Clinical validation of the Paediatric Pain Profile

Anne Hunt* RSCN PhD, Research Fellow, Royal College of Nursing Institute, Oxford; Institute of Child Health, London;
Ann Goldman FRCPCH, Consultant in Paediatric Palliative Care, Great Ormond Street Hospital NHS Trust, London;
Kate Seers RGN PhD, Head of Research;
Nicola Crichton PhD, Medical Statistician, Royal College of Nursing Institute, Oxford;
Kiki Mastroyannopoulou Dip Clin Psychol, Lecturer in Clinical Psychology, School of Medicine, Health Policy and Practice, University of East Anglia, Norwich;
Vivien Moffat RNLD MSc, Research Nurse, Royal Hospital for Sick Children, Lothian University Hospitals NHS Trust, Edinburgh;
Kate Oulton RN Child MA, Research Assistant, Royal College of Nursing Institute;
Michael Brady MRCGP, Medical Director, Martin House Hospice for Children, Wetherby, Yorkshire, UK.

**Correspondence to first author at* Royal College of Nursing Institute, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE.
E-mail: Anne.Hunt@rcn.org.uk

The Paediatric Pain Profile (PPP) is a 20-item behaviour rating scale designed to assess pain in children with severe neurological disability. We assessed the validity and reliability of the scale in 140 children (76 females, mean age 9 years 11 months, SD 4 years 7 months; range 1 to 18 years), unable to communicate through speech or augmentative communication. Parents used the PPP to rate retrospectively their child's behaviour when 'at their best' and when in pain. To assess interrater reliability, two raters concurrently observed and individually rated each child's behaviour. To assess construct validity and responsiveness of the scale, behaviour of 41 children was rated before and for four hours after administration of an 'as required' analgesic. Behaviour of 30 children was rated before surgery and for five days after. Children had significantly higher scores when reported to have pain than 'at their best' and scores increased in line with global evaluations of pain. Internal consistency ranged from 0.75 to 0.89 (Cronbach's alpha) and interrater reliability from 0.74 to 0.89 (intraclass correlation). Sensitivity (1.00) and specificity (0.91) were optimized at a cut-off of 14/60. PPP score was significantly greater before administration of the analgesic than after (paired-sample *t*-tests, $p<0.001$). Though there was no significant difference in mean pre- and postoperative scores, highest PPP score occurred in the first 24 hours after surgery in 14 (47%) children. Results suggest that the PPP is reliable and valid and has potential for use both clinically and in intervention research.

Pain can be difficult to assess in children, leaving them at risk of under-treatment. (Schecter 1989). The criterion standard in pain assessment is 'self-report' (Craig 1997, McIntosh 1997) but for some groups of children, such as infants and those with severe cognitive or neurological impairment, it is their behaviour rather than their verbal report which has to be interpreted to determine whether they have pain (Anand and Craig 1996). While it is possible for many neurologically and cognitively impaired children and adults to use forms of self-report despite their impairments (Ferrell et al. 1995, Parmelee 1996, Fanurik et al. 1998), those unable to report their pain will remain dependent on their caregivers' observational skills. Knowledge is now accumulating on the types of behaviours suggestive of pain in this group, and measurement tools are being developed to assist caregivers in assessing and monitoring their pain. This paper reports on the development of one of these: the Paediatric Pain Profile (PPP).

Several research teams including our own have been working to develop instruments to assess pain in children with physical and learning impairments. Giusiano et al. (1995) and Collignon et al. (2001) investigated pain behaviours in 100 individuals from 2 to 33 years (mean age 16 years) with multiple and profound disability, without speech or any means of communicating through symbols. The original 22 behavioural cues from which items for the scale were derived were elicited from physicians' reports of cues they considered to be indicative of pain during medical examination. The items included various facial expressions, cry (abrupt or spontaneous cry, cry during manipulation), movement and posture (increase in muscle tone and/or involuntary movements, taking up a posture that eases pain, protection of painful areas), and social (reduced interest in surroundings). Pain appeared to be detected by observing global behavioural changes in an individual rather than by the presence of a single sign (Giusiano et al. 1995). In addition, each combination of disabilities appeared to give a specific set of behaviours, for instance, behaviours associated with voluntary protection of painful areas were more likely to occur in individuals with a lesser degree of motor impairment.

McGrath et al. (1998) interviewed twenty parents or caregivers of cognitively impaired children aged from 6 to 29 years (mean 14 years 6 months) regarding the cues they took as indicative of pain in their children. The interviews focused on instances of short, sharp pain, such as immunization pain, and longer lasting pain, such as headache or injury. A list of 31 cues was elicited. While specific behaviours often differed from one child to another, classes of behaviours (vocal; eating/sleeping; social/personality; facial expression of pain; activity; body and limbs; and physiological) were common to almost all children.

The Non-Communicating Children's Pain Checklist (NCCPC) comprised of 30 items developed from the above study was tested in the home setting (Breau et al. 2000). Parents and caregivers prospectively assessed whether pain cues from the NCCPC were 'present' or 'absent' in four situations: acute pain, long-term pain, a non-painful but distressing situation, and when the child was calm. On average more than four times as many pain cues were indicated as present in painful than in calm situations. The total number of cues present did not differentiate between pain and distressed states but scores for the subscales 'eating/sleeping' and 'body /limb' were higher during the acute pain than during distressed episodes.

A subsequent version of the checklist (NCCPC-PV) was evaluated in the postoperative setting (Breau et al. 2002a). In this study, items relating to eating and sleeping were omitted and each of the remaining items were scored on a four point ordinal scale according to frequency of occurrence. Twenty-four children (aged 3–19 years) were each observed by one of their caregivers and one of the researchers for 10 minutes before and after surgery. Nurses when available also made assessments. Each observer independently completed the NCCPC-PV and gave a global rating of the intensity of the child's pain using a visual analog scale. The scale was internally consistent (Cronbach's alpha=0.91) and showed good interrater reliability (ICC 0.78 to 0.82). A repeated-measures ANOVA indicated that the total score and subscale scores were significantly higher after surgery and did not differ by observer. Moderate correlation (from 0.39 to 0.53) was observed preoperatively between scores on the NCCPC-PV and global assessments of the child's pain on the visual analog scale. Postoperatively, correlation between observers was more variable; in particular there was no correlation (0.09) between the visual analog scale scores of nurses and the NCCPC-PV scores of caregivers. A score of 11 on the NCCPC-PV provided 0.88 sensitivity and 0.81 specificity for classifying children rated on a verbal rating scale to have moderate to severe pain.

