Pain 76 (1998) 277–286

Research Papers

# Bedside application of the Neonatal Facial Coding System in pain assessment of premature neonates

Ruth Eckstein Grunau[a,b,c,*], Tim Oberlander[a,b], Liisa Holsti[c], Michael F. Whitfield[b,c]

[a]Centre for Community Child Health Research, B.C. Research Institute for Children's and Women's Health, Vancouver, BC V6H 3V4, Canada
[b]Department of Paediatrics, University of British Columbia, Vancouver, BC V6H 3V4, Canada
[c]B.C.'s Children's Hospital, Vancouver, BC V6H 3V4, Canada

Received 27 May 1997; revised version received 30 January 1998; accepted 17 February 1998

**Abstract**

Assessment of infant pain is a pressing concern, especially within the context of neonatal intensive care where infants may be exposed to prolonged and repeated pain during lengthy hospitalization. In the present study the feasibility of carrying out the complete Neonatal Facial Coding System (NFCS) in real time at bedside, specifically reliability, construct and concurrent validity, was evaluated in a tertiary level Neonatal Intensive Care Unit (NICU). Heel lance was used as a model of procedural pain, and observed with $n = 40$ infants at 32 weeks gestational age. Infant sleep/wake state, NFCS facial activity and specific hand movements were coded during baseline, unwrap, swab, heel lance, squeezing and recovery events. Heart rate was recorded continuously and digitally sampled using a custom designed computer system. Repeated measures analysis of variance (ANOVA) showed statistically significant differences across events for facial activity ($P < 0.0001$) and heart rate ($P < 0.0001$). Planned comparisons showed facial activity unchanged during baseline, swab and unwrap, then increased significantly during heel lance ($P < 0.0001$), increased further during squeezing ($P < 0.003$), then decreased during recovery ($P < 0.0001$). Systematic shifts in sleep/wake state were apparent. Rise in facial activity was consistent with increased heart rate, except that facial activity more closely paralleled initiation of the invasive event. Thus facial display was more specific to tissue damage compared with heart rate. Inter-observer reliability was high. Construct validity of the NFCS at bedside was demonstrated as invasive procedures were distinguished from tactile. While bedside coding of behavior does not permit raters to be blind to events, mechanical recording of heart rate allowed for an independent source of concurrent validation for bedside application of the NFCS scale. 

*Keywords:* Infants; Premature infant; Pain measurement; Facial expression; Heart rate

## 1. Introduction

Concerns about possible short- and long-term adverse consequences of early pain exposure in premature infants (Gottfried and Gaiter, 1985; Fitzgerald et al., 1988; Porter, 1989; Anand and McGrath, 1993; Grunau et al., 1994, 1998; Aynsley-Green, 1996; Johnston and Stevens, 1996; Anand et al., 1997) imply urgency to develop valid and reliable measures of infant pain for potential clinical application. In contrast to older children and adolescents, where pain experienced can be evaluated using self report, assessing pain in infants requires alternate, developmentally appropriate non-verbal measures. Recently many measures have been developed to quantify infant pain (Bours et al., 1996), however questions remain regarding reliability, validity, feasibility and ultimately clinical utility of these measures for application at bedside with individual infants.

Of the various behaviors available in the infant repertoire, especially for infants in the Neonatal Intensive Care Unit (NICU), facial activity appears to be the most relatively specific indicator of pain (Stevens et al., 1996), and is most salient to caregivers (Hadjistravropoulos et al., 1994). The Neonatal Facial Coding System (NFCS; Grunau

* Corresponding author. Rm. L408, 4480 Oak Street St., Vancouver, BC V6H 3V4, Canada. Tel.: +1 604 8753570; fax: +1 604 8753569; e-mail: grunau@unixg.ubc.ca

0304-3959/98/$19.00 
*PII* S0304-3959(98)00046-3

and Craig, 1987, 1990) has been used to study pain behavior of fullterm (Grunau and Craig, 1987; Grunau et al., 1990), preterm (Craig et al., 1993; Stevens et al., 1994; Johnston et al., 1995, 1996; Johnston and Stevens, 1996), and older infants (Johnston et al., 1993; Lilley et al., 1997). The NFCS was developed and validated using videotaping which allowed for intensive slow motion stop frame videocoding and playback, and with these methods, good reliability has been consistently demonstrated. Construct validity has been established as the NFCS discriminates tissue insult and non-tissue insult procedures, and differentiates infants receiving pharmacologic treatment (Benini et al., 1993; Scott et al., 1994; Taddio et al., 1997). Convergent validity has been demonstrated through comparison with comprehensive facial coding (Craig et al., 1994; Lilley et al., 1997). Moreover, the cluster of facial activity associated with pain is similar across infancy and in adults (Craig et al., 1992).

Facial actions of the NFCS have recently been used in studies at bedside. Guinsburg et al. (1997) applied the NFCS at bedside, with term and preterm infants, during blood collection or skin friction. These investigators simplified coding of mouth activity by combining two actions (mouth stretch vertical and mouth stretch horizontal) into one judgment of mouth stretch. Tongue protrusion was omitted, as it is counterindicative of pain in fullterm neonates (Grunau et al., 1990), resulting in eight face actions instead of 10. Rushforth and Levene (1994), Rushforth et al. (1995), and Ramenghi et al. (1996) have used four of the 10 NFCS face actions, namely the upper face actions of brow lower and eyes squeezed shut, plus naso-labial furrow and open mouth. Similarly, the Premature Infant Pain Profile (PIPP), which was designed to be used with videotaping or at bedside, incorporated three facial actions from the NFCS (Stevens et al., 1996). The rationale for using a limited subset of three or four face actions has been that the upper facial actions and open mouth are the most frequent responses common to almost all fullterm neonates undergoing acute invasive procedures. However, although other facial actions and tongue movements are not as universally observed, they may provide important information, especially about individual differences in pain expression. Furthermore, the studies which have identified a simple facial cluster have only involved acute procedural pain. Chronic, longer lasting, internally generated or post operative pain has received little attention. It is possible that subtle signs of ongoing discomfort will be missed if only those face actions common to most infants during acute procedural pain are measured. Therefore, it is important to evaluate the feasibility of applying the full NFCS at bedside, as restriction of the number of facial actions appears premature.

