Universität Basel
Fakultät für Psychologie

The consideration of individual contextual factors in neonatal pain assessment: Validation and revision of the Bernese Pain Scale for Neonates

Inauguraldissertation zur Erlangung der Würde eines Doktors der Philosophie vorgelegt der Fakultät für Psychologie der Universität Basel von

Karin Schenk

aus Röthenbach im Emmental

Basel, 2019

Universität Basel

Fakultät für Psychologie

Genehmigt von der Fakultät für Psychologie auf Antrag von

Prof. Dr. Alexander Grob

Prof. Dr. Sakari Lemola

PD Dr. Eva Cignacco

Datum des Doktoratsexamen: 11.12.2019

DekanIn der Fakultät für Psychologie

Universität
Basel

Fakultät für
Psychologie

Erklärung zur wissenschaftlichen Lauterkeit

Ich erkläre hiermit, dass die vorliegende Arbeit ohne die Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel selbstständig verfasst habe. Zu Hilfe genommene Quellen sind als solche gekennzeichnet. Die veröffentlichten oder zur Veröffentlichung in Zeitschriften eingereichten Manuskripte wurden in Zusammenarbeit mit den Koautoren erstellt und von keinem der Beteiligten an anderer Stelle publiziert, zur Publikation eingereicht, oder einer anderen Prüfungsbehörde als Qualifikationsarbeit vorgelegt. Es handelt sich dabei um folgende Manuskripte:

- Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. (2017). Individual contextual factors in the validation of the Bernese pain scale for neonates: protocol for a prospective observational study. *BMC Pediatrics, 17*(171). doi: 10.1186/s12887-017-0914-9
- Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., Nelle, M., & Cignacco, E. (2019). The influence of gestational age in the psychometric testing of the Bernese Pain Scale for Neonates. *BMC Pediatrics, 19*(20). doi: 10.1186/s12887-018-1380-8
- Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., Nelle, M., & Cignacco, E. The Bernese Pain Scale for Neonates-Revised accounts for individual contextual factors. (Manuscript submitted for publication)

Bern, März 2019

Karin Schenk

ACKNOWLEDGMENTS

Throughout the whole process of my dissertation I have received a great deal of valuable support and assistance. I am very grateful to the following people who supported me during this time:

Prof. Dr. Alexander Grob – for the opportunity to write my dissertation at your department, for your guidance and support, your confidence in my work, and for being my doctoral advisor.

Prof. Dr. Sakari Lemola – for your confidence in my work and for being my second doctoral advisor.

PD Dr. Eva Cignacco – for your guidance and continuous support during my dissertation, your valuable insights, expertise and effort you put into my research, and for being my external doctoral advisor.

Prof. Dr. Bonnie Stevens, Prof. Dr. Dirk Bassler, Prof. Dr. Sven Schulzke and PD Dr. Mathias Nelle – for your expertise, valuable feedback and confidence in my work, and for being co-authors of this project.

Dr. Reto Bürgin – for your expertise and support during data analyses, and for being a co-author of this project.

Lilian Stoffel – for your valuable support during the whole process of my dissertation, your expertise, and for being a co-author of this project. Seeing your commitment for the vulnerable patients of a NICU was consistently a source of motivation for my work.

Jean Anthony Grand-Guillaume-Perrenoud – for the academic exchange, for your support and encouraging words, and for your valuable comments when proofreading my dissertation.

Prof. Dr. Rebecca Pillai Riddell – for being my supervisor in Pain in Child Health (PICH) and for the possibility to be a member of PICH. This membership entitled me to participate in

several international PICH events which were of great benefit and inspiration for my research.

The participants of the D-Blocks – for their inputs, critical questions and valuable feedback.

All nurses, head nurses, and medical directors of the University Hospital NICUs in Basel, Bern and Zurich – this project would not have been possible without your support and efforts.

All research assistants of this project – data collection would not have been feasible without your help and support.

All other persons who made this study possible – especially the neonates who participated in this study and their parents, Janik Schneeberger for his support in technical acquisitions, and the New Media Center team of the University of Basel.

My colleagues from the University of Applied Sciences Bern – for the academic exchange and your supporting words and motivation.

My friends and family – for your patience, love and trust in me. I especially thank Bruno for his patience and support during hard times. I thank Joël for trusting in me at the very beginning of my academic period of life; without his support I would not be where I am. Finally, I would like to thank Lharampa Tenzin Kalden for his valuable teachings.

TABLE OF CONTENTS

ABSTRACT

Neonates are dependent on a caregiver to discover that they are in pain and to manage it. Numerous pain assessment scales have been developed, but pain assessment is challenging because neonates of different gestational ages (GAs) have widely varied pain responses. Individual contextual factors such as GA or health status may account for this variability in pain response. The aim of the present dissertation was the validation and revision of the Bernese Pain Scale for Neonates (BPSN) by testing its psychometric properties and analyzing the influence of individual contextual factors on the variability in pain response. The BPSN is a pain assessment tool that is widely used in Swiss neonatal intensive care units. In this prospective multisite validation study, 154 neonates between 24 2/7 and 41 4/7 weeks of gestation were videotaped during 1-5 routine capillary heel sticks in their first 14 days of life. For each heel stick, three video sequences were produced: baseline, heel stick, and recovery. Comprehensive psychometric testing was conducted to examine the BPSN's underlying factor structure, interrater reliability, concurrent and construct validity, sensitivity and specificity. Single and multiple linear mixed effects analyses were used to examine the influence of individual contextual factors on variability in pain response. The results of the psychometric testing indicated a significant reduction of the scale from nine to four items: crying, facial expression, posture and heart rate. This modified BPSN showed promising reliability and validity, especially when the cut-off that discriminates between no or low pain and moderate to severe pain is adjusted to increase with increasing GA. Apart from the GA, baseline behavioral state and ventilation status were the individual contextual factors which the revised BPSN should account for. The BPSN-Revised is a promising tool for acute procedural pain assessment in full-term and preterm neonates with different GAs. Future studies should test its validity, feasibility and clinical utility.

1. Introduction

Until the middle of the eighties the common assumption existed that neonates are not able to sense pain, due to their neurological immaturity (Unruh & McGrath, 2014). This assumption changed abruptly in 1985, when the mother of an extremely preterm neonate appealed to the public because her little son Jeffrey Lawson had experienced an operation on his open heart while he was only sedated, but not anesthetized (Lawson, 1986). This incident with its following popular outrage was an important turning point in neonatal pain management. Shortly after, Anand and Hickey (1987) showed that preterm and full-term neonates' nervous systems are developed sufficiently that they may have a sensation of pain. Since then, massive efforts have been undertaken to explore the underlying mechanisms of neonatal pain and to provide appropriate pain assessment and management strategies. Despite this progress, operationalization and assessment of neonatal pain remains a major challenge. The influence of individual contextual factors (e.g., neonate gestational age) on pain response poses one of these challenges.

During the past 30 years, there has also been an increase in preterm birth rate in most countries (Blencowe et al., 2012). Preterm birth is defined as birth of a neonate with a gestational age (GA) younger than 37 weeks. Preterm neonates can be subdivided based on their GA into extremely preterm (< 28 weeks), very preterm (28-31 6/7 weeks), and moderate to late preterm (32-36 6/7 weeks) neonates (World Health Organization, 2018). In Switzerland, the preterm birth rate averages 7% (Bundesamt für Statistik, 2018). Of these, about 14% are born with a GA younger than 32 weeks. Depending on their GA and state of health, preterm neonates spend their first postnatal days, weeks or months in a neonatal intensive care unit (NICU), where they are exposed to multiple painful diagnostic and therapeutic procedures (Carbajal et al., 2008; Cignacco, Hamers, et al., 2009; Stevens et al., 2011). A recent review indicated that preterm neonates experience 7.5 to 17.3 invasive

procedures each day on average during their first two weeks postnatal (Cruz, Fernandes, & Oliveira, 2016). The most immature infants often experience the highest number of painful procedures because of their longer stay in a NICU and their poorer state of health (Grunau, Oberlander, Whitfield, Fitzgerald, & Lee, 2001). The procedures are important for the survival and health of preterm neonates, but the results of several studies indicate that repeated painful experiences at this early age with an immature nervous system lead to negative short- and long-term consequences (e.g., Brummelte et al., 2012; Doesburg, 2013; Grunau et al., 2009; Ranger et al., 2013; Schneider et al., 2018; Vinall & Grunau, 2014). Therefore, reliable and valid pain assessment tools are crucial for an appropriate pain management and a healthy development of this vulnerable population group.

The aim of this dissertation was the validation and revision of the Bernese Pain Scale for Neonates (BPSN) by testing its psychometric properties (Cignacco et al., 2017; Schenk et al., 2019; see Appendixes II and III) and analyzing the effect of individual contextual factors on variability in neonates' pain response (Schenk et al, *submitted*; see Appendix IV). Chapter 2 summarizes the theoretical background regarding neonatal pain, Chapter 3 presents the research questions of this dissertation and Chapter 4 summarizes the method. Chapter 5 includes a synopsis of the study results and Chapter 6 contains a general discussion of the main findings and the conclusions.

2. Theoretical background

2.1 Conceptualization of pain

The International Association for the Study of Pain (2017) conceptualizes pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." Because some individuals (e.g., neonates) are not able to communicate their experiences verbally, this definition was updated in 2001 to include the proposition that "the inability to communicate verbally does not negate the possibility that an

individual is experiencing pain" (Hadjistavropoulos, Breau, & Craig, 2011). Williams and Craig (2016) recently recommended to define pain as "a distressing experience associated with actual or potential tissue damage with sensory, emotional, cognitive, and social components". This new definition considers all aspects of the biopsychosocial model of pain, which describes pain as complex interaction between biological, psychological and social characteristics (Craig, 2009).

The subjective pain experience and pain expression may be influenced by many factors. The social communication model of pain describes pain as a dynamic and interactive process between a person in pain and its caregiver (Craig, 2009). Sellam, Cignacco, Craig and Engberg (2011) adapted this model to preterm neonates in a NICU, where demographic and medical factors are particularly relevant and may have an immediate influence on neonates' pain experience and pain response. Cong, McGrath, Cusson and Zhang (2013) suggest in their conceptual framework of pain measurement that neonates' behavioral, physiological and biochemical pain response may be influenced by characteristics of the painful stimulus (e.g., type and duration of pain) as well as contextual factors (e.g., GA, number of painful experiences). Pain response as well as the measured pain level may also be influenced by characteristics of the observer (e.g., knowledge, attitude about pain) and the pain assessment tool used. In the present dissertation, neonates' pain response measured with the BPSN as well as the influence of individual contextual factors on variability in pain responses will be examined.

2.2 Individual contextual factors

A neonate's demographic characteristics (e.g., sex, age) and medical conditions (e.g., health status, medication) and his or her history of painful and stressful experiences might impact pain response (Cong et al., 2013; Lee & Stevens, 2014; Sellam et al., 2011). For instance, in neurologically impaired and very ill neonates, and in neonates on medications such as sedatives, pain may be

dampened or not observable (Hummel & van Dijk, 2006). Such individual contextual factors may account for the high variability in pain responses within and between neonates (Cignacco, Denhaerynck, Nelle, Bührer, & Engberg, 2009).

2.3 Physiology of pain

Whereas the concept of pain encompasses biological, emotional, cognitive, social and other contextual components, nociception refers to the neural process of transmission, processing and modulation of noxious stimuli at different levels of the nervous system. Potentially or actually noxious stimuli are detected by receptors of the peripheral somatosensory nervous system (nociceptor) and are transduced into electrical signals (Walker & Baccei, 2014). These nociceptive signals are transmitted to the dorsal horn where the processing of noxious information by the central nervous system (CNS) begins (Walker & Baccei, 2014). Nociceptive signals are processed through excitatory (increasing neural response to stimulation) and inhibitory (decreasing neural response) neural circuitry (Beggs, 2015). The conscious perception of pain happens in the brain, where nociceptive signals are combined with emotional and cognitive processes of the brain (Simons & Tibboel, 2006).

2.3.1 Nociception in preterm neonates

The neurophysiological system of extremely preterm neonates is already developed to such a degree that it is capable of transmitting nociceptive signals (Fabrizi, 2011; Fitzgerald, 2005). In preterm neonates, the excitatory circuitry in the dorsal horn is predominant and facilitates the activity-dependent development and re-organization of the CNS (Beggs, Torsney, Drew, & Fitzgerald, 2002; Fitzgerald, 2005); which are crucial processes in the maturation of neonates' CNS. The inhibitory circuitry is immature and leads to reduced sensory discrimination between tactile and noxious stimulation and poor spatial localization of noxious stimuli (Cornelissen et al., 2013; Walker,

Fitzgerald, & Hathway, 2015). The specific neuronal circuits that are necessary to distinguish between touch and noxious stimuli do not emerge until 35 to 37 weeks of gestation (Fabrizi, 2011).

Cortical neurons and networks are activated by noxious stimulation (Fitzgerald, 2015; Slater et al., 2006), but cortical responses to tactile and noxious stimulation are non-specific in very preterm neonates and become more specific with increasing age (Cornelissen et al., 2013; Green et al., 2019). Most of the brain regions that are involved in adults' pain are also activated in full-term neonates (Goksan et al., 2015), but the ability of cognitive pain processing (e.g., interpretation) emerges later in development (Ranger & Grunau, 2015). Due to this immaturity of the nociceptive system and CNS, preterm neonates may be more sensitive to pain than older neonates and adults (Grunau, 2013).

2.3.2 Consequences of early pain exposure

Repeated painful and stressful experiences during the critical period of nervous system plasticity have the potential to impact the development of the CNS in both the short- and long-term. Neonates' pain response may be impaired after only 20 painful procedures (Grunau et al., 2005), but the direction of this impairment is unclear. Some studies reported less intense pain responses in neonates subjected to frequent painful procedures (e.g., Grunau et al., 2001; Johnston & Stevens, 1996; Morison et al., 2003). Other study results suggested that repeated painful experiences may lead to increased pain response (hyperalgesia) or to pain responses without a painful stimulus (allodynia) (e.g., Fitzgerald, Millard, & McIntosh, 1989; Taddio, Shah, Gilbert-MacLeod, & Katz, 2002).

Painful experiences may impact the development of the CNS with consequences that persist into adulthood (Walker, Beggs, & Baccei, 2016). Adults who were injured during the neonatal period may show an increased response to painful stimulation of a previous injured region (local hyperalgesia; Beggs, Currie, Salter, Fitzgerald, & Walker, 2012) and a reduced response to painful stimulation on other parts of the body (global hypoalgesia; Walker et al., 2015). Furthermore, early

pain exposure is associated with long-term effects on the developing brain, such as reduced maturation of white matter and subcortical grey matter (Brummelte et al., 2012) and impaired structural and functional reorganization of the nervous system (Schneider et al., 2018). Further consequences of early pain exposure are for example an impaired cognitive and motor development (Bhutta & Anand, 2002; Grunau et al., 2009) and changes in the function of the stress-response system (Grunau et al., 2010). Accurate pain assessment and appropriate pain management are therefore fundamental for a healthy development of preterm neonates.

2.4 Pain assessment

Neonates depend on caregivers who detect and assess their suffering by observing behavioral and physiological pain responses. Because caregivers' judgments are subjective (Craig, 2009), pain assessment scales are used in the clinical setting to make the assessment more objective. Pain assessment scales can be classified as either unidimensional or multidimensional (Lee & Stevens, 2014). Unidimensional scales include behavioral pain indicators (e.g., facial expression, crying); multidimensional pain scales are a combination of behavioral and physiological indicators (e.g., changes in heart rate and oxygen saturation). Some pain assessment scales further include contextual factors such as GA or behavioral state (e.g., the Premature Infant Pain Profile-Revised [PIPP-R]; Stevens et al., 2014). Recently, more objective approaches for pain assessment have been investigated, such as measurement of skin conductance or heart rate variability (Cong et al., 2013). For a better understanding of the underlying mechanisms in neonatal pain and for an improvement in pain assessment, newer brain-oriented techniques such as functional magnetic resonance imaging (fMRI) are used (Fitzgerald, 2015; Ranger & Grunau, 2015). However, for a systematic clinical pain assessment, exclusively observable behavioral and physiological indicators are appropriate.

More than 40 pain assessment scales for neonates and infants have been developed in the past three decades (Cong et al., 2013). Most pain scales were designed for research purposes and are inappropriate for the clinical practice because their application requires too much time (Franck & Bruce, 2009). Of the pain scales developed for clinical pain assessment, only few have undergone rigorous psychometric testing (e.g., the Behavioral Infant Pain Profile [BIIP; Holsti & Grunau, 2007]; the Neonatal Pain, Agitation and Sedation Scale [N-PASS; Hummel et al., 2008; Hummel eta l., 2010]), and have been validated in extremely preterm neonates (AAP, Committee on Fetus and Newborn, & Section on Anesthesiology and Pain Medicine, 2016; Badr, 2013). In addition, the feasibility and clinical utility of validated pain scales have been rarely examined (Lee & Stevens, 2014). For this reason, none of the existing pain assessment scales are referred as being the gold standard (Lee & Stevens, 2014).

2.4.1 Challenges in pain assessment

Pain assessment in neonates is hindered by different reasons: (1) A lack of consensus exists in the international research community regarding the appropriate dimensionality of a pain scale (Pillai Riddell et al., 2016). Only low to moderate associations exist between behavioral and physiological indicators of pain (e.g., Holsti, Grunau, Oberlander, & Osiovich, 2008; Lucas-Thompson et al., 2008). In addition, no consistent association between behavioral and physiological pain indicators and nociception-specific brain activity has been detected so far (Pillai Riddell et al., 2016; Relland, Gehred, & Maitre, 2019). (2) The fact that unspecific physiological and behavioral indicators of pain may also be shown during non-painful, stressful experiences (e.g., agitation because of hunger or other factors) makes pain assessment more difficult (Hummel & van Dijk, 2006; Johnston, Fernandes, & Campbell-Yeo, 2011). (3) The absence of a pain reaction to a procedure that would normally be considered painful (e.g., heel stick) does not necessarily mean that the neonate

does not sense any pain (Johnston et al., 1999). In a study of Slater, Cantarella, Franck, Meek and Fitzgerald (2008) some infants showed no observable behavioral reaction to heel stick procedures, although their cortex was strongly activated. (4) Until today, no uniform definition exists for the different kinds of pain in newborns. Anand (2017) recently suggested that chronic pain in newborns should be defined as pain that lasts eight days or longer, with clearly different behavioral and physiological response patterns compared to acute pain. (5) The examination of the influence of individual contextual factors on variability in neonates' pain response is hindered because contextual factors may be strongly correlated with each other (Sellam et al., 2011). Extremely preterm neonates may, for example, have a longer stay in the NICU due to their poorer health status and may therefore experience more painful procedures compared to more mature neonates. Because neglecting the influence of individual contextual factors in pain assessment might lead to misjudgment of a painful state followed by a lack of effective pain management strategies (Hatfield & Ely, 2015), the consideration of relevant contextual factors in pain assessment has been recommended to enhance the accuracy of pain scales (e.g., AAP et al., 2016; Sellam, Engberg, Denhaerynck, Craig, & Cignacco, 2013).

2.4.2 Pain assessment and management in the clinical practice

The integration of pain assessment into clinical practice is a further challenge. Routine pain assessment in neonates has been strongly recommended (e.g., AAP et al., 2016; Hall & Anand, 2014), but the implementation and systematic use of valid and reliable pain assessment scales in daily practice have remained problematic (Avila-Alvarez et al., 2016; Cong et al., 2014; Polkki, Korhonen, & Laukkala, 2018; Stevens et al., 2012). In a recently conducted survey that captured the assessment of continuous pain in 18 European NICUs, pain assessment varied between 0% and 100% (Anand et

al., 2017). Only 31.8% of neonates received assessment of pain at least once during their stay in a NICU, and only 10% received daily assessments.

Pain assessment scales are also used to determine whether a pain-relieving intervention was successful or if further interventions are needed. Non-pharmacological pain-relieving interventions in a NICU are used for prevention and management of mild to moderate pain and include oral sucrose, non-nutritive sucking, swaddling, facilitated tucking, kangaroo (skin-to-skin) care and breastfeeding (Carter & Brunkhorst, 2017; Hatfield, Murphy, Karp, & Polomano, 2019). Pharmacological methods commonly used for pain management in NICUs include opioids (e.g., morphine, fentanyl), non-opioids (e.g., paracetamol, indomethacin), sedatives (e.g., propofol, midazolam), and ketamine (Carter & Brunkhorst, 2017; Hall, 2012).

2.5 The Bernese Pain Scale for Neonates

The BPSN was developed in 1996 by nurses of the University Hospital Bern, primarily for clinical use. The BPSN is a multidimensional pain assessment tool that includes seven subjective items (sleeping, crying, consolation, skin color, facial expression, posture and breathing) and two physiological items (heart rate and oxygen saturation; Cignacco, Mueller, Hamers, & Gessler, 2004). Since its first validation in the year 2004, the BPSN has been widely used in Swiss NICUs (Boettcher et al., 2012). The results of the first validation study suggested that the BPSN is a reliable and valid tool for assessing acute pain in preterm and full-term neonates (Cignacco et al., 2004). A limitation of this study was the small study population of only 12 neonates. In addition, feedback from clinical practice related to difficulties in pain assessment with the BPSN in very preterm neonates and the increasing scientific evidence that neonates' pain response is influenced by individual contextual factors suggested that a revalidation of the BPSN was required. Because the BPSN is already widely

used in the clinical setting, a modified version of the BPSN should be quickly adopted by the health professionals.

3. Research questions

The present dissertation had three overarching aims: (1) the validation of the BPSN using a large sample of neonates with different GAs; (2) the analysis of the influence of individual contextual factors on the variability in pain responses; and, (3) the revision of the BPSN according to the study results. Two sub-studies addressed the following research questions:

Sub-study 1 (Manuscript 2)

1. What are the psychometric properties of the BPSN?

Sub-study 2 (Manuscript 3)

2. Which individual contextual factors have an influence on variability in neonates' behavioral and physiological pain responses?
3. Which modifications are required to improve pain assessment with the BPSN based on the study results?

4. Method

4.1 Sample and Setting

This prospective multisite validation study with repeated measurement design was conducted in three university hospital NICUs in Switzerland (Basel, Bern, and Zurich). Preterm neonates born between 24 0/7 and 36 6/7 weeks of gestation were included if they were expected to undergo 2-5 routine capillary heel sticks in their first 14 days of life. Full-term neonates born between 37 0/7 and 42 0/7 weeks of gestation were included if they were expected to have at least two routine capillary heel sticks during their first 14 days of life.

4.2 Recruitment and data collection procedure

Neonates were recruited by consecutive sampling and then stratified according to GA at birth (Cignacco et al., 2017). After parents granted written informed consent, trained study assistants videotaped neonates during their next 1-5 routine capillary heel sticks. For each heel stick, three video sequences were produced: baseline, heel stick, and recovery phase. During each of the three phases, the study assistants recorded the neonate's highest heart rate and lowest oxygen saturation. Every neonate received a dose of 24% oral sucrose (0.2 ml/kg bodyweight) before the heel stick procedure as a pain-relieving intervention, in accordance with standards of care (Stevens, Yamada, Ohlsson, Haliburton, & Shorkey, 2016). Five nurses who were working in a NICU and were experienced in using the BPSN rated the behavioral pain expression using the BPSN and the PIPP-R. Individual contextual factors were retrospectively retrieved from patient charts or from observations made during video recording.

