© 2007 SNL All rights reserved

Selection of a tool to assess postoperative pain on a neonatal surgical unit

There is currently no gold standard for assessment of postoperative pain in neonates. Behavioural and physiological indicators have been described and incorporated into various postoperative pain assessment tools. A structured review of the literature was undertaken to find a reliable, valid and clinically useful method to assess postoperative pain in neonates.

Janet Burton
MBCHB
Evidence-based Medical Research Student,
Regional Neonatal Surgical Unit,
St Mary's Hospital, Manchester

Ralph MacKinnon
BSc (Hons), MBCHB, FRCA
Honorary Lecturer, University of Manchester and Consultant Paediatric Anaesthetist,
Regional Neonatal Surgical Unit,
St Mary's Hospital, Manchester
ralph.mackinnon@cmmc.nhs.uk

Keywords

postoperative pain; assessment tool; physiological indicators; behavioural indicators; surgery

Key points

Burton, J., MacKinnon, R. (2007) Selection of a tool to assess postoperative pain on a neonatal surgical unit. *Infant* 3(5): 188-96.

1. The Neonatal Facial Coding System (NFCS) and the Liverpool Infant Distress Scale (LIDS) were found to be the most reliable postoperative pain assessment tools, but need simplifying for clinical use.
2. Physiological variables were consistently highly reliable and favoured by nurses and studies of pain behaviour.
3. Behavioural indicators are more specific for pain than physiological variables, are valid indicators of pain and should be included in assessment.
4. Multidimensional tools appear to be more clinically useful.
5. The review found the Pain Assessment Tool (PAT) best meets the current needs of the neonatal surgical unit.

Postoperative pain assessment was reviewed as part of an ongoing update of evidence base behind management protocols at the regional neonatal surgical unit at St Mary's hospital, Manchester. This article reviews the current knowledge base and demonstrates evidence-based methodology to determine a neonatal postoperative pain assessment tool.

The main objective of this review is to establish what would be the most suitable method to assess postoperative pain in neonates. A systematic literature search and critical appraisal of the relevant literature was carried out to ascertain what methods have been developed, how effective they are in assessing pain, and to consider their suitability for clinical use on the neonatal surgical unit at St Mary's Hospital and similar units.

Background

Pain is defined by the International Association for the Study of Pain (IASP) as:

> 'An unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage'[1].

In recent decades it has been established that neonates have the capacity to feel pain and respond to it[2]. Prior to this, newborn infants undergoing surgical procedures would not receive any analgesia as it was assumed that they could not sense noxious stimuli and hence, not perceive pain[2,3]. Other explanations for withholding analgesics from neonates included fear of opioid side-effects and lack of knowledge[4]. Fortunately there is now an appreciation that analgesia is required for neonates undergoing surgery and other painful procedures. However for optimisation of postoperative analgesia, clinicians need to evaluate how much pain the neonate is experiencing.

Pain pathways in neonates

Pain pathways consist of a series of neurons that communicate noxious stimuli from the periphery to the brain. Noxious stimuli are detected by two main nociceptors. High-threshold mechanicoreceptors are triggered by heavy physical pressure to transmit impulses along small myelinated AΔ fibres[2]. Polymodal nociceptors are stimulated by pressure, heat and chemical stimuli and transmit along slow unmyelinated C fibres as do visceral nociceptors[2]. The pain fibres enter the spinal cord via the dorsal horn and synapse with ascending neurons. These ascending pathways transmit nociceptive stimuli to the thalamus reticular formation, periaqueductal grey matter and to the superior collicular nuclei[2]. There are also projections to areas of the cerebral cortex and the limbic system involved with the perception of pain[2].

Lack of myelination of neuronal pain pathways was previously argued as evidence that neonates did not have the capacity to perceive pain. More recent neuroanatomic research suggests central spinothalamic pathways and thalamocortical pathways are myelinated by 30 and 37 weeks' gestation, respectively[3,5]. Furthermore, the majority of noxious stimuli are transmitted via unmyelinated fibres in adults[3]. Nociceptors are present from 7 weeks' gestation in the fetus and an outborn neonate has at least an equal density of dermal nociceptors as an older child or adult[2,5]. In the dorsal horn the laminar structure, synaptic interconnections and neurotransmitter vesicles are present by 30 weeks' gestation[5]. The ascending pathways discussed above are present relatively early in fetal development and are complete by the end of the second trimester[2]. Moreover, cortical interconnections for higher centre processing of noxious stimuli are mature by 24 weeks' gestational age[2]. Hence, even premature neonates have the capacity to feel pain.

There is further development of the nociceptive pathways during the first weeks of life. This includes maturation of low-threshhold mechanoceptors and C fibres synaptic activity, and also increased levels of neurotransmitters in the dorsal horn[5]. Connections of descending inhibitory pathways with the dorsal horn are not established until 10-19 days after birth and levels of transmitters in these pathways (norepinephrine and serotonin) are initially low. Furthermore, local inhibitory mechanisms are not fully developed at birth[5] which may mean neonates experience more pain than older children. This indicates that pain responses in premature neonates may well differ from older infants, and care should be taken when interpreting pain research conducted on non-neonates.

The importance of preventing pain in neonates

Neonates in the intensive care environment are continually subjected to various kinds of stress[6]. Both acute procedural pain (such as heel pricks) and postoperative pain can be considered as stressors and hence are believed to have negative short and long-term consequences[5,7]. Neonates are thought to be particularly sensitive to stress, as it is a critical period in brain development[6]. Repeated painful experiences may result in hyperalgesia and allodynia in the short term, and long-term changes in pain threshold, possibly mediated by reorganisation of the nervous system[5,6,8]. Prolonged pain may also affect development by adversely affecting sleep-wake patterns, feeding behaviours, and parental bonding[6]. It is also suggested that physiological responses such as tachycardia and lowered oxygen saturations may result in cerebral haemorrhage secondary to overloading of immature cerebral vasculature[2]. Other complex metabolic disturbance may occur due to the hormonal response to pain, and this may adversely affect outcome[3,9]. Consideration of metabolic changes is particularly important in preterm neonates as they have smaller lipid stores and immature homeostatic mechanisms[9]. Effective management of postoperative pain does improve the morbidity and mortality outcomes in neonates[10].