The methodology used in all the bedside studies has been to score each NFCS face action as occurring or not occurring at any time during specified phases of a procedure. Guinsburg et al. (1997) scored the eight NFCS face actions as occurring or not occurring at any time during each of the following phases: at rest, foot warming, capillary puncture (or skin friction), and at 1 min, and 3 min after the procedure. In the studies of Levene and his colleagues (Rushforth and Levene, 1994; Rushforth et al., 1995; Ramenghi et al., 1996) each of the four face actions they used was scored as occurring or not-occurring during heel preparation as compared with heel lancing, and whether or not the infant cried. Very high inter-observer reliability has been reported for the face actions at bedside, namely 0.94 (Rushforth and Levene, 1994) using the conservative formula of Eckman and Friesen (1978) which controls for inflated agreement due to nonoccurrence, and 97.5% agreement (Guinsburg et al., 1997).

Infant state and specific postural and body activities of the Neonatal Individualized Developmental Care and Assessment Programme (NIDCAP) measurement system (Als, 1984) have been studied recently at the bedside in the NICU as potential acute pain indicators in premature infants (Grunau et al., under editorial review). Hand behaviors, namely finger splay and hand to mouth movements, were related to invasive procedural pain, in addition to global facial grimace and infant state. In the present study infant sleep waking state, finger splay and hand to mouth movements were further examined as potential pain indicators.

The main aims of this study were to evaluate the feasibility, inter-observer reliability, and construct and concurrent validity of the full NFCS at the bedside in real time. To achieve these goals, heel lance was used in this study as the model of an acute procedure involving tissue damage. Construct validity of the NFCS at bedside was assessed by comparing the responses at baseline, first contact by the laboratory technician, and cleansing the skin, to the heel lance and squeeze components of the blood collection procedure, and a recovery period. Concurrent validity was assessed by recording heart rate as a concomitant physiological signal and comparing the co-occurrence of changes in heart rate with changes in facial activity. Hand movements and infant sleep/waking state were also observed and recorded at bedside. Four questions were addressed: (1) Can health care providers learn the full NFCS in real time and apply it at bedside? (2) Can acceptable reliability be established in the NICU situation at bedside? (3) Is the real time bedside application valid as a behavioral measure of acute pain? (4) A further aim of this study was to examine the feasibility of recording infant state and hand movements in addition to the NFCS facial actions in real time at bedside.

## 2. Methods

### 2.1. Subjects

Written informed consent was obtained from the mother or other legal guardian according to a protocol approved by the Clinical Research Ethics Committee of the University of

British Columbia. A continuous series of $n = 42$ infants with birthweight ≤1500 g undergoing tertiary level III neonatal intensive care in B.C.'s Children's Hospital were recruited. All infants were observed at 32 weeks to 32 weeks 6 days post conceptional age (PCA). The criterion for inclusion was no major congenital anomaly. Infant state at baseline was examined and controlled by retaining only infants in States 1–5 (Als, 1984). One infant was in State 6 (crying) and one infant was in State 7 (prolonged respiratory pause >8 s) in the baseline period and their data were excluded, leaving $n = 40$ as the study sample. The only infant who had received opioid medication (morphine) administered within the last 24 h was the infant who was excluded due to being in State 7 in baseline. Subject characteristics are provided in Table 1.

### 2.2. Procedures

Infants were recruited in the NICU by a research nurse. Observations of the NFCS, infant sleep/wake state and hand movements were carried out by a coder at bedside in real time during routine blood collection, which was performed by a laboratory technician. There were three coders: a neonatal nurse with no previous background in behavioral coding, an occupational therapist who was NIDCAP certified (Als, 1984), and a social worker. Of the three coders, only the social worker had been trained in the original NFCS method of application to videotapes. Training of the coders was carried out initially using videotapes played in real time, followed by practice coding in the NICU at bedside until adequate inter-observer agreement (at least 85%) was achieved. The infants coded for training were not in the study. The behavior coders were blind to the purpose of the study and medical information about the infants. Reliability coding for the study was carried out on 15 infants (33% of the study sample).

The infant's NICU nurse applied the foot warmer, then the research nurse removed the electrodes from the bedside cardiac monitor cable, and switched them to the cable which fed into the study computer for heart rate data acquisition. Heart rate was recorded continuously during a 200 s baseline, during the blood collection and for a recovery period of 200 s following last contact. Infant sleep/wake state, NFCS and hand movements were rated during six events: 200 s baseline, first contact by the laboratory technician to remove the heel warmer (median 28 s), swabbing to cleanse the heel (median 17 s), lance (median 13 s), squeeze (median 184 s) and for a recovery period of 200 s following last contact. Variation in the time taken for each phase of the blood collection reflected outliers rather than typical time, thus median time was given here as representative of each phase of the coding period. More than one heel lance was required for 11 (28%) of the infants; responses were recorded only for the first lance and squeeze phase.