4.3 Measures

Neonates' pain response was measured with the BPSN (Cignacco et al., 2004) and the PIPP-R (Gibbins et al., 2014). The BPSN includes seven subjective items (sleeping, crying, consolation, skin color, facial expression, posture, breathing) and two physiological items (heart rate, oxygen saturation). Each item is rated on a 4-point Likert scale (0, 1, 2, and 3). The scores of 11 or more points indicate pain. In a first validation study the BPSN showed good construct validity among 12 neonates with GAs between 27 and 41 weeks (Cignacco et al., 2004).

The PIPP-R is a well validated multidimensional pain assessment tool for use with preterm and full-term neonates, widely used for research purposes and in clinical settings of North America (Gibbins et al., 2014; Stevens et al., 2014). The PIPP-R includes three behavioral items (brow bulge, eye squeeze, and naso-labial furrow), two physiological items (heart rate and oxygen saturation), and

two contextual factors (GA and baseline behavioral state). Each indicator is rated on a 4-point Likert scale (0, 1, 2, and 3). Contextual factors are only factored in if the infant's behavioral and physiological sub score is ≥ 1 (Stevens et al., 2014). Neonates with younger GA and neonates in a quiet and sleep state score the highest. Zero points indicate no or perhaps no response to pain, 1-6 points indicate low pain, 7-12 points indicate moderate pain, and ≥ 13 points indicate severe pain. The PIPP-R showed good construct validity among 202 full-term and preterm neonates with GAs as young as 26 weeks (Gibbins et al., 2014).

Individual contextual factors were determined based on the findings of a systematic review (Sellam et al., 2011). Three dimensions of individual contextual factors were collected: (1) demographic contextual factors including GA at birth, sex, birth weight, nationality, parity, and way of delivery; (2) medical contextual factors including the primary diagnoses (premature or term birth) and common comorbidities in preterm neonates, the number of comorbidities, the neonate's health status measured by the Clinical Risk Index for Babies (CRIB; Bührer, Grimmer, Metze, & Obladen, 2000), ventilation status, and medication; and, (3) experiences with previous painful and non-painful procedures (Cignacco et al., 2008). In addition, the following contextual factors were assessed for each measurement point: postnatal age (PNA), post-menstrual age (PMA; GA birth combined with PNA), weight, the duration of each heel stick procedure, the number of additional sucrose doses given during heel stick procedures, and the baseline behavioral state measured with the PIPP-R (Stevens et al., 2014).

4.4 Statistical analyses

In the **first sub-study** (Manuscript 2), the BPSN's psychometric properties were examined using the statistics programs SPSS (IBM© SPSS© Statistics Version 23.0, IBM Corp, Armonk, NY, USA) and R (R Core Team, 2017). Intraclass correlation coefficients (ICCs) were calculated to

determine interrater reliability of the seven subjective items. Multiple-group confirmatory factor analysis (CFA) was used to evaluate the extent to which individual items correlated with the unobservable pain construct and whether factor loadings were invariant across time and raters. The internal consistency was evaluated by calculating Cronbach's α. Construct validity was determined by comparing the level of pain scores between the three phases of the heel stick procedure (baseline, heel stick, and recovery) using linear mixed effects analyses. Pearson product-moment correlation coefficients were calculated to establish concurrent validity between the BPSN and the PIPP-R and the association between behavioral and physiological pain scores. Receiver-Operating Characteristic (ROC) curve analysis was used to evaluate the ability of the BPSN to detect pain in neonates and to determine the cut-off value that maximizes both sensitivity and specificity. Because the study sample was heterogeneous (neonates' GA at birth ranged from 24 2/7 to 41 4/7 weeks), the data was reanalyzed separately for the four GA-groups: extremely preterm neonates, very preterm neonates, moderate to late preterm neonates, and full-term neonates (World Health Organization, 2018). The CFA was not reanalyzed for different GA-groups separately because the subsamples were too small. GA was already considered in the linear mixed model analyses.

In the **second sub-study** (Manuscript 3), the influence of individual contextual factors on variability in neonates' behavioral and physiological pain responses was analyzed by conducting linear mixed effects analyses. The analysis was divided into two stages. First, the effect of each contextual factor on the level of pain scores was separately tested in simple linear mixed effects models. Second, all contextual factors that reached a *p*-value below 0.20 were included in a multiple linear mixed effects model and backward elimination of non-significant contextual factors was conducted.

5. Results

A total of 162 neonates was enrolled in the study; 8 neonates were excluded from data analyses because video sequences were missing or of poor quality. Mean GA at birth of the 154 neonates was 30.9 weeks (*SD* = 4.5).

The following chapters summarize the results of the two sub-studies. Sub-study 1 (Schenk et al., 2019) analyzed the psychometric properties of the BPSN. Sub-study 2 (Schenk et al., *submitted*) examined the influence of individual contextual factors on the variability in neonates' pain response and modified the BPSN according to the results of the two sub-studies.

5.1 Psychometric properties of the BPSN

5.1.1 Factor structure and reliability of the BPSN

The level of interrater agreement differed between the subjective items of the BPSN. During the heel stick phase of the five measurement points, interrater agreement was good to excellent for the items crying (*ICCs* = 0.905-0.945), facial expression (*ICCs* = 0.833-0.905), posture (*ICCs* = 0.722-0.860), consolation (*ICCs* = 0.634-0.805), and breathing (*ICCs* = 0.627-0.770). Interrater reliability was moderate to good for the item sleeping (*ICCs* = 0.532-0.646) and poor for the item skin color (*ICCs* = 0.189-0.285).

The results of the CFA showed that the items consolation, crying, facial expression, and posture had factor loadings higher than 0.30 for the subjective subscale. The factor loadings of the items breathing and skin color were low (range = -0.167-0.293), and loadings for the item sleeping varied widely between raters and measurement points (range = 0.096-0.982). Further analysis of the subjective subscale showed that a model including the three items crying, facial expression, and posture fit the data best. These three items showed within-rater invariance during the heel stick phase of the five measurement points, but no between-rater invariance.

The physiological items heart rate and oxygen saturation did not load on a common factor, nor did they correlate with each other (r = -0.028-0.106). Therefore, the physiological items' sensitivity to detect pain was analyzed by calculating linear mixed models. Scores of the item heart rate were on average 0.65 points higher during the heel stick phases than scores during the recovery phases (SE = 0.09, t-value = -7.38); scores of the item oxygen saturation were on average 0.26 points higher during the heel stick than during the recovery phases (SE = 0.12, t-value = -2.14).

Due to the results of the previous analyses, a first modification of the BPSN was conducted. This modified BPSN included a behavioral subscale (facial expression, crying, and posture) and the item heart rate as a physiological pain indicator. The next analyses were conducted with this modified version of the BPSN.

5.1.2 Validity of the modified BPSN

To determine construct validity of the modified BPSN, the level of pain scores of the behavioral subscale between the three phases was compared. Behavioral pain scores in the heel stick phases averaged 1.04 higher than pain scores in the baseline phases (SE = 0.07, t-value = 15.01), and 1.13 higher than pain scores in the recovery phases (SE = 0.07, t-value = 16.04). As mentioned previously, pain scores of the item heart rate were on average 0.65 points higher during the heel stick phases compared to the recovery phases (SE = 0.00, t-value = 7.38). GA at time of birth significantly affected behavioral pain scores (SE = 0.01, t = 5.49) and scores of the item heart rate (SE = 0.01, t = 6.15); pain scores increased with increasing GA. Concurrent validity between the modified BPSN and the PIPP-R ranged from r = 0.600-0.758 (Mdn = 0.688) among the five raters and measurement points and tended to increase as GA increases.

5.1.3 Sensitivity and specificity of the modified BPSN

The results of the sensitivity and specificity analyses indicated that cut-off points needed to increase along with GA to reach about 80% sensitivity and similarly high specificity; extremely preterm neonates require 0.5 points, very preterm neonates require 1.5 points, moderate to late preterm neonates require 2.5 points, and full-term neonates require 3.5 points.

5.1.4 Correlations between behavioral and physiological pain indicators

Correlations between the modified behavioral BPSN subscale and the item heart rate were low during the five heel stick phases (r = 0.102-0.379, Mdn = 0.235). There was no obvious difference between the correlation coefficients calculated for the four GA-groups separately.

5.2. Influence of individual contextual factors on pain response

Preterm neonates had about 0.72 points lower behavioral pain scores ($p < 0.001$) and about 0.23 points lower physiological pain scores ($p = 0.004$) than full-term neonates. Neonates in an active and awake state before the heel stick procedure showed the highest behavioral pain scores during the heel stick procedure. Neonates in an active and awake state scored about 0.28 points higher than neonates in a quiet and awake state ($p < 0.001$), about 0.16 points higher than neonates in an active and asleep state ($p = 0.006$) and about 0.50 points higher than neonates in a quiet and asleep state ($p < 0.001$). Neonates who received caffeine had about 0.30 points lower behavioral pain scores than neonates who did not receive caffeine during or shortly before the recorded heel stick procedure ($p < 0.001$). Finally, neonates who were mechanically ventilated during the recorded heel stick procedure had about 0.20 points lower physiological pain scores than neonates who were not ventilated ($p = 0.002$).

5.3 The Bernese Pain Scale for Neonates-Revised

The BPSN was revised based on the results of the two sub-studies. The BPSN-Revised (Appendix I) includes three behavioral items (crying, facial expression, and posture), one physiological item (heart rate), and three individual contextual factors (PMA, baseline behavioral state, and ventilation status).

6. Discussion

The main objective of this dissertation was a revised BPSN that accounts for relevant individual contextual factors. Therefore, the dissertation included a comprehensive psychometric testing of the BPSN (Manuscript 2) and the examination of the influence of individual contextual factors on variability in neonates' pain response (Manuscript 3).

6.1 Factor structure and reliability of the BPSN

The CFA showed that a model that includes the behavioral items crying, facial expression and posture fits the data best. Crying, facial expression and body movements are widely studied pain indicators and are considered as the most sensitive behavioral indicators for pain assessment in neonates (Anand, 2007; Hatfield & Ely, 2015). In the following sections, pros and cons of these three pain indicators are discussed and suggestions for improving the reliability of the BPSN items are made. The results of the CFA indicated that different raters assess pain differently, an assumption further supported by the results of the interrater reliability analysis. Therefore, improving the guidelines and training for applying the BPSN in the clinical practice may improve its reliability.

Facial expression is considered as the most reliable and sensitive indicator for pain assessment in both preterm and full-term neonates (Anand, 2007). Brow bulge, eye squeeze, nasolabial furrow, and vertical mouth stretch are facial expressions that neonates with different GAs show (Gibbins, Stevens, Beyene, et al., 2008). The BPSN's item facial expression assesses neonate's face more

generally. This general assessment may facilitate pain assessment in preterm neonates who wear CPAP masks or tapes that are used to fix a tube to the skin, which hide specific components of facial expression such as nasolabial furrow. On the other hand, the assessment of specific components of facial expression that neonates typically show when they sense pain makes pain assessment more precise and therefore more reliable. The inclusion of a description of these specific components in the guideline may enhance the BPSN's reliability.

Crying is included in several pain assessment scales (e.g., Hudson-Barr et al., 2002; Hummel, Puchalski, Creech, & Weiss, 2008; Merkel, Voepel-Lewis, Shayevitz, & Malviya, 1997), but the use of crying as a pain indictor has also been put into question. Some neonates have a limited ability to cry due to mechanical ventilation, inhibiting drugs or severe illness (Gibbins, Stevens, McGrath, et al., 2008; Hatfield & Ely, 2015). Furthermore, crying is not specific to pain, because neonates cry also when they are hungry or feeling unwell (Hatfield & Ely, 2015). However, preterm neonates with immature facial muscles are less able to communicate their pain through facial expressions, and therefore, crying may be an important first indication that alerts their caregiver (Craig, Korol, & Pillai, 2002; Johnston, Stevens, Craig, & Grunau, 1993). Because some neonates are not able to express an audible cry, "silent crying" should also be considered in pain assessment (Kostandy et al., 2008).

Specific or more general body movements are also included in several pain assessment scales (e.g., Carbajal, Paupe, Hoenn, Lenclen, & Olivier-Martin, 1997; Holsti & Grunau, 2007; Hudson-Barr et al., 2002). The BPSN item posture assesses body movement more generally by evaluating if the neonate's body is relaxed or tense. Holsti, Grunau, Oberlander and Whitfield (2004) found that early preterm neonates show specific body movements like flexing and extending their arms and legs, making fists, and finger splaying more often during a heel stick procedure. In addition, Morison et

al. (2003) found that neonates with lower GA at birth made more specific body movements but showed less facial expression at 32 weeks PMA. These results confirm that the consideration of body movement provides supplementary information that enhances accurate pain assessment among neonates with different GAs. The inclusion of a description of specific body movements observed in very preterm neonates in the guideline may also enhance the BPSN's reliability.

The item oxygen saturation was excluded from the BPSN because heart rate was more sensitive to pain. This conclusion is in line with the suggestion of the authors of another study that validated a Norwegian version of the PIPP-R (Vederhus, Eide, & Natvig, 2006).

6.1.2 Validity of the modified BPSN

The modified BPSN that includes the behavioral items crying, facial expression, and posture, and the physiological item heart rate showed promising construct validity and concurrent validity with the PIPP-R. Behavioral and physiological pain indicators were significantly higher during the heel stick phases compared to the baseline and recovery phases. Pain scores slightly increased with increasing GA.

6.1.3 Sensitivity and specificity of the modified BPSN

The results of the sensitivity and specificity analyses suggested that the cut-off that discriminates between no or low pain and moderate to high pain (as measured with the PIPP-R) had to increase with increasing GA. To reach a sensitivity and specificity of approximately 80%, extremely preterm neonates require a cut-off value of 0.5 points, while full-term neonates require 3.5 points (total overall scores = 12 points). Compared to the original BPSN's cut-off of 10.5 points that discriminates between pain and no pain (total overall score = 27 points), these cut-offs are much lower. In the present study, the means of the original BPSN total score varied widely between the five raters, but did not reach the cut-off value of 11 points that indicates a painful state. The oral

sucrose administered to each neonate before the heel stick may have lowered pain response, an effect which was already demonstrated in numerous studies (Stevens et al., 2016).

The ROC analysis showed also that the modified BPSN was least able, but still moderately good (Streiner, Norman, & Cairney, 2015) in discriminating between neonates who experience no or low pain and neonates who experience moderate to high pain in the lowest GA-group. Extremely preterm neonates' pain expression may be less obvious and weak due to their immature nervous system and muscles that prevent them from expressing a robust pain reaction (Gibbins, Stevens, Beyene, et al., 2008; Gibbins, Stevens, McGrath, et al., 2008; Morison et al., 2003). As mentioned previously, adding information about specific components in extremely preterm neonates' pain responses to the guideline may make pain assessment in this vulnerable population more accurate and reliable.

6.1.4 Correlations between behavioral and physiological pain indicators

Correlations between the modified behavioral subscale and the item heart rate were low. This result confirms previous findings (e.g., Välitalo et al., 2016; Vederhus et al., 2006). However, the consideration of both behavioral and physiological indicators in neonatal pain assessment is generally assumed to be most appropriate for the clinical setting because of the complex nature of pain (Lee & Stevens, 2014). Moreover, the results presented in the next chapter show that behavioral and physiological indicators are influenced by different individual contextual factors.

6.2 Influence of individual contextual factors on pain response

In the second sub-study, the influence of numerous individual contextual factors on variability in neonates' pain response was examined (Schenk et al., *submitted*). The relevance of significant contextual factors in pain assessment is discussed in the following chapters.

6.2.1 Gestational age

Preterm neonates had lower behavioral and physiological pain scores than full-term neonates. These results concur with the findings of the first sub-study; the younger the GA, the lower the cut-off should be that discriminates between no or low pain and moderate to high pain. Neonates with younger GA show less obvious and more inconsistent behavioral pain responses than more mature preterm or full-term neonates because they have an immature nervous system and less muscular strength, posture and body movements (Gibbins, Stevens, McGrath, et al., 2008; Johnston et al., 2011). Preterm infants may show less change in heart rate because they are in a state of constant autonomic arousal due to the repeated painful and stressful procedures they often experience during their NICU stay (Grunau et al., 2001). Therefore, neglecting neonate's age in pain assessment might lead to underestimation of a painful state in preterm neonates.

6.2.2 Baseline behavioral state

Neonates in a quiet (asleep or awake) behavioral state before the heel stick procedure had lower behavioral pain scores than neonates in an active (asleep or awake) state. This finding confirms the results of other studies (e.g., Ahn, 2006; Badr et al., 2010; Johnston et al., 1999). A lower pain response does not necessarily mean that a neonate senses less pain. Holditch-Davis, Brandon, and Schwartz (2003) showed that neonates' behavioral responses to environmental stimuli reflect also the sleep-wake state in which the responses occur. Preterm neonates spend up to 70% of their time in a sleep state (with active sleep as the major behavioral state) and the sleep-wake state changes with increasing PMA (Foreman, Thomas, & Blackburn, 2008; Werth et al., 2017). Therefore, the consideration of the baseline behavioral state in the assessment of acute procedural pain may enhance the accuracy of pain assessment.

6.2.3 Ventilation status

Neonates who were mechanically ventilated had lower physiological pain scores than neonates who were not ventilated during the recorded heel stick. There was also a tendency that neonates treated with CPAP had lower pain scores than neonates without any ventilation. These findings confirm the result of a previous study (Sellam et al., 2013), while other studies found no association between ventilation status and neonates' pain response (Grunau, Holsti, & Peters, 2006; Grunau et al., 2001; Johnston, Stevens, Yang, & Horton, 1996). When the effect of ventilation status was examined in a single mixed model, a significant association between ventilation status and behavioral pain scores was identified. Behavioral pain scores may be decreased because CPAP or tapes that are used to fix the tube hide neonate's face and impede pain assessment. Therefore, the consideration of ventilation status may reduce misjudgment of pain.

6.3 Bernese Pain Scale for Neonates-Revised

The BPSN was revised according to the results of the two sub-studies. While the behavioral and physiological items showed promising psychometric properties, the adding of the three contextual factors demands a further testing of the validity of the BPSN-Revised (BPSN-R). In addition, the BPSN was validated and revised for the assessment of acute procedural pain. Because the BPSN is used for routine pain assessment in the clinical setting, the BPSN-R should also be validated for different painful and stressful procedures and for different types of pain. Acute pain shows a clearly different behavioral and physiological response pattern compared to, for example, chronic pain (Anand, 2017). While the behavioral response to acute pain is reactive and reflexive, neonates' response to chronic pain may be hypo- or hyperreactive (Anand, 2017). Furthermore, in the assessment of pain that contains no clear beginning and end (e.g., prolonged or chronic pain), the

assessment of a baseline behavioral state may be problematic. Future validation studies should examine the feasibility and clinical utility of the BPSN-R.

6.4 Strengths and limitations

This study has several strengths. First, the study sample included full-term and preterm neonates that cover a wide range of GAs. This is advantageous compared to other studies mostly focusing on preterm neonates with higher GA or full-term infants, as this allows the generalization of the study results to the entire population in which the BPSN is applied. Second, neonates' pain response was measured repeatedly across the first 14 days of life. This allowed the consideration of neonates' development across time. Most studies that evaluated neonates' pain response used a cross-sectional design (Williams, Khattak, Garza, & Lasky, 2009). Third, the influence of numerous individual contextual factors that may have an influence on neonates' pain responses was examined. The use of multiple linear mixed effects analyses allowed the examination of each factor's independent contribution in explaining neonates' pain responses. Fourth, the influence of contextual factors on physiological and behavioral pain indicators was analyzed separately, because these indicators show only low correlations with each other.

There are several limitations that need to be considered when interpreting the study results. First, characteristics of the video sequences may have affected the reliability of the ratings (e.g., poor lighting conditions). Second, different nurses performed the heel sticks, and their individual characteristics may have influenced the neonates' pain response. Third, particularly during the baseline and recovery phases, where the scores of the items were low and therefore upper categories were almost or completely left empty, floor effects may have influenced study results. For example, a variety of extensions of the model specification was considered in the factor analysis, but they were discarded because of convergence problems likely related to floor effects. Fourth, hypothesis testing

may be compromised by measurement error caused by differences between the ratings of the five nurses (Hallgren, 2012). This possible influence was compensated by either including the raters in the model, or by conducting separate analyses for each rater and pooling the results afterwards.

6.5 Conclusions and outlook

This is the first study that examined the factor structure of the BPSN and tested its psychometric properties among a large sample of neonates that cover a wide range of GAs. The results of this dissertation suggested a significant reduction of the number of items in the original BPSN, leaving only three behavioral items (crying, facial expression, and posture) and the physiological item heart rate. The results further suggested to add the contextual factors PMA, behavioral state and ventilation status to the BPSN-R.

The BPSN-R is one of few pain scales that have undergone rigorous psychometric testing among full-term and preterm neonates with different GAs and across repeated measurement points. It is also one of few pain scales that considers preterm neonates' immaturity by using different cut-offs for different PMA-groups. The consideration of PMA and other contextual factors in the BPSN-R will contribute to higher accuracy of pain assessment and prevent misjudgment of a painful state. The results of this dissertation emphasize that neonates' pain response is influenced by more than the noxious stimulus itself. As many preterm neonates spend more than 14 days on a NICU, future studies should observe neonates during their entire NICU stay. Neonates with the same PMA, but born with different GAs may not show the same pain response because neonates with younger GA have spent a longer time outside the intrauterine environment and consequently have had different experiences (e.g., they may have had more painful experiences and longer maternal separation). In sum, the BPSN-R is a promising tool for acute procedural pain assessment in full-term and preterm neonates with different GAs. However, further testing of its validity, feasibility and clinical utility is warranted.

References

AAP, Committee on Fetus and Newborn, & Section on Anesthesiology (2016). Prevention and Management of Procedural Pain in the Neonate: An Update. *Pediatrics, 137*, 1-13.

Ahn, Y. (2006). The relationship between behavioral states and pain responses to various NICU procedures in premature infants. *Journal of Tropical Pediatrics, 52*, 201-205.

Anand, K. J. S. (2007). Pain Assessment in Preterm Neonates. *Pediatrics, 119*, 605-607.

Anand, K. J. S. (2017). Defining pain in newborns: need for a uniform taxonomy? *Acta Paediatr, 106*, 1438-1444.

Anand, K. J. S., Eriksson, M., Boyle, E. M., Avila-Alvarez, A., Andersen, R. D., Sarafidis, K., et al. (2017). Assessment of continuous pain in newborns admitted to NICUs in 18 European countries. *Acta Paediatr, 106*, 1248-1259.

Anand, K. J. S., & Hickey, P. R. (1987). Pain and its effects in the human neonate and fetus. *The New England Journal of Medicine, 317*, 1321-1329.

Avila-Alvarez, A., Carbajal, R., Courtois, E., Pertega-Diaz, S., Anand, K. J., Muniz-Garcia, J., et al. (2016). [Clinical assessment of pain in Spanish Neonatal Intensive Care Units]. *An Pediatr (Barc), 85*, 181-188.

Badr, L. K. (2013). Pain in Preamture Infants: What is Conclusive Evidence and What Is Not. *Newborn & Infant Nursing Reviews, 13*, 82-86.

Badr, L. K., Abdallah, B., Hawari, M., Sidani, S., Kassar, M., Nakad, P., et al. (2010). Determinants of Preamture Infant Pain Responses to Heel Sticks. *Pediatric nursing, 36*, 129-136.

Beggs, S. (2015). Long-Term Consequences of Neonatal Injury. *Can J Psychiatry, 60*, 176-180.

Beggs, S., Currie, G., Salter, M. W., Fitzgerald, M., & Walker, S. M. (2012). Priming of adult pain responses by neonatal pain experience: maintenance by central neuroimmune activity. *Brain, 135*, 404-417.