A further revised scale (NCCPC-R), with ordinal ratings as above but this time including the items relating to feeding and sleeping, was evaluated in home settings (Breau et al. 2002b). Caregivers of 71 children with severe cognitive impairments (aged 3–18 years) conducted observations of their children using the NCCPC-R during a time of pain and a time without pain. Fifty-five caregivers completed a second set of observations. The NCCPC-R was found to be internally consistent (Cronbach's alpha=0.93) and had moderate correlation (Pearson's r=0.46) with pain intensity ratings provided by caregivers. Sensitivity (0.84) and specificity (0.77) for pain were optimized at a cut-off of 7 or more out of a possible score of 90.

Stallard et al. (2002a) generated 205 pain cues from semi-structured interviews with parents of 30 children. Six of the cues were reported by 90% of caregivers as indicators of definite or severe pain in the child. These items were 'crying, with or without tears'; 'screaming, yelling, groaning or moaning'; 'screwed up or distressed looking face'; 'body appears stiff or tense'; 'difficult to comfort or console'; 'flinches or moves away if touched'. In a further study (Stallard et al. 2002b), caregivers of 49 children (mean age 10.1 years) used these six core cues and up to three additional cues personal to individual children to rate their child's behaviour during two one-hour observation periods for a week. Caregivers were asked to record whether they considered the child was in pain during this observation, and if so the severity of pain. Four items, 'screwed up or distressed looking face'; 'difficult to comfort or console'; 'body appears stiff or tense'; and 'screaming, yelling, groaning or moaning' resulted in 73.6% of pain episodes and 93.4% of non-pain episodes being classified in line with the caregiver's judgements of whether the child was in pain or not in pain.

Hunt and colleagues have developed the Paediatric Pain Profile (PPP), a behaviour rating scale to assist in assessing and monitoring pain in children with severe to profound neurological impairment. The degree of neurological impairment in this group can be so extreme as to be life-threatening, and pain problems, such as reflux oesophagitis or pain from a dislocated hip, can be recurrent and difficult to treat. Qualitative interviews with parents and health care professionals guided the development and design of the tool (Hunt 2001, Hunt et al. 2003). The aim has been to develop a tool that can be implemented as a parent-held record, able to be used at home, in respite, school, and hospital settings.

The study reported here is the final stage in the validation of the Paediatric Pain Profile before making it available for widespread use. The study follows on from three earlier studies in the development of the tool (Hunt 2001). In study 1, focused interviews with parents of 21 children with severe to profound neurological impairments and 26 health care professionals, together with a questionnaire survey completed by 121 parents elicited a list of 56 cues taken to be indicative of pain in this group. In study 2, an item analysis was performed through a second questionnaire sent to a different group of parents (n=46), this time recruited through schools for children with severe learning impairments rather than health-care settings. Guided by the results of that study, a small number of redundant items were eliminated and others were combined to form a 20-item scale. In study 3, the reduced 20-item scale was tested for its reliability and validity in the context of 29 children (mean age 9 years 8 months) who were video-filmed in their everyday settings during five specific daily activities. In addition, saliva samples were collected from children during the morning and on a reference day for estimation of saliva cortisol as an indicator of stress and potential correlate of pain. Three observers independently scored 142 episodes of film. These three observers' behaviour rating scores correlated with their own global pain ratings on a 0–10 numerical rating scale (Pearson's r=0.77 to 0.90) and with those of parents at the time of filming (Pearson's r=0.48 to 0.59). In addition, the behaviour rating score differentiated between different levels of pain on a 5-point verbal rating scale (none to very severe pain) and correlated with saliva cortisol. The behaviour rating scores of all video-raters differentiated between two groups of children, a group that had previously been described by parents as having frequently occurring pain of moderate or worse severity and a group in whom pain was reported to be a rare event. Following that study, minor adjustments were made to the wording of items to produce the 20-item PPP evaluated in this study (Appendix I). Each item is scored on a four point ordinal scale (0 to 3) according to the extent to which the behaviour occurs within a given time period. The score on each item is added to produce a score within the range 0 to 60.

A measurement tool needs to provide consistent measures across time (if the variable is unchanged) and when used by different individuals, i.e. it needs to be reliable. It also needs to measure the construct that it sets out to measure, that is, it needs to have validity (Streiner and Norman 1995, Bland and Altman 2002). While the 20-item PPP was found in the earlier study to be valid and reliable when behaviour was rated from video-film (Hunt 2001), the study being reported here (study 4) sought to provide evidence that it would be so when used in the clinical or 'bedside' setting. We sought to show that after observing a child over the same period of time there would be good agreement between two or more individuals' independent ratings of the child's behaviour (interrater reliability). For these purposes, good agreement was taken to be a reliability

coefficient of greater than 0.7 (Landis and Koch 1977). In addition, there should be a positive correlation between observers' PPP scores and their global judgements of the child's pain on a 5 point verbal rating scale (no pain to very severe pain). This would demonstrate face and concurrent validity of the tool. It is important too that the pain assessment tool should be responsive to change in pain, for example, after an analgesic is given for pain we would expect there to be a significant decrease in the PPP score. As well as demonstrating the scale's responsiveness, this would help to demonstrate that the scale measures the construct we aim to measure, that is 'pain', or at least 'pain behaviour'.

Methods

PARTICIPANTS

Children with severe neurological and cognitive impairments, unable to communicate through speech or any augmentative device, were recruited from five health care centres across the UK. The centres were two children's hospices, a health service respite care centre for children with severe learning disabilities* and two tertiary referral children's hospitals where children were recruited from the lists of children soon to undergo orthopaedic or gastrointestinal surgery. Ethical approval was obtained from the research ethics committees responsible for each participating centre. In addition to the first author who undertook baseline data collection for children recruited from one of the hospices and the respite centre, data collection was facilitated by the medical director at the second hospice and by two research nurses who facilitated the study at the tertiary referral hospitals. Before commencement of the study, health care staff from the centres were invited to attend sessions where they were introduced to the scale and had some practice scoring children on video-film.

Sample size for the study was based on a power calculation for proportion. Assuming, when assessing interrater agreement between observers, agreement on 70% of items, a sample of 131 children would be required to detect a difference of 15% (that is agreement on 55% of items rather than 70%) with a power of 0.95 and a significance level of 0.05.

The power calculation for the number of children to be recruited to subgroups was based on a comparison of means (paired-data). Information from earlier work suggested an average difference in PPP score of 21.4 between children 'at best' and when in pain, with a standard deviation of 12.7. A criterion which has been recommended for use in studies to indicate pain relief is a 50% reduction in the pain score (McQuay and Morre 1998). A sample of 50 would be sufficient to identify a 30% change in pain rating from before administration of the analgesic (T0) to one hour after (T1) with a power of 0.95 and a significance level of 0.05.