Table 1

Subject characteristics ($n = 40$)

| | Mean (SD) | Number (%) | Minimum | Maximum |
|---|---|---|---|---|
| Birthweight (g) | 1023.22 (289.12) | – | 500 | 1474 |
| Gestational age at birth (weeks) | 28.09 (2.16) | – | 24 | 32 |
| Age at observation (days) | 30.28 (14.31) | – | 5 | 56 |
| Male gender | – | 13 (33) | – | – |
| Ventilator support | – | 31 (78) | – | – |

Paper and pencil recording was used rather than handheld computer recording (Sackett, 1978; Hile, 1991) as the purpose of the study was to simulate the potential clinical situation as it exists currently. A grid was completed at bedside for each infant (Table 2). The research nurse cued the coder when the last 120 s of the baseline phase was beginning. When the coder completed the column of the grid for the baseline event, the blood collection procedure was begun. Sleep/wake state was rated first for each event, followed by NFCS and hand movements. One mark was placed in a box (see Table 2) if a face or hand action occurred at any time during each event, and left blank otherwise. Following each event there was sufficient time to score behaviors during the transition from one event to the next. For example, following the lance, which was the shortest event, the laboratory technician placed the lance on a tray and picked up the blood collection pipettes.

### 2.3. Measures

#### 2.3.1. Infant sleep/waking state

Infant state was coded as defined by Als (1984). Sleep/walking state was coded from 1 to 7: 1 = deep sleep, 2 = light sleep, 3 = drowsy, 4 = quiet awake, 5 = active awake, 6 = highly aroused, agitated, upset and/or crying, 7 = prolonged respiratory pause >8 s. Preterm infants at times appear to go into a transitory state of 'collapse' or 'withdrawal', characterized by muscular flacidity and prolonged pause in breathing which is captured in State 7.

#### 2.3.2. Facial activity

The 10 facial actions of the NFCS (Grunau and Craig, 1987, 1990) were coded: brow lowering, eyes squeezed shut, deepening of the naso-labial furrow, open lips, vertical mouth stretch, horizontal mouth stretch, taut tongue (cupping of the tongue), chin quiver (high frequency vibration of the chin), lip purse (tightening the muscles around the lips to form 'oo'), and one which has been found to be counter-indicative of pain in term infants: tongue protrusion. Each face action was coded as 1/0 (occurred/did not occur) during each event.

#### 2.3.3. Hand movements

Two hand movements, finger splay and hand to mouth, were coded as defined in the NIDCAP behavioral measurement system (Als, 1984). Finger splay was defined as back-

Table 2

Coding form for sleep/waking state, NFCS and hand movements at bedside

STUDY ID ________ **Behaviour Coder** ________ **Test Date** ________

| | | BASELINE record actual time | UNWRAP | SWAB | LANCE | SQUEEZE | POST |
|---|---|---|---|---|---|---|---|
| STATE | Deep sleep | | | | | | |
| | Light sleep | | | | | | |
| | Drowsy | | | | | | |
| | Quiet awake | | | | | | |
| | Active awake | | | | | | |
| | Agitated/Crying | | | | | | |
| | AA | | | | | | |
| | | | | | | | |
| FACIAL ACTION | Brow bulge | | | | | | |
| | Eye squeeze | | | | | | |
| | Naso-labial furrow | | | | | | |
| | Open lips | | | | | | |
| | Vertical mouth stretch | | | | | | |
| | Horizontal mouth | | | | | | |
| | Taut tongue | | | | | | |
| | Tongue protrusion | | | | | | |
| | Chin quiver | | | | | | |
| | Lip purse | | | | | | |
| | | | | | | | |
| HAND | Hand to mouth | | | | | | |
| | Finger splay | | | | | | |

ward extension of the fingers. Hand to mouth was defined as an active movement of either hand to the mouth. The hand had to touch the mouth, and was not coded if the hand was already on the mouth, or if the hand only touched the face around the mouth.

### 2.3.4. Heart rate

Continuous electrocardiographic (ECG) activity was recorded from a single lead of surface ECG (lead II) and digitally sampled at 360 Hz off line using a specially adapted computer acquisition system and custom physiologic signal processing software (Boston Medical Technologies, 1996). R waves were detected from the sampled ECG and used to form a smoothed instantaneous 4 Hz heart rate time series as described elsewhere (Berger et al., 1989). Epochs of mean heart rate were chosen to correspond with the times of the infant behavioral observations. Due to technical problems, heart rate was not available for one infant, and was missing for one event for two infants.

## 2.4. Data analysis

Repeated measures analysis of variance (ANOVA) was carried out on facial activity scores and separately on mean heart rate across the six events. A statistically significant ANOVA at $P < 0.05$ was followed by planned *t*-tests for paired comparisons to identify differences between specific events. With Bonferroni correction, $P < 0.01$ was the level set for significance for each comparison. Occurrence of finger splay and of hand to mouth movement across events was examined as differences in paired proportions, using the *z*-statistic reported with 95% confidence intervals (CI).

Table 3

Number of infants (%) displaying each NFCS face action across events

| | Baseline | Contact | Swab | Lance | Squeeze | Post |
|---|---|---|---|---|---|---|
| *NFCS face actions* | | | | | | |
| Eye squeeze | 2 (5) | 6 (15) | 7 (18) | 31 (78) | 33 (83) | 10 (25) |
| Brow | 1 (3) | 5 (13) | 8 (20) | 29 (73) | 32 (80) | 8 (20) |
| Naso-labial furrow | – | 4 (10) | 5 (13) | 26 (65) | 30 (75) | 6 (15) |
| Open mouth | 11 (28) | 12 (30) | 16 (40) | 25 (63) | 33 (83) | 26 (65) |
| Vertical stretch mouth | – | 1 (3) | 1 (3) | 9 (23) | 14 (35) | 6 (15) |
| Horizontal stretch mouth | – | 3 (8) | 3 (8) | 16 (40) | 22 (55) | 4 (10) |
| Chin quiver | 1 (3) | – | – | 3 (8) | 5 (13) | 2 (5) |
| Taut tongue | – | – | – | 5 (13) | 12 (30) | 2 (5) |
| Tongue protrusion | 3 (8) | 3 (8) | 8 (20) | 7 (18) | 10 (25) | 11 (28) |

## 3. Results

Overall occurrence of each NFCS face action and hand behavior was examined prior to data analysis, and the criterion for retention was set as 10% in at least one event. Lip purse was not observed at all and was dropped. All other behaviors met the criterion and were retained.