Modern thinking on management of postoperative pain in neonates centres on prevention rather than control[9]. Staff caring for neonates should anticipate pain, especially in the postoperative setting[11]. To optimise management accurate measures are needed not only to diagnose pain and to assess the need for analgesia, but also to monitor the efficacy of analgesic regimens and to evaluate outcomes[8].

Indicators of pain in neonates

The literature on neonatal pain contains a vast array of different indicators of pain that occur in neonates. These fall into three main categories: physiological responses (including heart rate, blood pressure, and respiratory rate); behavioural responses (including facial expression, movements/posture and verbalisation/cry); and biochemical responses (including increased catecholamine plasma levels).

It is widely accepted that as neonates cannot express their subjective experience of pain verbally, use of such indicators is the only way in which healthcare professionals can evaluate their pain[6]. In order to assess pain in neonates, staff need to understand how to interpret these responses objectively, and make clinical decisions regarding analgesia based upon them.

Behavioural indicators

Behaviours may be more specific to pain than physiological responses, however they are generally less objective and less quantifiable[2]. An important consideration with behavioural indicators is their degree of specificity for detecting the presence of pain as opposed to other states such as hunger or fear. Another important consideration is their applicability to all neonates returning to intensive care units following surgery. Ventilation, sedation, paralysis, and extreme illness/weakness are factors which would affect the assessment of neonatal behaviours[12].

Crying is a fundamental method of communication for the newborn and is not specific for pain. However, cry is reported to differ in both quantity (longer duration and frequency) and quality (higher pitch) when the neonate is responding to painful stimuli[2,3]. In research, cry is assessed using spectro-graphic equipment that would not be available at the bedside, hence accurate use of this indicator is more difficult in the clinical situation[13]. Cry is not useful in ventilated neonates, and postoperatively neonates may not have enough energy to cry[13]. Facial expression is described in several of the measures used to assess pain and is considered to be the most specific indicator[13]. Some use global facial changes such as 'grimace', whereas others score discrete facial actions such as brow bulge. It is important to consider the sleep/wake state and gestational age of the neonate when assessing facial expression however, as both these factors affect facial activity[14-16]. Studies on pain behaviour now claim neonates are capable of organised purposeful movements as opposed to older studies that reported diffuse body movements in response to acute pain[3]. However, much of this research was based upon the response to heel sticks and similar procedures. It may be that assessment of body movement in neonates following major surgery is not a reliable indicator as surgical pain may restrict gross body movements and neonates may actually 'splint' themselves to reduce this[9,13]. Pain disrupts the neonate's sleep/wake cycle and measures of this are sometimes incorporated into scales assessing pain[2]. Consolability is an indicator included in several scales despite little background as to its validity. However, it may prove to be a useful indicator in the clinical setting when assessed by experienced caregivers[13].

Physiological indicators

Much of the research on the physiological response to pain in neonates has been carried out in the acute pain setting (e.g. during tracheal intubation and heel sticks). Increases in heart rate, respiratory rate, blood pressure, palmar sweating/perspiration, intracranial pressure and pulmonary vascular resistance, and decreased oxygen saturation are widely acknowledged to occur in response to acute pain[2,3,5,12,17]. Increases of up to 42% above baseline systolic blood pressure and oxygen saturation decreases of 10-15% occur postoperatively in term and preterm neonates[9]. These changes are prevented with analgesia, hence supporting their association with pain perception[5,17]. Physiological indicators such as heart rate and blood pressure do have the advantage of objective assessment in the clinical setting and measurement postoperatively may be facilitated by standardised monitoring. Furthermore, in paralysed neonates, clinicians may have to rely on such indicators[3]. However, these indicators are non-specific for pain, vary between individuals, and are reflexive in nature, and therefore should be carefully interpreted within clinical context[2,11,23]. Factors such as blood loss, fluid intake and body temperature may limit the usefulness of physiological indicators for assessing neonates following surgery[13].

Biochemical indicators

The neuroendocrine response to stressful stimuli, such as pain, in neonates is characterised by increased levels of cortisol, catecholamines (norepinephrine and epinephrine), and glucagon, amongst other less measurable biochemical changes[9,18]. Such parameters are used in pain research to establish validity of other measures. However, these responses are non-specific for pain, expensive to measure, and are not suitable for routine assessment at the bedside[5,9]. Simple bedside tests such as urinanalysis can be utilised alongside a pain assessment scale to monitor the postoperative neonate for serious metabolic changes[9].

Measuring pain in neonates

Despite the development of several pain scales over recent years no single scale has emerged as a gold standard for clinical assessment of neonatal postoperative pain[19]. The lack of an objective and accurate measure of pain has been reported to be a source of stress in healthcare workers responsible for postoperative management of neonates[20]. Ideally the scale should be sensitive to severity of pain, valid, reliable, suitable for the unit population of neonates, and appropriate for clinical use by staff on the neonatal surgery unit[21]. So far validity and reliability of existing scales is not well established[12]. Acute pain management guidelines published in 1992 state that:

> 'caring for the child in pain requires frequent assessment and reassessment of the presence, amount, quality, and location of pain'[12].

The existing pain assessment scales are similar in the way that they include some combination of the pain indicators discussed above, yet vary in the way that these are assessed and recorded. Contextual indicators thought to modulate the pain response in neonates are also included in some scales. Such indicators include gestational age, severity of illness, and behavioural state[21].

A particular difficulty in selecting an appropriate postoperative pain assessment scale is that many are designed for use in acute procedural pain in the neonatal intensive care unit. Responses to the more prolonged type of pain that follows surgery are generally thought to differ from the more acute and often repeated procedural pain. It is important to recognise responses to postoperative pain change over time[16].

Another major factor to take into account when selecting a pain scale is the population for which it is intended. Few scales have been designed for or tested in neonates and even fewer for those born very prematurely or with low birth weight. Scales such as the Postoperative Pain Score[22] have only been applied to infants older than one month and require further research before being used in neonates.