Invitations to participate were distributed through the centres and parents were asked to send replies directly to the researchers. Approximately 70% of those invited agreed to participate.

ASSESSMENTS

Baseline assessments

Researchers visited the parents, usually at home, to discuss the study and take informed consent. A structured interview took place during which the parent's assessment of the child's communication, socialization, daily living, and motor skills was recorded using the Vineland Adaptive Behavior Scales (Sparrow et al. 1984). The Vineland Adaptive Behavior Scales were used both to describe the characteristics of the population and to screen for children who might be able to report their own pain. Research suggests that children need to be developmentally aged three or more before they have the capacity to reliably use a self-report scale to assist communication of their pain (Beyer and Wells 1989).

During the interview, the child's pain history was recorded. In addition, the parents retrospectively rated on the PPP scale their child's behaviour both when the child was 'at their best' or 'on a good day', and when they suffered any current or recurring pain. The child's most troublesome pain was labelled 'Pain A'. Other, usually less troublesome, pains were labelled 'Pain B' and 'Pain C'. The parents also made an assessment of the severity of these pains on a five point (0–4) verbal rating scale (no pain to very severe pain).

Tests were performed using SPSS (version 10.0.5) to assess the scale's concurrent and face validity: (a) paired *t*-tests were used to investigate differences between the child's PPP scores 'at best' and when in pain; (b) an analysis of variance (ANOVA) for linear relationship was calculated to assess the relationship between the PPP score and the verbal pain ratings; (c) the statistic Cronbach's alpha, a measure of the average correlation of each item with each of the other items, was calculated to determine the extent to which all the items in the scale measured the same construct. Items should moderately correlate with each other and each should correlate with the total score when the item itself is omitted (item–total correlation; Streiner and Norman 1995). Streiner and Norman (1995) suggest that alpha should be above 0.70 but less than 0.90. Alpha above 0.90 usually indicates that some items are redundant.

Assessment of interrater reliability

In order to assess interrater agreement, parents were asked to identify a co-rater, preferably a health care professional, who would watch the child over the same five-minute period as the parent and, without conferring about their judgements, complete the PPP scale and record the severity of the child's pain on the verbal rating scale. The child did not need to have pain when these assessments were made. The total PPP score was not calculated at this time but completed forms were placed in a pre-labelled envelope, sealed and returned to the investigators.

Interrater assessments were carried out for 111 of the 140 children participating. In 59 (53%) of the 111 cases, assessments were carried out together by a parent and a health care professional. To assess agreement on assessments, ICCs (Shrout and Fleiss model 1) were calculated for each of the items and for the total score (Shrout and Fleiss 1979, Bland and Altman 1996). Although the statistic assumes random allocation of raters to participants, such allocation is not realistic in this setting as parents assess their own child rather than being randomly allocated to assess one of many children.

Assessment of the scale's responsiveness to change in pain following administration of analgesics (construct validity)

Depending on their future treatment, children could fall into one or more of two subgroups: a short-acting analgesic group ($n=41$) and a group of children undergoing orthopaedic or gastrointestinal surgery ($n=30$).

In relation to the short acting analgesic group, if the pain

*US usage: mental retardation.

cues in the PPP are indicative of the construct 'pain' then the expectation would be that the PPP score should be reduced by analgesic administration. No child was prescribed treatment specifically for the purpose of the study. Parents and caregivers were asked to use the scale if during the weeks following the initial interview the child happened to receive an analgesic on an 'as required' basis. Assessments of pain were undertaken before and at hourly intervals for four hours following administration of the analgesic. Where possible, assessments were made by more than one observer in order to reassess inter-rater reliability in this setting. PPP scores were plotted and the area under the curve calculated (AUC-effect). This was compared with the area that would have pertained had no pain relief been achieved (AUC-no effect). The proportion of children in whom AUC-effect was 50% less than AUC-no effect was calculated (McQuay and Moore 1998).

In relation to the children who underwent orthopaedic or gastrointestinal surgery, it was hypothesized that the PPP score would be highest in the first 24 hours after surgery, reverting to preoperative level over the following four days. Parents and/or carers assessed pain before surgery and then two-hourly on the first postoperative day, four-hourly on the second day, and eight-hourly on postoperative days three to five. Analgesic medication given during the period was recorded from the treatment cards. Pain assessments were plotted and maximum and minimum pain scores and time of maximum and minimum pain scores identified. Average pain scores were calculated for each 24 hours postsurgery. Single peaks in PPP score of 20 or more were identified and associations between timing of these peaks and the administration of analgesic medication explored.

Table I: Diagnostic categories of 140 children recruited to study

Diagnostic Category	*n*	*%*
Cerebral palsy	60	42.9
Neurodegenerative disease	34	24.3
Congenital or chromosomal disorder	17	12.1
Developmental delay	13	9.3
Brain damage following infection or trauma	5	3.6
Diagnosis unknown or unrecorded	11	7.8

Results

DEMOGRAPHIC DATA AND PAIN HISTORY

One hundred and forty children aged 1 to 18 years (mean 9 years 11 months, SD 4 years 7 months) were recruited. Diagnostic categories are listed in Table I. Using the Vineland Adaptive Behavior Scales, developmental age equivalence ranged from less than one month to 32 months; the median and range are presented in Table II for each domain of the Vineland Adaptive Behavior Scales and Vineland Adaptive Behavior Composite Score. Motor developmental age of children ranged from less than one month to 51 months, with a median motor development age of just one month. Ninety per cent of children had a motor skills developmental age of 12 months or less. Fifty per cent of children (with raw motor skills score of 4 or less) were unable to raise themselves to sitting position and remain unsupported for at least one minute. Ninety per cent of children (with raw motor skills scores of 21 or less) were unable to walk as a primary means of getting around. As regards function, the great majority of children would fall in to level V of the Gross Motor Function Classification System for Cerebral Palsy (Palisano et al. 1997). None of the children had communicative ability above the developmental age of 29 months confirming that they would be unable to reliably self-report their pain, even with augmentative devices or symbols.

Parents were asked about their child's previous experience of pain. In 41% of cases the child's birth was reported as traumatic (either for the mother or child), 51% of children had spent time in neonatal intensive care, and 42% were reported to have had painful procedures as neonates. Eighty per cent of children had already had significant surgery, i.e. surgery that was additional to that often encountered by otherwise healthy children. Currently 35% of children were reported to have dislocated hip/s and 57% had some degree of spinal curvature. Seventy-six per cent of children were reported to have feeding difficulties with 60% being fed by gastrostomy or nasogastric tube.