Beggs, S., Torsney, C., Drew, L. J., & Fitzgerald, M. (2002). The postnatal reorganization of primary afferent input and dorsal horn cell receptive fields in the rat spinal cord is an activity-dependent process. *European Journal of Neuroscience, 16*, 1249-1258.

Bhutta, A. T., & Anand, K. J. S. (2002). Vulnerability of the developing brain. Neuronal mechanisms. *Clinics in Perinatology, 29*, 357-372.

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A., Narwal, R., et al. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet, 379*, 2162-2172.

Boettcher, M., Göttler, S., Stoffel, L., Schwab, K., Berger, S., & Mérat, M. (2012). Schmerzmanagement bei Kindern in der Schweiz. *Monatsschrift für Kinderheilkunde, 160*, 887-894.

Brummelte, S., Grunau, R. E., Chau, V., Poskitt, K. J., Brant, R., Vinall, J., et al. (2012). Procedural pain and brain development in premature newborns. *Ann Neurol., 71*, 385-396.

Bührer, C., Grimmer, I., Metze, B., & Obladen, M. (2000). The CRIB (Clinical Risk Index for Babies) score and neurodevelopmental impairment at one year corrected age in very low birth weight infants. *Intensive Care Medicine, 26*, 325-329.

Bundesamt für Statistik (2018). Gesundheit der Neugeborenen. Retrieved from https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/gesundheitszustand/gesundheit-neugeborenen.html (17 March 2019)

Carbajal, R., Paupe, A., Hoenn, E., Lenclen, R., & Olivier-Martin, M. (1997). DAN: une échelle comportementale d'évaluation de la douleur aiguë du nouveau-né. *Arch Pédiatr, 4*, 623-628.

Carbajal, R., Rousset, A., Danan, C., Coquery, S., Nolent, P., & Ducrocq, S. (2008). Epidemiology and Treatment of Painful Procedures in Neonates in Intensive Care Units. *JAMA, 300*, 60-70.

Carter, B. S., & Brunkhorst, J. (2017). Neonatal pain management. *Semin Perinatol, 41*, 111-116.

Cignacco, E., Denhaerynck, K., Nelle, M., Bührer, C., & Engberg, S. (2009). Variability in pain response to a non-pharmacological intervention across repeated routine pain exposure in preterm infants: a feasibility study. *Acta Paediatr, 98*, 842-846.

Cignacco, E., Hamers, J., Stoffel, L., Van Lingen, R. A., Schütz, N., Müller, R., et al. (2008). Routine procedures in NICUs: factors influencing pain assessment and ranking by pain intensity. *Swiss Med Weekly, 138*, 484-491.

Cignacco, E., Hamers, J., van Lingen, R. A., Stoffel, L., Büchi, S., Müller, R., et al. (2009). Neonatal procedural pain exposure and pain management in ventilated preterm infants during the first 14 days of life. *Swiss Med Weekly, 139*, 226-232.

Cignacco, E., Mueller, R., Hamers, J. P. H., & Gessler, P. (2004). Pain assessment in the neonates using the Bernese Pain Scale for Neonates. *Early Human Development, 78*, 125-131.

Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. (2017). Individual contextual factors in the validation of the Bernese pain scale for neonates: protocol for a prospective observational study. *BMC Pediatr, 17*, 171.

Cong, X., McGrath, J. M., Cusson, R. M., & Zhang, D. (2013). Pain Assessment and Measurement in Neonates: an updated Review. *Advances in Neonatal Care, 13*, 379-395.

Cong, X., McGrath, J. M., Delaney, C., Chen, H., Liang, S., Vazquez, V., et al. (2014). Neonatal nurses' perceptions of pain management: survey of the United States and China. *Pain Manag Nurs, 15*, 834-844.

Cornelissen, L., Fabrizi, L., Patten, D., Worley, A., Meek, J., Boyd, S., et al. (2013). Postnatal Temporal, Spatial and Modality Tuning of Nociceptive Cutaneous Flexion Refelxes in Human Infants. *PLoS ONE, 8*, e76470.

Craig, K. D. (2009). The Social Communication Model of Pain. *Canadian Psychology-Psychologie Canadienne, 50*, 22-32.

Craig, K. D., Korol, C. T., & Pillai, R. R. (2002). Challenges of judging pain in vulnerable infants. *Clin Perinatol, 29*, 445-457.

Cruz, M. D., Fernandes, A. M., & Oliveira, C. R. (2016). Epidemiology of painful procedures performed in neonates: A systematic review of observational studies. *Eur J Pain, 20*, 489-498.

Doesburg, S. M., Chau, C. M., Cheung, T. P. L., Moiseev, A., Ribary, U., Herdman, A. T., et al. (2013). Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. *Pain, 154,* 1946-1952.

Fabrizi, L., Slater, R., Worley, A., Meek, J., Boyd, S., Olhede, S., & Fitzgerald, M. (2011). A Shift in Sensory Processing that Enables the Developing Human Brain to Discriminate Touch from Pain. *Current Biology, 21*, 1552-1558.

Fitzgerald, C., Millard, C., & McIntosh, N. (1989). Cutaneous hypersensitivity following peripheral tissue damage in newborn infants and its reversal with topical anaesthesia *Pain, 39*, 31-36.

Fitzgerald, M. (2005). The development of nociceptive circuits. *Nature Reviews/Neuroscience, 6,* 507-520.

Fitzgerald, M. (2015). What do we really know about newborn infant pain? *Experimental Physiology, 100*, 1451-1457.

Foreman, S. W., Thomas, K. A., & Blackburn, S. T. (2008). Individual and gender differences matter in preterm infant state development. *J Obstet Gynecol Neonatal Nurs, 37*, 657-665.

Franck, L. S., & Bruce, E. (2009). Putting Pain Assessment into Practice: Why Is It so Painful? *Pain Research and Management, 14*, 13-20.

Gibbins, S., Stevens, B., Beyene, J., Chan, P. C., Bagg, M., & Asztalos, E. (2008). Pain behaviours in Extremely Low Gestational Age infants. *Early Human Development, 84*, 451-458.

Gibbins, S., Stevens, B., McGrath, P. J., Yamada, J., Beyene, J., Breau, L., et al. (2008). Comparison of Pain Responses in Infants of Different Gestational Ages. *Neonatology, 93*, 10-18.

Gibbins, S., Stevens, B. J., Yamada, J., Dionne, K., Campbell-Yeo, M., Lee, G., et al. (2014). Validation of the Premature Infant Pain Profile-Revised (PIPP-R). *Early Human Development, 90*, 189-193.

Goksan, S., Hartley, C., Ernery, F., Cockril, N., Roorun, R., Moultrie, F., et al. (2015). fMRI reveals neural activtiy overlap between adults and infant pain. *eLIFE, 4*, e06356.

Green, G., Hartley, C., Hoskin, A., Duff, E., Shriver, A., Wilkinson, D., et al. (2019). Behavioural discrimination of noxious stimuli in infants is dependent on brain maturation. *Pain, 160*, 493-500.

Grunau, R. E. (2013). Neonatal Pain in Very Preterm Infants: Long-Term Effects on Brain, Neurodevelopment and Pain Reactivity. *Rambam Maimonides Medical Journal, 4*, e0025.

Grunau, R. E., Holsti, L., Haley, D. W., Oberlander, T., Weinberg, J., Solimano, A., et al. (2005). Neonatal procedural pain exposure predicts lower cortisol and behavioral reactivity in preterm infants in the NICU. *Pain, 113*, 293-300.

Grunau, R. E., Holsti, L., & Peters, J. W. B. (2006). Long-term consequences of pain in human neonates. *Seminars in Fetal & Neonatal Medicine, 11*, 268-275.

Grunau, R. E., Oberlander, T. F., Whitfield, M. F., Fitzgerald, C., & Lee, S. K. (2001). Demographic and Therapeutic Determinants of Pain Reactivity in Very Low Birth Weight Neonates at 32 Weeks' Postconceptional Age. *Pediatrics, 107*, 105-112.

Grunau, R. E., Tu, M. T., Whitfield, M. F., Oberlander, T. F., Weinberg, J., Yu, W., et al. (2010). Cortisol, behavior, and heart rate reactivity to immunization pain at 4 months corrected age in infants born very preterm. *Clinical Journal of Pain, 26*, 698-704.

Grunau, R. E., Whitfield, M. F., Petrie-Thomas, J., Synnes, A. R., Cepeda, I. L., Keidar, A., et al. (2009). Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants. *Pain, 143*, 138-146.

Hadjistavropoulos, T., Breau, L., & Craig, K. D. (2011). Assessment of pain in adults and children with limited ability to communicate. In D. C. Turk & R. Melzack (Eds.), *Handbook of pain assessment, 3rd ed.* (pp. 260-282) New York: Guilford.

Hall, R. W. (2012). Anesthesia and Analgesiain the NICU. *Clin Perinatol, 39*, 239-254.

Hall, R. W., & Anand, K. J. (2014). Pain management in newborns. *Clin Perinatol, 41*, 895-924.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol, 8*, 23-34.

Hatfield, L. A., & Ely, E. A. (2015). Measurement of Acute Pain in Infants: A Review of Behavioral and Physiological Variables. *Biological Research for Nursing, 17*, 100-111.

Hatfield, L. A., Murphy, N., Karp, K., & Polomano, R. C. (2019). A Systematic Review of Behavioral and Environmental Interventions for Procedural Pain Management in Preterm Infants. *J Pediatr Nurs, 44*, 22-30.

Holditch-Davis, D., Brandon, D. H., & Schwartz, T. (2003). Development of behaviors in preterm infants: relation to sleeping and waking. *Nurs Res, 52*, 307-317.

Holsti, L., & Grunau, R. E. (2007). Initial Validation of the Behavioral Indicators of Infant Pain (BIIP). *Pain, 132*, 264-272.

Holsti, L., Grunau, R. E., Oberlander, T. F., & Osiovich, H. (2008). Is it painful or not? Discriminant validity of the behavioral indicators of infant pain (BIIP) scale. *Clinical Journal of Pain, 24*, 83-88.

Holsti, L., Grunau, R. E., Oberlander, T. F., & Whitfield, M. F. (2004). Specific Newborn Individualized Developmental Care and Assessment Program Movements Are Associated With Acute Pain in Preterm Infants in the Neonatal Intensive Care Unit. *Pediatrics, 114*, 65-72.

Hudson-Barr, D., Capper-Michel, B., Lambert, S., Palermo, T. M., Morbeto, K., & Lombardo, S. (2002). Validation of the Pain Assessment in Neonates (PAIN) scale with the Neonatal Infant Pain Scale (NIPS). *Neonatal Netw, 21*, 15-21.

Hummel, P., Puchalski, M., Creech, S. D., & Weiss, M. G. (2008). Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. *J Perinatol, 28*, 55-60.

Hummel, P., & van Dijk, M. (2006). Pain assessment: current status and challenges. *Semin Fetal Neonatal Med, 11*, 237-245.

International Association for the Study of Pain (2017). IASP Terminology. Retrieved from https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698 (17 March 2019)

Johnston, C. C., Fernandes, A. M., & Campbell-Yeo, M. (2011). Pain in neonates is different. *Pain, 152*, S65-S73.

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

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

Johnston, C. C., & Stevens, B. J. (1996). Experience in a Neonatal Intensive Care Unit Affects Pain Response. *Pediatrics, 98*, 925-930.

Johnston, C. C., Stevens, B. J., Franck, L. S., Jack, A., Stremler, R., & Platt, R. (1999). Factors Explaining Lack of Response to Heel Stick in Preterm Newborns. *JOGNN, 28*, 587-594.

Kostandy, R., Cong, X., Abouelfettoh, A., Bronson, C., Stankus, A., & Ludington, S. M. (2008). Effect of Kangaroo Care (skin contact) on crying response to pain in preterm neonates. *Pain Manag Nurs, 9*, 55-65.

Lawson, J. R. (1986). Letter. *Birth, 13*, 124-125.

Lee, G. Y., & Stevens, B. J. (2014). Neonatal and infant pain assessment. In P. McGrath, B. Stevens, S. M. Walker, & W. T. Zempsky (Eds.), *Oxford Textbook of Paediatric Pain* (pp. 353-369). Oxford: Oxford University Press.

Lucas-Thompson, R., Townsend, E. L., Gunnar, M. R., Georgieff, M. K., Guiang, S. F., Ciffuentes, R. F., et al. (2008). Developmental Changes in the Responses of Preterm Infants to a Painful Stressor. *Infant Behav Dev., 31*, 614-623.

Merkel, S. I., Voepel-Lewis, T., Shayevitz, J. R., & Malviya, S. (1997). The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatr Nurs, 23*, 293-297.

Morison, S. J., Holsti, L., Grunau, R. E., Whitfield, M. F., Oberlander, T. F., Chan, H. W. P., et al. (2003). Are there developmentally distinct motor indicators of pain in preterm infants? *Early Human Development, 72*, 131-146.

World Health Organization (2018). Preterm birth. Retrieved from https://www.who.int/en/news-room/fact-sheets/detail/preterm-birth (17 March 2019)

Pillai Riddell, R., Fitzgerald, M., Slater, R., Stevens, B., Johnston, C., & Campbell-Yeo, M. (2016). Using only behaviours to assess infant pain: a painful compromise? *Pain, 157*, 1579-1580.

Polkki, T., Korhonen, A., & Laukkala, H. (2018). Nurses' perceptions of pain assessment and management practices in neonates: a cross-sectional survey. *Scand J Caring Sci, 32*, 725-733.

Ranger, M., Chau, C. M. Y., Garg, A., Woodward, T. S., Beg, M. F., Bjornson, B., et al. (2013). Neonatal pain-related stress predicts cortical thickness at age 7 years in children born very preterm. *PLoS ONE, 8*, e76702.

Ranger, M., & Grunau, R. E. (2015). How do babies feel pain? *eLIFE, 4*, e07552.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/ (17 March 2019)

Relland, L. M., Gehred, A., & Maitre, N. L. (2019). Behavioral and physiological signs for pain assessment in preterm and term neonates during a nociception-specific response: a systematic review. *Pediatric Neurology, 90*, 13-23.

Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., et al. (2019). The influence of gestational age in the psychometric testing of the Bernese Pain Scale for Neonates. *BMC Pediatr, 19*, 20.

Schenk, K, Stoffel, L., Bürgin, R., Stevens, B., Bassler, D. Schulzke, S. et al. (submitted). The Bernese Pain Scale for Neonates-Revised accounts for individual contextual factors. *Manuscript submitted for publication*.

Schneider, J., Duerden, E. G., Guo, T., Ng, K., Hagmann, P., Bickle Graz, M., et al. (2018). Procedural pain and oral glucose in preterm neonates: brain development and sex-specific effects. *Pain, 159*, 515-525.

Sellam, G., Cignacco, E. L., Craig, K. D., & Engberg, S. (2011). Contextual factors influencing pain response to heelstick procedures in preterm infants: What do we know? A systematic review. *European Journal of Pain, 15*, 661.e661-661.e615.

Sellam, G., Engberg, S., Denhaerynck, K., Craig, K. D., & Cignacco, E. L. (2013). Contextual factors associated with pain response of preterm infants to heel-stick procedures. *European Journal of Pain, 17*, 255-263.

Simons, S. H. P., & Tibboel, D. (2006). Pain perception development and maturation. *Seminars in Fetal & Neonatal Medicine, 11*, 227-231.

Slater, R., Cantarella, A., Franck, L., Meek, J., & Fitzgerald, M. (2008). How Well Do Clinical Pain Assessment Tools Reflect Pain in Infants? *PLoS Medicine, 5*, 928-933.

Slater, R., Cantarella, A., Gallella, S., Worley, A., Boyd, S., Meek, J., et al. (2006). Cortical pain responses in human infants. *The Journal of Neurosciences, 26*, 3662-3666.

Stevens, B., Yamada, J., Ohlsson, A., Haliburton, S., & Shorkey, A. (2016). Sucrose for analgesia in newborn infants undergoing painful procedures (Review). *Cochrane Database of Systematic Review, 7*, CD001069.

Stevens, B. J., Abbott, L. K., Yamada, J., Harrison, D., Stinson, J., Taddio, A., et al. (2011). Epidemiology and management of painful procedures in children in Canadian hospitals. *CMAJ, 183*, E403-410.

Stevens, B. J., Gibbins, S., Yamada, R. N., Dionne, K., Lee, G., Johnston, C., et al. (2014). The Premature Infant Pain Profile-Revised (PIPP-R) Initial Validation and Feasibility. *Clinical Journal of Pain, 30*, 238-243.

Stevens, B. J., Harrison, D., Rashotte, J., Yamada, J., Abbott, L. K., Coburn, G., et al. (2012). Pain assessment and intensity in hospitalized children in Canada. *J Pain, 13*, 857-865.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales, a practical guide to their development and use* (5th ed.). Oxford: University Press.

Taddio, A., Shah, V., Gilbert-MacLeod, C., & Katz, J. (2002). Conditioning and Hyperalgesia in Newborns Exposed to Repeated Heel Lances. *JAMA: the journal of the American Medical Association, 288*, 857-861.

Unruh, A. M., & McGrath, P. J. (2014). History of pain in children. In P. J. McGrath, B. J. Stevens, S. M. Walker, & W. T. Zempsky (Eds.), *Oxford Textbook of Paediatric Pain* (pp. 3-11). Oxoford: University Press.

Välitalo, P. A. J., Van Dijk, M., Krekels, E. H. J., Gibbins, S., Simons, S. H. P., Tibboel, D., et al. (2016). Pain and distress caused by endotracheal suctioning in neonates is better quantified by behavioural than physiological items: a comparison based on item response theory modelling. *Pain, 157*, 1611-1617.

Vederhus, B. J., Eide, G. E., & Natvig, G. K. (2006). Psychometric testing of a Norwegian version of the Premature Infant Pain Profile: an acute pain assessment tool. A clinical validation study. *Int J Nurs Pract, 12*, 334-344.

Vinall, J., & Grunau, R. E. (2014). Impact of repeated procedural pain-related stress in infants born very preterm. *Pediatric Research, 75*, 584-587.

Walker, S. M., & Baccei, M. L. (2014). Nociceptive signalling in the periphery and spinal cord. In J. M. McGrath, B. J. Stevens, S. M. Walker, & W. T. Zempsky (Eds.), *Oxford Textbook of Paediatric Pain* (pp. 53-64). Oxford: Oxford University Press.

Walker, S. M., Beggs, S., & Baccei, M. L. (2016). Persistent changes in peripheral and spinal nociceptive processing after early tissue injury. *Exp Neurol*, 253-260.

Walker, S. M., Fitzgerald, M., & Hathway, G. J. (2015). Surgical injury in the neonatal rat alters the adult patttern of descending modulation from the rostroventral medulla. *Anesthesiology, 122*, 1391-1400.

Werth, J., Atallah, L., Andriessen, P., Long, X., Zwartkruis-Pelgrim, E., & Aarts, R. M. (2017). Unobtrusive sleep state measurements in preterm infants - A review. *Sleep Med Rev, 32*, 109-122.

Williams, A. C., & Craig, K. D. (2016). Updating the definition of pain. *Pain, 157*(11), 2420-2423.

Williams, A. L., Khattak, A. Z., Garza, C. N., & Lasky, R. E. (2009). The behavioral pain response to heelstick in preterm neonates studied longitudinally: Description, development, determinants, and components. *Early Human Development, 85*, 369-374.

Appendix I: The Bernese Pain Scale for Neonates-Revised

Pain indicators	0	1	2	3	Score
Crying	Not crying	Brief period of crying (less than 2 minutes)	Increased crying (more than 2 minutes)	Increased, shrill crying (more than 2 minutes)	
Facial expression	Face relaxed	Brief grimace	Increase grimace and trembling of chin	Permanent grimace of face and trembling of chin	
Posture	Body relaxed	Mainly relaxed, short bouts of tension	Frequent bouts of tension but relaxation possible	Permanently tense	
Heart rate (bpm) Baseline score: _____	Normal (Baseline)	Increase of 20 bpm or more over the baseline **with** return to baseline within 2 minutes	Increase of 20 bpm or more over baseline **without** return to baseline within 2 minutes	Increase of 30 bpm or more over baseline or more frequent episodes of bradycardia within 2 minutes	
				Subtotal →	
				Subtotal ≥ 1: Contextual factors need to be added!	
Contextual factors	**0**	**1**	**2**	**3**	**Score**
Postmenstrual age (GA + number of days since birth)	Full-term neonates (≥ 37 0/7 weeks)	Moderate to late preterm neonates (32 0/7 – 36 6/7 weeks)	Very preterm neonates (28 0/7 – 31 6/7 weeks)	Extremely preterm neonates (< 28 weeks)	
Behavioural state (baseline)	Active (awake or asleep)	Quiet (awake or asleep)			
Ventilation status	CPAP or no ventilation	Mechanical ventilation			
				Overall Total →	

Overall total = Subtotal of pain indicators + score of contextual factors (if subtotal of pain indicators ≥ 1).

0-4 points = **no pain** or no observable pain reaction
≥ 5 points = **pain**

Procedure:

1. Observation of the neonate during the **baseline phase** for 15 seconds: Assessment of the highest heart rate and behavioural state
2. Observation of the neonate **during the procedure** for **2 minutes**: Assessment of the three behavioural pain indicators (crying, facial expression, posture) and the highest heart rate.
3. Calculation of the sub-total based on the 4 pain indicators.
4. If **subtotal ≥ 1 point**, evaluation of the three contextual factors and calculation of the overall total. **Overall total = subtotal + contextual factors**

Appendix II: Manuscript 1

Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. (2017). Individual contextual factors in the validation of the Bernese pain scale for neonates: protocol for a prospective observational study. *BMC Pediatrics, 17*(171). doi: 10.1186/s12887-017-0914-9

Cignacco et al. BMC Pediatrics (2017) 17:171
DOI 10.1186/s12887-017-0914-9

BMC Pediatrics

STUDY PROTOCOL Open Access

Individual contextual factors in the validation of the Bernese pain scale for neonates: protocol for a prospective observational study

Eva Cignacco[1*], Karin Schenk[1], Bonnie Stevens[2], Liliane Stoffel[3], Dirk Bassler[4], Sven Schulzke[5] and Mathias Nelle[6]

Abstract

Background: The Bernese Pain Scale for Neonates (BPSN) is a multidimensional pain assessment tool that is already widely used in clinical settings in the German speaking areas of Europe. Recent findings indicate that pain responses in preterm neonates are influenced by individual contextual factors, such as gestational age (GA), gender and the number of painful procedures experienced. Currently, the BPSN does not consider individual contextual factors. Therefore, the aim of this study is the validation of the BPSN using a large sample of neonates with different GAs. Furthermore, the influence of individual contextual factors on the variability in pain reactions across GA groups will be explored. The results will be used for a modification of the BPSN to account for individual contextual factors in future clinical pain assessment in neonates.

Methods and design: This prospective multisite validation study with a repeated measures design will take place in three university hospital neonatal intensive care units (NICUs) in Switzerland (Bern, Basel and Zurich). To examine the impact of GA on pain responses and their variability, the infants will be stratified into six GA groups ranging from 24 0/7 to 42 0/7. Among preterm infants, 2–5 routine capillary heel sticks within the first 14 days of life, and among full-term infants, two heel sticks during the first days of life will be documented. For each heel stick, measurements will be video recorded for each of three phases: baseline, heel stick, and recovery. The infants' pain responses will be rated according to the BPSN by five nurses who are blinded as to the number of each heel stick and as to the measurement phases. Individual contextual factors of interest will be extracted from patient charts.

