The Clinical Journal of Pain
19:353–363 © 2003 Lippincott Williams & Wilkins, Inc., Philadelphia

Neonatal Facial Coding System for Assessing Postoperative Pain in Infants: Item Reduction is Valid and Feasible

*Jeroen W. B. Peters, PhD, †Hans M. Koot, PhD, ‡Ruth E. Grunau, PhD, §Josien de Boer, PhD, *Marieke J. van Druenen, MD, *Dick Tibboel, PhD, and ¶Hugo J. Duivenvoorden, PhD

**Department of Pediatric Surgery, Erasmus MC-Sophia, Rotterdam, The Netherlands; †Department of Developmental Psychology, Free University Amsterdam, Amsterdam, The Netherlands; ‡Centre for Community Child Health Research, British Columbia Research Institute for Children's and Women's Health, Vancouver, British Columbia, Canada; §Department of Child Health Division, TNO Prevention and Health, Leiden, The Netherlands; and ¶Department of Medical Psychology and Psychotherapy, NIHES, Erasmus MC, Rotterdam, The Netherlands*

Abstract:

Objective: The objectives of this study were to: (1) evaluate the validity of the Neonatal Facial Coding System (NFCS) for assessment of postoperative pain and (2) explore whether the number of NFCS facial actions could be reduced for assessing postoperative pain.

Design: Prospective, observational study.

Patients: Thirty-seven children (0–18 months old) undergoing major abdominal or thoracic surgery.

Outcome Measures: The outcome measures were the NFCS, COMFORT "behavior" scale, and a Visual Analog Scale (VAS), as well as heart rate, blood pressure, and catecholamine and morphine plasma concentrations. At 3-hour intervals during the first 24 hours after surgery, nurses recorded the children's heart rates and blood pressures and assigned COMFORT "behavior" and VAS scores. Simultaneously we videotaped the children's faces for NFCS coding. Plasma concentrations of catecholamine, morphine, and its metabolite M6G were determined just after surgery, and at 6, 12, and 24 hours postoperatively.

Results: All 10 NFCS items were combined into a single index of pain. This index was significantly associated with COMFORT "behavior" and VAS scores, and with heart rate and blood pressure, but not with catecholamine, morphine, or M6G plasma concentrations. Multidimensional scaling revealed that brow bulge, eye squeeze, nasolabial furrow, horizontal mouth stretch, and taut tongue could be combined into a reduced measure of pain. The remaining items were not interrelated. This reduced NFCS measure was also significantly associated with COMFORT "behavior" and VAS scores, and with heart rate and blood pressure, but not with the catecholamine, morphine, or M6G plasma concentrations.

Conclusion: This study demonstrates that the NFCS is a reliable, feasible, and valid tool for assessing postoperative pain. The reduction of the NFCS to 5 items increases the specificity for pain assessment without reducing the sensitivity and validity for detecting changes in pain.

Key Words: postoperative pain, infant newborn, pain measurement, facial action

Pain assessment during acute invasive procedures in infants has been extensively studied. In contrast, measuring postoperative pain or pain associated with prolonged neonatal diseases, such as necrotizing enterocolitis, meningitis, or osteomyelitis, has received little attention. The manifestation of observable behaviors

Received February 8, 2002; revised August 7, 2002; accepted August 14, 2002.

This study was supported by a research grant from NWO (Dutch Organisation for Scientific Research, 940-31-031).

Corresponding author: Jeroen Peters, Erasmus MC-Sophia, Department of Pediatric Surgery, P.O. Box 2060, 3000CB Rotterdam, The Netherlands. E-mail: j.peters@erasmusmc.nl.

may not be obvious postoperatively, or behavioral reactions may be diminished[1,2] or not continuously present.[3] Moreover, sedatives may mask behavioral expression of pain.

Of all the nonverbal acute pain indicators, facial expression appears to be the most prominent. Not only do facial expressions differentiate pain from anger and sadness,[4,5] for caregivers and nurses they are also more consistent and salient than cry.[6–8] Additionally, facial expression is more sensitive to noxious procedures than cry, body movements, or heart rate.[9,10] The importance of the face has been acknowledged in all multidimensional pain instruments. While there are few studies of facial response during postoperative pain, descriptions range from full "cry face" expression[11] to grimacing,[12,13] or a withdrawn or disinterested face with quivering chin or clenched jaws.[14]

The Neonatal Facial Coding System (NFCS)[15,16] has been extensively used to assess acute pain. This is an anatomically based measure in which the occurrences of 10 different facial actions are individually coded. It has been validated for use in premature neonates,[17–20] term-born neonates,[15,21] and infants up to 18 months of age.[22] This coding system is sensitive to acute pain, as it differentiates between noxious (eg, heel stick) and nonnoxious stimuli (eg, heel swab)[17,23,24] and also between infants receiving sucrose or morphine compared with controls during invasive procedures.[25–28] Its clinical value for assessing pain postoperatively has not yet been established.

Although the NFCS consists of 10 facial actions, some investigators suggested that the 3 most commonly observed facial actions (brow bulge, eye squeeze, and nasolabial furrow) suffice for pain assessment.[1,24,29,30] However, facial actions that are not universally observed may still provide important information about individual differences in pain expression.[23] Moreover, they may help in identifying subacute, chronic, or postoperative pain. The finding that tongue protrusion is associated with acute pain in preterm neonates ≤32 weeks of gestation[23] but not in term-born infants[21] emphasises the importance of evaluating the facial actions across development and in diverse situations.

First, we wished to evaluate the validity of the NFCS for assessing pain in the postoperative period. To this end we compared the NFCS with a modified version of the COMFORT scale (COMFORT "behavior" scale),[12,31] the Visual Analog Scale (VAS), and with several physiological measures (heart rate, blood pressure, catecholamine, and morphine plasma concentrations). Second, we wished to identify how many of the items included in the NFCS will suffice for valid assessment of postoperative pain in infants.

METHODS

Participants

This prospective, observational study was part of a larger double-blinded randomised clinical trial assessing the efficacy of continuous morphine administration and intermittent morphine administration after major surgery in neonates and young infants up to 3 years of age.[32]

For this study we included neonates aged 0–4 weeks with gestational age ≥35 weeks and birth weight ≥1500*g*, and infants up to 18 months, admitted to the ICU after thoracic or abdominal surgery. Exclusion criteria were (1) preoperative morphine or other opioids, (2) sedative drugs or muscle relaxants, (3) hepatic or renal dysfunction, (4) neurologic damage, (5) altered muscle tone, or (6) preoperative anemia (hematocrit less than 30%). The hospital's medical-ethical committee approved the study, and parental informed consent was obtained.

Measures

NFCS

The Neonatal Facial Coding System (NFCS) comprises 10 discrete facial actions: brow bulge, eye squeeze, nasolabial furrow, open lips, horizontal mouth stretch, vertical mouth stretch, taut tongue, lip purse, chin quiver, and tongue protrusion.[15,16] To compare the NFCS with the other measures, the total duration of occurrence of all 10 facial actions was used (NFCS total).