BASELINE ASSESSMENTS

In 133 (95%) cases, parents reported that their child had at least one current or recurring pain (Pain A: the 'most troublesome pain') and 56 (40%) reported a second pain (Pain B). Very few reported a third pain source. The reported sources of pain are listed in Table III.

In relation to Pain A, parents reported that 42% of children

Table II: Vineland Adaptive Behavior Scale (VABS) ratings and age equivalence in months of 140 children with severe neurological and learning impairment aged from 1 to 18 years

VABS domains and composite scores	*Raw score*		*Age equivalence (months)*	
	Median	*Range*	*Median*	*Range*
Domains				
Communication	11	0–53	7	<1–29
Daily living	6	0–62	8	<1–38
Socialization	16.5	0–72	6	<1–55
Motor skills	4	0–62	1	<1–51
Vineland Adaptive Behaviour Composite[a]	19	19–63	7	<1–32

[a]19 is lowest possible score.

(n=59) had pain that occurred all the time or on a daily basis. Consistent with findings in earlier phases in the scale's development (Hunt 2001), 20% of children (n=27) were reported to have pain that occurred daily and was rated as severe or very severe.

Face and concurrent validity

In calculating the PPP score for individual assessments any missing items or those items in which 'unable to assess' was ticked were scored as '0'. There was a normally distributed PPP score for all ratings. Out of a possible score of 60 (indicating the highest possible level of pain), mean PPP score for children 'at their best' was 11.1 SD 6.3. Many children (n=80, 57%) were rated as having pain even at their best. Mean PPP score for those who were considered to be pain free at their best (n=60, 43%) was 8.6 (SD 4.9). Mean PPP scores for Pains A and B were respectively 31.3 (SD 9.9) and 26.9 (SD 10.9); scores were significantly higher than those for children at their best, the differences being 20.2 (SD 9.3; 95% confidence interval [CI] 18.6, 21.8) and 15.4 (SD 10.9; 95% CI 12.6, 18.2) respectively. PPP scores for Pains A and B both increased in line with the parents' evaluations of the pain on the verbal rating scale (Figures 1a and 1b; ANOVA test for linear trend F=32.8, df 1, p<0.001 (Pain A); F=22.2, df 1, p<0.001 (Pain B). The data suggest that the PPP has both face and concurrent validity.

Internal consistency and item analysis

Cronbach's alpha statistic was calculated as a measure of the scale's internal consistency. Alpha was 0.75 for the child 'at best', 0.82 for Pain A and 0.86 for Pain B. Correlation of each separate item with the total pain score (item-total correlation) was between 0.3 and 0.7 for 75% of items for Pain A assessments and 85% of Pain B assessments. Only two items 'flexed inwards or drew legs up towards chest' and 'tended to touch or rub particular areas' had correlations less than 0.3 on both Pain A and B assessments. However, both these items had correlations with the total above 0.4 in the analgesic group where ratings were provided prospectively, rather than retrospectively.

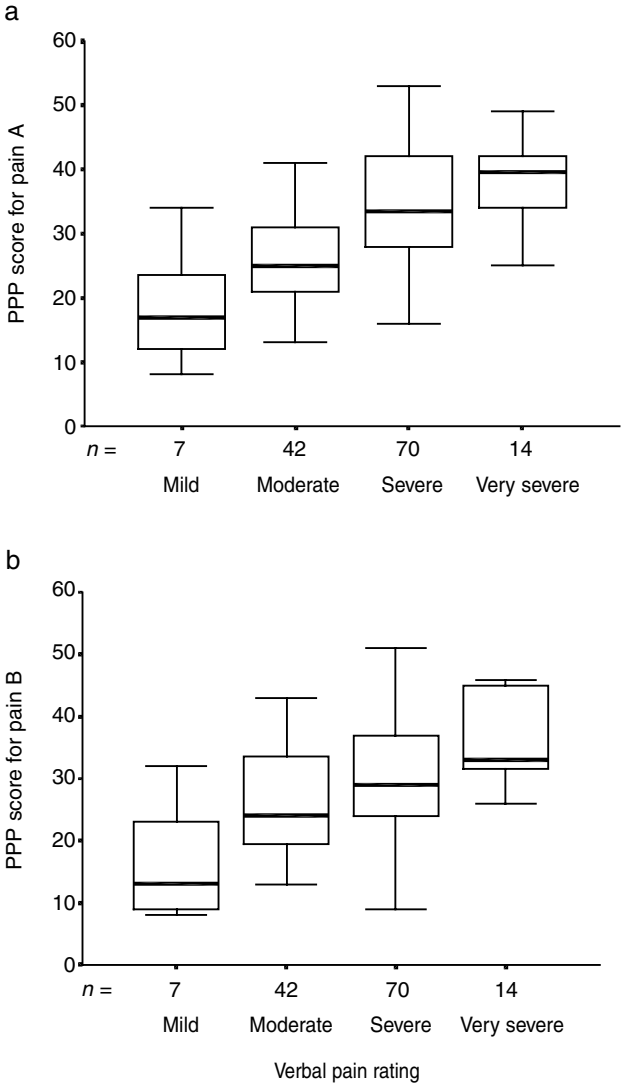
Differences in scores on Pain A and B from the 'at best' ratings were calculated for each item. All the items were endorsed for both Pain A and B, and all items used the full range of possible scores. The most heavily endorsed items were related to mood, in particular 'cheerful' and 'sociable or responsive' (both reverse scored), and items 'restless, agitated or distressed', 'cried/moaned/groaned/screamed or whimpered' and the facial expressions 'grimaced/screwed up face/screwed up eyes' and 'frowned/ had furrowed brow/looked worried'. Items less commonly used but which still scored highly for occasional children were related to particular movements:

Table III: Current or recurring pains of children described by their parents at baseline interview[a]

Pain sources	*Pain A*		*Pain B*		*Total[b]*
	n	%	*n*	%	%
Gastrointestinal	55	39.3	15	10.7	49.3
Musculoskeletal	34	24.3	11	7.9	31.4
Respiratory, dental, or ear	8	5.7	9	6.4	8.7
Headache	4	2.9	11	7.9	7.7
Muscle spasm	10	7.1	2	1.4	6.1
Seizures	2	1.4	5	3.6	3.6
Period pain	3	2.1	2	1.4	2.5
Skin	4	2.9	0	0	2
Urinary tract	3	2.1	0	0	1.5
Equipment	2	1.4	0	0	1
Other	5	3.6	2	0	2.6
Pain source not identified	4	2.9	2	0	2
No current pain	6	4.3	–	–	3.1

[a]Pain A is 'most troublesome pain'. 134 parents described at least one pain; 59 parents described two pains. Percentage is of total sample (n=140).