### 3.1. Reliability

For infant sleep/wake state, inter-observer agreement was 79% (kappa = 0.67). The conservative Facial Action Coding System reliability formula (Eckman and Friesen, 1978; Grunau and Craig, 1987) which eliminates inflated agreement due to non-occurrence was used for the NFCS for comparability with prior studies, and also applied to the hand movements. The reliability coefficients were 0.83 for the total NFCS and 0.86 for the NFCS excluding tongue protrusion (which was omitted from analyses of total facial activity as it is counterindicative of pain in fullterm neonates), 0.82 for hand to mouth and 0.70 for finger splay. NFCS reliability calculated on the same subset of four face actions (brow lower, eye squeeze, naso-labial furrow, open mouth) used by Rushforth and Levene (1994) was 0.91 in the present study.

### 3.2. Sleep/wake state

During baseline six infants (15%) were judged to be in deep sleep, 26 (65%) in light sleep, seven (18%) drowsy, and one (2%) quiet awake. Shifts in sleep/wake state were apparent over the events, with 43% of the infants crying (including non-vocalized cry displayed by intubated infants) during lance, and 53% crying during squeeze. One infant remained in deep sleep during lance, but no infant was in deep sleep during heel squeezing.

### 3.3. Facial activity

Occurrence of each facial action across events is presented in Table 3. In addition, the co-occurrence of face actions was examined. The cluster most frequently associated with invasive tissue damage, namely co-occurrence of brow lower, eye squeeze, naso-labial furrow and open mouth was not observed in any infants during baseline, and was seen in three infants (8%) during first contact, five infants (13%) during swab, 17 (43%) during lance, 27 (68%) during squeeze, and five (13%) during the recovery period. Mouth movements were examined separately: 78% of the infants displayed one or more mouth or tongue action during heel lance, and 85% during squeeze.

The individual pain face actions (excluding tongue protrusion, which in term infants is counterindicative of pain) were summed to provide a total NFCS score for each event, with a possible value from 0 to 8. Mean facial activity across events is displayed graphically in Fig. 1. Mean facial activity during baseline was 0.4 (SD = 0.7), unwrap 0.8 (SD = 1.4), swab 1.0 (SD = 1.5), lance 3.6 (SD = 2.0), squeeze 4.5 (SD = 2.0) and recovery 1.6 (SD = 2.0). Repeated measures ANOVA for facial activity showed a statistically significant main effect across events ($F_{5,195} = 58.26$; $P < 0.0001$). Planned comparisons on facial activity revealed no statistically significant difference between baseline and unwrap ($t = -1.58$, df = 39; $P < 0.12$), unwrap and swab ($t = 1.30$, df = 39; $P < 0.20$). Facial activity increased significantly from swab to lance ($t = 8.45$, df = 39; $P < 0.0001$), increased further from lance to squeeze ($t = -3.17$, df = 39; $P < 0.003$), and decreased from squeeze to recovery ($t = -8.32$, df = 39; $P < 0.0001$).

### 3.4. Heart rate

Mean heart rate across events is displayed graphically in Fig. 2. Mean heart rate (beats per minute, BPM) during baseline was 164.4 (SD = 14.0), unwrap 169.0 (SD = 14.6), swab 176.3 (SD = 16.0), lance 179.4 (SD = 15.8), squeeze 186.3 (SD = 18.6) and recovery 173.6 (SD = 16.1). Repeated measures ANOVA showed a significant main effect across events ($F_{5,180} = 44.02$, $P < 0.0001$). Planned comparisons indicated a significant increase between baseline and unwrap ($t = 4.03$, df = 38; $P < 0.0001$), and between unwrap and swab ($t = 5.32$, df = 38;