A prerequisite of any pain scale is that it does actually measure pain. Adults can confirm pain state verbally by self report; however for obvious reasons nobody can determine accurately what the neonate can feel. Hence, the items included on the scale have to be sensitive and specific for pain so that pain is diagnosed when it is present, and pain responses are distinguished from those of other states. When applying these measures in a clinical situation it is important to consider other factors which may influence both behavioural and physiological factors.

Putting a pain assessment tool into operation involves decisions on who does the assessment, when it is done and how the score is recorded. Nursing staff have a primary role in pain assessment and subsequent administration of analgesia. Guidelines recommend that assessment is carried out by the health care provider (i.e. the infant's own nurse) as part of routine observations of vital signs, and if possible the pain assessment scale should be incorporated into the bedside chart[11,12]. These guidelines also recommend that the frequency of assessment should depend on the type of surgery performed and on the severity of pain present. For major surgery, two-hourly assessments are suggested as a minimum on the day of surgery and the following day, and four-hourly thereafter[12].

Literature review

A structured search strategy, based on the Best BETS approach, was used to answer the question: "In neonates, what pain assessment tool should be used to improve postoperative pain management?" using the OVID interface.

Medline, Embase and CINAHL databases were all searched using terms related to neonates, pain measurement, and postoperative pain. All relevant papers were read and critically appraised using the critical appraisal checklists available on the Best BETS website (www.bestbets.org/index.html) and referring to Crombie's guide to critical appraisal[23,24]. The Landis and Koch categories of correlation were used throughout[26]. Altogether 80 papers were found through Medline, 61 through embase and 28 through CINAHL.

Results

Review of the available literature revealed numerous pain scales that have been developed for, or adapted for, use in neonates. Unfortunately many of these have only been tested in acute, procedural pain and would require further validation for use in the context of assessing the prolonged pain of surgery. A limitation of this review could be the exclusion of such studies from critical appraisal.

Studies of postoperative pain scales

These are summarised in **TABLE 1**. McNair et al[27] found that PIPP (Premature Infant Pain Profile) and CRIES (Crying, Required Oxygen, Increased Vital Signs, Expression and Sleeplessness) had some convergent validity as there was a fair or moderate correlation at most observations postoperatively. There were insufficient VAS (Visual Analogue Scale) observations for correlations with PIPP and CRIES to be valuable to this review. The finding that a slight increase in pain scores coincided with conversion from opioid to non-opioid analgesia may suggest discriminative validity. Gestational age of neonates and type of surgery did not have a significant effect and so PIPP and CRIES may be suitable for the whole neonatal surgical population.

Peters et al[28] tested the NFCS (Neonatal Facial Coding Systems) and found almost perfect inter-rater reliability for its items. COMFORT 'behaviour' reliability was lower. Convergent validity is suggested by a moderate association with COMFORT 'behaviour' and VAS. There is a low but significant association with the physiological variables (heart rate, blood pressure and heart rate variability [HRV]) which add to the validity. However, there was no association with catecholamine or morphine levels. Item reduction was feasible as validity of the NFCS subset (brow bulge, eye squeeze, nasolabial furrow, horizontal mouth stretch and taut tongue) was equal to that of the total set. Although this result increases the clinical utility of NFCS, the sample size was very

Study	Scale(s) and patient group	Outcomes and key results	Study weaknesses
McNair et al, Canada, 2004[27] **Diagnostic study level 2b**	PIPP, CRIES and VAS assessed on 51 neonates over 72 postoperative hours. Grouped by gestational age (GA) into 28-31weeks (6), 32-35 (10) and ≥36 (35), and by minor/major surgery.	*Convergent validity* – the degree of correlation of PIPP and CRIES varies over the postoperative period (Intraclass correlation range (ICC) = 0.6008-0.2523). *Effect of gestational age and type of surgery* – age and surgery group did not have a significant effect ($F_{151,2} = 1.37$, $p = 0.265$ and $F_{151,2} = 2.87$, $p = 0.973$). *Patterns of pain response over time* – consistent findings over 3 scales. Highest scores immediately after surgery. Decreased over 1st 12 hours, and increased slightly between 24 and 72 hours (coincides with opiod to non-opioid analgesia).	Inconsistent VAS assessment and small sample of VAS observations. Validation of CRIES questionable. PIPP used previously only for procedural pain. No gold standard comparator.
Peters et al, The Netherlands, 2003[28] **Diagnostic study level 2a**	NFCS tested on 11 neonates post thoracic or abdominal surgery. Compared with COMFORT 'behaviour', and VAS – HR, variability (HRV), BP, BP variability (BPV), catecholamines and morphine levels.	*Inter-rater reliability* – almost perfect for NFCS items Kappas = 0.84-1.0. Moderate to substantial for COMFORT 'behaviour' items, respiratory response (Kappa = 0.54) to alertness (Kappa = 0.74). *Validity* – significant association with VAS (standardised regression coefficient 0.31, p<0.001) and COMFORT 'behaviour' scores (0.48, p<0.001), and also with HR mean (0.34, p<0.001), BP mean (0.23, p<0.001), HRV (0.26, p<0.001), and BPV (0.15, p<0.007). No significant association between total NFCS score and adrenaline, noradrenaline, or morphine/M6G levels. COMFORT score most predictive of NCFS score. *Validity of item reduced NFCS* – A simpler 5 item NFCS was developed and associations equalled those of total NFCS (validity).	Small sample of neonates used. Neonates GA<35 weeks and LBW excluded. No gold standard comparator. With exception of NCFS score, assessments not independent. Potential for scoring bias.
van Dijk, The Netherlands, 2000[29] **Diagnostic study level 2b**	COMFORT tested in 56 neonates in a group of 0-3 year-olds undergoing abdominal or thoracic surgery.	*Inter-rater reliability* – almost perfect reliability for HR and MAP (Kappa = 0.93). Lowest for respiratory response (Kappa = 0.54). *Comparison of VAS scores pre and post COMFORT (2 minute assessment)* – correlation of COMFORT with VAS increased pre (Kappa = 0.64-0.73) and post (Kappa = 0.79-0.83). *Stability (test-retest)* – physiological variables reliably stable over time period (stability coefficients, MAP 0.89 and HR 0.82) COMFORT 'behaviour' (0.59) and VAS (0.58) less stable. *Convergent validity* – high correlations of behavioural items with VAS (0.89-0.96). Low correlations of HR and MAP with VAS (0.13-0.24, and 0.29-0.39 respectively).	Small sample. Excluded neonates ≥35 weeks GA, BW ≥1500g. No gold standard comparator. Validity of VAS questionable in neonates. No independence – potential for scoring bias. Unable to determine neonatal results.
Spence et al, Australia, 2005[30] **Diagnostic study level 2b**	PAT tested on 84 neonates up to 96 days post operation, as part of a study of 144 infants up to 182 days of age.	*Reliability* – almost perfect inter-rater reliability (IIC) = 0.85. Mean difference 0.17 on the scale of 1 to 10 (SD 1.73). Scale less reliable at higher pain scores (Spearman rho = 0.17, p < 0.05). *Convergent validity* – substantial correlation between PAT and CRIES (Pearson's r = 0.76, p< 0.001). Fair correlation between PAT and VAS (pearson's r = 0.38, p < 0.01). *Repeatability* – a difference in score of ≥4 indicates a true difference in pain state. *Comparing repeatability and validity between groups* – remained repeatable and valid when Levene's test used to test for equal variances between groups (surgical/non-surgical, term/preterm and ventilated/ non-ventilated).	Sample included non-postsurgical neonates and infants. No gold standard for comparator. CRIES validity questionable. Small sample of VAS observations (mother's). Diverse study population. Unable to determine postoperative neonates.
Horgan et al, UK, 1996[31] **Diagnostic study level 2b**	LIDS developed on 16 neonates up to 43 hours after surgery (6 major, 6 moderate and 4 minor).	*Inter-rater reliability* – almost perfect correlation between research nurse and clinical psychologist (Pearson's Correlation Coefficient = 0.95). Substantial to almost perfect correlation between research nurse and 4 trained assessors (0.74-0.88). 4 assessors scored consistently higher than research nurse. *Reproducibility* – almost perfect correlation between research nurse and 4 trained assessors independently scoring videoed segments (Pearson's Correlation Coefficient = 0.82-0.89). *Test/Re-test* – almost perfect correlation of initial and 2nd scores for research nurse and 4 assessors (Pearson's Correlation Coefficient = 0.81-0.96). *Validity* – scores greater in major surgery group up to 24 hours postoperatively, (especially 1-6 hours and	Small sample. Background characteristics not fully described. Levels of significance not given throughout. Inadequate detail of methodology. *Continued...*