COMFORT

The COMFORT "behavior" scale[12] is an adaptation of the original COMFORT scale.[31] The adapted version leaves out the physiological items "heart rate" (HR) and "mean arterial blood pressure" (BP) because in the postoperative pain situation they do not add information to that obtained from the behavioral items of the original COMFORT scale.[12] The item "Crying" was added for spontaneously breathing infants. The adapted COMFORT scale thus comprises the behavioral items alertness, calmness, muscle tone, physical movement, facial tension, and respiratory behavior/crying. Respiratory behavior is scored only in ventilated patients and Crying in nonventilated patients. The 6 items are scored on a 5-point scale, ranging from 1 to 5, with total score ranging from 6 to 30. The COMFORT "behavior" scale has been shown to be reliable and valid for the postoperative pain situation.[12]

VAS

The Visual Analog Scale (VAS) is a horizontal continuous 10-cm line with the anchors "no pain" at the left side and "extreme pain" at the right side. The VAS was scored by the nurses at the bedside. Congruent validity of nurses using a VAS has been demonstrated.[33,34]

Physiological and hormonal indices

Heart rate (HR) and blood pressure (BP) were measured with the Hewlett Packard Component Monitoring System (Böblingen, Germany). The raw data were used to calculate the level (mean) and variability (standard deviation) of HR and BP (HR mean, BP mean, HR var, and BP var, respectively). Concurrent validity of these physiological indicators has been established with the COMFORT "behavior" scale.[35]

Catecholamine responses

Catecholamine plasma concentrations of adrenaline and noradrenaline were measured by HPLC using the fluorometric assay method.[32]

Morphine

Plasma concentrations of morphine and its metabolite M6G were assessed using the fluorometric immunoassay method.

PROCEDURES

General

Anesthesia was standardized for all patients. After the induction of anesthesia, an arterial line was placed from which blood samples were drawn for catecholamine and morphine plasma concentration determination. Random postoperative analgesia consisted of either a continuous morphine infusion (10 $\mu g\ kg\ h^{-1}$) or intermittent morphine administration (30 $\mu g\ kg^{-1}$) at 3-hour intervals. If a child was judged to suffer moderate to severe pain (VAS $\geq$4), additional morphine was administered in bolus doses of 10 $\mu g\ kg^{-1}$. Endotracheally intubated patients had small plasters at the upper lip and at the upper part of the forehead (not covering the corrugator supercilii or the pyramidalis nasi) to fixate the tube.

Data collection

Each child was observed by a nurse during nine 120-second periods, the first on admission to the pediatric ICU and the remaining observations at 3-hour intervals up to 24 hours postoperatively. After each observation period the nurse assigned a COMFORT "behavior" and a VAS scores. During each observation the nurse registered HR and BP every 20 seconds. Video recordings were simultaneously carried out, providing a full view of the child's face. Arterial blood samples for catecholamine and morphine plasma concentrations were taken after surgery and at 6, 12, and 24 hours postoperatively.

Video recording and NFCS coding

Video recordings were carried out with a VHS color camera (Hitachi VM-S7200E) and videocassette recorder (Panasonic AG 5700). Each frame of the video recording was marked with a Vertical Interval Time Code (Adrienne Electronics Corporation, Las Vegas, NV), which allowed for coding of specific sections of the videotape. The Observer 3.0 Base Package for Windows[36] and the Observer 4.0 Software Package for video analysis[37] were used for video coding, which include remote control, stop action, and slow motion playback.

NFCS coding was carried out by 2 coders (JWBP and MJvD), each conducting about half of the sessions. Coder JWBP was trained before by REG. For each observation period separate video playbacks were used to code each facial action for presence or absence. The Observer system[36,37] was used in which computer keyboard entries marked the onset and offset of each facial movement. This enabled calculation of the total duration of occurrence of each facial action during 120 seconds. NFCS coders were blinded to the COMFORT "behavior" and VAS scores and the physiological parameters.

Reliability assessment

NFCS reliability coding was carried out on 10 randomly selected video recordings. Inter-observer reliability was calculated for each facial action using the conservative Facial Action Coding System formula.[15,38]

Reliability for the COMFORT "behavior" scale was assessed when nurses were trained to use this scale. After training, each nurse completed at bedside 10 COMFORT assessments with an experienced colleague. Inter-observer reliability was assessed by linearly weighted Cohen's κ.

Data analysis

Associations between NFCS, COMFORT "behavior", and VAS scores on the one hand and the physiological and hormonal indices and morphine plasma concentrations on the other hand were determined by Random Regression Modeling (RRM) for continuous data, using the SAS 6.12 program for Windows. RRM is a highly flexible approach for repeated measurements because it models changes across time at both group and individual levels. In contrast with MANOVA for repeated measures, RRM is not restricted to modeling time as a fixed effect as both the number of measurements across time and the moments of measurement may vary. RRM can handle missing data with the constraint that these are assumed to be at random. Compound symmetry is not required because the variance and covariance structure across time may also have other kinds of error structure. In this study the error is assumed to be unstructured, which allows variances and correlations to diverge across time. RRM estimates the longitudinal trend (of any kind) for every child on the basis of individual data, augmented by time trend. The following clinical variables were entered in the regression model as covariates: time, morphine administration condition (CM and IM), age, sex, and mechanical ventilation.

With respect to the second aim of this study, namely, determining the minimum number of NFCS facial actions required for valid assessment of postoperative pain, we carried out a 3-step procedure. First, we assessed the complexity of the structure of facial actions using Multidimensional Scaling with SPSS 9.0 for Windows. Distances ≥1.0 between stimuli coordinates were regarded as substantial. For the purpose of this analysis each observational event of 120 seconds was divided into 24 equally spaced intervals of 5 seconds in which the coder identified the occurrence of the facial actions. Intervals of 5 seconds were considered suitable for the postoperative or chronic pain situation. An S-stress value of 0.15 was regarded as good fit.

Secondly, on the basis of the outcome of multidimensional scaling, RRM was carried out again with the same covariates as in the previous RRM analyses. Stepwise backward elimination (one by one; *P* out >0.10) served to obtain only the relevant facial variables. Conversely, none of covariates was removed from the model.

Thirdly, comparisons were made between the models found in step 2 and those with NFCS total. Three measures of model fit were used for model comparisons, ie, −2 Residual Log Likelihood ratio, Akaike's information criterion (AIC),[39] and Schwarz's Information Criterion (SIC).[40] Likelihood ratio tests were used to assess whether item reduction would lead to worse model fits, as reflected by −2 Residual Log Likelihood ratios.[41] With respect to AIC and SIC, larger values indicate a better fit.

To achieve normal distributions, transformations (natural logarithm) were carried out on VAS, HR var, BP var, adrenaline, noradrenaline, and morphine plus M6G. For the sake of comparability and interpretability of the different regression coefficients, all independent variables were transformed into *z* scores. All significance testing was at the 0.05 level (2-sided).

