the 'hot-burning' item on SF-MPQ with $\rho = 0.23$, but this was not statistically significant at the controlled error rate of $P = 0.003$ for the multiple comparisons. Grouped mean comparisons for the sensory, affective and total PRI, VAS, PPI and total PDI difference scores revealed no significant differences between high and low anxiety

subjects ($t < 1.68, P > 0.10$), nor between high and low experience subjects ($t < 1.28, P > 0.20$).

When asked 'Comparing the two paper questionnaires with the *same* two computer questionnaires, which were EASIER to understand and complete?', more subjects chose the electronic (39%) over the paper (24%) mode, while 37% rated no difference. Responses to the similar question on 'PREFERRED' format more substantially favored the electronic (48%) over paper (24%) questionnaires, with 28% of subjects reporting no preference.

4. Discussion

The results of this study suggest equivalent patterns and distributions of responses to paper (P) and electronic (E) versions of two widely used pain assessment questionnaires among a sample of chronic pain patients. There was high concordance between responses to the two testing modes for the SF-MPQ and moderate concordance for the PDI, with no significant mean differences. The E questionnaires were highly accepted, and in many cases preferred, by patients with chronic pain.

Our findings are consistent with those of other randomized, crossover comparisons of P and E questionnaires for health assessment. Prior studies have shown uniform psychometric properties between modes of administration, with high correlations and no significant mean differences (Kleinman et al., 2001; Pouwer et al., 1998). The high rates of perceived ease and acceptance we found for the E questionnaires are also consistent (Bock et al., 1999; Fawdry, 1989; Pouwer et al., 1998). Although almost half of our subjects reported no formal computer training, and many used computers infrequently, nearly half preferred the E questionnaires. The high preference ratings were mirrored in frequent comments by subjects indicating enjoyment of the novelty and challenge of the computer task.

For a minority of subjects there were notable differences in P and E responses. It is important to consider whether these response differences were due to random measurement error, the mode of administration (P vs. E) or other variable(s). One possible factor is time or activity-related changes in pain. Our subjects spent the time between the two test administrations sitting and working at computer stations. However, the lack of a time effect (first to second testing) for pain or disability ratings suggests this was not a factor here. A second possibility is order effects. Order of presentation had minimal effect on our subjects' responses, with a trend to providing higher ratings for some questions when asked at the beginning of the evaluation (in either P or E format). A possible rationale would be higher generalized anxiety levels at the start of the evaluation process. Further study will be required to determine the impact of sequencing and timing on E responses. Another possibility is the test-retest reliabilities of the questionnaires. Available data suggest that these are only fair for the PDI, and good to excellent for the SF-MPQ. The observed response correlations between modes are generally consistent with known reliabilities of the scale scores for clinical pain ratings. Further research with other highly reliable questionnaires will help clarify the impact of computer administration on observed response differences.

A potentially related factor is variability in patient ratings of chronic pain across time. Without significant change in a painful condition, it is conceivable that a person can report his pain with a slightly different perspective or focus at different points in time. In addressing this issue relative to test-retest variations in McGill Pain Questionnaire ratings, Melzack and Katz (2001) suggested 'many clinical pains show fluctuations in quality over time, yet still represent the 'same' pains to the persons who experience them.' (Melzack and Katz, 2001, p.42). This might be more likely when subjects are asked to focus on some aspects of their pain experiences between ratings, as in the current study. This phenomenon is supported by chronic pain data from electronic diary assessments showing stability of pain report over weeks, but high variability between and within days (Peters et al., 2000). In some patients this variability is linear, while in others no discernible pattern is evident.

The majority of observed differences in P vs. E ratings were small for the SF-MPQ and PDI. Although these differences were not related to reported levels of computer anxiety or experience, the reported levels of anxiety were surprisingly low in a sample of patients with limited computer experience. It is possible that unfamiliarity or lower proficiency for computer use could have accounted for response differences by some subjects. For example, subjects with less proficiency could make data entry errors more commonly, or have greater difficulty following on-screen questionnaire instructions as compared to those on paper. Differences were also not related to reported pain levels at the time of the assessments. However, variables that we did not assess, such as medication use, could account for some of the within-subjects differences.

Given these potential influences on observed results, future comparative studies could be enhanced by sampling both P and E responses on repeated occasions within and between days, comparing reliabilities and averaged scores. Additionally, computer proficiency could be assessed through a standardized task assessing comprehension of on-screen instructions and response precision. The potential effects of visual presentation style and novelty on acceptance and preference for E questionnaires could be assessed through comparison of responses to questions presented in different formats and/or color schemes. The relationship of medication use at the time of the testing to response patterns could also be evaluated.