Discussion: Understanding and considering the influence of individual contextual factors on pain responses in a revised version of the BPSN will help the clinical staff to more appropriately assess pain in neonates, particularly preterm neonates hospitalized in NICUs. Pain assessment is a first step toward appropriate and efficient pain management, which itself is an important factor in later motor and cognitive development in this vulnerable patient population.

Trial registration: The study is registered in the database of Clinical Trial gov. Study ID-number: NCT 02749461. Registration date: 12 April 2016.

Keywords: Pain assessment, Premature infants, Contextual factors, Diagnostic

* Correspondence: eva.cignacco@bfh.ch
[1]Health Department, Midwifery Discipline, Bern University of Applied Sciences, Murtenstrasse 10, 3008 Bern, Switzerland
Full list of author information is available at the end of the article

Background

In order to ensure their survival, premature born infants hospitalized in a neonatal intensive care unit (NICU) are subjected to many painful diagnostic and therapeutic procedures [1–3]. Although there have been efforts in recent years to quantify, and most importantly, reduce the number of procedural exposures to pain in preterm infants, procedural acute pain remains a challenge in the NICU setting [3–5]. Often, these treatment interventions take place during a crucial period in the development of the nociceptive and central nervous systems [6–8]. There is more and more alarming evidence that repeated painful stimuli at this early age may induce both structural and functional reorganization of the nervous system [7, 9–13] and result in an altered pain response [14–16]. As a consequence, the motor and cognitive development of premature infants may be impaired [9, 13, 17–22]. In premature infants requiring intensive care, the frequency of exposure to pain and systematic implementation of preventive pain measures are therefore of key importance for their later development [4, 5]. Accurate pain measurement is the first step toward effective pain management.

Pain assessment in neonates

Clinical pain assessment in neonates, particularly those delivered preterm, is highly challenging [4, 23]. In the clinical setting, their pain responses have to be observed and assessed using behavioral and physiological indicators, which can vary across premature infants depending on their physiological and neurological development stages [23]. Behavioral indicators used as pain assessment tools include body movements, facial expressions and crying [24]. Some pain assessments also include behavior status indicators, e.g., sleep-wake state [25, 26]. Physiological responses to pain include, for instance, changes in heart rate, respiratory rate, blood pressure, oxygen saturation, vagal tone, and peripheral blood flow [25, 27]. Recently, researchers have begun to investigate more objective approaches to pain assessment, such as measurement of heart rate variability, skin conductance and cortisol as a biomarker of stress [23, 25]. To better understand and assess neonatal pain responses at cortical level, newer brain-oriented techniques, such as electroencephalography (EEG) [28, 29] and functional magnetic resonance imaging (fMRI) [30, 31], are used [11, 32–34]. However, for systematic clinical pain assessment, exclusively observable indicators need to be considered.

Because of the complex nature of pain, multidimensional pain measures that include behavioral and physiological indicators are generally assumed to be most appropriate for the clinical setting [23]. Although most infants show both types of pain response indicators, the correlation between these two indicators is often low [25, 35]. Moreover, no consistent associations between behavioral, physiological and cortical measures of pain have been detected so far [36]. In the face of inconclusive associations between different indicators of pain, the validity of existing multidimensional tools and their choices of indicators are currently being questioned, and, to date, no universally accepted gold standard exists for neonatal pain assessment [23].

More than 40 pain assessment scales for premature and full-term infants exist to date [25, 37]. The majority were designed for research purposes and are inappropriate for routine clinical procedures (e.g., because they require extended observation periods) [25, 38]. Furthermore, only a few have undergone extensive psychometric testing and are both reliable and valid [25, 39]. Of the pain assessment scales compiled for clinical application, few have been validated in premature infants and even fewer consider individual contextual factors, e.g. gestational age (GA) and health status [23, 40].

The Bernese pain scale for neonates

The Bernese Pain Scale for Neonates (BPSN; [41]) was developed by nurses of the University Hospital of Berne primarily for clinical use. Since its development in 1996, it has been widely used for bedside pain assessment in NICUs in the German speaking areas of Europe. Several hospitals in Switzerland have fully integrated the BPSN into their daily routine.

The BPSN is a 9-item multidimensional pain assessment tool that includes behavioral and physiological indicators. The instrument consists of seven subjective (alertness, crying, consolation, skin color, facial expression, posture, and changes in respiratory rate) and two physiological (i.e. objective) (changes in heart rate and oxygen saturation) indicators. Each item is rated on a four point Likert scale (0, 1, 2, and 3). Higher scores indicate greater pain-related distress, and a total score of 11 or higher is considered to indicate pain.

In the year 2004, the BPSN was validated to differentiate between pain and non-pain status in neonates between 27 and 41 weeks of gestation [41]. The results suggested that the BPSN is a valid and reliable pain assessment instrument for assessing acute pain in term and preterm neonates. A shortcoming of this first validation study of the BPSN is the small study population of 12 infants. Furthermore, increasing evidence indicates that pain reactions of neonates are probably influenced by more than noxious stimulation alone; individual contextual factors might also impact pain reactivity [40, 42–44]. Currently, the BPSN focuses entirely on physiological and behavioral indicators.

Individual contextual factors

Individual contextual factors encompass individual infant characteristics (e.g., GA, gender, health status, and

weight), previous pain experience, or the duration of hospitalization [23, 44]. The variability in pain responses between and within premature infants as well as the low association between behavioral and physiological pain responses may be explained by the influence of individual contextual factors [35, 42, 45, 46].

Neonatal age is the most commonly examined individual contextual factor associated with neonatal pain response [44]. Premature neonates generally seem more sensitive to painful stimulation than full-term newborns. In addition to having low reflex thresholds [47, 48], newborns lack the inhibitory control that mature brain structures would exert [49]. As a result, premature neonates display diffuse responses to noxious stimuli rather than more complex affective reactions [50]. Moreover, the association between behavioral and physiological stress responses may differ depending on GA [35]. Although older GA infants displayed a positive association between the extent of behavioral pain reaction and heart rate levels, Lucas-Thompson et al. (2008) found no association between physiological and behavioral responses in the youngest GA infants. Despite the high variability in behavioral and physiologic pain responses in premature neonates, their responses are less intense [42, 45, 51, 52].

The results of several studies suggest that facial expression in response to pain increases with GA [45, 52–55]. This difference is manly influenced by the older infants' increased facial expressiveness, which results from their more developed nervous system and facial muscles [53, 54]. In contrast, several studies have reported no significant relationship between GA and facial expression in response to pain [44, 56]. However, the consideration of reduced facial movement in response to pain in premature neonates is important. Using pain assessment scales which rely only on facial expressions may lead clinicians to the incorrect conclusion that younger premature infants do not feel or feel less pain [57]. In addition, the presence of endotracheal tubes in premature neonates impedes using facial reaction and crying as indicators of pain because endotracheal tubes are typically secured by taping them to the skin of the face [52, 54, 57]. Therefore, the consideration of other behavioral pain indicators encoded in specific body movements (e.g., hand on face), may provide further information about pain in premature infants with extremely low GA [52, 56, 58].

Several studies have examined the influence of previous pain exposure on reaction to pain, but the findings do not provide a clear answer [44]. Some studies report that infants subjected to frequent painful procedures during their hospitalization display less intense behavioral responses to heel sticks than those who have undergone fewer procedures [46, 52, 59]. The dampened pain responses in very premature neonates may be a sign of exhaustion or a state of passivity resulting from the numerous procedures they experience during their stay in a NICU [43, 60, 61]. Contrary to those findings, other studies suggest that repeated exposure to pain may lead either to increased pain response (hyperalgesia) or to pain responses without painful stimulus (allodynia) [15, 62].

Few studies have investigated the influence of other contextual factors (e.g., gender, health status) on pain reactions in neonates, and of those that have, the results are inconsistent [44]. This might be explained by methodological limitations (e.g. the comparison of different GA groups and the use of a variety of pain assessment tools) [44]. One challenge in examining the influence of contextual factors on pain response is the associations between the individual factors [44]; for example, extremely low GA infants have a longer stay in a NICU and are exposed to a higher number of painful procedures than more mature infants. Due to the fact that contextual factors can lead to underestimation or misjudgment of pain severity [54, 63–65], further research is needed to better understand the factors that influence pain responses in neonates. Relevant contextual factors should also be considered in future pain assessment.

Study aims

The aim of this observation study is the validation of the BPSN, using a large sample of neonates spanning a full range of GAs. The validation will involve the detection of the underlying structure of the data and the examination of the concurrent validity of the BPSN with the Premature Infant Pain Profile-Revised (PIPP-R; [26]), construct validity, interrater reliability, specificity and sensitivity. Furthermore, the variability of pain reactions over time related to behavioral and physiological patterns will be analyzed and the relationship between behavioral and physiological indicators examined. In addition, the influence of contextual factors on the variability of pain reactions across GA groups will be explored. Finally, the results of this analysis will be used for modification of the BPSN, to account for individual contextual factors in future clinical pain assessment in neonates.

Based on a previous validation study of the BPSN [41], we hypothesize that the BPSN will be a valid and reliable pain assessment tool for premature and term infants. In addition, we expect that the impact of single contextual factors on infants' pain reaction will be described and considered for future pain assessment. In particular, we anticipate finding a difference in pain reaction depending on GA. Moreover, we hypothesize that behavioral and physiological indicators will show low association across time and that this low association may be explained by the influence of individual contextual factors.

Methods

This prospective multisite validation study focuses on psychometric testing of the BPSN and involves repeated

measurement design. The study will take place in three university hospital NICUs in Switzerland (Basel, Bern and Zurich).

In total, 150 preterm and healthy-term infants hospitalized in a NICU will be included. Consecutive sampling will be used to recruit subjects and the infants will be stratified according to GA at birth (Fig. 1). Stratification is based on the assumption that premature neonates with a lower GA will show a higher variability in pain responses, due to their neurological immaturity, than will premature neonates with a higher GA and full-term infants [42]. Therefore, larger sample sizes of premature infants with GAs between 24 0/7 and 29 6/7 weeks (n = 102) will be included, compared to the samples of those with GAs between 30 0/7 and 42 0/7 weeks (n = 48).

Inclusion and exclusion criteria

Premature infants born between 24 0/7 and 36 6/7 weeks of gestation will be included if they are expected to undergo 2–5 routine capillary heel sticks during the first 14 days of life. Full-term infants born between 37 0/7 and 42 0/7 weeks of gestation will be included if they are expected to have at least 2 routine capillary blood samplings during their first days of life. Furthermore, signed consent is needed from the infant's parents, who have to understand either German or French.

Infants will be excluded if they have suffered a high-grade intraventricular hemorrhage (grades III and IV), if they have a severe life-threatening malformation or suffer from any condition involving partial or total loss of sensitivity, if they have had an arterial cord pH < 7.15, if they have had surgery for any reason, or if they have a congenital malformation affecting brain circulation and/or cardiovascular system. Infants treated with continuous positive airway pressure (CPAP) or mechanical ventilation will be included if they meet the other inclusion and exclusion criteria.

Recruitment and data collection procedures

In each study center, a trained study assistant will identify potentially eligible infants and inform the parents about the study both verbally and via printed information material. Interested parents will receive the information material and a copy of the informed consent form to read. A member of the research team will answer any parental questions about the study. No study procedures will be performed until a signed informed consent form is obtained from the child's parents.

After written consent has been received, the neonate will be videotaped (using a HC-V757 high-definition camcorder manufactured by Panasonic, Osaka, Japan) during his or her next 2–5 routine capillary heel sticks. Before each heel stick procedure, every infant will receive a dose of 24% oral sucrose (0.2 ml/kg bodyweight) as a pain relieving intervention in accordance with standards of care [66]. Video sequences and physiological variables will be recorded continuously from 2 to 3 min before the beginning of the heel stick procedure (baseline phase), through the heel stick (heel stick phase) and until 2–3 min after the heel stick (recovery phase). Therefore, three rating sequences will be produced for each heel stick. The camera operator will begin each video sequence by focusing on the face of the neonate for at least one minute to allow adequate assessment of facial activity and cry. Then, the infant's body will be recorded for another minute. For healthy-term infants, six video sequences per infant will be produced, resulting in 96 videos (2 heel sticks * 3 phases * 16 n). For premature neonates, 2010 video sequences (5 heel sticks * 3

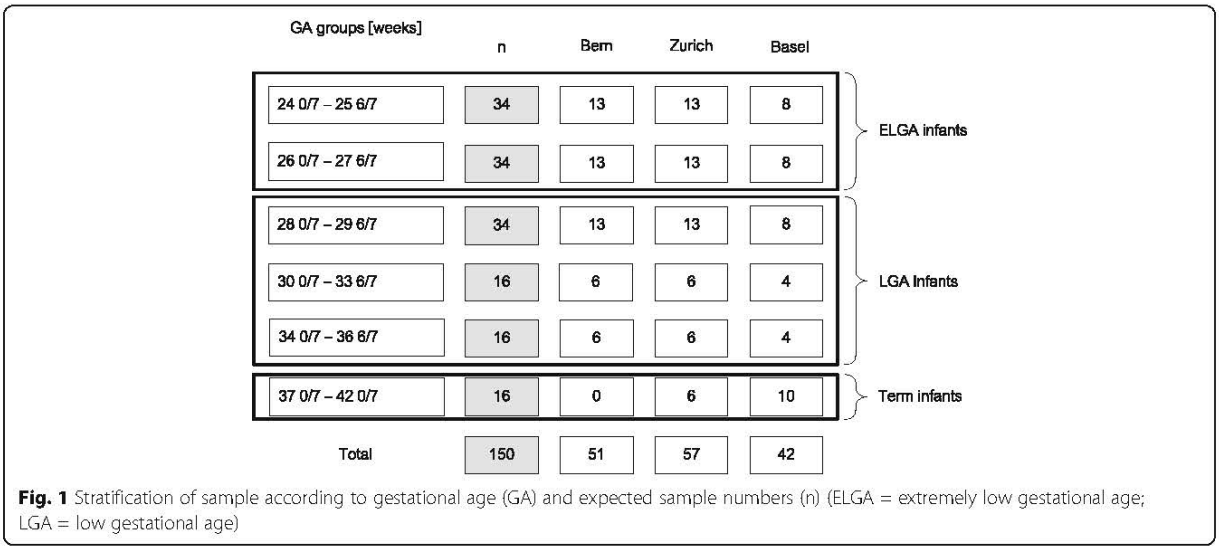

Fig. 1 Stratification of sample according to gestational age (GA) and expected sample numbers (n) (ELGA = extremely low gestational age; LGA = low gestational age)

phases * 134 n) will be produced. This will lead to a total of 2106 video sequences, all of which will be filmed by trained study collaborators. Each video sequence will be checked for quality, and digitally elaborated by trained study assistants using Final Cut Pro X (Apple Inc., Cupertino, CA, USA) video editing software. To preserve rater blindness, any information that could indicate the heel stick phase to the raters will be eliminated. Data quality and completeness of the video sequences will be controlled continuously by the doctoral student before uploading each video sequence onto a web-based rating tool. The web-based rating tool has been developed specially for the study and includes a randomizing generator. Uploaded sequences are randomized related to sequence number, phases and presentation order. Five trained nurses who are presently working in a NICU and are experienced users of the BPSN will retrieve the randomized sequences from the web-based platform and will rate the behavioral pain reaction by means of the BPSN and the PIPP-R.

Individual contextual factors will be retrieved retrospectively from patient charts by trained study assistants. All extracted data will be entered into secuTrial®, a web-based data capture system (InterActive Systems, Berlin, Germany). Five percent of the patient charts will be audited by the doctoral student to detect and correct discrepancies. Emerging questions and inconsistencies during the overall data collection process will be continuously discussed to ensure the quality of ongoing data extraction.

Measures

To establish concurrent validity, neonates' pain expression is measured by the BPSN [41] and the PIPP-R [26]. The BPSN measures 9 indicators. The two physiological indicators will be captured on an ongoing basis from the neonate's routine continuous monitoring records (heart rate and oxygen saturation) during the video recording. The six subjective indicators (sleeping state, crying, consolation, skin color, facial expression, posture, and breathing) will be rated by five independent and blinded video raters on a 4 point Likert scale. The raters are blinded towards the phase of the video sequence they are looking at (baseline, heel stick, and recovery). The PIPP-R, which is widely used in North America for assessing acute pain in neonates, measures five indicators of which two are physiological (heart rate and oxygen saturation). The three behavioral indicators (brow bulge, eye squeeze, and naso-labial furrow) will also be assessed by the five raters. Each indicator of the PIPP-R is numerically rated on a Likert scale from 0 to 3 points, with higher ratings reflecting the rater's impression of more intense pain responses. Additionally, the PIPP-R accounts for GA and baseline behavioral states as contextual factors. According to the instructions of the authors, these contextual factors need only be scored if there are changes in any of the behavioral or physiological items [26]. Neonates with the youngest GAs and those in quiet sleep receive the highest scores for these indicators. The PIPP-R scores will be used as a standard reference in this study.

Based on the findings of a systematic review [44], the following individual contextual factors will be retrieved from patient charts: demographic contextual factors, including GA at birth, gender, birth weight, nationality, parity and way of delivery; the primary diagnosis and the most common comorbidities in preterm neonates, including bronchopulmonary dysplasia, necrotizing enterocolitis, respiratory distress syndrome, patent ductus arteriosus, septic events, cardiac events and respiratory events; the health status at time of birth measured by the Clinical Risk Index for Babies (CRIB; [67]). For the time of each heel stick, the following individual contextual factors will be retrieved: postnatal age; post-menstrual age (GA at birth combined with postnatal age); weight; CPAP or mechanical ventilation at the time of the heel stick procedure; medication administered (sedatives, opioids, non-opioids, steroids, caffeine, antibiotics and catecholamines) from birth and between the recorded heel stick procedures; number of previous painful (e.g., heel stick) and non-painful (e.g., diaper change) interventions from birth and between the recorded heel stick procedures (painful and non-painful interventions were defined in a previous study [68]); number of painful and non-painful procedures in the past 24 h; time since the last painful and non-painful interventions; and, finally, type of last painful and non-painful interventions. The duration of each heel stick and the number of additional sucrose doses given during the heel stick procedures will be registered while video recording.

Data analyses

Data will be analyzed using SPSS (IBM© SPSS© Statistics Version 23.0, IBM Corp, Armonk, NY, USA) and Stata (Stata/MP 13.1, StataCorp LP, Lakeway Drive, USA). Initially, an exploratory analysis will be conducted to describe the data and uncover any anomalies that may impact the validity of the data analysis. Methods for handling missing data will be applied after considering the volume and pattern of missing data. Descriptive statistics including measures of central tendency and dispersion will be used to characterize the individual variables and to determine the distribution of the data.

Several data analyses will be used for the validation of the BPSN. An exploratory factor analysis will be performed to analyze the underlying structure of the data. Cronbach's Alpha and item-total correlations will be conducted to analyze the reliability of the scale. Furthermore,

construct validity will be examined by comparing mean measurements at each of the three rated phases (baseline, heel stick and recovery). The analysis will be performed for the total sum score of the BPSN as well as for the physiological items and the behavioral items alone. In order to determine the concurrent validity of the BPSN with the PIPP-R, the total sum scores of the two tools will be correlated. Intra-class correlation (ICC) will be used to determine interrater reliability across the 2–5 heel sticks. To test sensitivity and specificity in the BPSN, a receiver operating characteristic (ROC) curve analysis will be performed using the PIPP-R as reference value. Furthermore, the pain and non-pain cut-off values of the two instruments will be compared.

To explore and depict both temporal variability of pain reactivity between measurements of each subject, and variability between corresponding measurements of all subjects, linear mixed modeling will be applied to the behavioral and physiological data on pain reactivity. Additionally, individual contextual factors will be added to these models to test for associations with the BPSN scores. As contextual factors are highly dependent on organizational procedures, the possible confounding effect of the participating sites will also be taken into account.

In addition to analyzing the total sum scores of the BPSN, the separate physiological and behavioral subscores will be tested both against the total scores and against one another. Pearson correlation will be used as a descriptive indication of the strength of associations, while linear mixed modeling will be used to test the associations themselves.

Sample size and power

The target sample size of 150 neonates is indicated on a power analysis of the hypothesized association between the BPSN and GAs at baseline. This analysis is based on the data from a descriptive-explorative analysis (n = 23) and a previous study (n = 71; [69]), i.e., assuming an alpha of 0.05, a beta of 0.80, with at least three baseline heel sticks conducted per study infant (taking into account both intra- and inter-infant variability). Because an attrition rate of 10–15% is anticipated, approximately 170 infants will be enrolled in the study.

Discussion

The BPSN is already widely used in clinical settings in the German speaking areas of Europe. Pain assessment with the BPSN requires only two to three minutes of observation. Despite its practical application, another advantage of the BPSN is its consideration of various aspects of behavioral pain responses. Because of less intense facial reactions in premature neonates and the frequent presence of artificial respiration in this patient population, the consideration of various behavioral indicators of pain may provide further information for appropriate pain assessment. In addition, the repeated measurement design in this study will facilitate consideration of the development of pain responses across time.

The validation of the BPSN on a large sample of neonates with different gestational age and the consideration of the influence of individual contextual factors on pain reactivity should lead to a higher accuracy of routine pain assessment. A revised version of the BPSN may help the clinical staff to prevent and minimize the pain endured by neonates, particularly preterm neonates in NICUs. For preterm infants requiring intensive care, appropriate and efficient pain management is an important factor in later motor and cognitive development. This study will hopefully contribute to a more accurate pain assessment tool and to the prevention of negative long-term outcomes in this vulnerable patient population.

Abbreviations

BPSN: Bernese Pain Scale for Neonates; GA: Gestational age; NICU: Neonatal intensive care unit; PIPP-R: Premature Infant Pain Profile-Revised

Acknowledgements

We are grateful to all parents who have already consented to include their infant in the study. We acknowledge the financial contribution of the Swiss National Science Foundation. Furthermore, we thank all nurses, head nurses, and medical directors of the University Hospital NICUs in Basel, Bern, and Zurich for their support and efforts to make this study possible. Moreover, we thank Janik Schneeberger for the development of the web-based rating tool and his support in all technical acquisitions. We would further like to thank the Department of Clinical Research team of the University of Bern for the development of the web-based data capture system and the New Media Centre team of the University of Basel for their professional support in the video-data elaboration support. Furthermore, we thank Heather Murray for her editing work. Finally, the data collection in this study would not be feasible without the help and support of research assistants.

Funding

This research is funded by the Swiss National Science Foundation (SNF 320030_159573).

Availability of data and materials

Not applicable.

Authors' contributions

EC conceived and designed the study. The doctoral student KS will be responsible for all tasks of the data collection process as well as data entry, management and analysis. Furthermore, she will also report and disseminate the outcomes through peer-reviewed journals and conferences. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The study is approved by the Ethics Committee Bern (2015–238), the Ethics Committee northwest/central Switzerland EKNZ (2015–385) and the Ethics Committee Zurich (2015–563).

Written informed consent will be obtained from the parents according to the protocol approved by the ethics committees. In this study, no infant will be exposed to additional painful situations and no heel sticks will be performed solely for research purposes. Furthermore, the current standard of care in pain prevention will be upheld. All infants will receive oral sucrose before each heel stick.

Consent for publication

Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Health Department, Midwifery Discipline, Bern University of Applied Sciences, Murtenstrasse 10, 3008 Bern, Switzerland. [2]Lawrence S. Bloomberg Faculty of Nursing and Faculties of Medicine and Dentistry, University of Toronto, Toronto, Canada. [3]Neonatalogy, Children's Hospital, University Hospital of Bern, Bern, Switzerland. [4]Department of Neonatology, University Hospital Zurich and University of Zurich, Zurich, Switzerland. [5]Department of Neonatology, University of Basel Children's Hospital (UKBB), Basel, Switzerland. [6]Department of Neonatology, Children's University Hospital, Bern, Switzerland.