RESULTS

Subject characteristics

The background characteristics of the 37 children are presented in Table 1. Twenty-eight (76%) were younger than 6 months, and 27 (73%) had undergone an abdominal operation. Postoperative mechanical ventilation was necessary in 16 (43%) children for at least 24 hours and in 3 (8%) children for about 12–24 hours. Sixty video recordings were missing: due to night time (n = 39), admission at the ICU (n = 9), and not codeable (n = 12).

TABLE 1. *Background characteristics of the children (N = 37)*

	n	%
Age group		
Neonates (0–4 weeks)	11	30
Young infants (4 weeks–6 months)	17	46
Infants (6–18 months)	9	24
Sex		
Male	19	51
Female	18	49
Site of surgery		
Abdominal high/low	27	73
Thoracic (with or without abdominal)	9	24
Superficial	1	3
Surgical procedures		
Stoma closure	9	24
Miscellaneous (diaphragmatic paresis, tumour, teratoma cyst, choledochal atresia, Hirschsprung's disease)	6	16
Tracheo-oesophagal atresia	4	11
Lobectomy	4	11
Septic gastro-intestinal (intussusception, perforation, ileus)	4	11
Intestinal atresia/malrotation	3	8
Congenital diaphragmatic hernia	3	8
Gastroschisis	1	3
Nephrectomy	1	3
Nissen fundoplication	1	3
Colon interposition	1	3
Postoperative mechanical ventilation		
None	18	49
12–24 hours	3	8
>24 hours	16	43

Reliability of the behavioral scales

Inter-observer agreement for the NFCS ranged from 0.84 for vertical mouth stretch to 1.0 for open lips. Cohen's κ for the items on the COMFORT "behavior" scale ranged between 0.54 for respiratory response and 0.74 for alertness, which can be considered satisfactory.[12]

Association between the NFCS and other pain indices

RRM showed that NFCS total was significantly associated with VAS ($P < 0.001$) and COMFORT "behavior" scores ($P < 0.001$) and also with HR mean ($P < 0.001$), BP mean ($P < 0.001$), HR var ($P < 0.001$), and BP var ($P = 0.007$). No significant associations were found between the NFCS total scores on the one hand and adrenaline, noradrenaline, morphine, or M6G plasma concentrations on the other hand.

The standardised regression coefficients in Table 2 show that the prediction of NFCS total was highest for COMFORT "behavior". This degree of predictability is somewhat better than for HR mean or VAS, ie, 1.4 and 1.6 times, respectively, and is much better than for BP var, ie, 3.2 times.

TABLE 2. *Overview of NFCS total associations*

	NFCS total β (SE)	NFCS total β (SE)
	LN (VAS)	COMFORT
Time	−0.25 (0.06)***	−0.03 (0.07)
Morphine	−0.07 (0.09)	−0.02 (0.08)
Age (1)	0.17 (0.12)	0.09 (0.12)
Age (2)	0.22 (0.11)*	0.13 (0.10)
Sex	−0.10 (0.09)	−0.05 (0.08)
Mech vent	0.21 (0.10)¶	0.18 (0.10)¶
NFCS	0.31 (0.05)***	0.48 (0.05)***
	HR mean	BP mean
Time	0.00 (0.07)	0.07 (0.05)
Morphine	−0.07 (0.12)	0.00 (0.09)
Age (1)	−0.36 (0.17)*	−0.53 (0.13)***
Age (2)	−0.27 (0.15)¶	−0.03 (0.11)
Sex	−0.20 (0.12)¶	−0.14 (0.09)
Mech vent	−0.08 (0.12)	0.27 (0.09)**
NFCS	0.34 (0.04)***	0.23 (0.03)***
	LN (HR var)	LN (BP var)
Time	0.02 (0.06)	−0.02 (0.08)
Morphine	−0.01 (0.07)	0.01 (0.10)
Age (1)	0.01 (0.10)	0.00 (0.15)
Age (2)	0.20 (0.09)*	0.16 (0.13)
Sex	−0.07 (0.07)	−0.11 (0.10)
Mech vent	0.27 (0.09)**	0.24 (0.12)*
NFCS	0.26 (0.06)***	0.15 (0.05)**

Note. This table presents all standardized regression coefficients and standard errors [β (SE)] from each RRM model in which NFCS total was entered.

Abbreviations: LN (VAS), logarithmic transformed VAS; LN (HR var), logarithmic transformed HR var; LN (BP var), logarithmic transformed BP var; Time, time postoperative; Morphine, continuous or intermittent morphine; Age (1), 0–4 weeks; Age (2), 4 weeks–6 months; Mech vent, spontaneous breathing or mechanical ventilation; ¶$P \le 0.10$; *$P \le 0.05$; **$P < 0.01$; ***$P < 0.001$.

The clinical relevance of these associations is depicted in Figure 1. This figure presents unstandardized coefficients as well as all co-variates significantly related to the outcome measures. When the NFCS total increased from minimum to maximum activity score (ie, maximum NFCS total score was 600), the increase in VAS scores ranged between 1.8 and 3.0. COMFORT "behavior" scores increased 11 points, HR mean 37 beats, and BP mean 20 mm Hg. The increase in HR var ranged between 0.6 and 0.7, and the increase in BP var ranged between 2.7 and 3.6.

Figure 1 further indicates that VAS scores in the early postoperative period were higher than those scores 24 hours after surgery ($P < 0.001$). VAS scores were also higher in young infants than in neonates or infants ($P = 0.044$). BP mean in neonates was lower than that in (young) infants ($P < 0.001$) and higher in mechanically ventilated children compared with the spontaneous breathing ones ($P = 0.005$). Mechanically ventilated children had also higher BP var and HR var values than the spontaneous breathing ones ($P = 0.045$ and 0.002). With respect to HR var, these values were also higher in young infants than in neonates and infants ($P = 0.021$).

Empirical complexity of the NFCS

Multidimensional Scaling revealed a 4-dimensional configuration. Five of the 10 facial actions appeared to cluster homogeneously and could consequently be grouped together; these are brow bulge, eye squeeze, nasolabial furrow, horizontal mouth stretch, and taut tongue. These 5 substantially interrelated items were combined into a single index (NFCS subset). The other items were not interrelated (see Table 3). The NFCS subset score was derived from summating the duration of presence of the observed 5 facial actions. As the other items turned out to be independent, they were analyzed as individual variables.

Association between item-reduced NFCS and other pain indices

The NFCS subset and the other 5 independent facial variables were entered into RRM. The findings were similar to those in the previous section (see Table 4). Of these variables, only the NFCS subset significantly estimated VAS ($P < 0.001$), COMFORT "behavior" ($P < 0.001$), HR mean ($P < 0.001$), BP mean ($P < 0.001$), HR var ($P < 0.001$), and BP var ($P = 0.006$), respectively. With respect to the 5 independent facial actions, open lips were related to VAS ($P = 0.013$) and to HR mean ($P = 0.007$). Vertical mouth stretch was related to BP mean ($P = 0.050$). In contrast, chin quiver was negatively related to COMFORT "behavior" ($P = 0.022$) (see Table 4). Adrenaline, noradrenaline, morphine, and M6G plasma concentrations were not related to either the NFCS subset or any of the independent NFCS items.