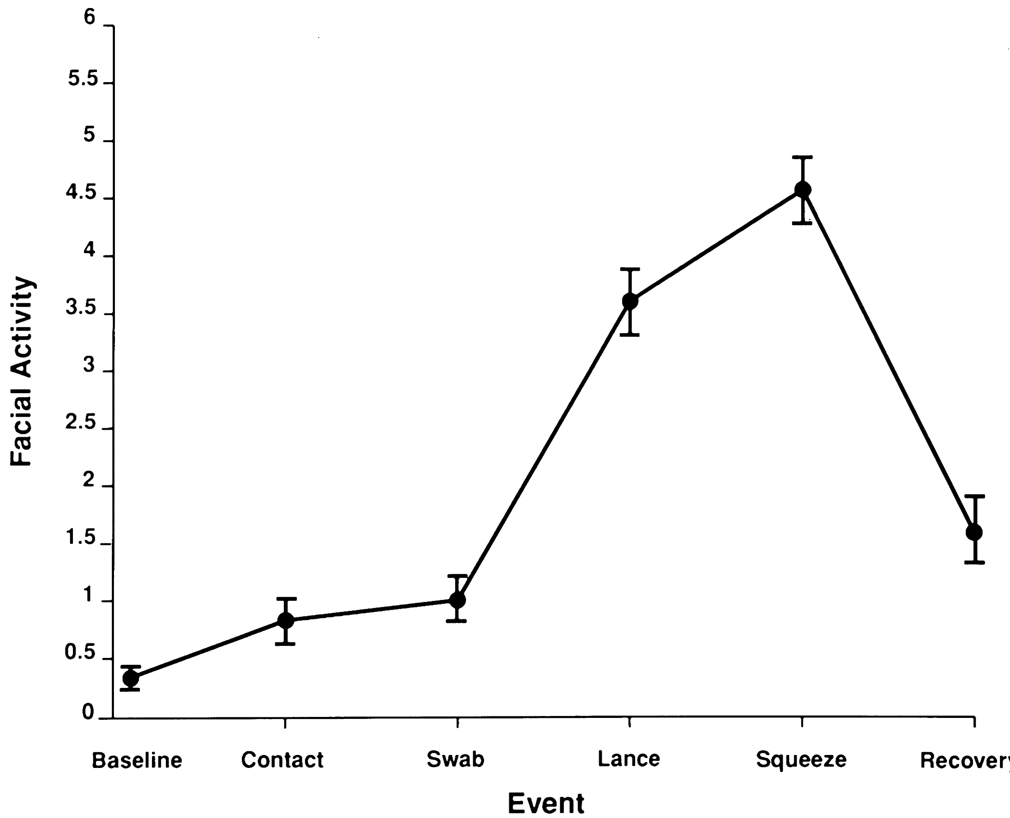


Fig. 1. Mean (±SEM) NFCS facial activity score across blood collection events.

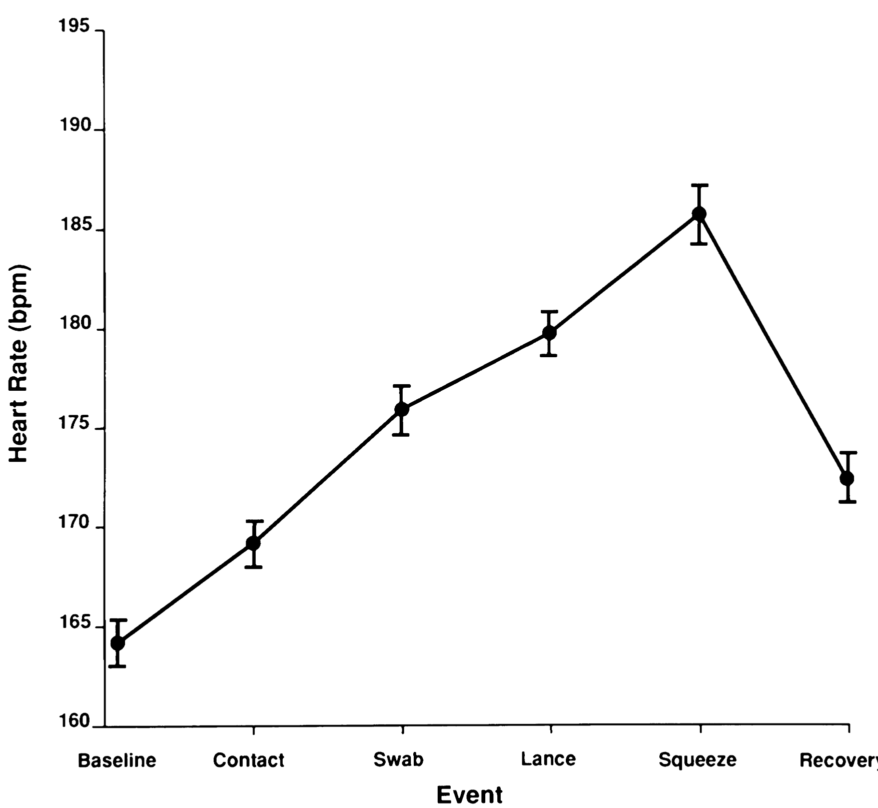


Fig. 2. Mean ( ± SEM) heart rate across blood collection events.

Table 4

Number of infants (%) displaying each hand movement across events

| | Baseline | Contact | Swab | Lance | Squeeze | Post |
|---|---|---|---|---|---|---|
| Finger splay | 2 (5) | 11 (28) | 5 (13) | 11 (28) | 23 (58) | 10 (25) |
| Hand to mouth | 3 (8) | 3 (8) | 2 (5) | 6 (15) | 15 (38) | 12 (30) |

$P < 0.0001$). Heart rate increased significantly from swab to lance ($t = 4.41$, df = 38; $P < 0.0001$), increased further from lance to squeeze ($t = -5.00$, df = 37; $P < 0.0001$), and decreased from squeeze to recovery ($t = -7.39$, df = 36; $P < 0.0001$).

### 3.5. Finger splay

Occurrence of finger splay across events is presented in Table 4. The proportion of infants who displayed finger splay increased significantly from baseline to first contact ($z = -2.56$, difference = –0.23, CI = –0.39 to –0.07; $P < 0.02$), then fell during swab but not significantly. The change from swab to lance was not significant. The increase from lance to squeeze was significant ($z = -2.73$, difference = 0.30, CI = 0.10–0.50; $P < 0.01$), and so was the decline from squeeze to recovery at 0.25 ($z = 3.3$, difference = 0.33, CI = 0.15–0.51; $P < 0.001$).

### 3.6. Hand to mouth

Occurrence of hand to mouth across events is presented in Table 4. The proportion of infants who displayed hand to mouth movement did not differ during baseline, first contact and swab. The increase from swab to lance was significant ($z = -2.5$, difference = 0.10, CI = –0.02–0.22; $P < 0.02$), as was the further increase during squeeze ($z = -2.0$, difference = –0.23, CI = $-$ 0.39 to –0.07; $P < 0.05$). The occurrence remained high during recovery at 0.30, which did not differ from squeeze.

## 4. Discussion

The full NFCS was found to be applicable at bedside in the NICU, confirming the findings of Guinsburg et al. (1997) who applied eight of the 10 face actions in a similar manner. In the present study, health care providers from varying backgrounds (neonatal nurse, occupational therapist and social worker) learned the system in real time and applied it in the clinical setting.