Received: 29 December 2016 Accepted: 29 June 2017
Published online: 19 July 2017

References

1. Carbajal R, Rousset A, Danan C, Coquery S, Nolent P, Ducrocq S, et al. Epidemiology and treatment of painful procedures in neonates in intensive care units. JAMA. 2008;300(1):60–70.
2. Cignacco E, Hamers J, van Lingen RA, Stoffel L, Büchi S, Müller R, et al. Neonatal procedural pain exposure and pain management in ventilated preterm infants during the first 14 days of life. Swiss Med Wkly. 2009; 139(15–16):226–32.
3. Johnston C, Barrington KJ, Taddio A, Carbajal R, Filion F. Pain in Canadian NICUs: have we improved over the past 12 years? Clin J Pain. 2011;27(3):225–32.
4. Amercian Academy of Pediatrics, Committee on Fetus and Newborn, Section on Anesthesiology and Pain Medicine. Prevention and Management of Procedural Pain in the Neonate: An Update. Pediatrics. 2016;137(2): e20154271.
5. Amercian Academy of Pediatrics, Committee on Fetus and Newborn, Section on Surgery, and Section on Anesthesiology and Pain Medicine, Canadian Paediatric Society, Fetus and Newborn Committee. Prevention and Management of Pain in the neonate: an update. Pediatrics. 2006; 118(5):2231–41.
6. Bhutta AT, Anand KJS. Vulnerability of the developing brain: neuronal mechanisms. Clin Perinatol. 2002;29:357–72.
7. Brummelte S, Grunau RE, Chau V, Poskitt KJ, Brant R, Vinall J, et al. Procedural pain and brain development in premature newborns. Ann Neurol. 2012;71(3):385–96. doi:10.1002/ana.22267.
8. Anand KJ, Carr DB. The neuroanatomy, neurophysiology, and neurochemistry of pain, stress, and analgesia in newborns and children. Pediatr Clin N Am. 1989;36(4):795–822.
9. Anand KJS, Scalzo FM. Can adverse neonatal experiences Alter brain development and subsequent behavior? Biol Neonate. 2000;77(2):69–82.
10. Zwicker JG, Grunau RE, Adams E, Chau V, Brant R, Poskitt KJ, et al. Score for neonatal acute physiology-II and neonatal pain predict corticospinal tract development in premature newborns. Pediatr Neurol. 2013;48:123–9.
11. Smith GC, Gutovich J, Smyser C, Pineda R, Newnham C, Tjoeng TH, et al. NICU stress is Asscociated with brain development in preterm infants. Ann Neurol. 2011;70(4):541–9. doi:10.1002/ana.22545.
12. Ranger M, Grunau RE. Early repetitive pain in preterm infants in relation to the developing brain. Pain Management. 2014;4(1):57–67. doi:10.2217/pmt.13.61.
13. Vinall J, Grunau RE. Impact of repeated procedural pain-related stress in infants born very preterm. Pediatr Res. 2014;75(5):584–7. doi:10.1038/pr.2014.16.
14. Taddio A, Katz J. The effects of early pain experience in neonates on pain responses in infancy and childhood. Paediatric Drugs. 2005;7(4):245–57.
15. Taddio A, Shah V, Gilbert-MacLeod C, Katz J. Conditioning and hyperalgesia in newborns exposed to repeated heel lances. JAMA. 2002;288(7):857–61.
16. Hermann C, Hohmeister J, Demirakca S, Zohsel K, Flor H. Long-term alteration of pain sensitivity in school-aged children with early pain experiences. Pain. 2006;125:278–85.
17. Vinall J, Miller SP, Bjornson BH, Fitzpatrick KPV, Poskitt KJ, Brant R, et al. Invasive procedures in preterm children: brain and cognitive development at school age. Pediatrics. 2014;133(3):412–21. doi:10.1542/peds.2013-1863.
18. Doesburg SM, Chau CM, Cheung TPL, Moiseev A, Ribary U, Herdman AT, et al. Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. Pain. 2013;154(10):1946–52. doi:10.1016/j.pain.2013.04.009.
19. Grunau RE, Whitfield MF, Petrie-Thomas J, Synnes AR, Cepeda IL, Keidar A, et al. Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants. Pain. 2009; 143(1–2):138–46. doi:10.1016/j.pain.2009.02.014.
20. Aarnoudse-Moens GSH, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. Pediatrics. 2009;124(2):717–28.
21. Grunau RE. Neonatal pain in very preterm infants: long-term effects on brain, neurodevelopment and pain reactivity. Rambam Maimonides medical journal. 2013;4:e0025.
22. Ranger M, Chau CMY, Garg A, Woodward TS, Beg MF, Bjornson B, et al. Neonatal pain-related stress predicts cortical thickness at age 7 years in children born very preterm. PLoS One. 2013;8(10):e76702. doi:10.1371/journal.pone.0076702.
23. Lee GY, Stevens BJ. Neonatal and infant pain assessment. In: McGrath PJ, Stevens BJ, Walker SM, Zempsky WT, editors. Oxford textbook of Paediatric pain. Oxford: Oxford University Press; 2014. p. 353–69.
24. Stevens BJ, Pillai Ridell RR, Oberlander TE, Gibbins S. Assessment of pain in neonates and infants. In: Anand KJS, Stevens BJ, McGrath PJ, editors. Pain in neonates and infants. 3rd ed. Edinburgh: Elsevier; 2007. p. 67–90.
25. Cong X, McGrath JM, Cusson RM, Zhang D. Pain assessment and measurement in neonates: an updated review. Advances in Neonatal Care. 2013;13(6):379–95.
26. Stevens BJ, Gibbins S, Yamada RN, Dionne K, Lee G, Johnston C, et al. The premature infant pain profile-revised (PIPP-R) initial validation and feasibility. Clin J Pain. 2014;30(3):238–43.
27. Brummelte S, Oberlander TF, Craig KD. Biomarkers of pain: physiological indices of pain reactivity in infants and children. In: McGrath PJ, Stevens BJ, Walker SM, Zempsky WT, editors. Oxford textbook of Paediatric pain. Oxford: Oxford University Press; 2014. p. 391–400.
28. Maimon N, Grunau RE, Cepeda IL, Friger M, Selnovik L, Gilat S, et al. Electroencephalographic activity in response to procedural pain in preterm infants born at 28 and 33 weeks gestational age. Clin J Pain. 2013;29(12):1044–9.
29. Slater R, Fabrizi L, Worley A, Meek J, Boyd S, Fitzgerald M. Premature infants display increased noxious-evoked neuronal activity in the brain compared to healthy age-matched term-born infants. NeuroImage. 2010;52:583–9.
30. Williams G, Fabrizi L, Meek J, Jackson D, Tracey I, Robertson N, et al. Functional magnetic resonance imaging can be used to explore tactile and nociceptive processing in the infant brain. Acta Paediatr. 2015;104:158–66.
31. Ranger M, Grunau RE. How do babies feel pain? elife. 2015;4:e07552.
32. Fitzgerald M. What do we really know about newborn infant pain? Exp Physiol. 2015;100(12):1451–7.
33. Holsti L, Grunau RE, Shany E. Assessing pain in preterm infants in the neonatal intensive care unit: moving to a 'brain-oriented' approach. Pain management. 2011;1(2):171–9. doi:10.2217/pmt.10.19.
34. Slater R, Cantarella A, Franck L, Meek J, Fitzgerald M. How well do clinical pain assessment tools reflect pain in infants? PLoS Med. 2008;5(6):e129.
35. Lucas-Thompson R, Townsend EL, Gunnar MR, Georgieff MK, Guiang SF, Ciffuentes RF, et al. Developmental changes in the responses of preterm infants to a painful stressor. Infant Behav Dev. 2008;31(4):614–23. doi:10.1016/j.infbeh.2008.07.004.
36. Pillai Riddell R, Fitzgerald M, Slater R, Stevens B, Johnston C, Campbell-Yeo M. Using only behaviours to assess infant pain: a painful compromise? Pain. 2016;157(8):1579–80.
37. Maxwell LG, Malavolta CP, Fraga MV. Assessment of pain in the neonate. Clin Perinatol. 2013;40(3):457–69.
38. Franck LS, Bruce E. Putting pain assessment into practice: why is it so painful? Pain Research and Management. 2009;14(1):13–20.
39. Badr LK. Pain in premature infants: what is conclusive evidence and what is not. Newborn & Infant Nursing Reviews. 2013;13:82–6.
40. Sellam G, Engberg S, Denhaerynck K, Craig KD, Cignacco EL. Contextual factors associated with pain response of preterm infants to heel-stick procedures. Eur J Pain. 2013;17:255–63.
41. Cignacco E, Mueller R, Hamers JPH, Gessler P. Pain assessment in the neonates using the Bernese pain scale for neonates. Early Hum Dev. 2004;78:125–31.
42. Cignacco E, Denhaerynck K, Nelle M, Bührer C, Engberg S. Variability in pain response to a non-pharmacological intervention across repeated routine pain exposure in preterm infants: a feasibility study. Acta Paediatr. 2009; 98(5):842–6.

43. Johnston CC, Stevens BJ, Franck LS, Jack A, Stremler R, Platt R. Factors explaining lack of response to heel stick in preterm newborns. JOGNN. 1999;28(7):587–94.
44. Sellam G, Cignacco EL, Craig KD, Engberg S. Contextual factors influencing pain response to heelstick procedures in preterm infants: What do we know? A systematic review. Eur J Pain. 2011;15:661.e1–661.e15.
45. Craig KD, Whitfield MF, Grunau RVE, Linton J, Hadjistavropoulos HD. Pain in the preterm neonate: behavioural and physiological indices. Pain. 1993;52:287–99.
46. Johnston CC, Stevens B. Experience in a neonatal intensive care unit affects pain response. Pediatrics. 1996;98:925–30.
47. Andrews K, Fitzgerald M. The cutaneous withdrawal reflex in human neonates: sensitization, receptive fields, and the effects of contralateral stimualtion. Pain. 1994;56:95–101.
48. Johnston CC, Fernandes AM, Campbell-Yeo M. Pain in neonates is different. Pain. 2011;152(3):S65–73.
49. Fitzgerald M. The development of nociceptive circuits. Nat Rev Neurosci. 2005;6(7):507–20.
50. Beggs S, Fitzgerald M. Development of peripheral and spinal nociceptive systems. In: KJS A, Stevens BJ, PJ MG, editors. Pain in Neonates and Infants. 3rd ed. Edinburgh: Elsevier; 2007. p. 11–24.
51. Johnston CC, Stevens BJ, Yang F, Horton L. Differential response to pain by very premature neonates. Pain. 1995;61(3):471–9.
52. Morison SJ, Holsti L, Grunau RE, Whitfield MF, Oberlander TF, Chan HWP, et al. Are there developmentally distinct motor indicators of pain in preterm infants? Early Hum Dev. 2003;72:131–46.
53. Gibbins S, Stevens B. The influence of gestational age on the efficacy and short-term safety of sucrose for procedural pain relief. Advances in Neonatal Care. 2003;3(5):241–9.
54. Gibbins S, Stevens B, McGrath PJ, Yamada J, Beyene J, Breau L, et al. Comparison of pain responses in infants of different gestational ages. Neonatology. 2008;93:10–8.
55. Slater R, Cantarella A, Yoxen J, Patten D, Potts H, Meek J, et al. Latency to facial expression change following noxious stimulation in infants is dependent on postmenstrual age. Pain. 2009;146:177–82.
56. Gibbins S, Stevens B, Beyene J, Chan PC, Bagg M, Asztalos E. Pain behaviours in extremely low gestational age infants. Early Hum Dev. 2008;84:451–8.
57. Holsti L, Grunau RE. Initial validation of the behavioral indicators of infant pain (BIIP). Pain. 2007;132(3):264–72.
58. Holsti L, Grunau RE, Oberlander TF, Whitfield MF. Specific newborn individualized developmental care and assessment program movements are associated with acute pain in preterm infants in the neonatal intensive care unit. Pediatrics. 2004;114(1):65–72.
59. Grunau RE, Oberlander TF, Whitfield MF, Fitzgerald C, Lee SK. Demographic and therapeutic determinants of pain reactivity in very low birth weight neonates at 32 Weeks' postconceptional age. Pediatrics. 2001;107(1):105–12.
60. Als H, Lawhon G, Brown E, Gibes R, Duffy FH, McAnulty G, et al. Individualized behavioral and environmental care for the very low birth weight preterm infant at high risk for bronchopulmonary dysplasia: neonatal intensive care unit and developmental outcome. Pediatrics. 1986;78:1123–32.
61. Als H, Lawhon G, Duffy FH. Individualized developmental Care for the Very low-Birth-Weight Preterm Infant. JAMA. 1994;272(11):853–8.
62. Fitzgerald C, Millard C, McIntosh N. Cutaneous hypersensitivity following peripheral tissue damage in newborn infants and its reversal with topical anaesthesia. Pain. 1989;39(1):31–6.
63. Grunau RE, Holsti L, Haley DW, Oberlander T, Weinberg J, Solimano A, et al. Neonatal procedural pain exposure predicts lower cortisol and behavioral reactivity in preterm infants in the NICU. Pain. 2005;113(3):293–300.
64. Evans JC, McCartney EM, Lawhon G, Galloway J. Longitudinal comparison of preterm pain responses to repeated heelsticks. Pediatr Nurs. 2005;31:216–21.
65. Badr LK, Abdallah B, Hawari M, Sidani S, Kassar M, Nakad P, et al. Determinants of premature infant pain responses to heel sticks. Pediatr Nurs. 2010;36:129–36.
66. Stevens B, Yamada J, Ohlsson A, Haliburton S, Shorkey A. Sucrose for analgesia in newborn infants undergoing painful procedures. Cochrane Database Syst Rev. 2016;7 doi:10.1002/14651858.CD001069.pub.5.
67. Bührer C, Grimmer I, Metze B, Obladen M. The CRIB (clinical risk index for babies) score and neurodevelopmental impairment at one year corrected age in very low birth weight infants. Intensive Care Med. 2000;26(3):325–9.
68. Cignacco E, Hamers J, Stoffel L, Van Lingen RA, Schütz N, Müller R, et al. Routine procedures in NICUs: factors influencing pain assessment and ranking by pain intensity. Swiss Med Wkly. 2008;138:484–91.
69. Cignacco EL, Sellam G, Stoffel L, Gerull R, Nelle M, Anand KJS, et al. Oral sucrose and "facilitated tucking" for repeated pain relief in Preterms: a randomized controlled trial. Pediatrics. 2012;129(2):299–308.

Appendix III: Manuscript 2

Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., Nelle, M., & Cignacco, E. (2019). The influence of gestational age in the psychometric testing of the Bernese Pain Scale for Neonates. *BMC Pediatrics, 19*(20). doi: 10.1186/s12887-018-1380-8

Schenk et al. BMC Pediatrics (2019) 19:20
https://doi.org/10.1186/s12887-018-1380-8

BMC Pediatrics

RESEARCH ARTICLE Open Access

The influence of gestational age in the psychometric testing of the Bernese Pain Scale for Neonates

Karin Schenk[1*], Liliane Stoffel[2], Reto Bürgin[1], Bonnie Stevens[3], Dirk Bassler[4], Sven Schulzke[5], Mathias Nelle[2] and Eva Cignacco[1]

Abstract

Background: Assessing pain in neonates is challenging because full-term and preterm neonates of different gestational ages (GAs) have widely varied reactions to pain. We validated the Bernese Pain Scale for Neonates (BPSN) by testing its use among a large sample of neonates that represented all GAs.

Methods: In this prospective multisite validation study, we assessed 154 neonates between 24 2/7 and 41 4/7 weeks GA, based on the results of 1–5 capillary heel sticks in their first 14 days of life. From each heel stick, we produced three video sequences: baseline; heel stick; and, recovery. Five blinded nurses rated neonates' pain responses according to the BPSN. The underlying factor structure of the BPSN, interrater reliability, concurrent validity with the Premature Infant Pain Profile-Revised (PIPP-R), construct validity, sensitivity and specificity, and the relationship between behavioural and physiological indicators were explored. We considered GA and gender as individual contextual factors.

Results: The factor analyses resulted in a model where the following behaviours best fit the data: crying; facial expression; and, posture. Pain scores for these behavioural items increased on average more than 1 point during the heel stick phases compared to the baseline and recovery phases ($p < 0.001$). Among physiological items, heart rate was more sensitive to pain than oxygen saturation. Heart rate averaged 0.646 points higher during the heel stick than the recovery phases ($p < 0.001$). GA increased along with pain scores: for every additional week of gestation, the average increase of behavioural pain score was 0.063 points ($SE = 0.01$, $t = 5.49$); average heart rate increased 0.042 points ($SE = 0.01$, $t = 6.15$). Sensitivity and specificity analyses indicated that the cut-off should increase with GA. Modified BPSN showed good concurrent validity with the PIPP-R ($r = 0.600–0.758$, $p < 0.001$). Correlations between the modified behavioural subscale and the item heart rate were low ($r = 0.102–0.379$).

Conclusions: The modified BPSN that includes facial expression, crying, posture, and heart rate is a reliable and valid tool for assessing acute pain in full-term and preterm neonates, but our results suggest that adding different cut-off points for different GA-groups will improve the BPSN's clinical usefulness.

Trial registration: The study was retrospectively registered in the database of Clinical Trial gov. Study ID-number: NCT 02749461. Registration date: 12 April 2016.

Keywords: Pain assessment, Neonates, Premature infants, Psychometric testing, Contextual factors, Gestational age, Reliability, Validity

* Correspondence: karin.schenk@bfh.ch
[1]Division of Midwifery, Department of Health Professions, Bern University of Applied Sciences, Murtenstrasse 10, 3008 Bern, Switzerland
Full list of author information is available at the end of the article

Background

Acute painful status in preverbal infants is assessed and interpreted by observing measurable behavioural and physiological indicators. An infant who undergoes an invasive procedure may react to pain that is not caused solely by the painful stimulus [1, 2]. Incorporating individual contextual factors, like gestational age (GA) and gender, into pain assessment tools might make them more accurate [3, 4]. The physiological and behavioural dimensions of pain in neonates are measured by several multidimensional pain assessment tools developed over the last three decades [4–6], but experts agree that behavioural, physiological and cortical measures of pain do not converge to reliably depict and assess the phenomenon of pain in such a vulnerable population [7, 8]. Discrepancies and low-to-moderate associations between behavioural (e.g., facial expression) and physiological (e.g., changes in heart rate) indicators of pain [9–12] have sparked ongoing debate about the appropriate dimensionality of pain scales [7]. Infants may also display nonspecific physiological and behavioural pain indicators during stressful experiences that are not painful, which makes it more challenging to accurately assess pain in neonates [13, 14].

Many pain assessment tools are used in neonatal intensive care unit (NICU) settings. Most add behavioural and physiological indicators to a summary score that is then measured against a cut-off that separates pain from no pain [4]. Rigorous psychometric testing has been applied only to a few [15] (e.g., the Premature Infant Pain Profile [16]). Most were validated for a specific GA in tests that assessed acute pain in full-term and healthy preterm infants with higher GA [4]. However, neurodevelopment and the associated ability to react to painful stimulus varies greatly among early and late preterm infants and full-term neonates: neonates with lower GA express less behavioural pain than more mature neonates [17–22]. In neurologically impaired and very ill neonates, and in neonates on medications (e.g., sedatives), pain may be faintly expressed, or not at all [13, 23].

The Bernese Pain Scale for Neonates (BPSN) is a multidimensional pain assessment tool that includes seven subjective items (sleeping, crying, consolation, skin colour, facial expression, posture, and breathing) and two physiological items (changes in heart rate and oxygen saturation) [24]. The BPSN has been used by clinicians since 2001; 46% of Swiss NICUs rely on this tool to assess pain in neonates [25]. The results of the first validation study in the year 2004 suggested that the BPSN is a valid and reliable scale for assessing acute pain in full-term and preterm neonates with different GAs [24]. However, clinical experts have said the tool is less useful for assessing pain in extremely preterm neonates who, for example, always score very low. This feedback and the increasing scientific evidence which indicates that neonates' pain reaction is influenced by individual contextual factors [1] have motivated us to re-evaluate the tool with sophisticated psychometric tests to assess its accuracy across all GAs.

This study is the first part of a comprehensive BPSN validation and extension study, designed to develop a modified version of the BPSN that includes relevant individual contextual factors in pain assessment. In this first part, we evaluated the BPSN with psychometric tests. The second part of the study will explore the influence of individual contextual factors (e.g., medication, or number of previous painful experiences) on variability in pain reactions across repeated measurement points.

We used psychometric tests to determine the applicability of the BPSN across neonates who ranged from 24 to 42 weeks of GA. We evaluated interrater reliability, the underlying factor structure of the BPSN, and the internal consistency of the scale. We also assessed concurrent validity with the Premature Infant Pain Profile-Revised (PIPP-R; [26]), construct validity, specificity and sensitivity, and determined the relationship between behavioural and physiological indicators of pain. GA groups and gender were considered as individual contextual factors.

Based on the results of the first validation study of the BPSN [24], we hypothesized that the BPSN is a valid and reliable tool for assessing pain in preterm and full-term neonates. Due to feedback from clinical experts concerning difficulties in pain assessment in extremely preterm neonates and the increasing scientific evidence that indicates neonates' pain reaction is influenced by individual contextual factors [1], we assumed that we will find a difference in pain reaction depending especially on neonates' GA. Furthermore, we hypothesized only a low-to-moderate association between behavioural and physiological indicators of pain.

Methods

Sample and settings

This was a prospective multisite validation study with repeated measurement design. It was conducted in three university hospital NICUs in Switzerland (Basel, Bern and Zurich). The study was approved by the Ethics Committee Bern, the Ethics Committee northwest/central Switzerland, and the Ethics Committee Zurich. Recruitment and data collection were ongoing, from January 1 to December 31, 2016. Data collection was extended in Bern until January 31, 2017, because we needed to recruit more extremely premature neonates. We included premature neonates born between 24 0/7 and 36 6/7 weeks of gestation, if they were expected to undergo 2–5 routine capillary heel sticks in their first 14 days of life. We included full-term neonates born between 37 0/7 and 42 0/7 weeks of gestation, if they were

expected to have at least two routine capillary heel sticks during their first 14 days of life. We needed parental permission to include preterm and full-term neonates. We excluded neonates if they had had a high-grade intraventricular haemorrhage (grades III and IV), if they had a severe life-threatening malformation or suffered from any condition that caused partial or total loss of sensitivity, if they had an arterial cord pH < 7.15 at birth, if they had surgery for any reason, or if they had a congenital malformation that affected brain circulation and/or cardiovascular system.

Recruitment and data collection procedures

Neonates were recruited by consecutive sampling and then stratified according to GA at birth [27]. Trained study assistants in each study centre identified potentially eligible neonates and informed their parents of the aim and purpose of the study. After parents granted written informed consent, trained study assistants videotaped neonates (using a HC-V757 high-definition camcorder manufactured by Panasonic, Osaka, Japan) during their next 1–5 routine capillary heel sticks. For each heel stick, we produced three video sequences: baseline, heel stick, and recovery phases. Each video sequence began by focusing on the face of the neonate for at least 1 minute to allow adequate assessment of facial activity and cry. Thereafter, the infant's body was recorded for at least 1 minute. Bedside nurses were asked not to handle the neonates before the baseline phase was recorded, to avoid additional distress that could change the measurement. During the heel stick procedure, the neonates were lying in their incubator (or crib) and the position of the infants was unchanged for the video recording. The baseline phase was recorded 2 to 3 min before the beginning of the heel stick procedure. Afterwards, the bedside nurse warmed the neonate's heel and gave the infant a dose of 24% oral sucrose (0.2 ml/kg bodyweight) to relieve pain [28]. When the nurse disinfected the neonate's heel, the recording of the heel stick phase began. First, the neonate's face was recorded, until the nurse finished the heel stick procedure, which lasted at least a minute. Then the infant's body was recorded for at least one more minute. The recovery phase began immediately after the heel stick phase was recorded. During each phase of the heel stick procedure, our study assistants recorded the infant's highest heart rate and lowest oxygen saturation measurement from the infant's monitors, which tracked this data continuously.