[b]Both pains A and B were gastrointestinal in one child, and musculoskeletal in another. Only one of these instances is taken in to account in total percentage of children reported to have each class of pain.



Figures 1a and 1b: *Box plots of Paediatric Pain Profile (PPP) scores for pain A and Pain B plotted against verbal rating scores. There was a linear increase in PPP score in line with verbal rating. F=32.8, df 1, p<0.001 (Pain A); F=22.2, df 1, p<0.001 (Pain B) (ANOVA test for linear trend).*

'pulled away or flinched when touched', 'had involuntary or stereotypical movements etc', 'bit self or banged head', 'flexed inwards or drew legs up to chest', and 'tended to touch or rub particular areas'.

Interrater reliability

The behaviour of 111 children was assessed concurrently by two raters. Of the first raters (mainly parents), 11 (10%) did not record a global pain assessment at this time, 74 rated the child as having no pain, 7 as mild pain, 11 as moderate pain, four as severe pain, and two as very severe pain. Of the second rater, one did not record a global assessment, 81 rated the child as having no pain, 19 as mild pain, 4 as moderate, and 5 as having severe pain.

Interrater reliability (using ICC) for individual items varied from 0.14 to 0.76 (mean 0.51) with reliability of the total PPP score being 0.74. Correlation between raters on the total score was 0.75. In contrast, the reliability of the single item verbal rating scale was 0.46 (ICC) and correlation between raters 0.56. Mean difference between raters on PPP score was 1 (SD 5.8). The difference was not significant (95% CI –2.04, 0.14; t=–1.721, df 110, p=0.088; paired t-test). The difference in PPP score between each of the raters was less than 5 in 66% of cases and less than 10 in 88% of cases.

Assessments were identified where ratings had been performed simultaneously by a parent and a health-care professional (n=54). In this case interrater reliability (ICC) of the total PPP score was 0.73 and correlation between the raters on total score was 0.70 (Pearson's r, p<0.001). A paired t-test was performed to assess the difference between parent and professional ratings. Mean score of parents was 8.5 (SD 7.9) and mean score of professionals was 9.9 (SD 8.2; mean difference 1.4, SD 6.3). The scores were not significantly different (95% CI –3.081, 0.341; t=–1.606, df 53, p=0.114). The difference in PPP score between each of the raters was less than 5 in 57% of cases and less than 10 in 87% of cases. In four cases, substantial differences (from 14 to 17) in scores were seen. In two of these, parents scored higher than professionals, in the other two the direction was reversed.

Both professionals and parents have suggested that knowing the child is important in recognizing their pain behaviours (Carter et al. 2002, Hunt et al. 2003). A one-way ANOVA was used to assess whether the difference in PPP scores between parents and professionals was affected by how well the professional thought they knew the child (rated on a 0 to 4 scale, 'not at all' to 'very well'). The results suggest that familiarity with the child did not influence the extent of agreement with the parent on PPP score (F=0.390, df 3, p=0.760).

Sensitivity and specificity of the scale

There is currently no criterion standard for pain in this group of children, hence for the purpose of estimating the sensitivity and specificity of the scale, sensitivity and specificity were calculated using as a marker for the presence of significant pain an average score on the verbal rating scale of 2 or more (moderate pain or worse) at the time when interrater reliability was assessed (96 observations). Using this criterion, 11 of the 96 children were assessed as having significant pain at the time this assessment was made. Sensitivity (1.00) and specificity (0.91) were optimized at a cut-off on the PPP of 14 (Figure 2). No child rated as having pain of moderate or greater severity had average PPP scores of less than 14. Only two children (2%) had scores above the cut-off of 14 and were rated by both observers as having no pain.

RESPONSIVENESS OF SCALE/CONSTRUCT VALIDITY: SUB-GROUP ANALYSES

Short-acting analgesic group

Pain was assessed before and after the administration of an 'as required' analgesic in 41 children. It is usual practice in assessing the effectiveness of an analgesic for only those patients whose pain is rated as moderate or worse at baseline to be studied (McQuay and Moore 1998), therefore, two children whose pain was rated as mild were omitted from this analysis. In addition, one or more of the five, hourly assessments was missing in five cases and these children were also excluded from the analysis of the four hour post analgesic profile but not from the reliability analysis. Pain assessments were analyzed on the 34 children whose pain was rated on the verbal rating scale before the administration of analgesic

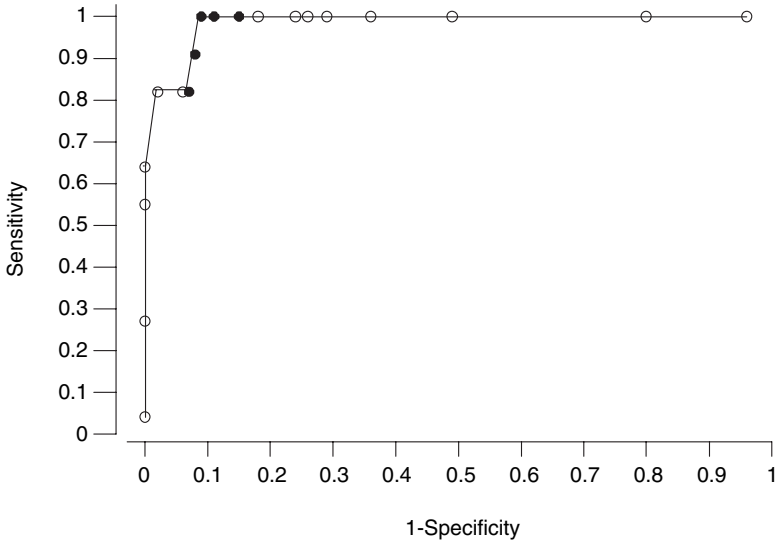


Figure 2: *ROC curve of sensitivity and 1-specificity for average PPP scores (n=111) of two raters. Presence of significant pain was taken to be an average rating of two or more (moderate pain or worse) on a verbal rating scale. ● indicates cut-points between 12 and 16 on the PPP scale (i.e. indicates areas of particularly high sensitivity and specificity). At a cut-point of 14, sensitivity was 1 and specificity 0.91. No child with verbal rating of moderate or more had a PPP score of less than 14.*

as 'moderate' to 'very severe'.

The analgesics administered were paracetamol ($n=19$), codeine ($n=4$), or a non-steroidal anti-inflammatory drug (either ibuprofen or diclofenac; $n=5$). In 6 cases the analgesic was not recorded. Reported pains were classified by investigators as gastrointestinal ($n=8$), postoperative ($n=6$), musculoskeletal ($n=4$), respiratory, dental or oral pain ($n=4$), and headache ($n=2$). The source of pain was either not known or not recorded in the remaining 10 children.