TABLE 1 Most relevant studies of postoperative pain scales. For the full table see the Best BETS website[46].

Study	Scale(s) and patient group	Outcomes and key results	Study weaknesses
		preoperatively). Individual scores consistent with clinical events (diagnosis/condition, analgesia administration/ duration of action).	
Horgan et al, Liverpool, UK, 2002[32] **Diagnostic study level 2b**	Further testing of LIDS on 31 neonates post operation. (13 major, 11 moderate and 7 minor surgery). Control group of 10 'normal' neonates.	*Reliability* – high internal consistency. Cronbach's alphas of 0.86, 0.84, 0.88, 0.94, 0.90, 0.87, 0.93 over 1-6 and 18 hours respectively. *Validity* – LIDS significantly lower after analgesia (t, 10 = 3.67, $p = 0.004$, $n = 11$). Control group had significantly lower scores than minor surgery group (ANOVA $F = 4.66$, $p = 0.05$). Significant decrease in scores over time (ANOVA $F = 8.11$, $p = 0.000$). Interaction not significant ($F6,66 = 0.34$). Comparing control and major and moderate groups: Significant main effect of time (ANOVA $F = 1.9$, $p = 0.004$). Significant difference between groups ($F2,26 = 11.96$, $p = 0.000$) and no significant interaction (F 22,286 = 1.04, $p = 0.41$). Control scores were significantly lower than moderate and major surgery groups (post-hoc Turkey tests).	Small sample size. Background characteristics not fully described. Few observations of pre- and post-analgesia ($n = 11$). Questionable use of related t test (non-parametric data). Minor surgery group data only up to 18 hours. Researcher not strictly blinded to analgesic status. Control neonates observed after delivery – distress noted.
Van Dijk et al, The Netherlands, 2001[33] **Diagnostic study level 2b**	COMFORT 'behaviour' score: 66 neonates in a group of 0-3 year-olds undergoing major abdominal or thoracic surgery, studied COMFORT 'behaviour' items and separate physiological indicators (HR, HRV, MAP, MAP variability (MAPV)).	*Age group differences* – VAS, MAP, HRV and MAPV all significantly lower in neonates than older age groups *Within-subject correlations of COMFORT 'behaviour' scores with physiological indicators* - on average, increase in COMFORT 'behaviour' score associated with increase in physiological indicators. Highest correlation with COMFORT 'behaviour' was for MAPV, $r = 0.49$. And lowest for HRV, $r = 0.37$. *Between subject variability in COMFORT 'behaviour' scores and physiological indicators* – median correlation for all physiological indicators was moderate but massive range (-0.81-0.98). Correlation between HR, MAP, HRV and MAPV low to moderate (0.04 to 0.45). *Behaviour and physiology correlations (Regression analysis)* – correlation poorer with lower age (i.e neonates): $r_{\text{HR and COMFORT}}$ (0.19) and $r_{\text{MAP and COMFORT}}$ (0.28). Correlation with morphine dose: COMFORT 'behaviour' with MAP and MAPV, $r = 0.18$. Increased correlation of behavioural and physiological indicators when scores high (i.e. most pain).	Not all results specifically for neonate group. Neonates sample excluded neonates <35 weeks or <1500g.
Büttner, Germany, 2000[34] **Diagnostic study level 3a**	Development and testing of Children's and Infants Posoperative Pain Scale (CHIPPS) postoperatively on 0-5 year-olds including neonates.	*Identifying suitable parameters and testing internal validity in infant group* – set of 13 'suitable' behavioural items reduced in series of steps to 5. Factor analysis of CHIPPS (5 behavioral and 2 physiological) gave 2 factorial solution (1 behavioural and 1 physiological) and internal consistency not reached (Cronbach's $\alpha = 0.22$, not significant). *Selectivity and sensitivity* – selectivity of behavioural items good in infants (corrected item-scale correlation = 0.86-0.93, $p < 0.005$). No significant difference between sensitivity of different combinations of behavioural items *Inter-rater reliability* – substantial correlation coefficients for items in <36 month-olds ($r = 0.64$-0.77, except crying, $r = 0.07$).	Number of neonates included unknown. Combined with 1-12 month-old infants. Used subjective assessment of analgesic demand as external criterion. Some stages of data analysis not given. Only 1st hour post extubation studied. Cannot always determine results for Infant group Results of construct validity testing not clear. Concurrent validity not tested on the infant group.
Krechel et al, US, 1995[36] **Diagnostic study level 3b**	CRIES tested on 24 infants post operation GA range 32-60 weeks, (mean 44 weeks). Compared with OPS and VAS.	*Inter-rater reliability* – substantial reliability for CRIES (Spearman Rank Correlation Coefficient $r = 0.72$, $p < 0.0001$, $n = 680$) and for OPS ($r = 0.73$, $p<0.0001$, $n = 659$). Almost perfect agreement on VAS (94%). *Convergent validity* – CRIES correlates well with OPS (Spearman-rank correlation coefficient = 0.73 ($p < 0.0001$) Lower correlation of VAS with OPS and CRIES = 0.49 ($p < 0.0001$, n = >1300 for both).	Small sample size. Not all neonates at time of participation. No gold standard comparator – OPS used but not previously validated for neonates and VAS. Observations *Continued...*