Each video sequence was checked for quality and digitally elaborated by trained study assistants in Final Cut Pro X [29] video editing software. We removed any information that could have revealed the heel stick phase to the raters to ensure continued blindness. The video sequences were uploaded onto a web-based rating tool developed for our study. Uploaded sequences were randomized by sequence number, phase, and presentation order. Five nurses who were working in a NICU and were experienced in using the BPSN (*Mean* = 8.3 years of experience, *SD* = 6.1, *Range* = 3.5–15 years) retrieved the video sequences from the web-based platform and independently rated the behavioural pain expression of the neonates using the BPSN and the PIPP-R. The nurses were trained to use and score the PIPP-R.

Measures

Pain reaction was measured with the BPSN [24] and the PIPP-R [26]. Each of the nine items of the BPSN is rated on a 4-point Likert scale (0, 1, 2, and 3), and then the scores are summed. On the BPSN total score, which includes seven subjective items (i.e., sleeping, crying, consolation, skin colour, facial expression, posture, and breathing), and two physiological items (i.e., changes in heart rate and oxygen saturation), the scores of 11 or more points indicate pain (BPSN total scores range from 0 to 27). In a first validation study in the year 2004 [24], the BPSN showed good construct validity among neonates with GAs between 27 and 41 weeks ($n = 12$); BPSN scores were significantly higher during painful ($M = 15.96$, $SD = 5.7$) compared to non-painful ($M = 2.32$, $SD = 1.6$, $p < 0.001$) situations. Furthermore, the correlations between the BPSN and the Visual Analog Scale (VAS; $r = 0.855$, $p < 0.0001$) and the PIPP ($r = 0.907$, $p < 0.0001$) were high, as well as the interrater ($r = 0.86–0.97$) and intrarater reliability ($r = 0.98–0.99$) of the BPSN [24]. In our study, five independent blinded raters watched the videos to rate the seven subjective items. Both physiological indicators were captured from the neonate's monitoring records during video recordings. Because the raw data on heart rate, oxygen saturation and breathing rate in the baseline phase was used to calculate differences during the heel stick and recovery phases, we set the baseline scores of these items to zero, and retrospectively converted the raw data between baseline, heel stick, and recovery phase into BPSN scores that ranged between 0 and 3.

The PIPP-R is a well validated pain assessment tool for use with premature and full-term neonates, widely used in North America in clinics and for research [16, 26, 30, 31]. The PIPP-R includes three behavioural indicators (brow bulge, eye squeeze, and naso-labial furrow) and two physiological indicators (heart rate and oxygen saturation). Each indicator is rated on a 4-point Likert scale (0, 1, 2, and 3). The PIPP-R accounts for GA and baseline behavioural state as contextual factors. Neonates with younger GAs and neonates in quiet sleep state score the highest, but they are only factored in if the infant's behavioural and physiological sub score is ≥1 [26]. Zero points indicate no pain or perhaps no response to pain, 1–6 points indicate

low pain, 7–12 points indicate moderate pain, and ≥ 13 severe pain. Total PIPP-R scores range from 0 to 21 for neonates with GA < 28 weeks in a quiet and sleep baseline behavioural state, and from 0 to 15 for full-term neonates in an active and awake baseline behavioural state [26]. The PIPP-R shows beginning construct validity [30]; PIPP-R scores were significantly higher during painful (M = 6.7, SD = 3.0) compared to non-painful (M = 4.8, SD = 2.9; $p < 0.001$) procedures among full-term and preterm neonates with GAs as young as 26 weeks of gestation (n = 202). In addition, the PIPP-R showed good interrater reliability between nurses and pain experts (R^2 = 0.87–0.92; $p < 0.001$), and nurses reported that the PIPP-R is a feasible and appropriate pain assessment tool [30]. In our study, both physiological indicators were captured from the neonate's monitoring records and converted into PIPP-R scale values like the physiological indicators of the BPSN. The behavioural indicators and behavioural state were rated from the videos by the same five independent raters. We calculated interrater reliability of the three behavioural items with a two-way random-effects, absolute agreement, single measure model that ranged from 0.750 to 0.842 (Mdn = 0.803) in the heel stick phases of the five measurement points.

We retrieved individual contextual factors retrospectively from patient charts [27] and will publish a separate paper describing their influence on the variability of pain reaction across repeated measurement points.

Sample size and power

Our target sample size of 150 neonates was based on an a priori power analysis of the hypothesized association between the BPSN and GAs at baseline. That analysis was based on data from a previous study (n = 71; [32]) and a descriptive-explorative analysis (n = 23); it assumed a Type I error probability of 5%, a power of 80%, and at least three documented baseline heel sticks per study infant.

Data analysis

Factor analyses explored the structure of the BPSN and measurement invariance. Psychometric tests examined interrater reliability, internal consistency, construct validity, concurrent validity with the PIPP-R [30], association between behavioural and physiological items, and sensitivity and specificity. Because the sample was heterogeneous, we also conducted analyses for different GA-groups. We used the statistics programs SPSS [33] and R [34] for all analyses. Space restriction limit us to reporting mainly our results from the heel stick phases. In this comprehensive validation study, we did multiple testing of outcome data arising from individual neonates. Correction of p-values with Bonferroni adjustment [35] would not have rendered findings non-significant. Therefore, all p-values are presented uncorrected for multiple testing unless otherwise specified. A p-value < 0.05 was considered statistically significant.

Preliminary analyses

Exploratory analyses described the data and looked for anomalies that could reduce the validity of the data analysis. We used descriptive and frequency statistics to describe sample characteristics and each rater's pain scores.

Missing values

We analysed the ratings of the 1'817 video sequences for the volume and pattern of missing data, since single items of the BPSN and the PIPP-R could be rated "non-evaluable". Because it is impossible to compute BPSN and PIPP-R sum scores when an item was not rated, we used multiple imputation [36] and the R-package *partykit* [37] to derive those scores by replacing the values of non-rated items with random substitutes generated from conditional inference regression trees [38]. We generated five data sets, so there were five variants on the BPSN and PIPP-R sum scores.

Interrater reliability

Intraclass correlation coefficients (ICCs) and their 95% confidence intervals were calculated to determine interrater reliability of the seven subjective BPSN-items [39, 40]. Since pain reaction of a neonate is rated by a single nurse in the clinical setting, and pain level scores were central to our outcome, we assessed interrater reliability with a two-way random-effects, absolute agreement, single measure model [41]. ICC coefficients were also calculated with a two-way random-effects, absolute agreement, average measure model, to generate more information about the reliability of the mean ratings provided by the five raters [40]. Each phase of the five measurement points was analysed separately, resulting in 120 ICC coefficients (8 rating scores * 3 phases * 5 measurement points) per model.

Factor analyses

Measurement construct

Multiple group longitudinal confirmatory factor analysis [42] was used to evaluate the extent to which individual items correlated with the unobservable pain construct, the predictive performance of the construct, and whether factor loadings were invariant across time and raters. The R-package *lavaan* [43] was used for this analysis. Full maximum likelihood estimates were based on the assumption that data were missing at random.

Model specification

Figures 1 and 2 show the structures of our confirmatory factor analysis (CFA) models for the subjective and physiological subscales. For item selection, we used only data from the heel stick phases of the five measurement points. Measurement invariance tests were based on data from all phases (baseline, heel stick, and recovery) and all measurement points (t1-t5).

The longitudinal structure of the data was accounted for by implementing covariances between factors (Fig. 3, structure of the subjective subscale). The covariance structure of factors for the physiological subscale or additional phases or measurement points was implemented as shown.

For the subjective subscale, we stacked the data records of raters, and used the rater as a grouping variable. This specification of this model made it impossible to model covariances between values of the same child measured by different raters. We chose this specification because it did allow us to test invariance of model parameters within and across raters.

Analytical procedure

We selected items to improve the fit of the CFA model. At estimation, to remove inconsistent items, we restricted loadings of a given item to a common value across raters and measurement points. For both subscales, we estimated several model configurations with at least two items, resulting, for the subjective subscale with 7 items, in 120 models. For the physiological subscale, we used only one model since it included only two items. Selecting the final model was a three-step process. First, we excluded several models with loadings < 0.3 and also excluded models with root mean square errors of approximation (RMSEA) > 0.06, Comparative Fit Indices (CFI; [44]) < 0.95 and Tucker-Lewis Indices < 0.95 (TLI; [45]). The minimal loading size of 0.3 was inspired by Brown [46], and the combinations of cut-offs for the RMSEA, CFI and TLI were inspired by Hu and Bentler [47, 48]. Second, we chose from the remaining models those with the highest number of parameters because we wanted to keep as many appropriate items as possible. Third, we planned to select the model with the highest CFI if Step 2 left us with more than one candidate, but this step turned out to be unnecessary. We found no suitable factor model for the physiological subscale and therefore, we used regression analysis to pick the item most sensitive to pain.

We continued factor analysis by examining measurement invariance across time points within-raters and overall measurement invariance. Only loading (weak) invariance was considered, because other parameters like intercepts and variances could be expected to vary over time and phases. Measurement invariance was examined with Satorra and Bentler's likelihood ratio test [49] and tests based on the RMSEA, CFI and TLI that used Cheung and Rensvold's critical values [50].

Reliability and validity of the modified BPSN

The results of our factor analyses showed that only the behavioural items crying, facial expression, and posture had consistently high factor loadings over time. The physiological items heart rate and oxygen saturation did not load on a common factor and did not correlate with each other. Further analyses showed that the item heart rate was more sensitive to pain than oxygen saturation. We thus decided to exclude the items sleeping, consolation, skin colour, breathing, and oxygen saturation from the BPSN. In following examinations, we used a modified version of the BPSN that included facial expression, crying, and posture, as a behavioural subscale, and heart rate as an additional physiological indicator. Because the results of the measurement invariance analyses showed that the measurement construct measured with the modified behavioural subscale works differently for different raters, we accounted for differences between the raters by either including the raters in the model, or by

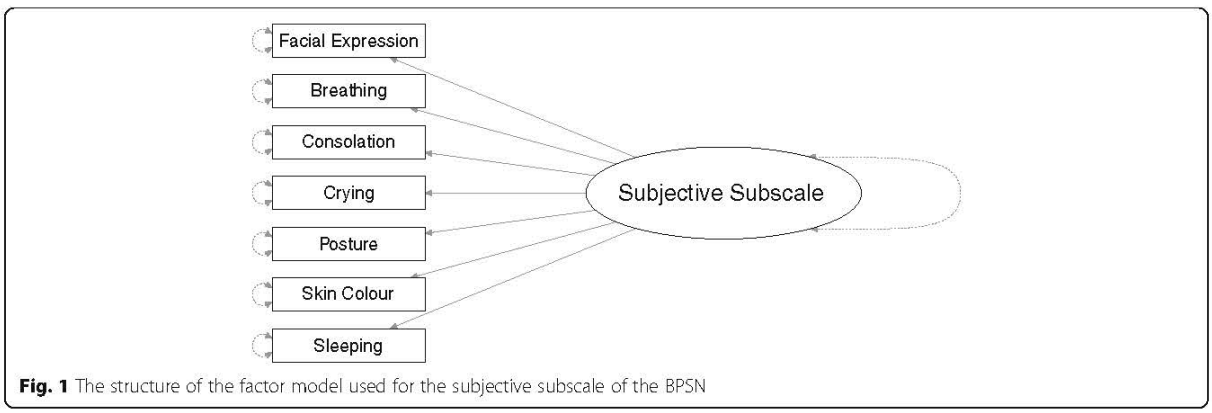

Fig. 1 The structure of the factor model used for the subjective subscale of the BPSN

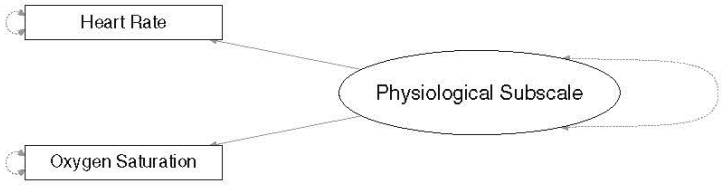

Fig. 2 The structure of the factor model used for the physiological subscale of the BPSN

conducting separate analyses for each rater and then pooling the results.

Internal consistency and corrected item-total correlation

We evaluated the internal consistency of the modified version of the behavioural subscale that included items facial expression, crying and posture by calculating Cronbach's α. We calculated corrected item-total correlations to analyse correlations between single items and the behavioural subscale. In addition, we calculated the resulting Cronbach's Alpha when an individual item is removed from the scale (Cronbach's Alpha if Item Deleted) [51]. Data from each rater were analysed separately, resulting in 75 analyses (5 raters * 3 phases * 5 measurement points), and then we used *cocron* [52], a web interface, to statistically compare the Cronbach's Alpha coefficients calculated for each rater.

Correlations between behavioural and physiological indicators of pain

Pearson product-moment correlation coefficients were calculated to establish the association between the modified behavioural subscale of the BPSN and heart rate. Data from each rater were analysed separately, resulting in 50 analyses (5 raters * 2 phases * 5 measurement points). Afterwards, for each phase we examined at each measurement point whether the correlation coefficients calculated for the five raters were statistically different, using the χ^2-statistics of Steiger [53].

Construct validity

We compared the level of pain scores between the three phases (baseline, heel stick and recovery) to determine construct validity of the BPSN. We analysed the modified behavioural subscale and heart rate in a linear mixed effect analysis that used the R-package *lme4* [54]. Linear mixed effect analysis allowed us to control variance created by multiple measurement points per subject [55]. The three phases, five measurement points, GA at time of birth, and gender were fixed effects in the model. Neonates and raters were random intercepts. Likelihood Ratio Tests tested the effect of the three phases on the level of pain scores [55].

Concurrent validity

Pearson product-moment correlation coefficients were calculated to establish concurrent validity between the modified total scores of the BPSN (facial expression, crying, posture, heart rate) and the PIPP-R. Separate analysis were performed for the data of each rater, resulting in 75 analyses (5 raters * 3 phases * 5 measurement points), and afterwards, we examined for each phase at each measurement point if the correlation coefficients calculated for the five raters were not statistically different, again using the χ^2-test of Steiger [53].

Specificity and sensitivity analysis

A Receiver-Operating Characteristic (ROC) curve analysis was used to evaluate the ability of the modified BPSN total score to detect pain in neonates and to determine the cut-off value that maximized both sensitivity and specificity [56]. The PIPP-R was the reference value that allowed us to determine sensitivity and specificity; PIPP-R values of ≤6 characterized neonates as experiencing no or low pain; values ≥7 characterized neonates as experiencing moderate to severe pain. We tested whether the area under the curve (AUC) was greater than 0.5 and calculated sensitivity and specificity of the

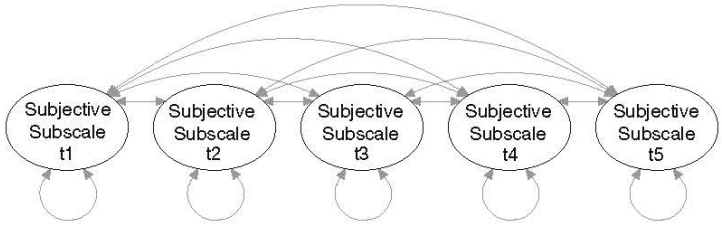

Fig. 3 Specified covariances between factors

BPSN by using the cut-off values the ROC curve suggested. We performed this analysis separately for the heel stick phases of the five measurement points and the five raters, resulting in 25 ROC curves analysis (5 raters * 5 measurement points), and we averaged the values calculated for each rater.

Secondary analyses by GA-groups

Infants that ranged from 24 2/7 to 42 5/7 GA at time of birth were included in the primary analyses. Because the sample was heterogenous, we reanalysed the data separately for four GA-groups [57]: extremely preterm neonates (24 0/7–27 6/7 weeks GA); very preterm neonates (28 0/7–31 6/7 weeks GA); moderate to late preterm neonates (32 0/7–36 6/7 weeks GA); and, full-term neonates (37 0/7–42 6/7 weeks GA). Analyses remained the same with exception of the factor and linear mixed model analyses. We could not reanalyse the factor analysis for different GA-groups separately because the sub-samples were too small. In the linear mixed model analyses, GA was already considered as a fixed effect. We did not use Bonferroni adjustment in this subgroup analyses because we exploratively analysed if there were any obvious differences between the four GA-groups.

Results

Missing data and sample characteristics

We enrolled a total of 162 neonates in the study; 8 were excluded from data analysis because video sequences were missing or of poor quality. Figure 4 illustrates the flow of recruitment and data collection.

For the five raters, ≤ 1.0% data was missing for the BPSN items sleeping, crying, consolation, skin colour and posture; for facial expression, 0.1 to 4.0% (*Mdn* = 0.8%) data was missing, and for breathing, 0.3 to 8.7% (*Mdn* = 1.9%) was missing. For the PIPP-R, 0.5 to 3.3% (*Mdn* = 1.0%) of data was missing for brow bulge, 0.4 to 3.6% (*Mdn* = 0.7%) for eye squeeze, 0.6 to 28.3% (*Mdn* = 4.3%) for naso-labial furrow, and 0.1 to 0.9% (*Mdn* = 0.4%) for behavioural state. Less than 1% of data was missing for the physiological items heart rate and oxygen saturation.

Mean GA at birth of the total sample was 30.85 (*SD* = 4.5) weeks and ranged from 24.29 to 41.57. Demographic and medical characteristics of the sample are summarized in Table 1.

Results of descriptive and preliminary analysis

Means of the BPSN total-scale, subjective subscale, and items are summarized in Table 2. Physiological items are

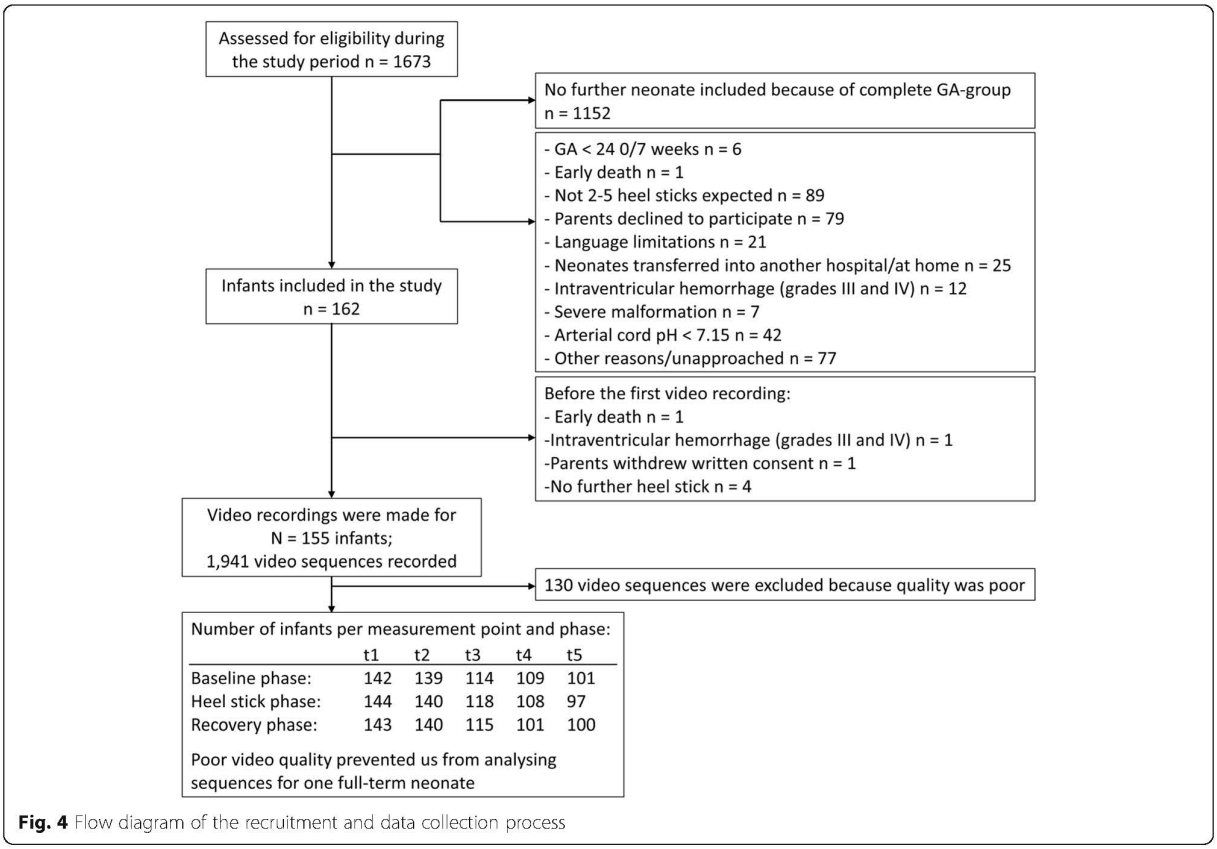

Fig. 4 Flow diagram of the recruitment and data collection process

Table 1 Demographic and medical characteristics of the total sample and the four gestational age groups

	Total Sample	Gestational age groups			
		Extremely preterm neonates	Very preterm neonates	Moderate to late preterm neonates	Full-term neonates
Sample, n (%)	154 (100)	50 (32.5)	45 (29.2)	38 (24.7)	21 (13.6)
Sex, n (%)					
- Male	87 (56.5)	31 (62.0)	23 (51.1)	20 (52.6)	13 (61.9)
GA at birth in weeks, mean (SD)	30.85 (4.5)	26.23 (1.2)	29.44 (1.0)	34.21 (1.0)	38. 81 (1.3)
Birth weight in grams, mean (SD)	1630.10 (934.3)	851.40 (196.4)	1285.11 (328.2)	2093.68 (377.5)	3384.52 (811.6)
Number of comorbidities, mean (SD)	5.70 (4.4)	10.06 (4.2)	5.44 (2.4)	2.66 (1.4)	1.38 (1.1)
CRIB score, mean (SD)	3.76 (3.9)	7.50 (3.7)	3.24 (2.8)	1.05 (1.7)	0.86 (1.6)
Way of delivery, n (%)					
- Vaginal-spontan	36 (23.4)	10 (20.0)	4 (8.9)	13 (34.2)	9 (42.9)
- Vaginal-operativ	4 (2.6)	0 (0)	1 (2.2)	2 (5.3)	1 (4.8)
- Planned c-section	23 (14.9)	3 (6.0)	8 (17.8)	7 (18.4)	5 (23.8)
- Emergency c-cection	91 (59.1)	37 (74.0)	32 (71.1)	16 (42.1)	6 (28.6)
Number of birth, mean (SD)					
- Single	104 (67.5)	43 (86.0)	20 (44.4)	21 (55.3)	20 (95.2)
- One of twins	44 (28.6)	4 (8.0)	22 (48.9)	17 (44.7)	1 (4.8)
- One of triplet	6 (3.9)	3 (6.0)	3 (6.7)	0 (0)	0 (0)
Day of life at first measure point, mean (SD)	3.95 (2.0)	4.80 (2.2)	3.56 (1.9)	3.18 (1.0)	4.19 (2.6)

Note. CRIB Clinical Risk Index for Babies

not included in this table because they were captured from the neonates' monitoring records during video recordings and the raw data was retrospectively converted into BPSN scores between 0 and 3. The mean scores for heart rate ranged from 0.47 to 0.76 (*Mdn* = 0.72) during the five heel stick phases, and from 0.03 to 0.11 (*Mdn* = 0.09) during the five recovery phases. The mean scores for oxygen saturation ranged from 0.77 to 1.25 (*Mdn* = 0.86) during the five heel stick phases, and from 0.51 to 0.71 (*Mdn* = 0.61) during the five recovery phases.