Inter-observer reliability was high for the NFCS pain facial actions (0.86) using the conservative FACS reliability formula (Eckman and Friesen, 1978), consistent with studies of preterm infants using videocoding (0.89; Craig et al., 1993). Reliability calculated on the subset of four face actions used by Rushforth and Levene (1994) was 0.91 in the present study, comparable with their report of 0.94. Reliability of 97.5% on eight NFCS face actions, reported by Guinsburg et al. (1997) was higher as they used percentage agreement, which included agreement due to non-occurrence. Reliability reported at bedside has been surprisingly high, probably due to an important difference in the time periods coded at bedside compared with videotapes. Microanalytic videotape coding assesses inter-observer agreement over much briefer intervals (e.g. 2 s), and thus requires a very high degree of coding precision. In bedside coding, on the other hand, an action is coded if it occurs at any point during a specified event. The time period involved is far longer at bedside than in videotape coding, thus the basis for comparison of inter-observer agreement is quite different. Furthermore, in the first author's experience, an infant's face seems relatively easier to score at bedside than from videotape, as the coder's view is three dimensional, the coder can move to accommodate changes in the infant's position (e.g. head turning, obstructing the lens with arm movement), and also is not affected by extraneous factors such as recording quality (e.g. variation in lighting due to the angle of the camera).

While bedside coding does not provide for raters to be blind to events, mechanical recording of heart rate allowed for an independent source of validation of the scale. Rise in facial activity was consistent with increased heart rate, except that facial activity more closely paralleled initiation of the invasive event. Heart rate increased significantly from baseline during first contact, and increased further during skin swabbing. In contrast, total facial activity did not change significantly during the initial tactile events, and thus was relatively more specific to tissue damage compared with heart rate. This was consistent with previous conclusions that physiological response to tissue damaging stimuli indicates global distress rather than pain (McIntosh et al., 1993; Stevens et al., 1995). Further, the raters were naive to the purpose of the study, and to previous literature on the NFCS, yet the relative occurrence of the individual face actions was comparable to studies of preterm infants using videotaping (Craig et al., 1993).

Examination of the individual face actions suggested it is reasonable to drop lip purse from the NFCS as it has not been found to be related to pain behavior in any studies of fullterm or preterm neonates. However, using a truncated set of three or four face actions, while capturing the most generalized features of facial change across tactile to pain events, misses subtleties which may be indicative of a fuller range of pain sensation or expression. Vertical and/or horizontal stretch mouth were mainly observed in tissue damaging events of heel lance followed by squeezing to carry out blood collection. Chin quiver, while low incidence, was almost exclusively seen in the pain events, and during initial recovery.

Tongue protrusion appeared to be a ‘no pain’ response in fullterm neonates, and thus considered useful as a fairly conclusive counterindicator to pain (Grunau et al., 1990). In fullterm newborns tongue protrusion to initial tactile contact was viewed as perhaps anticipatory to feeding. In the present study, preterm infants at 32 weeks gestation showed a very different pattern. Tongue protrusion was highest during the squeeze and recovery periods. The squeeze event was the most aversive event in the present study, based on the highest frequency of the pain face cluster of high brow bulge, eye squeeze, naso-labial furrow and open mouth. The unexpected pattern of high tongue protrusion during nociception in this study needs further investigation to determine whether this may indicate, for example, immaturity of the nervous system (Andrews and Fitzgerald, 1997), or possibly an extensor preterm stress response (Als, 1984).

Two hand movements were observed, as they were found to be related to procedural pain in a study of routine procedures carried out as part of NICU care (Grunau et al., under editorial review). In the present study, finger splay appeared to be a sensitive but non-specific indicator of distress, as occurrence increased significantly to initial tactile contact for 28% and to invasive events for 58% of the infants. In contrast, hand to mouth movement was mainly evident in the squeeze event, persisting into recovery, but was only displayed by slightly more than one-third of infants. Hand to mouth actions have been interpreted as attempts at self-soothing (Als, 1984), and may reflect relative developmental maturity of some infants.

In the present study, heart rate increased significantly with noxious stimuli and paralleled changes in facial activity, sleep/wake state and hand movements, consistent with previous studies in preterm infants during blood collection (Craig et al., 1993; Johnston et al., 1995). Further, these responses to noxious stimuli were over and above those elicited by non-noxious stimuli (positioning and handling). However, considerable variability in heart rate response has been reported among premature and ill infants in a NICU setting following acute pain (Porter, 1989; McIntosh et al., 1993). While increased heart rate has been described as reflecting a central nervous system stress response that is primarily sympathetically mediated (Lewis et al., 1989), both sympathetic and parasympathetic activity may be involved. Study of low- and high-frequency power spectra show promise as measures which in combination may reflect the neonatal autonomic stress response more comprehensively (Oberlander et al., 1995; Andrews and Fitzgerald, 1997). In the present study, mean heart rate was used only to provide an independent source of concurrent validation for the behavioral coding.

The NFCS at bedside, together with noting shifts in physiological responses and sleep waking state, appears to be a useful approach to multidimensional bedside assessment of pain in infants. However, while this study demonstrated that health practitioners of diverse backgrounds coded these behaviors reliably, and there was validity to these ratings for procedural pain, this only established bedside feasibility and does not imply that the scale has clinical utility. Clinical utility applies to the needs of the individual patient, especially to the basis for decisions regarding medical and nursing management of pain. While it remains to be seen whether patients would benefit from NFCS training of nurses, physicians, and other health care practitioners such as laboratory technicians or physiotherapists, there is unpublished data which suggests this may be a useful approach. Pomietto (personal communication) found that nurses trained in the NFCS showed higher reliability in bedside clinical pain assessment than nurses who had not received the training. The NFCS requires more training than, for example, application of a visual analogue scale (VAS) or the Neonatal Infant Pain Scale (NIPS; Lawrence et al., 1993). Guinsburg et al. (1997) compared the NIPS and the NFCS and reported similar results with term infants. In contrast with preterm infants the NFCS, but not the NIPS, differentiated between lance and skin friction. Therefore they concluded that the NFCS assessed pain in premature infants more specifically than the NIPS. It is now clear that pain must be monitored systematically and continuously in hospitalized infants. Further studies, particularly of ongoing discomfort and distress, in addition to studies of a wider variety of procedural pain, are needed. Whether the NFCS will be the best way to monitor pain for clinical purposes is still an open question.