TABLE 1 Continued

Study	Scale(s) and patient group	Outcomes and key results	Study weaknesses
		Discriminant validity – pain scores declined following analgesia. Wilcoxon Sign Rank Test- decline of 3.4 units for OPS ($p < 0.0001$, n = 77) and 3.0 units for CRIES ($p < 0.0001$, n = 74). *Nurse preference for tools* – 73% preferred CRIES, 24% OPS, 3% none.	not independent – independence assumed in statistical tests used. Institutional loyalty bias in preference of CRIES.
Schade et al, US, 1996[39] **Diagnostic study level 2b**	POPS, RIPS and NAPI tested on neonates in a group of non-verbal children within 48 hours of surgery (30% minor). 57% of neonates born < 38 weeks GA.	*Inter-rater reliability* (n = 201) – massive range of reliability in neonates (ICC = 0.39-0.87 for RIPS, ICC = 0.40-0.74 for NAPI). *Internal consistency and homogeneity* – Cronbach's alphas satisfactory for all 3 scales RIPS 0.87 to 0.93, NAPI 0.73 to 0.88 and POPS 0.91-0.95 (no improvement of alphas with item removal). Most item-item correlations between 0.30 to 0.70 (desirable) for RIPS and NAPI. For POPS several correlations < 0.30 or > 0.70 – suggests item redundancy. *Frequency of item usage* – items used least were consolability, response to touch, and sleep. *Discriminant validity* – no scale identified significant differences between pain and no pain observations for neonates (n = 6). *Sensitivity and specificity (RIPS 'column rate, RIPS calculated rate, NAPI and POPS)* – sensitivity was low for all scales (0.23-0.32). Specificity was high for all (0.84-0.90). Positive predictive values (0.55-0.61). Negative predictive values (0.65-0.66). *Clinical usefulness (likert scales (0-5) on 6 dimensions)* – composite scores: RIPS 4.4, NAPI 4.6, POPS 3.5.	Lack of background data of neonates (unknown sample size). Reliability of POPS not given for neonates. Unable to determine all results for neonates. Hourly assessment over 4 hours done but possible variability as 4 hour period could be any time during first 48 hour postoperative period.
Hodgkinson K, Australia, 1994[41] **Diagnostic study level 4**	PAT tested on 20 neonates undergoing surgery. (GA ≥ 27 weeks, postnatal age 3-42 days).	*PAT score observations* – scores in 7/17 infants > 10 on return to unit (need for analgesia). 6 babies scored > 5 (need for comfort measures) and 8 scored > 10 (need for comfort and analgesia. *Validity* – in general, scores coincided with interventions to alleviate discomfort/relieve pain – nurses' perception of pain. 15 (88.23%) scored highest in 1st 4 hours. PAT scores 'fell quickly' in 2 receiving bolus and infusion, compared to 6 just with infusion where scores high for 4-6 hours. *Clinical utility* – 7 nurses of pain management group found PAT 'not difficult to use' but explanatory notes needed refinement.	Small sample (pilot study). Some >1 month postnatal age. 3/20 excluded as indicators not suitable (paralysed). No gold standard comparator. Assessments brief as assessors often caring for other infants. Results not explicit – data not available to draw own conclusions.

TABLE 1 Continued

small and excluded premature neonates.

Van Dijk[29] tested the COMFORT score and found almost perfect inter-rater reliability for the physiological items but much lower reliability for the behavioural items. Heart rate and MAP were also much more stable over time. Establishment of convergent validity was attempted using VAS and almost perfect correlation was found for COMFORT 'behaviour', but heart rate and MAP correlated poorly with VAS. This may indicate only that scoring of VAS is not based on physiological variables. An improvement of correlation when VAS was assessed after the two minute COMFORT illustrates the lack of objectivity of the VAS. Use of VAS to test validity is a recurrent weakness in studies of postoperative pain assessment.

Van Dijk[33] later looked at the association of the behavioural components of COMFORT with physiological variables HR, HRV, MAP and mean arterial pressure variability. Neonates were found to have lower physiological values and lower associations between these and COMFORT 'behaviour'. Although increased COMFORT 'behaviour' score was, on average, associated with increased physiological indicators, the correlations were limited and the range of between subject variability was enormous. Correlations between behaviour and some physiological variables were affected by other determinants. This suggested that the association between behavioural and physiological indicators of pain is: lower in neonates; increased with more severe pain; and is affected by morphine levels. The results of both studies may not directly apply to the neonatal surgical population as the study group included 0-3 year-old children, excluded preterm and low birthweight neonates, and the results were combined for all age groups.