Interrater reliability

We derived the results of our interrater reliability analyses by calculating two-way random-effects, absolute agreement models. The results are summarized in Table 3. We again excluded heart rate and oxygen saturation. Interrater agreement for the items crying, consolation, facial expression, and posture tended to decrease across the five measurement points.

Factor analyses

Item selection

First, we used all items and heel stick phases of the five measurement points to estimate the multiple group confirmatory factor models for the subjective and physiological subscale. No parameter restrictions were applied, so that loadings could vary across measurement points and raters. To compare the loadings of all items, we restricted factor variance to 1. Figure 5 shows the estimated factor loadings of the model for the subjective subscale and Fig. 6 for the physiological subscale. For the subjective subscale, loadings for breathing (range = – 0.167-0.110) and skin colour (range = – 0.034-0.293) are low, while loadings for sleeping vary widely between raters (range = 0.096–0.982). Loadings of the remaining items, consolation, crying, facial expression, and posture, seem consistent, but they tend to decrease over time. Rater D's loadings often conflict with other raters and vary over time.

For the physiological subscale, two loadings exceed by far a value of 1, indicating poor fit between model and data. Additional analyses showed no association between heart rate and oxygen saturation. Pearson product-moment correlations between heart rate and oxygen saturation ranged from $r = -0.028$ to 0.106 (*Mdn* = 0.017; $p > 0.05$) during the heel stick phases of the five measurement points. Large loadings are probably numerical artefacts and should not be over-interpreted. Because the physiological items did not load on a common factor or correlate with each other, we discarded all but one of the physiological items based on their sensitivity to pain. We analysed the sensitivity to pain of heart rate and oxygen saturation by calculating linear mixed effect models (see next section).

Table 2 Means of the Bernese Pain Scale for Neonates total-scale and the subjective subscale and items

		Rater A	Rater B	Rater C	Rater D	Rater E
	Phase	Means t1-t5	Means t1-t5	Means t1-t5	Means t1-t5	Means t1-t5
		Range (Median)	Range (Median)	Range (Median)	Range (Median)	Range (Median)
BPSN total-scale	Baseline	0.89–1.14 (1.06)	1.99–2.47 (2.21)	1.31–1.51 (1.38)	4.44–5.15 (4.98)	4.66–4.97 (4.80)
N = 81–142	Heel Stick	4.03–4.77 (4.14)	5.98–6.98 (6.33)	4.57–5.41 (4.87)	8.15–9.53 (8.29)	8.00–9.07 (8.52)
	Recovery	1.84–2.30 (2.19)	3.08–3.40 (3.22)	2.37–2.67 (2.46)	5.27–6.27 (6.06)	5.37–5.99 (5.65)
Subjective subscale	Baseline	0.89–1.14 (1.06)	1.99–2.47 (2.21)	1.31–1.51 (1.38)	4.44–5.15 (4.98)	4.66–4.97 (4.80)
N = 82–142	Heel Stick	2.51–2.82 (2.68)	4.64–4.96 (4.73)	3.00–3.35 (3.31)	6.59–7.47 (6.84)	6.65–7.04 (6.90)
	Recovery	1.17–1.63 (1.45)	2.39–2.76 (2.51)	1.70–1.97 (1.77)	4.59–5.60 (5.28)	4.66–5.26 (4.94)
Sleeping	Baseline	0.23–0.28 (0.23)	0.39–0.43 (0.41)	0.42–0.51 (0.47)	1.04–1.28 (1.19)	0.89–1.10 (1.05)
N = 95–143	Heel Stick	0.39–0.45 (0.42)	0.75–0.91 (0.89)	0.55–0.63 (0.60)	1.19–1.29 (1.23)	1.35–1.46 (1.41)
	Recovery	0.20–0.32 (0.30)	0.40–0.49 (0.41)	0.41–0.51 (0.42)	1.02–1.31 (1.19)	0.89–1.08 (1.06)
Crying	Baseline	0.02–0.06 (0.06)	0.04–0.09 (0.07)	0.04–0.10 (0.06)	0.06–0.11 (0.09)	0.07–0.12 (0.09)
N = 96–143	Heel Stick	0.21–0.30 (0.24)	0.30–0.43 (0.36)	0.31–0.42 (0.37)	0.35–0.47 (0.42)	0.36–0.48 (0.43)
	Recovery	0.02–0.06 (0.03)	0.03–0.10 (0.06)	0.03–0.11 (0.07)	0.05–0.11 (0.06)	0.04–0.12 (0.09)
Consolation	Baseline	0.02–0.06 (0.05)	0.05–0.10 (0.09)	0.04–0.12 (0.07)	0.77–1.07 (0.97)	0.03–0.12 (0.08)
N = 96–143	Heel Stick	0.21–0.32 (0.21)	0.31–0.48 (0.43)	0.28–0.43 (0.33)	1.19–1.48 (1.26)	0.35–0.55 (0.46)
	Recovery	0.00–0.07 (0.02)	0.03–0.13 (0.06)	0.01–0.15 (0.09)	0.68–0.99 (0.85)	0.02–0.14 (0.11)
Skin colour	Baseline	0.02–0.06 (0.04)	1.00–1.27 (1.11)	0.02–0.06 (0.03)	0.86–1.06 (0.97)	1.51–1.67 (1.61)
N = 96–143	Heel Stick	0.05–0.08 (0.07)	1.19–1.29 (1.26)	0.03–0.05 (0.03)	0.99–1.36 (1.07)	1.55–1.79 (1.69)
	Recovery	0.00–0.06 (0.04)	1.05–1.18 (1.13)	0.02–0.04 (0.03)	0.89–1.09 (1.04)	1.48–1.69 (1.53)
Facial expression	Baseline	0.16–0.29 (0.24)	0.17–0.29 (0.19)	0.22–0.32 (0.25)	0.73–0.86 (0.75)	0.83–0.89 (0.87)
N = 95–143	Heel Stick	0.50–0.64 (0.61)	0.61–0.69 (0.64)	0.60–0.65 (0.63)	1.01–1.13 (1.06)	1.08–1.18 (1.12)
	Recovery	0.19–0.33 (0.24)	0.09–0.19 (0.17)	1.16–0.26 (0.23)	0.62–0.79 (0.69)	0.80–0.89 (0.87)
Posture	Baseline	0.33–0.49 (0.40)	0.27–0.36 (0.30)	0.45–0.49 (0.48)	0.93–1.04 (0.99)	1.15–1.29 (1.19)
N = 97–143	Heel Stick	0.55–0.67 (0.60)	0.69–0.80 (0.78)	0.57–0.71 (0.70)	1.17–1.24 (1.20)	1.38–1.45 (1.41)
	Recovery	0.32–0.43 (0.34)	0.20–0.34 (0.32)	0.37–0.46 (0.41)	0.80–0.94 (0.87)	1.06–1.20 (1.19)
Breathing	Heel Stick	0.47–0.57 (0.50)	0.32–0.65 (0.54)	0.61–0.72 (0.65)	0.50–0.69 (0.64)	0.39–0.62 (0.47)
N = 84–142	Recovery	0.35–0.54 (0.45)	0.31–0.46 (0.40)	0.40–0.63 (0.58)	0.49–0.64 (0.53)	0.31–0.58 (0.41)
Raw Scores	Baseline	26.7–27.9 (27.6)	25.7–26.9 (25.8)	27.8–29.5 (28.1)	26.0–26.9 (26.6)	28.4–30.1 (29.5)
Breathing	Heel Stick	28.6–29.2 (28.5)	26.1–27.8 (27.0)	28.2–29.9 (28.9)	27.2–28.3 (27.5)	29.4–30.4 (30.0)
N = 91–142	Recovery	27.0–28.7 (27.7)	25.3–27.1 (26.2)	27.4–29.3 (28.3)	26.4–27.4 (26.6)	28.9–30.1 (29.7)

Note. N = number of neonates included in the analysis. This number varies because of differences in the amount of missing data between the raters at each measurement point and differences in the number of neonates included at each point of measurement

We selected items of the subjective subscale by estimating several configural models with at least two items. In contrast to the model presented in Fig. 5, we restricted factor loadings of a given item to a common value across time points and raters. We excluded models with factor loadings < 0.3, a RMSEA > 0.06 and CFI and TLI < 0.95. This left us with four models, from which we selected the model with the highest number of items. Our final model included only the items crying, facial expression and posture. Table 4 compares model fit indices of the baseline model with all items to the final model with only crying, facial expression, and posture.

Table 3 Intraclass Correlation Coefficients and their 95% confident intervals for the single items of the Bernese Pain Scale for Neonates

	Heel Stick Phase 1 ICC [95%CI]	Heel Stick Phase 2 ICC [95%CI]	Heel Stick Phase 3 ICC [95%CI]	Heel Stick Phase 4 ICC [95%CI]	Heel Stick Phase 5 ICC [95%CI]
Sleeping					
N	135	139	117	105	93
Single measures	0.215 [0.13–0.31]	0.267 [0.18–0.36]	0.211 [0.13–0.30]	0.185 [0.11–0.28]	0.221 [0.13–0.33]
Average measures	0.578 [0.43–0.69]	0.646 [0.52–0.74]	0.572 [0.43–0.69]	0.532 [0.37–0.66]	0.586 [0.43–0.71]
Crying					
N	138	140	117	107	94
Single measures	0.773 [0.72–0.82]	0.694 [0.63–0.76]	0.721 [0.65–0.78]	0.719 [0.65–0.78]	0.655 [0.57–0.73]
Average measures	0.945 [0.93–0.96]	0.919 [0.89–0.94]	0.928 [0.90–0.95]	0.927 [0.90–0.95]	0.905 [0.87–0.93]
Consolation					
N	140	140	117	108	94
Single measures	0.453 [0.31–0.58]	0.381 [0.22–0.53]	0.420 [0.27–0.55]	0.319 [0.16–0.48]	0.257 [0.11–0.41]
Average measures	0.805 [0.69–0.87]	0.755 [0.58–0.85]	0.784 [0.65–0.86]	0.701 [0.48–0.82]	0.634 [0.38–0.78]
Skin colour					
N	141	138	115	108	96
Single measures	0.074 [0.02–0.14]	0.049 [0.03–0.37]	0.073 [0.02–0.15]	0.045 [0.00–0.10]	0.072 [0.01–0.15]
Average measures	0.285 [0.09–0.45]	0.205 [0.03–0.37]	0.284 [0.08–0.46]	0.189 [0.01–0.36]	0.280 [0.06–0.47]
Facial expression					
N	135	130	112	102	92
Single measures	0.655 [0.53–0.75]	0.555 [0.43–0.66]	0.558 [0.45–0.66]	0.500 [0.37–0.62]	0.514 [0.37–0.64]
Average measures	0.905 [0.85–0.94]	0.862 [0.79–0.91]	0.863 [0.80–0.91]	0.833 [0.75–0.89]	0.841 [0.74–0.90]
Posture					
N	141	139	117	108	97
Single measures	0.551 [0.38–0.68]	0.487 [0.31–0.63]	0.536 [0.38–0.66]	0.400 [0.25–0.54]	0.342 [0.21–0.48]
Average measures	0.860 [0.75–0.92]	0.826 [0.69–0.89]	0.852 [0.75–0.91]	0.769 [0.62–0.85]	0.722 [0.57–0.82]
Breathing					
N	119	111	100	95	82
Single measures	0.252 [0.17–0.34]	0.348 [0.26–0.44]	0.334 [0.24–0.44]	0.348 [0.25–0.45]	0.402 [0.30–0.51]
Average measures	0.627 [0.51–0.72]	0.727 [0.64–0.80]	0.715 [0.62–0.79]	0.727 [0.63–0.81]	0.770 [0.68–0.84]
Raw Scores Breathing					
N	128	123	107	106	91
Single measures	0.636 [0.56–0.71]	0.632 [0.56–0.71]	0.674 [0.59–0.75]	0.610 [0.53–0.69]	0.630 [0.54–0.71]
Average measures	0.897 [0.87–0.92]	0.896 [0.86–0.92]	0.912 [0.88–0.94]	0.887 [0.85–0.92]	0.895 [0.86–0.93]

Note. ICC = Intraclass Correlation Coefficients, calculated with two-way random-effects, absolute agreement models; [95% CI] = 95% confident intervals of the ICCs

This improves the CFI and the TLI indices from about 0.8 to 0.95.

Physiological items' sensitivity to pain

Because the factor analysis indicated that the physiological items heart rate and oxygen saturation do not fit the data well, we next examined these items for their sensitivity to pain. We calculated linear mixed models that included the variables phases, measurement points, GA at time of birth, and gender as fixed effects, and neonates as random intercept. We used Likelihood Ratio Tests to compare a model without the heel stick and recovery phases to a model that included the phases. There was a significant effect of phase on heart rate ($\chi^2(5) = 172.91$, $p < 0.001$). Heart rate scores during the recovery phases were, on average, 0.646 point lower than scores during the heel stick phases ($SE = 0.09$, t-value = − 7.383). Phase also significantly affected oxygen saturation ($\chi^2(5) = 33.658$, $p < 0.001$). Oxygen saturation scores were, on average, 0.258 points lower during the recovery phases than during the heel stick phases ($SE = 0.12$, t-value = − 2.136). We

Fig. 5 Factor loadings of the baseline factor models for the subjective subscale

thus decided to use only heart rate for the physiological subscale.

Measurement invariance

Measurement invariance was examined only for the subjective subscale, since the physiological subscale contained one item. In this analysis, we re-estimated the final model that included crying, facial expression and posture. We used different parameter restrictions: (Free) = all parameters are free; (WRLInv) = within-rater loadings invariance was assumed by restricting loadings of items across time but not across raters; (OLInv) = overall loadings invariance was assumed by restricting loadings across time and across raters. We already applied the OLInv assumption to select items. We next asked if the restricted models fit the data as well as the unrestricted models, and whether factor loadings are (partially) invariant. We performed the same analysis but used only data from the heel stick phase of the five measurement points. Then we used data from all phases and measurement points. Table 5 shows differences between fit indices of the unrestricted and restricted models, including the likelihood ratio test. At a 5% significance level, the zero hypothesis of equal fit or loadings invariance is not rejected for within-rater invariance when we used only data from the heel stick phases, but it was otherwise rejected, most sharply for overall loading invariance (OLInv).

Differences between the fit indices RMSEA, CFI and TLI yield different test results. Using the 1% level rejection areas [50] for the RMSEA, measurement invariance is rejected when the difference is > 0.013, for the CFI, it is rejected when it is < -0.0085, and, for the TLI, when it is < -0.0078. Accordingly, within-rater loadings invariance (WRLInv) is never rejected, but overall measurement invariance (OLInv) is always rejected with CFI and TLI, and never with RMSEA.

The tests strongly suggest that the pain measurement construct under consideration works differently for different raters. For within-rater invariance, invariance is not rejected during the heel stick phases; for all data, it is rejected by the χ^2-test but not by RMSEA, CFI and TLI. We may assume approximate invariance, while keeping in mind the results.

Reliability and validity of the modified BPSN

Our factor analysis and analysis of the physiological items' sensitivity to pain led us to adopt a modified version of the BPSN for our next analyses. The modified BPSN includes a behavioural subscale (facial expression, crying, and posture) and adds heart rate as a pain indicator.

Cronbach's alpha and corrected item-Total correlation

Cronbach's Alpha, corrected item-total correlation coefficients and the resulting Alpha when an individual item is removed from the scale (Alpha if Item Deleted) for the modified behavioural subscale are summarized in Table 6. During the heel stick phases of the five measurement points, Cronbach's Alpha coefficients of the five raters

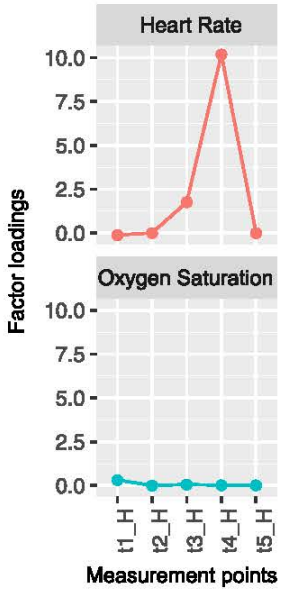

Fig. 6 Factor loadings of the baseline factor models for the physiological subscale

differed significantly ($p < 0.01$). Internal consistency of the behavioural subscale tended to decrease over time.

Correlations between behavioural and physiological indicators of pain

We examined the associations between behavioural and physiological indicators of pain with the modified behavioural subscale of the BPSN including the items crying, facial expression, and posture, and the physiological item heart rate. See Table 7 for the correlation coefficients of these analyses. At measurement point 3, the correlation coefficients differed significantly between the five raters ($p = 0.008$), while the correlation coefficients were approximately the same during the other measurement points ($p > 0.05$). When we considered a Bonferroni adjusted p-value ($p < 0.05/10$), none of the correlation coefficients would differ significantly between the five raters.

Construct validity

To determine construct validity of the BPSN, we compared levels of pain scores of the modified behavioural subscale between the three phases. The residual variance of this analysis was $\sigma^2 = 1.708$ ($SD = 1.307$); variances of the random effects were $\sigma^2 = 0.354$ ($SD = 0.595$) for neonates and $\sigma^2 = 0.391$ ($SD = 0.625$) for raters. Phases significantly affected the level of behavioural pain scores ($\chi^2(10) = 864.18$, $p < 0.001$). Behavioural pain scores in the heel stick phases averaged 1.04 higher than pain scores in the baseline phases, and 1.13 higher than pain scores in the recovery phases. More results are summarized in Table 8. The same analysis was performed for the item heart rate (Table 8). The residual variance of this analysis was $\sigma^2 = 0.588$ ($SD = 0.767$) and variance of the random effect neonates was $\sigma^2 = 0.037$ ($SD = 0.191$). GA at time of birth significantly affected behavioural pain scores ($SE = 0.01$, $t = 5.488$) and heart rate ($SE = 0.01$, $t = 6.145$). Gender had no effect on behavioural pain scores ($SE = 0.10$, $t = -0.170$) or on heart rate ($SE = 0.05$, $t = 0.051$).

Concurrent validity

We examined the concurrent validity between the modified total score of the BPSN and the PIPP-R. See Table 9 for the correlation coefficients of these analyses. The correlation coefficients of the five raters were the same in about half of the cases. They differed significantly at measurement point 1 ($p = 0.010$) and measurement point 4 ($p = 0.045$). With a Bonferroni adjusted p-value ($p < 0.05/15$), none of the correlation coefficients differed significantly between the five raters.

Sensitivity and specificity

The results of the ROC analyses to examine sensitivity and specificity of the modified BPSN total score (including crying, facial expression, posture, and heart rate) are shown in Table 10. During the heel stick phases of the five measurement points, a cut-off of 1.5 points fits best to reach a sensitivity of approximately 80% and a specificity of similar accuracy.

Results of the psychometric testing of the BPSN separated by GA-groups

Interrater reliability

ICCs coefficients of the four different GA-groups are summarized in Table 11. Interrater reliability of the items facial expression, posture and consolation tended to improve as GA increases.

Table 4 Fit indices of the Baseline and Final Models differ by item inclusions

Model	df	χ^2	AIC	RMSEA	CFI	TLI	SRMR
Baseline (7 subjective items)	2918	4985	36,875	0.068	0.807	0.803	0.135
Final (Crying, Facial expression, Posture)	472	648	13,575	0.049	0.961	0.957	0.111

Note. Model indices: df = degrees of freedom; AIC = Akaike Information Criterion; RMSEA = root mean squared error of approximation; CFI = Bentler's Comparative Fit Index; TLI = Tucker-Lewis Indices; SRMR = standardized root mean square residual

Table 5 Difference statistics for measurement invariance testing

Model	Restriction	df	Δ RMSEA	Δ CFI	Δ TLI	$\Delta\chi^2$	Δ df	p ($\Delta\chi^2$)
Heel stick phases of measurement points t1-t5	WRLInv	440	0	0.000	0.002	39	40	0.531
	OLInv	448	0	− 0.015	− 0.015	115	48	0.000
All phases and measurement points	WRLInv	4340	0	−0.007	− 0.006	123	40	0.000
	OLInv	4348	0	−0.024	− 0.025	343	48	0.000

Note. WRLInv = within-rater loadings invariance; OLInv = overall loadings invariance; df = degrees of freedom; RMSEA = root mean squared error of approximation; CFI = Bentler's Comparative Fit Index; TLI = Tucker-Lewis Indices; $\Delta\chi^2$ = Satorra-Bentler 2010 χ^2-test statistic

Internal consistency of the modified behavioural BPSN subscale

Cronbach's Alpha calculated separately for the four GA-groups, are summarized in Table 12. Most Cronbach's Alpha coefficients were in the range of acceptable to excellent [58] during the heel stick phases of the five measurement points.

Correlations between behavioural and physiological indicators of pain

During the heel stick phases of the five measurement points and among the five raters, correlations between the modified behavioural subscale of the BPSN and the item heart rate ranged from $r = -0.173$-0.577 ($Mdn = 0.196$) among extremely preterm neonates, from $r = 0.024$–0.480 ($Mdn = 0.329$) among very preterm neonates, from $r = -0.174$-0.442 ($Mdn = 0.172$) among moderate to late preterm neonates, and from $r = -0.044$ to 0.402 ($Mdn = 0.236$) among full-term neonates.

Concurrent validity

During the heel stick phases of the five measurement points and among the five raters, correlations between the total scale of the modified BPSN and the PIPP-R ranged from $r = 0.560$–0.775 ($Mdn = 0.683$) among extremely preterm neonates, from $r = 0.582$–0.875 ($Mdn = 0.750$) among very preterm neonates, from $r = 0.603$–0.860 ($Mdn = 0.769$) among moderate to late preterm neonates, and from $r = 0.757$–0.898 ($Mdn = 0.808$) among full-term neonates.

Sensitivity and specificity

The results of the ROC analyses to examine sensitivity and specificity of the modified BPSN total scale separately for each GA-group are provided in Table 13. We found cut-off points needed to increase along with GA to reach about 80% sensitivity and similarly high specificity.

Discussion

After rigorous statistical testing, we significantly reduced the number of items in the original BPSN, leaving only three behavioural items: facial expression, crying, and posture. We included only one physiological item, heart rate, in the new version. Psychometric properties of these four items indicate convincing validity across all GA groups, but GA should be considered in pain assessment because different GA-groups require different cut-off points.

Factor structure and reliability of the BPSN

The factor analysis showed that a model that includes the items crying, facial expression, and posture fits the data best. In fact, facial expression, crying, and body movement are widely studied indicators for pain assessment in neonates and are considered the most sensitive behavioural indicators of pain [4, 59, 60].

Facial expression is considered the most reliable and sensitive indicator for pain assessment in both preterm and full-term neonates [4]. Facial expressions extremely preterm neonates are likely to show include brow bulge, eye squeeze, nasolabial furrow, and vertical mouth stretch [20]. The BPSN more generally assesses facial expression, which aids in assessing preterm infants who wear CPAP masks and tapes to fix tubes to the skin, which can make it difficult to assess specific components of expression, like nasolabial furrow. The PIPP-R item nasolabial furrow was the least frequently rated item in our study, often because it was obscured by CPAP masks or tapes.