Mean PPP score was 26 (SD 8.5) pre-dose, then 11 (SD 10.4) at 1 hour, 8 (SD 7.6) at 2 hours, 7.7 (SD 7.3) at 3 hours, and 8.9 (SD 8.5) at 4 hours post-dose (Figure 3). Mean PPP score pre-dose was significantly greater than all later assessments (paired-sample *t*-tests, $p<0.001$).

The area under the time-analgesic effect curve (AUC-effect) was calculated for the PPP scores of each child (mean AUC-effect 44.2, SD 27.1) and compared with an area under the curve for the period if no change from the pre-analgesic score had occurred (mean AUC-no effect 104.2, SD 34). Treatment resulted in reduction in PPP score by 50% or more in 23 (68%) children. Five children (15%) had less than 30% reduction in their pain.

Internal consistency and item analysis: Cronbach's alpha was calculated for the scale at time 0 (before administration of the analgesic). Two items 'reluctant to eat/difficulty in feeding' and 'had disturbed sleep' were removed from this analysis as in only 10 cases had the pain coincided with these activities. Cronbach's alpha for the scale excluding these two items was 0.89. All items had correlations above 0.3 with the total score and alpha was not improved by removal of any items.

Interrater reliability: Interrater reliability was reassessed in this analgesic subgroup. Observations from two raters were available at some point in the four-hour assessment period in 18 cases. In the case of the first rater (parent or informal carer), six children were rated as having no pain, four as having mild pain, six as moderate, one as severe, and one as very severe pain. Of the second raters (nurse or other professional carer), five children were rated as having no pain, one as mild, eight as moderate, three as severe, and one as having very severe pain. Interrater reliability of the verbal rating scale was 0.63 (ICC). Reliability of the individual items in the PPP varied from 0.19 to 0.84 (mean 0.64). For the total PPP score, reliability was excellent at 0.87. The difference in PPP score between parent and professional of 0.5 (SD 4.6) was not significant (95% CI –2.80, 1.80; $t=-0.46$, df 17, $p<0.652$). Difference in score was less than 5 in 56% of cases and less than 10 in 100% of cases.

Sensitivity and specificity: The cut-off of 14 was re-examined in this subgroup for the assessments made before administration of the analgesic (at time 0). Thirty-nine children were rated as having moderate pain or worse, and two as having mild pain. Only two children rated as having moderate or worse pain had PPP scores below the cut-off point. Both children whose pain was rated as mild had PPP scores below the cut-off of 14. In this setting, with a cut-off point at 14 for moderate or worse pain, the scale had sensitivity of 0.95 and specificity of 1.

POSTOPERATIVE GROUP

Following surgery, children received concurrently a number of different analgesic treatments and regimens. These included administration of analgesics as intravenous and/or epidural infusions, regular interval dosing, and 'as required' dosing of analgesics. In some cases, analgesics prescribed on an 'as required' basis appeared to be administered on a regular basis.

Twenty-eight pain assessments were carried out over a period of five days in each of 30 children who had either gastrointestinal ($n=14$) or orthopaedic surgery ($n=16$). The mean maximum PPP score during the period was 24.8 (SD 8.6; range 9 to 41) and mean time of maximum pain score was 44 hours but varied from 2 to 120 hours. While we had hypothesized that PPP scores would be highest on the first postoperative day, average pain scores did not differ across postoperative days, being within a range varying from 8.3 (SD 5) on the first day, to 7.3 (SD 5.2) on the second day. The highest pain score occurred in the first 24 hours after surgery in the case of 14 children, on the second day in four children, third day in five children, fourth day in two children, and fifth day in five children. The modal time for highest pain score (4 patients) was at 56 hours after surgery, possibly in line with withdrawal of continuous epidural and intravenous analgesics.

Thirty-two peaks in PPP score of 20 or above were identified in 17 (57%) children. The number of peaks per child in this subgroup varied from 1 to 11 (median 3). Relationships between the timing of these peaks and administration of analgesics were explored. In 13 (41%) of the 32 peaks it was possible to detect a probable relationship with analgesic treatments. Either a fall in the PPP score occurred within an hour after the administration of the analgesic, or a drug appeared to have been given within an hour of the peak. In the remaining 19 cases the information was inconclusive.

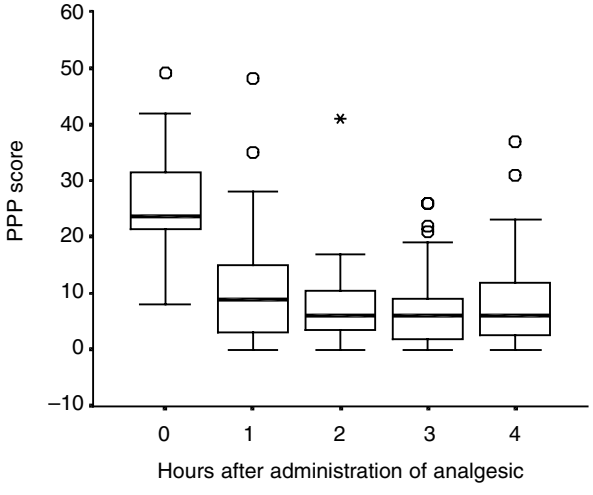


Figure 3: *Boxplot of PPP scores plotted against time after administration of an 'as required' analgesic. PPP score before administration was significantly higher than all later assessments (paired-sample t-tests, $p<0.001$).*
*○ and * denote outliers (* extreme).*

Discussion

In the earlier studies (studies 1 to 3; Hunt 2001) items for the scale were generated and underwent preliminary assessments of face, concurrent, and criterion validity, as well as internal consistency and external reliability. In the current study (study 4) the 20-item PPP was retested in clinical and domestic settings for its internal consistency, interrater reliability, and face, concurrent, and construct validity.

The PPP has been found to have excellent internal consistency. In each of the instances where it was assessed, the statistic Cronbach's alpha fell between 0.7 and 0.9 as recommended (Streiner and Norman 1995). Although interrater reliability for individual items in the scale was variable, the reliability of the summated scale (from 0.73 to 0.87) was very good. Indeed, the PPP score was found in this study to have higher interrater reliability than the verbal pain rating scale. No significant difference in the total PPP score was detected between raters. There were, however, a small number of cases (7%) where differences between raters were substantial. Such differences in perceptions, even though occurring in a minority of cases, are important as they can cause significant distress for child, parents, and professional carers and make it more difficult to implement appropriate treatments. Further study, perhaps using 'thinking aloud' methods might be useful in exploring such differences in perceptions.