Because only the NFCS subset provided results similar to those of the NFCS total, the other 5 independent items were removed from the model. Likewise, RRM demonstrated that the predictability of NFCS subset was again the highest for COMFORT "behavior". The degree of predictability was better than that for HR mean or VAS, ie, 1.4 and 1.6 times, respectively, and was much better than for BP var, ie, 3.4 times (see Table 4).

The clinical relevance of these findings is depicted in Figure 2. It demonstrates that the association between NFCS subset, COMFORT "behavior", and VAS, HR mean, BP mean, HR var, and BP var equalled that of the NFCS total. For example, when facial activity increased from minimum to maximum activity (ie, maximum NFCS subset score was 500), the predicted increase in VAS scores ranged between 2.5 and 3.7. The COMFORT "behavior" scores increased 12 points, HR mean 29 beats, and BP mean 19 mm Hg. The increase in

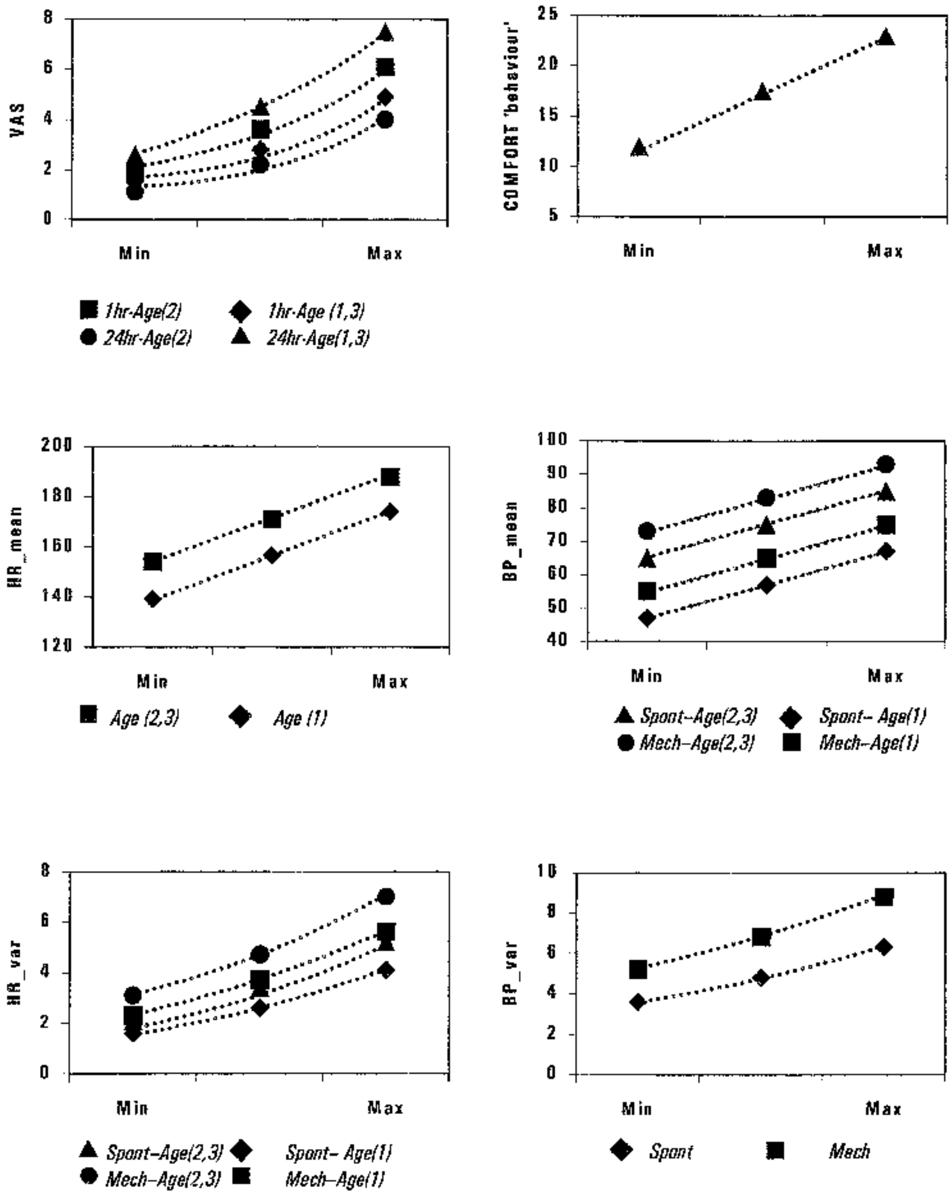

FIGURE 1. Relation between NFCS and other pain indices. Only significant estimates (ie, NFCS total and covariates) are included. Min, NFCS subset score equals 0; Max, NFCS subset score equals 600; Spont, spontaneous breathing; Mech, mechanical ventilation; 1 hr, 1 hour after surgery; 24 hr, 24 hours after surgery; Age (1), 0–4 weeks; Age (2), 4 weeks to 6 months; Age (3), 6–18 months.

HR var ranged between 3.0 and 3.7; the increase in BP var was 3.0.

In addition, VAS scores in the early postoperative period were higher than those 24 hours after surgery ($P < 0.001$). VAS scores in young infants also exceeded those in neonates and infants ($P = 0.049$). HR mean was lowest in neonates compared with (young) infants ($P = 0.013$). HR var in young infants was higher than that in neonates or infants ($P = 0.026$) and in mechanically ventilated children ($P = 0.014$). BP mean was lowest in neonates and highest in mechanically ventilated children than in spontaneous breathing children ($P < 0.001$ and $P = 0.013$, respectively).

Performance of item-reduced NFCS

Comparison between the standardised regression coefficients of Tables 2 and 4 demonstrated that the predictability of the NFCS subset coefficients equalled that of the NFCS total. This was supported by the likelihood ratio tests, demonstrating no significant differences between the –2 Residual Log Likelihood ratios. AIC and SIC also showed that there were no significant differences between

TABLE 3. *Coordinates of each facial action for the four dimensions*

	Dimensions			
	1	2	3	4
Brow bulge	−0.14	−0.03	−0.03	0.03
Eye squeeze	0.03	−0.01	0.05	−0.12
Nasolabial furrow	−0.08	0.04	0.13	0.05
Open lips	−1.84	0.57	−0.35	−0.78
Horizontal mouth stretch	0.07	0.10	0.27	−0.12
Vertical mouth stretch	−0.06	−0.20	0.14	2.46
Taut tongue	0.13	−0.10	0.25	0.15
Tongue protrusion	−2.92	2.43	0.28	0.33
Lip purse	−2.87	−0.81	−2.07	0.31
Chin quiver	−3.27	−1.46	1.74	0.29

Note. Outcome of multidimensional scaling; distances ≥1.0 between the stimuli coordinates are defined as substantial.

these values of model fit (see Table 5). These findings imply that reducing the NFCS to the content of the NFCS subset did not induce substantial loss of information.