The PAT (Pain Assessment Tool)[41] (**TABLE 2**) was developed and tested originally by Hodgkinson et al in 1994. In this pilot study, validity was vaguely suggested by pain-relieving interventions by nurses coinciding with score boundaries hypothesised to indicate need for intervention. Furthermore, the vast majority scored highest in the four hours after surgery, potentially the period of most pain. Scores in neonates given a morphine bolus fell quickly, also supporting some validity. The tool was indicated to be clinically useful although the need for further testing was highlighted.

Spence et al (2005)[30] tested the PAT score further and found almost perfect inter-rater reliability. Repeatability was tested and a score difference of 4 or more would indicate a true difference in pain severity. Convergent validity was established comparing PAT with CRIES. Validity and repeatability were not affected by gestational age, ventilation or having surgery. Results for the postoperative neonates could not be separated from other infants.

Horgan et al[31,32] developed and tested the LIDS (Liverpool Infant Distress Scale) score over two studies. Reliability, reproducibility and test/retest were established in the first study with almost perfect correlations and high internal consistency was found in the second. Validity was suggested by scores being higher in the major surgery group in the 24 hours after surgery, and by reflecting clinical events in individual neonates. Further validity testing in the second study confirmed significant differences between all surgical groups compared to a control of normal neonates. Scores were significantly lower after analgesia (although assessors were not blinded to analgesic status), and scores decreased over time. There was a lack of background data for the small samples of neonates and also some lack of detail of methods used. Clinical utility was not good, however, as LIDS required 10 minute observations.

Büttner et al[34] developed CHIPPS (Children's and Infant's Postoperative Pain Scores) by stepwise reduction of behavioural items found in the literature to five key indicators (cry, facial expression, posture of trunk and of the legs and motor restlessness) suitable for 0-5 year-old children. Heart rate and respiratory rate were added although internal consistency was not reached. Selectivity of the behavioural items was good, and substantial inter-rater reliability found for all except crying (poor reliability). Validity testing was not done in the infant group and the number of neonates in the study is unknown. The CHIPPS scale appears to need more testing if it is to be used in the neonatal population.

Krechel et al[36] tested the CRIES score and found substantial inter-rater reliability. However, the observations were not strictly independent, and this condition is assumed in the Spearman rank test of correlation. Discriminant validity is suggested by a decrease of 3 units on the

Parameters	0	1	2
Posture/tone		Extended Digits widespread Shoulders raised off bed	Flexed and/or tense Fists clenched Trunk guarding Limbs drawn to midline Head and shoulders resist posturing
Cry	No		Yes When disturbed Doesn't settle after handling Loud Whimpering Whining
Sleep pattern	Relaxed		Agitated or withdrawn Wakes with startle Easily woken Restless Squirming No clear sleep/wake patterm Eye aversion "shut out"
Expression		Frown Shallow furrows Eyes lightly closed	Grimace Deep furrows Eyes tightly closed Pupils dilated
Colour	Pink, well perfused		Pale/dusky/flushed Palmar sweating
Respirations		Tachypnoea At rest	Apnoea At rest or with handling
Heart rate		Tachycardia At rest	Fluctuating Spontaneous or at rest
Oxygen saturation	Normal		Desaturation with or without handling
Blood pressure	Normal		Hypo-/hypertension at rest
Nurse's perception	No pain perceived		Baby perceived to be in pain

Note: Infants are assessed and scores obtained every 2-4 hours. An infant with a score > 5 requires comfort measures: > 10 requires an analgesia dose adjustment.

TABLE 2 The Pain Assessment Tool (PAT) scoring system.

CRIES score following analgesia. Testing of convergent validity by comparing CRIES to OPS (Objective Pain Scale) showed substantial correlation but it should be noted that OPS has not been validated for use in neonates. CRIES was the most preferred assessment tool although as it was developed at that unit, this choice may be affected by institutional bias. The small sample size used and inclusion of non-neonates may limit the extent to which these findings can be generalised.

Schade et al[39] compared three different pain scores POPS (Postoperative Pain Score), RIPS (Riley Infant Pain Scale) and NAPI (Nursing Assessment of Pain Intensity). Inter-rater reliability varied massively for both RIPS and NAPI (data not known for POPS) and so could not be categorised. Internal consistency was satisfactory for all three scales, but item reduction was indicated only for POPS. Validity was not established for any scale in neonates. All three scales had low sensitivity for pain and high specificity, meaning that neonates in pain may not be diagnosed as being so by these scales. NAPI was most clinically useful. Although sample size estimates were met, the sample size of neonates was not given, and hence, applicability of these results is unknown.

Studies of nurses' assessment of postoperative pain

In a survey of 26 nurses from a neonatal surgical unit, Charlton[37] found that the categories of behaviours nurses believe to change with pain were cry, vital signs, posture, facial expression, movements, tone and colour. The list of the characteristics of these behaviours believed to indicate pain was extensive. The most frequently described characteristics were: raised pulse rate, respirations or blood pressure; high pitch of cry; rigidity; drawing up of knees; and furrowed facial expression. These findings may be relevant to future development or adaptation of a pain assessment tool for use in neonates. The findings also document some of the indicators nurses use currently to assess neonatal pain.

A second study by Charlton[38] looked at nurses' preferences of indicators of pain in practice. In neonates heart rate was by far the most frequently used indicator of pain and 'vital signs' was ranked as the most preferred group of indicators. The results also suggest that nurses put unequal weighting on behavioural indicators with tone used more than crying, response to handling and spontaneous movement. Facial expression and sleep pattern were used even less. Although this work is based purely on subjective preferences of nurses, it may be valuable when assessing the clinical usefulness of existing pain scales. It also challenges the equal weighting of different components of existing scales.

Hultgren[42] conducted a questionnaire looking at how NICU nurses in assess pain and make medication decisions. As before, vital signs were the indicators regarded as most important in neonatal postoperative pain assessment. Cry and restlessness/agitation also ranked highly.