While it is important to recognize that limiting the NFCS to a small subset of face actions is probably premature in the study of development of pain behaviors, individual differences, prolonged or chronic pain, as we know too little as yet about these areas, there are many purposes for which using a limited subset can be recommended. Demonstration that four NFCS face actions (plus cry) judged at bedside were sufficient to discriminate between tactile and invasive handling (Rushforth and Levene, 1994) was valuable for clinical research (Rushforth et al., 1995; Ramenghi et al., 1996), and offers a relatively inexpensive alternative to the full NFCS. Furthermore, for situations where scoring the lower face (and hearing cry) are precluded, use of only three face actions is recommended, as it may be more reliable than global coding of ‘grimace’. The results of this study and Guinsburg et al. (1997) showed that an array of NFCS face actions were readily observable and reliably scored in real time, thus videotaping is not required for detailed facial coding. However, while clinical studies at bedside can utilize the full NFCS, state, and hand movements, these numerous behaviors can only be recorded phase by phase through a procedure, or at intervals selected by time sampling. Video recording is still necessary for fine-grained second to second precision coding. One of the strengths of the NFCS appears to be flexibility for specific purposes. For clinical research, the choice of how many face actions are needed, and whether to code at bedside or use video recording depend on the study aims.

In summary, the full NFCS can be applied readily at bedside, and construct and convergent validity were demonstrated. A set of eight NFCS face actions was identified as an easily applicable, relatively specific measure of reaction to procedural pain. Tongue protrusion was erratic at 32 weeks postconceptional age, and unlike studies of fullterm infants, was unexpectedly high during nociceptive events. Thus it is important to retain this NFCS action for research studies in the development of pain behaviors in preterm infants. Infant state appears to be a useful additional global indicator of biobehavioral shifts. While hand movements from the NIDCAP were reliably recorded at bedside, finger splay appears to be a generalized distress signal, and hand to mouth movement an attempt to self-soothe, rather than pain cues per se.

Distress appears to be identifiable in infants as a matrix of behavioral and physiological responses which can be viewed as a set of complementary cues. Moreover, pain appears to be on the continuum of distress, particularly for extremely immature infants for whom many of the apparently non-invasive tactile and handling procedures in the NICU may be perceived as painful. A specific ‘pain probe’ is unlikely to be found. From the body of literature which is available on infant behavioral responses to acute pain, facial activity appears to be the most consistent and relatively specific behavioral cue available. In multivariate assessment of pain, together with heart rate, the NFCS face actions, state and finger splay may provide a set of behaviors useful to observing ongoing or long lasting pain such as postoperative pain, in addition to acute procedural pain, however validity for this purpose, and clinical utility remain to be evaluated.

## Acknowledgements

This study was supported by a grant from the British Columbia Medical Services Foundation. We would like to thank Colleen Fitzgerald, Ann-Louise Ellwood, Jacqueline Smit Alex, Marlene Jones, and the staff of the Special Care Nursery, B.C. Children’s Hospital, for their invaluable contributions.

## References

Als, H., Manual for the Naturalistic Observation of Newborn Behavior (Preterm and Fullterm Infants), The Children’s Hospital, Boston, MA, 1984.

Anand, K.J.S. and McGrath (Eds.), Pain in Neonates, Pain Research and Clinical Management, Vol. 5, Elsevier, New York, 1993.

Anand, K.J.S., Grunau, R.E. and Oberlander, T., Developmental character and long-term consequences of pain in infants and children, Child Adolesc. Psychiatr. Clin. North Am., 6 (1997) 703–724.

Andrews, K.A. and Fitzgerald, M., Biological barriers to paediatric pain management, Clin. J. Pain, 13 (1997) 138–143.

Aynsley-Green, A., Pain and stress in infancy and childhood – where to now?, Paediatr. Anaesth., 6 (1996) 167–172.

Benini, F., Johnston, C.C., Faucher, D. and Aranda, J.V., Topical anesthesia during circumcision in newborn infants, J. Am. Med. Assoc., 270 (1993) 850–853.

Berger, R.D., Saul, J.P. and Cohen, R.J., Transfer function analysis of autonomic regulation. I. Canine atrial rate response, Am. J. Physiol., 256 (1989) H142–152.

Boston Medical Technologies, HRView Software, Boston Medical Technologies, Brighton, MA, 1996.

Bours, G.J.J.W., Abu-Saad, H.H., Hamers, J.P.H. and Van Dongen, R.T.M., Pain Assessment in Neonates, Department of Nursing Science, University of Limburg, Maastricht, The Netherlands, 1996.

Craig, K.D., Prkachin, K.M. and Grunau, R.V.E., The facial expression of pain. In: D.C. Turk and R. Melzack (Eds.), Handbook of Pain Assessment, Guilford, New York, 1992, pp. 257–276.

Craig, K.D., Whitfield, M.F., Grunau, R.V.E., Linton, J. and Hadjistravropoulos, H.D., Pain in the preterm neonate: behavioral and physiological indices, Pain, 52 (1993) 287–299.

Craig, K.D., Hadjistravropoulos, H.D., Grunau, R.V.E. and Whitfield, M.F., A comparison of two measures of facial activity during pain in the newborn child, J. Pediatr. Psychol., 19 (1994) 305–318.

Eckman, P. and Friesen, W.V., Manual for the Facial Action Coding System, Consulting Psychologist’s Press, Palo Alto, CA, 1978.

Fitzgerald, M., Millard, C. and MacIntosh, N., Hyperalgesia in premature infants, Lancet, 1 (1988) 292.