DISCUSSION

The NFCS was associated with other behavioral and physiological indicators of pain in neonates and infants, indicating that it is sensitive to changes in postoperative pain. In addition, the original NFCS can be reduced to a subset of 5 core facial actions, ie, brow bulge, eye squeeze, nasolabial furrow, horizontal mouth stretch, and taut tongue, without unfavourable consequences for its sensitivity.

Psychometric qualities of the (item-reduced) NFCS

Our findings demonstrate that the NFCS has good psychometric properties in the postoperative pain situation. Good inter-rater reliability was obtained easily and was as high for postoperative pain as demonstrated for acute pain.[15,23]

Furthermore, we established content validity of the NFCS for postoperative pain assessment, especially of the reduced version, as brow bulge, eye squeeze, nasolabial furrow, horizontal mouth stretch, and taut tongue together form a facial display of pain. This finding is supported by many commonalities between this configuration and those facial actions widely accepted as indices of acute pain in neonates and infants,[15,38,42] of postoperative pain in children aged 1–6 years,[43] and of acute or chronic pain in adults.[44–46] These 5 core facial actions carry the bulk of information and are robust across postoperative and acute pain situations in infancy.[9,17,18,38]

Furthermore, our findings suggest that the other 5 facial actions do not supply extra information either because of low specificity or sensitivity, or both. Open lips were not interrelated with any of the 5 core facial actions and was only associated with VAS and HR mean. In contrast, in studies of acute pain, open lips were part of the core facial actions. However, these studies only used data obtained during pain and not during baseline or the recovery period. This procedure carries the risk that facial actions with low specificity, such as open lips,[29,43] are not excluded from the NFCS. In our study, therefore, all observational periods were used to delete those facial actions that lack sensitivity and specificity.

With respect to the other independent facial actions, chin quiver was found to be negatively related with increasing COMFORT "behavior" scores, suggesting no sensitivity or specificity for postoperative pain assessment. We also found that lip purse was not related to any of the postoperative pain-related indicators. Consistent with previous studies,[17,21,23,38] we dropped lip purse because of low occurrence. Tongue protrusion was also not related to any of the postoperative pain-related indices. This facial action, however, has been observed as pain related in premature infants only.[23] In term-born infants, it has not been observed during pain.[21]

The role of vertical mouth stretch is not clear yet. We omitted this facial action from the NFCS as there was no relation with the 5 core facial actions or any of the postoperative pain-related indices, except BP mean. In (premature) neonates this facial action has been associated with acute pain, although its sensitivity is small compared with that of the upper facial actions[15,17,18,23,29,38] and similar to that in healthy term-born neonates.[15,21] For infants aged 2–28 months, horizontal mouth stretch as well as vertical mouth stretch were eliminated from the NFCS because of infrequent occurrence.[22] While some studies in adults found vertical stretching of the mouth during pain,[44,45] others did not.[46,47] It has been suggested that vertical mouth stretch is more likely to occur during severe pain.[42,47,48] In our study, however, severe pain was observed only infrequently, probably because all children received morphine.

For the above reasons we concluded that open lips, chin quiver, lip purse, tongue protrusion, and vertical mouth stretch can be deleted from the NFCS when assessing acute or postoperative pain in infants with birth weight above 1500*g,* up to 18 months. However, for infants born at lower birth weight undergoing neonatal intensive care, specific NFCS items remain to be evaluated.

Association of the item-reduced NFCS with other pain-related indices

Congruent validity is another requisite of instrument development. We demonstrated validity, showing that the NFCS scores were associated with the behavioral and

TABLE 4. *Overview of associations of the NFCS_subset with and without the 5 independent NFCS items*

	NFCS_subset and independent NFCS items β (SE)	NFCS_subset β (SE)	NFCS_subset and independent NFCS items β (SE)	NFCS_subset β (SE)
	LN(VAS)		COMFORT	
Time	−0.28 (0.05)***	−0.27 (0.06)***	−0.06 (0.06)	−0.05 (0.07)
Morphine	−0.07 (0.09)	−0.10 (0.08)	−0.07 (0.07)	−0.07 (0.07)
Age (1)	0.18 (0.13)	0.13 (0.12)	0.02 (0.11)	0.01 (0.11)
Age (2)	0.25 (0.11)*	0.21 (0.10)*	0.14 (0.09)	0.11 (0.09)
Sex	−0.11 (0.09)	−0.09 (0.08)	−0.05 (0.07)	−0.04 (0.07)
Mech vent	0.20 (0.10)¶	0.16 (0.10)	0.07 (0.09)	0.08 (0.09)
NFCS	0.25 (0.05)***	0.29 (0.05)***	0.46 (0.05)***	0.48 (0.05)***
OL	0.14 (0.06)*		—	
VMS	—		—	
TP	—		—	
CQ	−0.10 (0.06)¶		−0.12 (0.05)*	
LP	—		—	
	HR_mean		BP_mean	
Time	−0.01 (0.06)	−0.02 (0.06)	0.06 (0.05)¶	0.05 (0.05)
Morphine	−0.07 (0.12)	−0.11 (0.12)	−0.02 (0.09)	0.03 (0.09)
Age (1)	−0.36 (0.17)*	−0.41 (0.17)*	−0.57 (0.12)***	−0.57 (0.12)***
Age (2)	−0.25 (0.15)¶	−0.27 (0.15)¶	−0.05 (0.11)	−0.05 (0.11)
Sex	−0.20 (0.12)¶	−0.20 (0.12)¶	−0.14 (0.09)	−0.13 (0.09)
Mech vent	−0.09 (0.12)	−0.14 (0.12)	0.23 (0.09)*	0.23 (0.09)*
NFCS	0.29 (0.04)***	0.32 (0.04)***	0.20 (0.04)***	0.23 (0.03)***
OL	0.15 (0.05)**		—	
VMS	—		0.07 (0.04)*	
TP	—		—	
CQ	—		—	
LP	—		—	
	LN(HR_var)		LN(BP_var)	
Time	0.00 (0.06)	0.00 (0.06)	−0.02 (0.08)	−0.02 (0.08)
Morphine	−0.03 (0.07)	−0.03 (0.07)	0.00 (0.10)	−0.01 (0.10)
Age (1)	−0.04 (0.10)	−0.04 (0.10)	−0.04 (0.15)	−0.03 (0.15)
Age (2)	0.19 (0.08)*	0.19 (0.08)*	0.14 (0.13)	0.16 (0.13)
Sex	−0.06 (0.07)	−0.06 (0.07)	−0.10 (0.10)	−0.10 (0.10)
Mech vent	0.21 (0.08)*	0.21 (0.08)*	0.20 (0.12)¶	0.21 (0.12)¶
NFCS	0.27 (0.06)***	0.27 (0.06)***	0.14 (0.05)**	0.14 (0.05)**
OL	—		—	
VMS	—		—	
TP	—		0.11 (0.06)¶	
CQ	—		—	
LP	—		—	

Note. The columns NFCS subset and independent NFCS items present the standardized regression coefficients and standard errors [β (SE)] provided after backward elimination (*P* out 0.10). The columns NFCS subset present the β (SE) from each RRM model in which only NFCS subset was entered.