Hudson-Barr et al[43] also studied how nurses assess pain to make medication decisions. Using medication activity to classify video snippets of postoperative neonates and infants, nurses were better at correctly diagnosing 'no-pain' situations than pain/need for analgesia. The finding that nurses' agreement with the pharmacological standard was only moderate may support the need for an objective score in clinical practice. Although numerous behaviours were discussed, facial expression was most frequently mentioned and considered to be most important. The number of neonates in the small sample is not given, so the relevance to the neonatal population cannot be determined.

Studies of responses to neonatal pain

Warnock et al[35] did an ethological study to describe distress and post-distress behaviours in neonates and found that although many of the behaviours commonly occurred in all distress and post-distress events, some were exclusive to circumcision or post circumcision. This indicates validity of some behaviours to distinguishing between different degrees of pain/distress. However, only four neonates were studied, so findings cannot be generalised.

Mills[40] conducted a qualitative study of pain behaviours following tissue trauma including surgery. Features of commonly used pain indicators were identified that appear to distinguish between severities of pain and differ with gestational age. For example, cry was found to be more intermittent and of shorter duration in neonates than older infants. Such findings are useful when selecting appropriate measures of pain, especially when they are not designed for neonates.

Côte[44] did an ethological study describing 'distress' and 'no distress' states and the behaviours that characterise them. Using these, episodes of acute distress were more numerous in the neonate that had the least analgesia postoperatively. Although acute distress was present during pain-inducing procedures, it was also observed in the absence of external stimuli, and so the assumption can be made that behaviours can detect postoperative pain. Reduced crying and frowning, and drowsiness were noted post analgesia. These findings support the validity of using behaviours to differentiate pain/distress from other states in postoperative neonates. Furthermore, increases in pulse rate were noted in the neonates with most episodes of acute distress, suggesting that this indicator may be helpful in conjunction with behavioural items in an objective pain scale. Findings cannot be generalised as the sample was very small.

Postoperative pain assessment in practice and in future research

A finding of this review was that many of the studies on measurement of postoperative pain in neonates are of insufficient quality to validate the pain scales for clinical use. A weakness in several studies is the combination of neonates with older infants in study groups and subsequent analysis of combined data. The literature clearly highlights that, although neonates have the capacity to respond to noxious stimuli, ongoing maturation of the pain pathways and cognitive development inevitably mean that responses in neonates will differ to those of older infants. Furthermore, many of the studies excluded very preterm neonates. As the population of the neonatal surgical unit includes such babies, consideration of the validity of these scales in very preterm neonates is important. Many postoperative neonates are ventilated, and few tools consider this. The Distress Scale for Ventilated Newborn Infants (DSVNI)[45] is designed for assessing distress behaviours and physiological changes expressed during procedures, and will require further testing before use for postoperative pain. Future research should be directed at validating a clinically useful and reliable gold standard tool for assessment of pain in postoperative neonates. This opinion is supported by pain management guidelines[12].

Conclusions

Through critical appraisal of the papers found in a structured search, it is apparent that there is a current lack of a gold standard pain assessment method for postoperative neonates.

Studies of the responses to postoperative pain in neonates support the validity of using behavioural and physiological indicators for assessment. The consistent finding that nurses use vital signs to assess postoperative pain cannot be ignored. Multidimensional tools include both types of indicator and despite excellent reliability of unidimensional scales, such as NFCS, they appear to be more clinically useful.

Testing of CHIPPS was inadequate for the neonatal population. PIPP had some validity and has potential for further development, but reliability was not tested. The reliability, validity and internal consistency of LIDS were established, but unfortunately the scale needs simplifying before it could be used as a routine bedside tool. Only the physiological items of COMFORT had high reliability and validation was inadequate. The testing of reliability of CRIES was questionable, but nevertheless found to be substantial. CRIES was validated against PIPP, PAT, and less usefully, against OPS. Some discriminative validity was established. The tool is appropriate for clinical use and has been tested on neonates down to 28 weeks' gestation. PAT was found to have high reliability and repeatability was calculated. Validity was established against CRIES and discriminative validity suggested during initial testing. PAT is also suitable for clinical use and has been used for neonates down to 27 weeks' gestational age.

Although no tool has emerged as having perfect reliability, validity, and clinical feasibility, PAT appears to best meet the requirements of the neonatal surgical unit. However, further validation and reliability testing is necessary.