An additional benefit of the E questionnaires is the ability to restrict missing data entries. Our E questionnaires required subjects to complete all items before advancing to the next form or end point. This provided a higher rate of data completion than for the P questionnaires. Incomplete

responses are a known limitation of P questionnaires, often requiring additional staffing for compliance monitoring and facilitation. Recent research has shown compliance with P diaries for pain assessment to be poor, and much less than reported compliance by participants (Stone et al., 2003). Substantially higher compliance rates were found for E diaries with enhanced compliance features, such as active prompting and compliance feedback.

Many opportunities exist for use of validated E versions of pain assessment questionnaires. These include web-based clinical and research applications, and the rapidly growing field of telemedicine. Electronic pain assessment has recently been incorporated in an internet-based clinical trial for osteoarthritis of the knee (McAlindon et al., 2003), and has potential benefits for pain assessment in telemedicine interventions such as self-regulation training for chronic pain (Appel et al., 2002). Significant growth in these areas of pain research and practice is anticipated.

There are several limitations to this study. As previously noted, the effects of test-retest reliabilities could not be isolated as each format was completed only once by each subject, and comparable data for same day retesting with the SF-MPQ and PDI are not available. For the SF-MPQ, the 101-point VAS scale could not be validated in this study because the sensitivity of our E administration was limited by technical factors. Equivalence of a 5 cm electronic pain VAS on PDA to the 10 cm paper version has previously been demonstrated through psychophysical techniques (Jamison et al., 2002). Additionally, our measure of computer anxiety and experience might have been biased by inadequate sensitivity. This was an unfortunate tradeoff in our effort to avoid excessive assessment burden on patients undergoing full clinical evaluations. Future research should include more comprehensive measures of these important variables (e.g. Heinssen et al., 1987). While maximizing consistency of pain and associated perceptions, the short delay between questionnaire administrations might have produced a memory effect in our results. This is likely to have been minimized by the distractor tasks. However, further research should evaluate the influence of a longer interval with a more neutral (non-pain related) task.

In summary, our results suggest that the SF-MPQ and PDI perform equally well when administered in paper or touch-screen computer format. High agreement was demonstrated for P and E SF-MPQ responses, and moderate agreement for PDI responses. The two questionnaires were found to have highly similar response distributions in E vs. P formats, similar internal consistencies and internal structures, and the E forms were rated as comparable or preferred by the majority of participants. Observed differences in responses to the two formats were not related to reported computer anxiety or experience. These findings support the validity of these E questionnaires, and along with observed benefits for patient enjoyment and compliance, suggest strong potential for their expanded use in chronic pain assessment.

Acknowledgements

The authors wish to thank Beth Hall and Melissa Bunnell for assistance in conducting this study.

References

American Psychological Association, 1986, Guidelines for computer based tests and interpretations. Washington: DC: APA; 1986.

Appel PR, Bleiberg J, Noiseux J. Self-regulation training for chronic pain: can it be done effectively by telemedicine? Telemedicine J E-Health 2002;8:361–8.

Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med 1990;20:337–40.

Bock B, Niaura R, Fontes A, Bock F. Acceptability of computer assessments among ethnically diverse, low-income smokers. Am J Health Promotion 1999;13:299–304.

Booth-Kewley S, Edwards JE, Rosenfeld P. Impression management, social desirability, and computer administration of attitude questionnaires: does the computer make a difference? J Appl Psychol 1992;77:562–6.

Burkhardt CS, Bjelle A. A Swedish version of the short-form McGill Pain Questionnaire. Scand J Rheumatol 1993;23:77–81.

Dudgeon D, Raubertas RF, Rosenthal SN. The short-form McGill Pain Questionnaire in chronic cancer pain. J Pain Symp Man 1993;8:191–5.

Fawdry R. The development and evaluation of a computerised antenatal questionnaire. Int J Gyn Obst 1989;28:229–35.

Gaudron JP, Vignoli E. Assessing computer anxiety with the interaction model of anxiety: development and validation of the computer anxiety trait subscale. Comp Hum Behav 2002;18:315–25.

Georgoudis G, Oldham JA, Watson PJ. Reliability and sensitivity measures of the Greek version of the short form of the McGill Pain Questionnaire. Eur J Pain 2001;5:109–18.

Heinssen Jr RK Gl, ass CR, Knight L. Assessing computer anxiety: development and validation of the Computer Anxiety Rating Scale. Comput Hum Behav 1987;3:49–59.