Abbreviations: LN (VAS), logarithmic transformed VAS; LN (HR var), logarithmic transformed HR var; LN (BP var), logarithmic transformed BP var; OL, open lips; VMS, vertical mouth stretch; TP, tongue protrusion; CQ, chin quiver; LP, lip purse; Time, time postoperative; Morphine, continuous or intermittent morphine; Age (1), 0–4 weeks; Age (2), 4 weeks–6 months; Mech vent, spontaneous breathing or mechanical ventilation; ¶$P \leq 0.10$; *$P \leq 0.05$; **$P < 0.01$; ***$P < 0.001$.

physiological (ie, heart rate and blood pressure) indices postoperatively. Reducing the NFCS to 5 items did not affect any of these associations, suggesting that the congruent validity of the reduced version equals that of the complete NFCS.

As both the NFCS and the COMFORT "behavior" scale measure facial behavior, we expected that they would correlate more highly with each other than with nonfacial measures. This expectation was confirmed. We also expected that the association between NFCS and VAS would equal that between NFCS and COMFORT "behavior", because the COMFORT "behavior" and VAS previously proved to be highly interrelated ($R \geq 0.89$).[12] Moreover, nurses have reported that they rely heavily on facial action when estimating pain.[8] In our study, however, the association of NFCS with VAS was lower than that with COMFORT "behavior". This suggests that nurses did not rely mainly on facial activity but rather on body movements, possibly because they had been using the COMFORT scale.

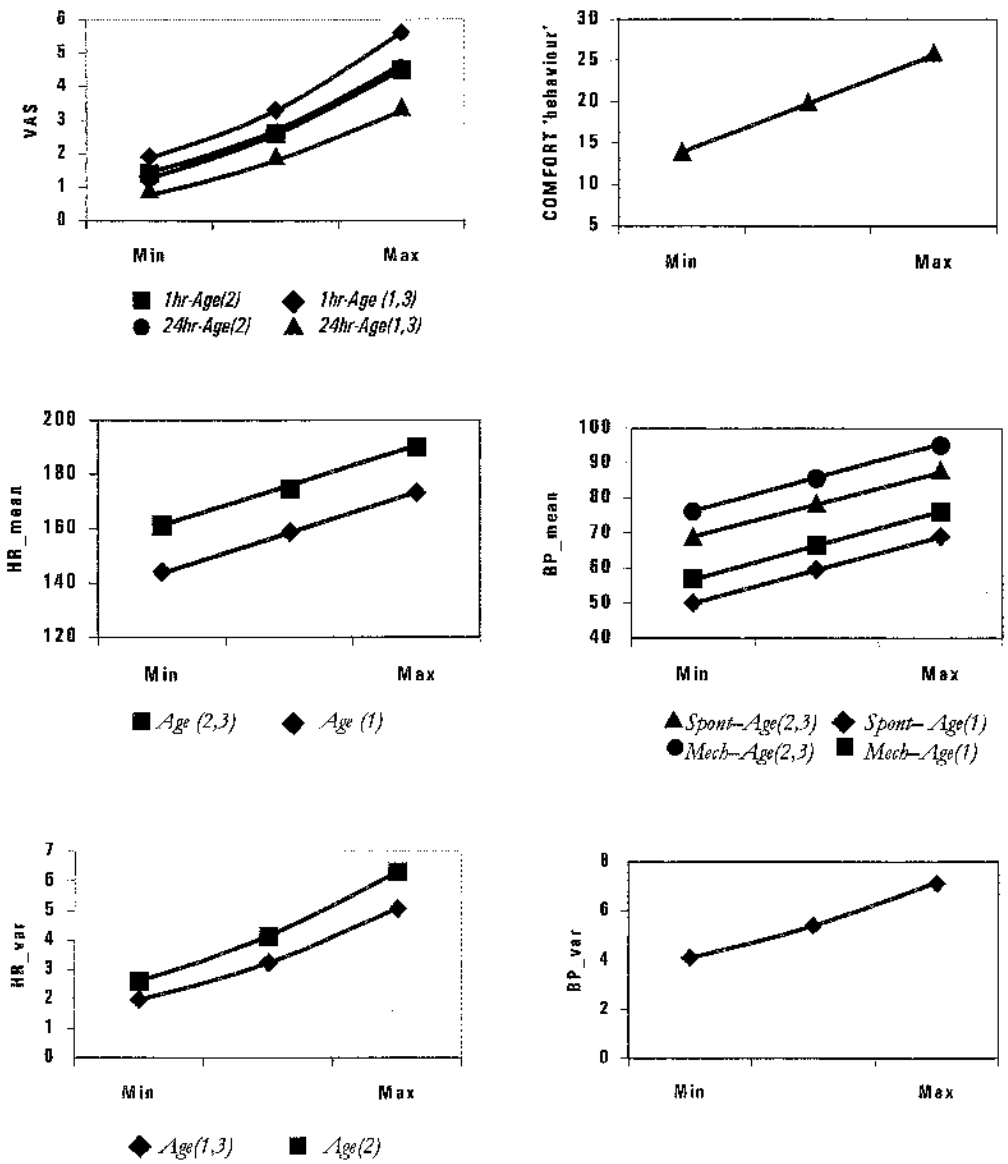

FIGURE 2. Relation between item-reduced NFCS and other pain indices. Only significant estimates (ie, NFCS subset and covariates) are included. Min, NFCS subset score equals 0; Max, NFCS subset score equals 500; Spont, spontaneous breathing; Mech, mechanical ventilation; 1 hr, 1 hour after surgery; 24 hr, 24 hours after surgery; Age (1), 0–4 weeks; Age (2), 4 weeks to 6 months; Age (3), 6–18 months.

In addition, we also found that, independent of the amount of facial activity, VAS scores decreased over time and were lower in mechanically ventilated children as compared with spontaneously breathing children. These findings suggest that apart from overt behavior, nurses also take into account additional information when assigning VAS scores.[49–51] Further research is needed to examine which indicators they attend to under various conditions.

In this study the association of the NFCS with heart rate or blood pressure was lower than that between NFCS and COMFORT "behavior". This was consistent with dissociations between behavioral and physiological reactivity found in other studies.[35,52] However, the strength of associations between behavioral pain indicators and heart rate or blood pressure increases with increasing pain scores.[35] Severe pain was uncommon in this study.