References

1. **Merskey H. et al.** Pain Terms: A list with definitions and notes on usage. *Pain* 1979; **6**(3): 249-52.
2. **Scanlon J.W.** Appreciating neonatal pain. *Adv Pediatrics* 1991; **38**: 317-33.
3. **Truog R., Anand K.J.S.** Management of pain in the postoperative neonate. *Neonatal Surgery* 1989; **16**(1): 61-78.
4. **Goldman A., Lloyd-Thomas A.R.** Pain management in children. *Br Med Bulletin* 1991; **47**(3): 676-89.
5. **Walker S.M.** Acute pain management in pediatric patients. *Int Anaesthesiology Clin* 1997; **35**(2): 105-30.
6. **Stevens B.J., Franck L.S.** Assessment and management of pain in neonates. *Paediatric Drugs* 2001; **3**(7): 539-58.
7. **Stevens B.J., Johnson C.C., Gruneau R.V.** Issues of assessment of pain and discomfort in neonates. *JOGNN* 1995; **24**: 849-55.
8. **Walker S.M., Horward R.F.** Neonatal pain. *Pain Reviews* 2002; **9**: 69-79.
9. **Bell S.G.** The National Pain Management Guideline: Implications for neonatal intensive care. *Neonatal Network* 1994; **13**(3): 9-17.
10. **Anand K.J., Hickley P.R.** Halothane-morphine compared with high-dose sufentanil for anaesthesia and postoperative analgesia in neonatal cardiac surgery. *NEJM* 1992; **326**: 1-9.
11. **Committee on Psychosocial Aspects of Child and Family Health.** The assessment and management of acute pain in infants, children and adolescents. *Pediatrics* 2001; **108**(3): 793-97.
12. **Agency for Health Care Policy and Research.** Acute pain management: Operative or medical procedures and trauma, part 2. *Clin Pharmacy* 1992; **11**: 391-414.
13. **van Dijk M., Peters J.W.B., Bouwmeester, Tibboel D.** Are postoperative pain instruments useful for specific groups of vunerable infants? *Clin Perinatology* 2002; **29**: 469-91.
14. **Grunau R., Craig K.** Pain expression in neonates: Facial action and cry. *Pain* 1987; **28**(3): 395-410.
15. **Craig K. et al.** Pain in the preterm neonate: Behavioural and physiological indices. *Pain* 1993; 52(3): 287-99.
16. **Choonara I.** Pain in neonates, assessment and management. *Sem Neonatol* 1998; **3**: 137-42.
17. **Anand K.J.S., Hickley P.R.** Pain and its effects in the human neonate and fetus. *NEJM* 1987; **317**: 1321-29.
18. **Frank L.S., Miaskowski C.** Measurement of neonatal responses to painful stimuli: A research review. *Pain Symptom Management* 1997; **14**: 343-78.
19. **Frank L.S.** Some pain, some gain: Reflections on the past two decades of neonatal pain research. *Neonatal Network* 2002; 21: 37-41.
20. **Nagy S.** A comparison of the effects of patient's pain on nurses working in burns and neonatal intensive care units. *Neonatal Network* 1998; 5-11.
21. **Abu-Saad H.H., Bours G.J.J.W., Stevens B., Hamers J.P.H.** Assessment of pain in the neonate. *Sem Perinatology* 1998; **22**(5): 402-16.
22. **Barrier G., Attia J., Mayer M.N., Amiel-Tison C.L., Shnider S.M.** Measurement of post-operative pain and narcotic administration in infants using a new clinical scoring system. *Intens Care Med* 1989; **15**: S37-S39.
23. **bestbets.org.** Manchester: Accident and Emergency Department, Manchester Royal Infirmary; c2000-2002 [cited July 18, 2006] Available from: http://www.bestbets.org/index.html
24. **Crombie I.K.** The Pocket Guide to Critical Appraisal. 12th impression. London: BMJ. 2004.
25. **cebm.net: Oxford:** Oxford Centre for Evidence-based Medicine Levels of Evidence cMay 2001 [cited July18,2006] Available from http://www.cebm.net/levels of evidence.asp
26. **Landis J.R., Koch G.G.** The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159-74.
27. **McNair C., Ballantyne, Dionne K. et al.** Postoperative pain assessment in the neonatal intensive care unit. *Arch Dis Child Fetal Neonatal Ed* 2004; **89**: 537-41.
28. **Peters J.W., Koot H.M., Grunau R.E. et al.** Neonatal Facial Coding System for assessing postoperative pain in infants: Item reduction is valid and feasible. *Clin Pain* 2003; **19**: 353-63.
29. **van Dijk M., de Boer J.B., Koot H.M. et al.** The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain* 2000; **84**(2-3): 367-77.
30. **Spence K., Gillies D., Harrison D., Johnston L., Nagy S.** A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. *JOGNN* 2005; **34**(1): 80-86.
31. **Horgan M., Choonara I.** Measuring pain in neonates: An objective score. *Paediatric Nursing* 1996; **8**(10): 24-27.
32. **Horgan M.F., Glenn S., Choonara I.** Further development of the Liverpool infant Distress Scale. *J Child Health Care* 2002; **6**(2): 96-106.
33. **van Dijk M., de Boer J.B., Koot H.M. et al.** The association between physiological and behavioural pain measures in 0- to 3-year-old Infants after major surgery. *J Pain Symptom Management* 2001; **22**(1): 600-09.
34. **Büttner W., Finke W.** Analysis of behavioural and physiological parameters for the assessment of postoperative analgesic demand in newborns, infants and young children: A comprehensive report on seven consecutive studies. *Paediatric Anaesthesia* 2000; **10**: 303-18.
35. **Warnock F., Sandrin D.** Comprehensive description of newborn distress behaviour in response to acute pain (newborn male circumcision). *Pain* 2004; **107**(3): 242-55.
36. **Krechel S.W., Bildner J.** CRIES: A new neonatal postoperative pain measurement score. Initial testing of validity and reliability. *Paediatric Anaesthesia* 1995; **5**: 53-61.
37. **Charlton A.J.** Assessment of postoperative pain in neonates: A survey of nursing practice. *JNN* 1999; **5**(5): 21-23.
38. **Charlton A.J.** Pain indicators in postoperative neonates: A study of nurses' preferences. *JNN* 2000; **6**(2): 41-42.
39. **Schade J.G., Joyce B.A., Gerkensmeyer J., Keck J.F.** Comparison of three preverbal scales for post-operative pain assessment in a diverse pediatric sample. *J Pain Symptom Management* 1996; **12**(6): 348-59.
40. **Mills N.M.** Pain behaviour in infants and toddlers. *J Pain Symptom Management* 1989; **4**(4): 184-90.
41. **Hodgkinson K., Bear M., Thorn J., van Blaricum S.** Measuring pain in neonates: Evaluating an instrument and developing a common language. *Aus J Adv Nursing* 1994; **12**(1): 17-22.
42. **Hultgren M.S.** Assessment of postoperative pain in critically ill infants. *Progress Cardiovascular Nursing* 1990; **5**(3): 104-12.
43. **Hudson-Barr, Duffey, Holditch-Davis D., Funk S., Frauman A.** Pediatric nurses use of behaviours to make medication administration decisions in infants recovering from surgery. *Res Nursing Health* 1998; **21**: 3-13.
44. **Côte J.J., Morse J.M., James S.G.** The pain response of the postoperative newborn. *J Adv Nursing* 1991; **16**: 378-87.
45. **Sparshott M.M.** The development of a clinical distress scale for ventilated newborn infants: Identification of pain and distress based on validated behavioural scores. *JNN* 1996; **2**(2): 5-11.
46. **Best Bets** http://www.bestbets.org/database /search.html