It is often stated by parents and professional carers that 'knowing the child' is necessary for identifying pain in children with profound disability (Hunt et al. 2003). However, when using the PPP the extent of the clinician's familiarity with the child did not appear to affect significantly the level of interrater agreement with parents, strengthening the potential value of the scale in all settings. Breau et al. (2002a) reported similar findings. It appears that the subjectivity of the judgement may be reduced when using a behaviour rating scale to assess pain. Further study might investigate whether this could be particularly so when items for a scale are generated by parents and carers who are particularly familiar with the population.

In addition to good levels of interrater reliability, this study has demonstrated that the PPP has face, concurrent, and construct validity with PPP scores increasing in line with the observer's global rating of the child's pain and decreasing when analgesics were administered for pain. It is of note that each of the items in the PPP overlaps with one or more of the other scales cited in the introduction, providing support for the content validity of each.

LIMITATIONS OF THE STUDY

In the absence of any criterion-standard measures for pain in this group of children, we have, in assessing the concurrent validity of the PPP, compared the PPP score with the observers' global impressions of the children's pain on a verbal rating scale. However, several studies have identified differences between the ratings of a child's pain by parents and professional carers and the self-reports of the children themselves (Chambers et al. 1998, Cleary et al. 2002). Because the validity of proxy ratings of other's pain is open to doubt, using such ratings as a means of validating an alternative pain measure is not ideal. However, finding appropriate criteria against which to evaluate the scale is challenging. In a previous study, in addition to correlating PPP scores with global pain judgements of observers, we have correlated the PPP scores with saliva cortisol concentrations and found there to be a positive correlation, though the relationship did not appear to be consistent across children (Hunt 2001). Further investigation is intended.

As anticipated, PPP score fell following the administration of a short acting analgesic. The study would have been strengthened had it been possible to blind raters to the administration and timing of analgesic treatments, but this would have proven difficult for both practical and ethical reasons.

While the majority of children appeared to benefit from their analgesic treatment, a proportion had poor response. As these assessments were often carried out in the children's homes and we have limited knowledge of the circumstances, we can only suggest possible reasons. In several cases some reduction in pain is observed at either one, two, or three hours post-analgesic only for pain to increase again at four hours, suggesting that the dose may be insufficient or the dosing interval too short. In addition, the nature of the treatment given may not always have been optimal for the type of pain experienced, for instance in one case paracetamol was given for pain reported to be due to 'reflux/sickness'. In a further case, a child, known before the onset of a neurodegenerative condition to suffer from migraine characterized by vomiting, sleepiness, and withdrawal from social interactions, had low PPP scores across the interval. It is possible that, in the absence of such contextual knowledge, the scale would not be optimal for picking up change when pain behaviour is very limited and unspecific.

Analysis of data from the postoperative group was complicated by the number and variety of analgesic medications administered. Observing relationships retrospectively between analgesic administration and change in PPP score was often inconclusive. This was particularly so on days when, in an effort not to overburden staff, pain assessments using the PPP were carried out only eight-hourly. In addition, while other studies have detected an increase in pain score in the first 24 hours after surgery (Breau et al. 2002a) the generally good level of pain management for children in this study made it more difficult for us to demonstrate clear relationships between a potentially painful event (surgery) and the PPP score. Although the short acting analgesic group appears to have been the best model for testing the validity of the scale, the PPP did, however, appear to be a feasible and acceptable instrument for assessing and monitoring postsurgery pain.

IMPLEMENTATION OF PAEDIATRIC PAIN PROFILE AND FURTHER WORK

The PPP currently consists of three sets of recordings: (a) a retrospective rating by the parent of the child's behaviour when the child is at his or her best; (b) retrospective ratings of the child's behaviour during any current or recurring pains; and (c) future/prospective ratings. The retrospective assessments, preferably completed during a discussion between parent and a clinician who works regularly with the family, provide both screening for the child's current state, a reference for future assessments, and the opportunity for communication between parent and clinician about the child's current pain problems. The Pain Profile would then be held and maintained by the parents in partnership with their professional caregivers. The Profile can accompany the child wherever it is needed, for instance on visits by the child to hospital, general practitioner, school, or short breaks (respite care). Completing

the scale should not be over-burdensome in practice, taking just 2–3 minutes to complete. However, the challenge lies not so much in completing the scale but in the quality of the observation. Use of the tool should increase the awareness and alertness of both parents and clinicians to the child's pain cues. A score of 14 or above would suggest significant pain that needs to be addressed. In relation to the use of ranges and cut-off points, it should be noted that individual children would inevitably differ in their capacity and tendency to show some behaviours in the scale. With continuing use of the scale, a pattern of behaviour unique to an individual child may become apparent and a cut-off point recognized that is applicable to that child.

Further research will be undertaken to evaluate the acceptability, feasibility, and the usefulness of the tool in practice settings with children with severe to profound neurological disabilities. It is possible that the tool may be useful for children who have less severe disabilities than our population. However, it is also possible that a proportion of the items would be less appropriate for these children. Further validation would be required. The tool could also be useful for adults with a similar degree of disability and further research in this area is planned.

In conclusion, the PPP appears a valid and reliable tool for assessing pain in children with severe neurological and cognitive impairments. While it has been difficult in the past to judge the effectiveness of treatments for pain in this group, with the development of this tool it becomes easier to monitor the impact of therapeutic interventions. The PPP provides a tool that can be used both clinically and in intervention research.

DOI: 10.1017/S0012162204000039

Accepted for publication 13th October 2003.

Acknowledgements
We wish to thank all children, parents, clinicians, and participating centres for their contributions to the study. The participating centres were Great Ormond Street Hospital for Children NHS Trust, London; the Hospital for Sick Children, Edinburgh; Martin House Hospice for Children, Boston Spa; Helen House Hospice for Children, Oxford; and the Oxfordshire Learning Disability NHS Trust. This study was funded by The Health Foundation. Earlier studies in the series were funded by Action Research with full support from the National Lottery Charities Board.

References

Anand KJS, Craig KD. (1996) New perspectives on the definition of pain. *Pain* **67:** 3–6.

Beyer JE, Wells N. (1989) The Assessment of Pain in Children. *Pediatr Clin North Am* **36:** 837–854.

Bland JM, Altman DG. (1996) Measurement error [corrected and republished article originally printed in *BMJ* (1996) 312: 1654]. *BMJ* **313:** 744.

Bland JM, Altman DG. (2002) Statistics Notes: Validating scales and indexes. *BMJ* **324:** 606–607.

Breau LM, Finley GA, McGrath PJ, Camfield CS. (2002a) Validation of the Non-communicating Children's Pain Checklist-Postoperative Version. *Anesthesiology* **96:** 528–535.