We found no association between facial responsiveness and catecholamine responses. The latter were thought to be good measures for evaluating congruent validity because they are blunted by appropriate analgesia during surgery[53,54] and in the ICU.[27] Our findings are consistent with other studies demonstrating no clear

TABLE 5. *Performance of facial activity*

	NFCS total				NFCS subset			
	−2 RLL	df	AIK	SIC	−2 RLL	df	AIK	SIC
LN (VAS)	573	7	−290	−297	576	7	−292	−299
COMFORT "behavior"	559	7	−283	−290	551	7	−279	−286
HR mean	538	7	−273	−280	544	7	−276	−282
BP mean	376	7	−192	−199	373	7	−190	−197
LN (HR var)	572	7	−290	−297	572	7	−290	−296
LN (BP var)	624	7	−316	−323	623	7	−315	−322

Note. As a measure of performance, ie, −2 Residual Log Likelihood ratio (−2 RLL), Akaike's Information Criterion (AIC), and Schwarz's Bayesian Criterion (SIC) were used.

Abbreviations: LN (VAS), logarithmic transformed VAS; LN (HR var), logarithmic transformed HR var; LN (BP var), logarithmic transformed BP var.

association between behavioral reactivity and stress hormones.[32,55–57] This discrepancy can be explained by the fact that catecholamine responses are only sensitive for severe pain[27,53,54] but not for minor pain following inoculation[55,56] or when children are given pain relief.[32,57] Catecholamine responses, therefore, appear to be useful only in cases of severe pain. Even then, it remains necessary to control for aspects other than pain.[58,59]

Clinical implications of the item-reduced NFCS

An important clinical question concerns the added value of the NFCS compared with multidimensional postoperative pain measures. The NFCS facial actions are well defined, in contrast to the global and sometimes vague descriptions of pain-related behaviors in most multidimensional pain assessment instruments. Secondly, the NFCS has greater specificity than multidimensional pain measures, as the facial configuration of pain can be distinguished from emotional states.[4] For example, when anger or sadness is apparent, NFCS scores will be lower as the pain face includes different facial actions than facial configurations of anger or sadness. Multidimensional pain measures appear to lack this ability to differentiate. Thirdly, in the postoperative situation the NFCS appears to be more sensitive to morphine than COMFORT or VAS[60] and in the acute pain situation in comparison to the NIPS.[27]

In this study the facial actions were painstakingly coded using fine-grained video analysis afterwards. In others studies the NFCS has also been used in real time at bedside to assess acute pain,[20,23,27] with inter-observer reliability as high as that obtained from video analyses.[23] This finding combined with our findings suggest that the NFCS might also be used at bedside to assess postoperative pain. Moreover, the proposed item reduction would improve the utility of this scale.

In summary, this is the first study demonstrating that the facial actions that characterize acute pain are also valid indicators of postoperative pain. Content and congruent validity for the postoperative period were established, as well as a first phase of construct validity. This study is consistent with previous findings, demonstrating good reliability and feasibility of the NFCS for the assessment of pain in infants and young children. Reducing the NFCS to 5 items appears to increase the specificity for pain assessment without reducing the sensitivity and validity for detecting changes in pain.

REFERENCES

1. Johnston CC, Stevens BJ. Experience in a neonatal intensive care unit affects pain response. *Pediatrics*. 1996;98:925–930.
2. Grunau RE, Oberlander TF, Whitfield MF, et al. Demographic and therapeutic determinants of pain reactivity in very low birth weight neonates at 32 weeks' postconceptional Age. *Pediatrics*. 2001;107: 105–112.
3. Beyer JE, McGrath PJ, Berde CB. Discordance between self-report and behavioral pain measures in children aged 3–7 years after surgery. *J Pain Symptom Manage*. 1990;5:350–356.
4. Izard CE, Hembree EA, Dougherty LM, et al. Changes in facial expressions of 2- to 19-month-old infants following acute pain. *Dev Psychol*. 1983;19:418–426.
5. Izard CE, Hembree EA, Huebner RR. Infants' emotion expressions to acute pain: developmental change and stability of individual differences. *Dev Psychol*. 1987;23:105–113.
6. Johnston CC, Strada ME. Acute pain response in infants: a multidimensional description. *Pain*. 1986;24:373–382.
7. Craig KD, Grunau RE, Aquan-Assee J. Judgement of pain in newborns: facial activity and cry as determinants. *Can J Behav Sci*. 1988;20:442–451.
8. Goodenough B, Addicoat L, Champion GD, et al. Pain in 4- to 6-year-old children receiving intramuscular injections: a comparison of the Faces Pain Scale with other self-report and behavioral measures. *Clin J Pain*. 1997;13:60–73.
9. Hadjistavropoulos HD, Craig KD, Grunau RV, et al. Judging pain in newborns: facial and cry determinants. *J Pediatr Psychol*. 1994; 19:485–491.
10. Hadjistavropoulos HD, Craig KD, Grunau RVE, et al. Judging pain in infants: behavioral, contextual, and developmental determinants. *Pain*. 1997;73:319–324.
11. Schade JG, Joyce BA, Gerkensmeyer J, et al. Comparison of three preverbal scales for postoperative pain assessment in a diverse pediatric sample. *J Pain Symptom Manage*. 1996;12:348–359.
12. van Dijk M, de Boer JB, Koot HM, et al. The reliability, stability and validity of the COMFORT scale as a postoperative pain instrument in 0- to 3-year-old infants. *Pain*. 2000;84:367–377.
13. Krechel SW, Bildner J. CRIES: a new neonatal postoperative pain measurement score. Initial testing of validity and reliability. *Paediatr Anaesth*. 1995;5:53–61.
14. Merkel SI, Voepel-Lewis T, Shayevitz JR, et al. The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatr Nurs*. 1997;23:293–297.
15. Grunau RVE, Craig KD. Pain expression in neonates: facial action and cry. *Pain*. 1987;28:395–410.
16. Grunau RVE, Craig KD. Facial activity as a measure of neonatal pain expression. In: Tyler DC, Krane EJ, eds. *Advances in Pain Research and Therapy*. Vol 15. New York: Raven Press; 1990: 147–155.
17. Craig KD, Whitfield MF, Grunau RVE, et al. Pain in the preterm neonate: behavioral and physiological indices. *Pain*. 1993;52:287–299.