Gottfried, A.W. and Gaiter, J.L. (Eds.), Infant Stress Under Intensive Care: Environmental Neonatology, University Park Press, Baltimore, MD, 1985.

Grunau, R.V.E. and Craig, K.D., Pain expression in neonates: facial action and cry, Pain, 28 (1987) 395–410.

Grunau, R.V.E. and Craig, K.D., Facial activity as a measure of neonatal pain expression. In: D.C. Tyler and E.J. Krane (Eds.), Advances in Pain Research and Therapy, Vol. 15, Raven Press, New York, 1990, pp. 147–155.

Grunau, R.V.E., Johnston, C.C. and Craig, K.D., Neonatal facial and cry responses to invasive and non-invasive procedures, Pain, 42 (1990) 295–305.

Grunau, R.V.E., Whitfield, M.F. and Petrie, J.H., Pain sensitivity and temperament in extremely low-birth-weight premature toddlers and preterm and full-term controls, Pain, 58 (1994) 341–346.

Grunau, R.E., Whitfield, M.F. and Petrie, J.H., Children’s judgments about pain at age 8–10 years: do extremely low birthweight (≤1000 g) children differ from full birthweight peers?, J. Child Psychol. Psychiatry, 39 (1998) 587–594.

Guinsburg, R., Berenguel, R.C., de Cassia Xavier, R., de Almeida, M.F.B. and Kopelman, B.I., Are behavioral scales suitable for preterm and term neonatal pain assessment? In: T.S. Jensen, J.A. Turner and Z. Wiesenfeld-Hallin (Eds.), Proceedings of the 8th World Congress on Pain, Progress in Pain Research and Management, Vol. 8, IASP Press, Seattle, WA, 1997, pp. 893–901.

Hadjistravropoulos, H.D., Craig, K.D., Grunau, R.V.E. and Whitfield, M.F., Judging pain in newborns: facial and cry determinants, J. Pediatr. Psychol., 19 (1994) 485–491.

Hile, M.G., Hand-held behavioral observations: the observer, Behav. Assess., 13 (1991) 187–196.

Johnston, C.C. and Stevens, B.J., Experience in a neonatal intensive care unit affects pain response, Pediatrics, 98 (1996) 925.

Johnston, C.C., Stevens, B., Craig, K.D. and Grunau, R.V.E., Developmental changes in pain expression in premature, full-term, two and four month old infants, Pain, 52 (1993) 201–208.

Johnston, C.C., Stevens, B., Yang, F. and Horton, L., Differential response to pain by very premature babies, Pain, 61 (1995) 471–479.

Johnston, C.C., Stevens, B., Yang, F. and Horton, L., Developmental changes response to heelstick in preterm infants: a prospective cohort study, Dev. Med. Child Neurol., 38 (1996) 438–445.

Lawrence, J., Alcock, D., McGrath, P., Kay, J., MacMurray, S.B. and

Dulberg, C., The development of a tool to assess neonatal pain, Neonatal Network, 12 (1993) 59–66.

Lewis, M., Worobey, J. and Thomas, D., Behavioral features of early reactivity: antecedents and consequences. In: M. Lewis and J. Worobey (Eds.), Infant Stress and Coping, Jossey-Bass, San Francisco, CA, 1989, pp. 33–46.

Lilley, C.M., Craig, K.D. and Grunau, R.E., The expression of pain in infants and toddlers: developmental changes in facial action, Pain, 72 (1997) 161–170.

McIntosh, N., Van Veen, L. and Brameyer, H., The pain of heel prick and its measurement in preterm infants, Pain, 52 (1993) 71–74.

Oberlander, T.F., Berde, C.B., Lam, K.H., Rappaport, L.A. and Saul, J.P., Infants tolerate spinal anaesthesia with minimal overall autonomic changes: analysis of heart rate variability in former premature infants undergoing hernia repair, Anesth. Analg., 80 (1995) 20–27.

Porter, F., Pain in the newborn, Clin. Perinatol., 16 (1989) 549–564.

Ramenghi, L.A., Wood, C.M., Griffith, G.C. and Levene, M.I., Reduction of pain response in premature infants using intraoral sucrose, Arch. Dis. Child., 74 (1996) F126–F128.

Rushforth, J.A. and Levene, M.I., Behavioral response to pain in healthy neonates, Arch. Dis. Child., 70 (1994) F174–F176.

Rushforth, J.A., Griffiths, G., Thorpe, H. and Levene, M.I., Can topical lidocaine reduce behavioural response to heelstick?, Arch. Dis. Child., 72 (1995) F49–F51.

Sackett, G.P., Measurement in observational research. In: G.P. Sackett (Ed.), Observing Behavior, Vol. II: Data Collection and Analysis Methods, University Park Press, Baltimore, MD, 1978, pp. 25–43.

Scott, C.S., Riggs, K.W., Ling, E., Grunau, R.V.E., Craig, K. and Solimano, A., Morphine pharmacokinetics and pain assessment in premature neonates, Pediatr. Res., 35 (1994) 254(A).

Stevens, B., Johnston, C.C. and Horton, L., Factors that influence the behavioral pain responses of premature infants, Pain, 59 (1994) 101–109.

Stevens, B.J., Johnston, C.C. and Grunau, R.V.E., Issues of assessment of pain and discomfort in neonates, J. Obstet. Gynecol. Neonatal Nurs., 24 (1995) 849–855.

Stevens, B., Johnston, C.C., Petryshen, P. and Taddio, A., The premature infant pain profile, Clin. J. Pain, 12 (1996) 13–22.

Taddio, A., Stevens, B., Craig, K., Rastogi, P., Ben-David, S., Shennan, A., Mulligan, P. and Koren, G., Efficacy and safety of lidocaine-prilocaine cream for pain during circumcision, N. Eng. J. Med., 336 (1997) 1197–1201.