Breau LM, McGrath PJ, Camfield C, Rosmus C, Finley GA. (2000) Preliminary validation of an observational pain checklist for persons with cognitive impairments and inability to communicate verbally. *Dev Med Child Neurol* **42:** 609–616.

Breau LM, McGrath PJ, Camfield CS, Finley GA. (2002b) Psychometric properties of the non-communicating children's pain checklist-revised. *Pain* **99:** 349–357.

Carter B, McArthur E, Cunliffe M. (2002) Dealing with uncertainty: parental assessment of pain in their children with profound special needs. *J Adv Nurs* **38:** 449–457.

Chambers CT, Reid GJ, Craig KD, McGrath PJ, Finley GA. (1998) Agreement between child and parent reports of pain. *Clin J Pain* **14:** 336–342.

Cleary AG, Ramanan AV, Baildam E, Birch A, Sills JA, Davidson JE. (2002) Nitrous oxide analgesia during intra-articular injection for juvenile idiopathic arthritis. *Arch Dis Child* **86:** 416–418.

Collignon P, Giusiano B. (2001) Validation of a pain evaluation scale for patients with severe cerebral palsy. *Eur J Pain* **5:** 433–442.

Craig KD. (1997) On knowing another's pain. In: Ekman P, Rosenberg E, editors. *What the Face Reveals. Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* Oxford: Oxford University Press. p 187–180.

Fanurik D, Koh JL, Harrison RD, Conrad TM, Tomerlin C. (1998) Pain assessment in children with cognitive impairment. An exploration of self-report skills. *Clin Nurs Res* **7:** 103–119; discussion 120–124.

Ferrell BA, Ferrell BR, Rivera L. (1995) Pain in cognitively impaired nursing home patients. *J Pain Symptom Manage* **10:** 591–598.

Giusiano B, Jimeno MT, Collignon P, Chau Y. (1995) Utilization of neural network in the elaboration of an evaluation scale for pain in cerebral palsy. *Methods Inf Med* **34:** 498–502.

Hunt A, Mastroyannopoulou K, Goldman A, Seers K. (2003) Not knowing – the problem of pain in children with severe neurological impairment. *Int J Nurs Stud* **40:** 171–183.

Hunt AM. (2001) Towards an understanding of pain in the child with severe neurological impairment. Development of a behaviour rating scale for assessing pain. (PhD thesis). Manchester: Royal College of Nursing Institute. University of Manchester.

Landis JR, Koch GG. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33:** 117–129.

McGrath PJ, Rosmus C, Canfield C, Campbell MA, Hennigar A. (1998) Behaviours caregivers use to determine pain in non-verbal, cognitively impaired individuals. *Dev Med Child Neurol* **40:** 340–343.

McIntosh N. (1997) Pain in the newborn, a possible new starting point. *Eur J Pediatr* **156:** 173–177.

McQuay HJ, Moore RA. (1998) *An Evidence-Based Resource for Pain Relief.* Oxford: Oxford University Press.

Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. (1997) Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* **39:** 214–223.

Parmelee PA. (1996) Pain in cognitively impaired older persons. *Clin Geriatr Med* **12:** 473–487.

Schecter NL. (1989) The undertreatment of pain in children: an overview. *Pediatr Clin North Am* **36:** 781–793.

Shrout PE, Fleiss J. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* **86:** 420–428.

Sparrow S, Balla D, Cicchetti D. (1984) *Vineland Adaptive Behavior Scales, Interview Edition: Survey Form Manual*. Circle Pines, Minn: American Guidance Service.

Stallard P, Williams L, Velleman R, Lenton S, McGrath PJ. (2002a) Brief report: behaviors identified by caregivers to detect pain in noncommunicating children. *J Pediatr Psychol* **27:** 209–214.

Stallard P, Williams L, Velleman R, Lenton S, McGrath PJ, Taylor G. (2002b) The development and evaluation of the pain indicator for communicatively impaired children (PICIC). *Pain* **98:** 145–149.

Streiner DL, Norman GR. (1995) *Health Measurement Scale*s. Oxford: Oxford University Press.

Appendix I: The 20-item Paediatric Pain Profile

	During the last... (first name)...	*Not at all*	*A little*	*Quite a lot*	*A great deal*	*Unable to assess*
1	Was cheerful (reverse scored)	[ ]	[ ]	[ ]	[ ]	[ ]
2	Was sociable or responsive (reverse scored)	[ ]	[ ]	[ ]	[ ]	[ ]
3	Appeared withdrawn or depressed	[ ]	[ ]	[ ]	[ ]	[ ]
4	Cried/moaned/groaned/screamed or whimpered	[ ]	[ ]	[ ]	[ ]	[ ]
5	Was hard to console or comfort	[ ]	[ ]	[ ]	[ ]	[ ]
6	Bit self or banged head	[ ]	[ ]	[ ]	[ ]	[ ]
7	Was reluctant to eat/difficult to feed (includes nasogastric and gastrostomy feeding)	[ ]	[ ]	[ ]	[ ]	[ ]
8	Had disturbed sleep	[ ]	[ ]	[ ]	[ ]	[ ]
9	Grimaced/screwed up face/screwed up eyes	[ ]	[ ]	[ ]	[ ]	[ ]
10	Frowned/had furrowed brow/looked worried	[ ]	[ ]	[ ]	[ ]	[ ]
11	Looked frightened (with eyes wide open)	[ ]	[ ]	[ ]	[ ]	[ ]
12	Ground teeth or made mouthing movements	[ ]	[ ]	[ ]	[ ]	[ ]
13	Was restless/agitated or distressed	[ ]	[ ]	[ ]	[ ]	[ ]
14	Tensed/stiffened or spasmed	[ ]	[ ]	[ ]	[ ]	[ ]
15	Flexed inwards or drew legs up towards chest	[ ]	[ ]	[ ]	[ ]	[ ]
16	Tended to touch or rub particular areas	[ ]	[ ]	[ ]	[ ]	[ ]
17	Resisted being moved	[ ]	[ ]	[ ]	[ ]	[ ]
18	Pulled away or flinched when touched	[ ]	[ ]	[ ]	[ ]	[ ]
19	Twisted and turned/tossed head/writhed or arched back	[ ]	[ ]	[ ]	[ ]	[ ]
20	Had involuntary or stereotypical movements/was jumpy/startled or had seizures	[ ]	[ ]	[ ]	[ ]	[ ]

Of the 56 items originally included in Study 2, eight have been excluded from the version above: 'appears exhausted', 'appears tired', 'skin seems clammy or sweaty', 'skin feels hot', 'looks pale', 'looks red', and 'breathes at unusual/abnormal rate and rhythm'.