18. Stevens BJ, Johnston CC, Horton L. Factors that influence the behavioral pain responses of premature infants. *Pain*. 1994;59: 101–109.
19. Johnston CC, Stevens BJ, Yang F, et al. Differential response to pain by very premature neonates. *Pain*. 1995;61:471–479.
20. Guinsburg R, de Araujo Peres C, Branco de Almeida MF, et al. Differences in pain expression between male and female newborn infants. *Pain*. 2000;85:127–133.
21. Grunau RVE, Johnston CC, Craig KD. Neonatal facial and cry responses to invasive and non-invasive procedures. *Pain*. 1990;42: 295–305.
22. Lilley CM, Craig KD, Grunau RVE. The expression of pain in infants and toddlers: developmental changes in facial action. *Pain*. 1997;72:161–170.
23. Grunau RE, Oberlander T, Holsti L, et al. Bedside application of the Neonatal Facial Coding System in pain assessment of premature neonates. *Pain*. 1998;76:277–286.
24. Johnston CC, Stevens B, Yang F, et al. Developmental changes in response to heelstick in preterm infants: a prospective cohort study. *Dev Med Child Neurol*. 1996;38:438–445.
25. Ramenghi LA, Evans DJ, Levene MI. Sucrose analgesia: absorptive mechanism or taste perception? *Arch Dis Child Fetal Neonatal Ed*. 1999;80:F146–F147.
26. Johnston CC, Collinge JM, Henderson SJ, et al. A cross-sectional survey of pain and pharmacological analgesia in Canadian neonatal intensive care units. *Clin J Pain*. 1997;13:308–312.
27. Guinsburg R, Kopelman BI, Anand KJ, et al. Physiological, hormonal, and behavioral responses to a single fentanyl dose in intubated and ventilated preterm neonates. *J Pediatr*. 1998;132:954–959.
28. Scott CS, Riggs KW, Ling EW, et al. Morphine pharmacokinetics and pain assessment in premature newborns. *J Pediatr*. 1999;135: 423–429.
29. Stevens BJ, Johnston CC, Petryshen P, et al. Premature Infant Pain Profile: development and initial validation. *Clin J Pain*. 1996;12: 13–22.
30. Taddio A, Katz J, Ilersich AL, et al. Effect of neonatal circumcision on pain response during subsequent routine vaccination. *Lancet*. 1997;349:599–603.
31. Ambuel B, Hamlett KW, Marx CM, et al. Assessing distress in pediatric intensive care environments: the COMFORT scale. *J Pediatr Psychol*. 1992;17:95–109.
32. Bouwmeester NJ, Anand KJS, van Dijk M, et al. Hormonal and metabolic stress responses after major surgery in children aged 0–3 years: a double blind, randomized trial comparing the effects of continuous versus intermittent morphine. *Br J Anaesth*. 2001;87: 390–399.
33. Tarbell SE, Cohen TI, Marsh JL. The toddler–preschooler postoperative pain scale: an observational scale for measuring postoperative pain in children aged 1–5. Preliminary report. *Pain*. 1992;50: 273–280.
34. McGrath PJ, Johnson G, Goodman JT, et al. CHEOPS: a behavioral scale for rating postoperative pain in children. In: Fields HL, Dubner R, Cervero F, eds. *Advances in Pain Research and Therapy.* Vol. 9. New York: Raven Press; 1985:395–402.
35. van Dijk M, de Boer JB, Koot HM, et al. Association between physiological and behavioral pain measures in 0- to 3-year-old children after major surgery. *J Pain Symptom Manage*. 2001;22: 600–609.
36. *The Observer Base Package for Windows* [computer program]. Version 3. Wageningen, The Netherlands: Noldus Information Technology; 1997.
37. *The Observer Video-pro. Support Package for Video Analysis* [computer program]. Version 4. Wageningen, The Netherlands: Noldus Information Technology; 1997.
38. Craig KD, Hadjistavropoulos HD, Grunau RVE. A comparison of two measures of facial activity during pain in the newborn child. *J Pediatr Psychol*. 1994;19:305–318.
39. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Control*. 1974;AC-19:716–723.
40. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978; 6:461–464.
41. Verbeke G, Molenberghs G. *Linear Mixed Models in Practice*. New York: Springer; 1997.
42. Craig KD, Grunau RVE. Neonatal pain perception and behavioral measurement. In: Anand KJS, McGrath PJ, eds. *Pain in Neonates*. Amsterdam: Elsevier, 1993:67–106.
43. Gilbert CA, Lilley CM, Craig KD, et al. Postoperative pain expression in preschool children: validation of the child facial coding system. *Clin J Pain*. 1999;15:192–200.
44. Prkachin KM. Dissociating spontaneous and deliberate expressions of pain: signal detection analyses. *Pain*. 1992;51:57–65.
45. LeResche L, Dworkin SF. Facial expression accompanying pain. *Soc Sci Med*. 1984;19:1325–1330.
46. Craig KD, Patrick CJ. Facial expression during induced pain. *J Pers Soc Psychol*. 1985;48:1080–1091.
47. Patrick CJ, Craig KD, Prkachin KM. Observer judgments of acute pain: facial action determinants. *J Pers Soc Psychol*. 1986;50: 1291–1298.
48. Craig KD, Hyde SA, Patrick CJ. Genuine, suppressed and faked facial behavior during exacerbation of chronic low back pain. *Pain*. 1991;46:161–171.
49. Hamers JP, Abu-Saad HH, Halfens RJ, et al. Factors influencing nurses' pain assessment and interventions in children. *J Adv Nurs*. 1994;20:853–860.
50. Seymour E, Fuller BF, Pedersen-Gallegos L, et al. Modes of thought, feeling, and action in infant pain assessment by pediatric nurses. *J Pediatr Nurs*. 1997;12:32–50.
51. Shapiro CR. Nurses' judgements of pain in term and preterm newborns. *J Obstet Gynecol Neonatal Nurs*. 1993;22:41–47.
52. Morison SJ, Grunau RE, Oberlander TF, et al. Relations between behavioral and cardiac autonomic reactivity to acute pain in preterm neonates. *Clin J Pain*. 2001;17:350–358.
53. Anand KJS, Sippell WG, Aynsley-Green A. Randomised trial of fentanyl anaesthesia in preterm babies undergoing surgery: effects on the stress response. *Lancet*. 1987;31:243–248.
54. Anand KJS, Hickey PR. Halothane–morphine compared with high-dose sufentanil for anesthesia and postoperative analgesia in neonatal cardiac surgery. *New Engl J Med*. 1992;326:1–9.
55. Lewis M, Thomas D. Cortisol release in infants in response to inoculation. *Child Dev*. 1990;61:50–59.
56. Lewis M, Ramsay DS. Developmental change in infants' responses to stress. *Child Dev*. 1995;66:657–670.
57. Quinn MW, Wild J, Dean HG, et al. Randomised double-blind controlled trial of effect of morphine on catecholamine concentrations in ventilated pre-term babies. *Lancet*. 1993;342:324–327.
58. Cabal LA, Siassi B, Artal R, et al. Cardiovascular and catecholamine changes after administration of pancuronium in distressed neonates. *Pediatrics*. 1985;75:284–287.
59. Stopfkuchen H, Racke K, Schworer H, et al. Effects of dopamine infusion on plasma catecholamines in preterm and term newborn infants. *Eur J Pediatr*. 1991;150:503–506.
60. Peters JWB, Koot HM, de Boer J, et al. Postoperative pain assessment in infants: validation of the Neonatal Facial Coding System. *The 5th International Symposium on Paediatric Pain 2000,* 18–21 June 2000; London, United Kingdom.