Hufford MR, Shields AL. Electronic diaries. Appl Clin Trials 2002;46–56.

Igbaria M, Iivari J. The effects of self-efficacy on computer usage. Int J Manage Sci 1995;23:587–605.

Jacob MC, Kerns RD. Assessment of the psychosocial context of the experience of chronic pain. In: Turk DC, Melzack R, editors. Handbook of pain assessment. New York: Guilford Press; 2001. p. 362–84.

Jamison RN, Gracely RH, Raymond SA, Levine JG, Marino G, Herrmann TJ, Daly M, Fram D, Katz NP. Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers. Pain 2002;99:341–7.

Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, Katz NP. Electronic diaries for monitoring chronic pain: 1-year validation study. Pain 2001;91:277–85.

Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, editors. Handbook of pain assessment. New York: Guilford Press; 1992. p. 15–34.

Kleinman L, Leidy NK, Crawley J, Bonomi A, Schoenfeld P. A comparative trial of paper-and-pencil versus computer administration of the Quality of Life in Reflux and Dyspepsia (QOLRAD) questionnaire. Med Care 2001;39:181–9.

Love A, Leboeuf DC, Crisp TC. Chiropractic chronic low back pain sufferers and self-report assessment methods: Part I. A reliability study

of the visual analogue scale, the pain drawing and the McGill Pain Questionnaire. J Manipul Physiol Therap 1989;12:21–5.

Lucas RW, Mullin PJ, Luna CBX, McInroy DC. Psychiatrists and a computer as interrogators of patients with alcohol-related illness: a comparison. Br J Psychiatry 1977;131:160.

McAlindon T, Formica M, Kabbara K, LaValley M, Lehmer M. Conducting clinical trials over the internet: feasibility study. Br Med J 2003;327:484–7.

Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30: 191–7.

Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975;1:277–99.

Melzack R, Katz J. The McGill Pain Questionnaire: appraisal and current status. In: Turk DC, Melzack R, editors. Handbook of pain assessment. New York: Guilford Press; 2001. p. 35–52.

Paperny DM, Aono JY, Lehman RM, Hammar SL, Risser J. Computer-assisted detection and intervention in adolescent high-risk health behaviors. J Pediatr 1990;116:456–62.

Peters ML, Sorbi MJ, Kruise DA, Kerssens JJ, Verhaak PFM, Bensing JM. Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. Pain 2000;84: 181–92.

Pollard CA. Preliminary validity study of the Pain Disability Index. Percept Mot Skills 1984;59:974.

Pouwer F, Snoek FJ, van der Ploeg HM, Heine RJ, Brand AN. A comparison of the standard and the computerized versions of the Well-being Questionnaire (WBQ) and the Diabetes Treatment Satisfaction Questionnaire (DTSQ). Qual Life Res 1998;7:33–8.

Price DD, Riley I.I.I.JL, Wade JB. Psychophysical approaches to measurement of the dimensions and stages of pain. In: Turk DC, Melzack R, editors. Handbook of pain assesssment, 2nd ed. ; 2001. p. 53–75.

Scudds RA, Fishbain DA, Scudds RJ. Concurrent validity of an electronic descriptive pain scale. Clin Rehab 2003;17:206–8.

Skinner HA, Allen BA. Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug and tobacco use. J Consul Clin Psychol 1983;51:267–75.

Swanston M, Abraham C, Macrae WA, Walker A, Rushmer R, Elder L, Methven H. Pain assessment with interactive computer animation. Pain 1993;53:347–51.

Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003; 24:182–99.

Tabachnick BG, Fidell LS. In: Tabachnick BG, Fidell LS, editors. Using multivariate statistics, 2nd ed. New York: Harper Collins; 1989.

Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. Pain 1990;40:171–82.

Tseng HM, Macleod HA, Wright P. Computer anxiety and measurement mood change. Comp Hum Behav 1997;13:305–16.

Watson S, Calam R, Jimmieson P. Can computers help in assessing children's postoperative pain? Initial validation of a computer-assisted interview. Eur J Anaesthes 2002;19:510–6.

Wilkie DJ, Judge MKM, Berry DL, Dell J, Zong S, Gilespie R. Usability of a computerized PAIN *Report* It in the general public with pain and people with cancer pain. J Pain Symptom Manage 2003;25:213–24.

Wilson FR, Genco KT, Yager GG. Assessing the equivalence of paper-and-pencil vs. computerized test: demonstration of a promising methodology. Comp Hum Behav 1985;1:265–75.