Crying is a common pain response in neonates and is included in several pain scales (e.g., [27, 61–63]), but some have questioned crying as an indicator of pain because it cannot be assessed in some neonates [21, 59]. Mechanical ventilation, inhibiting drugs, severe illness, and other reasons may limit the ability to cry. Although crying is not specific to pain [59], it may be the first indication a caregiver has that an infant is in pain [64]. Preterm neonates with immature facial muscles are less able to communicate their pain through facial expressions, so crying can alert their caregivers [17].

Several pain assessment tools include one or more items that assess body movements (e.g., [9, 61, 65, 66]. Holsti, Grunau, Oberlander and Whitfield [67] analysed behavioural pain reaction of early preterm neonates with the Newborn Individualized Development Care and Assessment Program (NIDCAP). They found that neonates flexed and extended their arms and legs, put their hands

Table 6 Cronbach's Alpha, Corrected Item-Total Correlation and Alpha if Item Deleted calculated for the modified behavioural subscale of the Bernese Pain Scale for Neonates

	Heel Stick Phase 1		Heel Stick Phase 2		Heel Stick Phase 3		Heel Stick Phase 4		Heel Stick Phase 5	
	Median (Range)		Median (Range)		Median (Range)		Median (Range)		Median (Range)	
Cronbach's α	0.876 (0.841–0.922)		0.848 (0.778–0.885)		0.845 (0.762–0.893)		0.815 (0.725–0.884)		0.825 (0.669–0.852)	
	r_{cor}*	α**	r_{cor}*	α**	r_{cor}*	α**	r_{cor}*	α**	r_{cor}*	α**
Crying	0.772	0.850	0.693	0.831	0.689	0.822	0.678	0.768	0.957	0.783
	(0.680–0.875)	(0.801–0.916)	(0.572–0.752)	(0.711–0.881)	(0.559–0.777)	(0.705–0.881)	(0.427–0.772)	(0.689–0.880)	(0.297–0.698)	(0.761–0.846)
Facial expression	0.817	0.773	0.774	0.736	0.781	0.703	0.668	0.696	0.778	0.653
	(0.793–0.907)	(0.700–0.837)	(0.704–0.854)	(0.598–0.783)	(0.699–0.867)	(0.550–0.781)	(0.656–0.845)	(0.444–0.778)	(0.663–0.851)	(0.296–0.682)
Posture	0.741	0.829	0.737	0.790	0.683	0.809	0.694	0.753	0.647	0.781
	(0.679–0.849)	(0.786–0.894)	(0.569–0.794)	(0.718–0.846)	(0.547–0.776)	(0.695–0.866)	(0.567–0.793)	(0.617–0.822)	(0.521–0.687)	(0.519–0.840)

Note. Median = Median of the coefficients calculated for each rater separately; Range = Range of the five coefficients calculated for each rater; *r_{cor} = Corrected Item-Total Correlation; **α = Cronbach's Alpha if Item deleted; Number of observations per measurement point $N = 94–143$

Table 7 Pearson product-moment correlation coefficients of the correlations between the modified behavioural Bernese Pain Scale for Neonates-subscale and heart rate

	Heel Stick Phase 1	Heel Stick Phase 2	Heel Stick Phase 3	Heel Stick Phase 4	Heel Stick Phase 5
N	144	140	118	109	97
Median (Range)	0.316* (0.237–0.329*)	0.235 (0.183–0.285)	0.234 (0.102–0.327*)	0.188 (0.155–0.251)	0.305 (0.223–0.379*)

Note. Median = Median of the Pearson product-moment correlation calculated for each rater separately; * Bonferroni adjusted *p*-value < 0.001

Table 8 Results of the linear mixed modelling analysis for the modified behavioural Bernese Pain Scale for Neonates-subscale and heart rate

	Behavioural Subscale		
Likelihood Ratio Test	χ^2	**df**	***p*-value**
Phases	864.18	10	< 0.001
Fixed effects	**Estimated coefficients**	**Std. Error**	***t*-value**
Intercept	0.265	0.458	0.579
Baseline phase	−1.041	0.069	−15.008
Recovery phase	−1.134	0.069	−16.040
Measurement point 2	− 0.130	0.070	−1.852
Measurement point 3	0.079	0.074	1.077
Measurement point 4	− 0.078	0.076	−1.038
Measurement point 5	0.097	0.079	1.238
GA at time of birth	0.063	0.012	5.488
Gender (female)	−0.017	0.101	−0.170
Measurement point 2 * Baseline	0.251	0.099	2.538
Measurement point 2 * Recovery	0.309	0.098	3.140
Measurement point 3 * Baseline	0.147	0.104	1.419
Measurement point 3 * Recovery	0.071	0.103	0.682
Measurement point 4 * Baseline	0.308	0.106	2.919
Measurement point 4 * Recovery	0.354	0.105	3.359
Measurement point 5 * Baseline	0.062	0.109	0.569
Measurement point 5 * Recovery	0.067	0.109	0.614
	Item Heart rate		
Likelihood Ratio Test	χ^2	**df**	***p*-value**
Phases	172.91	5	< 0.001
Fixed effects	**Estimated coefficients**	**Std. Error**	***t*-values**
Intercept	−0.563	0.221	−2.547
Recovery phase	−0.646	0.088	−7.383
Measurement point 2	0.023	0.089	0.260
Measurement point 3	−0.199	0.093	−2.139
Measurement point 4	−0.141	0.095	−1.477
Measurement point 5	0.117	0.099	1.183
GA at time of birth	0.042	0.007	6.145
Gender (female)	0.003	0.054	0.051
Measurement point 2 * Recovery	0.021	0.126	0.167
Measurement point 3 * Recovery	0.206	0.131	1.578
Measurement point 4 * Recovery	0.155	0.133	1.160
Measurement point 5 * Recovery	−0.032	0.138	−0.231

Note. χ^2 = Chi-square value; df = degrees of freedom; *N* = 154. Bonferroni adjusted *p*-value < 0.025

Table 9 Pearson product-moment correlation coefficients of the correlations between the total scores of the modified Bernese Pain Scale for Neonates and the PIPP-R

	Heel Stick Phase 1	Heel Stick Phase 2	Heel Stick Phase 3	Heel Stick Phase 4	Heel Stick Phase 5
N	144	140	118	109	97
Median (Range)	0.697** (0.652**-0.758**)	0.709** (0.662**-0.735**)	0.688** (0.649-**0.723**)	0.666** (0.636**-0.735**)	0.648** (0.600**-0.711**)

Note. Correlation coefficients were calculated for the heel stick phases of the five measurement points (t1-t5); Median = Median of the Pearson product-moment correlation coefficients that were calculated separately for each rater; **$p < 0.01$; * $p < 0.05$

on their faces, fisted, and finger splayed more often during the heel stick procedure. Morison et al. [68] found neonates with lower GA at birth made more specific body movements but had less facial expression at 32 weeks post-conceptional age, which suggests assessing body movements could provide useful supplementary information about preterm neonates. The BPSN more generally assesses body movement by evaluating a neonate's posture on a 4-point Likert-scale, ranging from relaxed body to permanent tension. Our results suggest that posture is a sensitive indicator for assessing pain across GA-groups.

We found that heart rate and oxygen saturation did not load on a common physiological factor or correlate with each other. Because heart rate was more sensitive to pain and more strongly associated with the three behavioural indicators of pain, we included heart rate in the new version of the BPSN. The results of our analyses confirm previous findings that correlations between behavioural and physiological indicators of pain were low [69–71], behavioural indicators were more sensitive to pain than physiological indicators [69, 72], and heart rate was more sensitive to pain than oxygen saturation [71].

Table 10 Results of the ROC analyses for the modified Bernese Pain Scale for Neonates total score

		Cut-off points			AUC
Heel stick phase	N	0.5	1.5	2.5	[95% CI]
t1	144				
Sensitivity		0.926	**0.853**	0.724	0.863
Specificity		0.515	**0.662**	0.857	[0.800–0.926]
t2	140				
Sensitivity		0.908	**0.811**	0.667	0.825
Specificity		0.442	**0.597**	0.805	[0.756–0.894]
t3	118				
Sensitivity		0.874	**0.769**	0.631	0.812
Specificity		0.424	**0.672**	0.858	[0.734–0.890]
t4	109				
Sensitivity		0.870	**0.750**	0.574	0.812
Specificity		0.484	**0.685**	0.876	[0.730–0.894]
t5	97				
Sensitivity		0.869	**0.794**	0.646	0.812
Specificity		0.383	**0.670**	0.879	[0.722–0.902]

Note. The PIPP-R was the reference value, with a cut-off point of 6.5 that discriminated between no/low pain (≤ 6 points) and moderate to high pain (≥ 7 points); AUC = Area under the curve; [95% CI] = 95% confidence intervals of the AUC; the results were originally computed separately for each rater and aggregated assuming normal distribution of the parameters; bold-set font = cut-offs with sensitivity and specificity nearest 80%

Though factor loadings of crying, facial expression, and posture did not vary within raters during the heel stick phases, they did vary between raters. This result suggests that different raters assess pain differently, an assumption further supported by the results of our interrater reliability analysis. There was good to excellent interrater agreement on crying, but agreement on facial expression and posture ranged from poor to good [73], depending on the measurement point and the model to calculate ICCs. The differences in interrater reliability could be explained by differences in the way raters defined the items. Crying may be a more objective and reliable item than facial expression or posture because it considers duration. Improving the guidelines and training for applying the BPSN may improve interrater agreement.

The first validation study of the BPSN [24] used Cronbach's Alpha reliability coefficient to calculate interrater reliability, and found interrater reliability of the subjective subscale of the BPSN ($r = 0.77–0.97$) was high. Cronbach's Alpha determines if the ratings of two or more persons are consistent, but it does not measure absolute agreement [74]. Since the cut-off differentiates between a painful and non-painful state, agreement between nurses and other caregivers about an infant's level of pain is crucial. We thus decided to use the more stringent absolute agreement model to calculate interrater reliability.

Interrater agreement and factor loadings of the items crying, facial expression, consolation, and posture tended to decrease over time. Cronbach's Alpha and corrected item-total correlations of the items crying, facial expression, and posture tended to decrease too. This accords with the results of another study that showed high within-subject variability among preterm neonates' pain reaction across repeated measurement points [75]. Interrater reliability was high during the heel sticks 1–3 and decreased during heel sticks 4–5. These findings cannot be explained by rater fatigue, because the video sequences were analysed in random order. The variability in pain reactions might be explained by the

Table 11 Intraclass Correlation Coefficients for the subjective Bernese Pain Scale for Neonates-items calculated with two-way random-effects, absolute agreement models

	Extremely Preterm Neonates	Very Preterm Neonates	Moderate to Late Preterm Neonates	Full-term Neonates
	Heel Stick Phases t1-t5 Range (Median)	Heel Stick Phases t1-t5 Range (Median)	Heel Stick Phases t1-t5 Range (Median)	Heel Stick Phases t1-t5 Range
Sleeping				
N	41–47	32–44	20–34	14–20
Single measures	0.175–0.310 (0.260)	0.145–0.356 (0.198)	0.090–0.289 (0.160)	0.155–0.225
Average measures	0.515–0.692 (0.637)	0.459–0.734 (0.553)	0.330–0.670 (0.487)	0.478–0.592
Crying				
N	40–47	33–44	21–35	14–20
Single measures	0.622–0.794 (0.701)	0.538–0.786 (0.716)	0.564–0.783 (0.702)	0.619–0.680
Average measures	0.892–0.951 (0.921)	0.854–0.948 (0.926)	0.866–0.948 (0.922)	0.890–0.914
Consolation				
N	40–47	33–44	21–35	14–20
Single measures	0.227–0.281 (0.257)	0.216–0.565 (0.390)	0.374–0.598 (0.469)	0.389–0.684
Average measures	0.595–0.661 (0.634)	0.579–0.866 (0.761)	0.749–0.881 (0.815)	0.761–0.915
Skin colour				
N	41–48	34–44	21–36	13–19
Single measures	0.010–0.058 (0.051)	0.002–0.104 (0.062)	0.057–0.166 (0.069)	0.071–0.080
Average measures	0.049–0.236 (0.211)	0.011–0.367 (0.248)	0.230–0.498 (0.271)	0.276–0.302
Facial expression				
N	41–46	31–40	20–34	13–19
Single measures	0.392–0.514 (0.436)	0.498–0.698 (0.526)	0.438–0.748 (0.601)	0.616–0.817
Average measures	0.763–0.841 (0.794)	0.832–0.921 (0.847)	0.796–0.937 (0.883)	0.889–0.957
Posture				
N	42–48	34–44	21–35	14–20
Single measures	0.333–0.479 (0.420)	0.369–0.501 (0.472)	0.286–0.685 (0.519)	0.576–0.795
Average measures	0.714–0.821 (0.783)	0.745–0.834 (0.817)	0.667–0.916 (0.839)	0.872–0.951
Breathing				
N	36–41	29–37	17–35	9–14
Single measures	0.019–0.378 (0.287)	0.313–0.507 (0.371)	0.158–0.419 (0.314)	0.171–0.317
Average measures	0.090–0.752 (0.669)	0.695–0.837 (0746)	0.485–0.783 (0.696)	0.508–0.699
Breathing Raw Scores				
N	39–45	32–40	19–35	11–14
Single measures	0.508–0.680 (0.618)	0.530–0.637 (0.587)	0.655–0.780 (0.681)	0.558–0.664
Average measures	0.838–0.914 (0.890)	0.850–0.898 (0.876)	0.905–0.947 (0.914)	0.863–0.908

Note. N = Number of observations per measurement point

influence of individual contextual factors and needs to be investigated [1, 2, 20, 21].

Validity of the modified BPSN

The modified BPSN that includes crying, facial expression, posture, and heart rate showed good construct validity and concurrent validity with the PIPP-R. Pain scores on the behavioural subscale averaged more than one point higher during the heel stick than during the baseline and recovery phases. Pain scores on heart rate averaged 0.65 points higher during the heel stick phase than during the recovery phase. Neonates' GA at time of birth influenced their pain scores. With every additional week of GA, pain scores on the behavioural subscale (crying, facial expression, posture) increased about 0.063 points. If we apply this result on our study sample with a wide range of GAs (24 2/7–42 5/7 weeks of GA), behavioural pain reaction of the neonate with

Table 12 Cronbach's Alpha for the modified behavioural Bernese Pain Scale for Neonates-subscale, separated by GA-groups

	Cronbach's Alpha				
	Heel Stick Phase 1 Median (Range)	Heel Stick Phase 2 Median (Range)	Heel Stick Phase 3 Median (Range)	Heel Stick Phase 4 Median (Range)	Heel Stick Phase 5 Median (Range)
Extremely preterm neonates	0.819 (0.813–0.894)	0.821 (0.695–0.862)	0.760 (0.720–0.883)	0.796 (0.690–0.841)	0.830 (0.691–0.840)
N = 42–48					
Very preterm neonates	0.908 (0.833–0.915)	0.835 (0.787–0.868)	0.800 (0.705–0.878)	0.794 (0.624–0.902)	0.824 (0.708–0.841)
N = 32–44					
Moderate to late preterm neonates	0.836 (0.736–0.932)	0.863 (0.724–0.930)	0.892 (0.844–0.924)	0.872 (0.765–0.896)	0.774 (0.576–0.871)
N = 20–36					
Full-term neonates	0.909 (0.906–0.964)	0.832 (0.813–0.932)			
N = 13–20					

Note. Median = Median of the coefficients calculated for each rater separately; Range = Range of the five coefficients calculated for each rater

the highest GA was about 1.13 points higher than pain reaction of the neonate with the lowest GA. Heart rate of the neonate with the highest GA was also about 0.76 points higher than heart rate of the neonate with the lowest GA. Like other studies that analysed the relationship between gender and pain reaction in neonates (e.g., [76–78]), we found gender had no effect on the level of pain scores.

Sensitivity and specificity of the modified BPSN

The results of the sensitivity and specificity analyses suggest that a cut-off of 1.5 points (total overall score = 12 points) would discriminate between no to low pain and moderate to high pain (measured with the PIPP-R). For the original BPSN scale, the cut-off was much higher, at 10.5 points (total overall score = 27 points). We found that the mean of the BPSN total scale that included nine

Table 13 Results of the ROC analyses for the modified Bernese Pain Scale for Neonates total score, separated for GA-groups

	Heel Stick Phases of Measurement Points t1-t5			
	AUC Range (Median)	Cut-off points	Sensitivity Range (Median)	Specificity Range (Median)
Extremely Preterm Neonates	0.707–0.878 (0.801)			
N = 42–48		**0.5**	**0.734–0.875 (0.839)**	**0.398–0.562 (0.538)**
		1.5	0.637–0.765 (0.666)	0.691–0.853 (0.713)
		2.5	0.410–0.594 (0.494)	0.901–0.970 (0.945)
Very Preterm Neonates	0.810–0.930 (0.852)			
N = 34–44		0.5	0.849–0.970 (0.905)	0.284–0.606 (0.439)
		1.5	**0.745–0.901 (0.811)**	**0.638–0.728 (0.680)**
		2.5	0.596–0.785 (0.648)	0.864–0.977 (0.902)
Moderate to Late Preterm Neonates	0.874–0.941 (0.927)			
N = 21–37		1.5	0.900–0.990 (0.970)	0.564–0.660 (0.581)
		2.5	**0.763–0.950 (0.897)**	**0.705–0.832 (0.787)**
		3.5	0.532–0.763 (0.675)	0.879–0.975 (0.933)
Full-term Neonates	0.893–0.906			
N = 14–20		2.5	0.942–0.959	0.419–0.664
		3.5	**0.807–0.888**	**0.808–0.824**
		4.5	0.714–0.831	0.836–0.896
		5.5	0.423–0.751	0.892–0.969

Note. The PIPP-R was the reference value, with a cut-off point of 6.5 that discriminated between no/low pain (≤ 6 points) and moderate to high pain (≥ 7 points); AUC = Area under the curve; [95% CI] = 95% confidence intervals of the AUC; the results were originally computed separately for each rater and aggregated assuming normal distribution of the parameter; Range = heel stick phases of measurement points t1-t5; bold-set font = cut-offs with sensitivity and specificity nearest 80%

items varied widely and depended on the rater, but it did not reach the cut-off value of 11 points during the heel stick phases of the five measurement points. The preliminary dose of oral sucrose administered to neonates before each heel stick may have lowered pain scores in our study [28]. In the first validation study of the BPSN, neonates received no pain relieving intervention before the heel stick, and BPSN total scores increased significantly during the heel stick, averaging 15.96 points ($SD = 5.7$) [24]. The relief provided by sucrose should be factored into the decision about a new cut-off value for the modified BPSN.

Comparison of different GA-groups

Neonates with younger GA at birth had lower pain scores than more mature infants. The results of the separate sensitivity and specificity analyses for the four GA-groups indicated as GA increases, so should the cut-off of the BPSN that discriminates between no to low pain and moderate to high pain (measured with the PIPP-R). To reach a sensitivity and specificity of approximately 80%, extremely preterm neonates require a cut-off value of 0.5 points, very preterm neonates require 1.5 points, moderate to late preterm neonates require 2.5 points, and full-term neonates require 3.5 points. Our ROC analysis showed that the modified BPSN was least able, but still moderately good [41], to discriminate between neonates who experience no or low pain and neonates who experience moderate to high pain in the group of extremely preterm neonates and increases with increasing GA. Extremely preterm neonates' pain expression may be less apparent because their immature nervous system and facial muscles prevent them from expressing a robust pain reaction [20, 21, 60, 68]. Understanding the difficulty this poses for accurate pain assessment in extremely preterm neonates could be helpful when establishing cut-off values for the BPSN. Based on our study results, we recommend differentiating between GA-groups and establishing cut-off values based on GA. The PIPP-R already includes GA in pain assessment; the younger the GA, the more points PIPP-R adds to the pain score [26].

The other analyses we conducted separately for the four GA-groups showed that concurrent validity of the modified BPSN total score with the PIPP-R was highest for full-term neonates ($r = 0.814–0.834$) and lowest, but still good, for extremely preterm neonates ($r = 0.631–0.710$). Interrater agreement on facial expression and posture tended to improve as GA increased.

Limitations

This study is limited, first, by our decision to rate neonates' pain expression from video sequences. Characteristics of the videos may have affected the reliability of the ratings (e.g., poor lighting conditions, quality of the raters' screen, position of the neonate, several assistants for video recording). Second, different nurses performed the heel sticks, and their individual characteristics may have influenced neonates' pain reaction. Third, particularly during the baseline and recovery phases, where the scores of the items were low, floor effects may have influenced our study results. For example, we considered a variety of extensions of the model specification in our factor analysis but discarded them because of convergence problems likely related to floor effects, when upper categories were almost or completely left empty. Treating the rating scores as numeric did not resolve floor effect problems, or rather the opposite [79], but allowed to obtain results. Floor effects may also have lowered interrater agreement, especially during the baseline and recovery phases. Fourth, our later hypothesis testing may be compromised by measurement error caused by low interrater agreement [40]. We compensated for this possible problem by either including the raters in the model, or by conducting separate analyses for each rater and then pooling the results. Fifth, pain reaction was measured during the heel stick, so our results cannot be generalized to other acute painful procedures or more persistent or chronic pain. The BPSN is used for routine pain assessment in NICUs and should therefore be sensitive to repeated and more prolonged and chronic pain, so future validation studies should assess and compare the level of pain scores during different painful situations.

Conclusions

The modified version of the BPSN that includes facial expression, crying, posture, and heart rate is a promising tool for assessing acute pain in full-term and preterm neonates across gestational ages, but our results suggest that adding different cut-off points for different GA-groups will improve the BPSN's clinical usefulness.

Abbreviations

AUC: Area under the curve; BPSN: Bernese Pain Scale for Neonates; CFA: Confirmatory factor analysis; CFI: Comparative Fit Indices; CPAP: Continuous positive airway pressure; GA: Gestational age; ICC: Intraclass Correlation Coefficient; *M*: Mean; *Mdn*: Median; NICU: Neonatal intensive care unit; NIDCAP: Newborn Individualized Development Care and Assessment Program; OLInv: Overall loadings invariance; PIPP-R: Premature Infant Pain Profile-Revised; RMSEA: Root mean square errors of approximation; ROC: Receiver-operating characteristic; *SD*: standard deviation; TLI: Tucker-Lewis Indices; WRLInv: Within-rater loadings invariance

Acknowledgements

We are grateful to all parents who granted permission to include their infant in the study. We thank the Swiss National Science Foundation for their financial contribution. Further thanks to all nurses, head nurses, and medical directors of the University Hospital NICUs in Basel, Bern, and Zurich for their support and efforts to make this study possible. We thank Janik Schneeberger for developing the web-based rating tool and supporting us in all technical acquisitions. We would also like to thank the Department of Clinical Research team of the University of Bern for developing the web-based data capture system, and the New Media Centre team of the University of Basel for their

professional support for video-data elaboration. We thank Dr. Kali Tal for her editing work. Finally, we thank the research assistants, without whose help and support the data collection in this study would not have been feasible.

Funding
This research was funded by the Swiss National Science Foundation (SNF 320030_159573).

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
EC conceived of and designed the study and was the primary investigator. DB, SS, MN, and BS co-authored the study proposal and supported EC in designing the study. DB, SS, and MN provided access to the research field. The doctoral student KS was responsible for all tasks of the data collection process and data entry, management and analysis. She was also responsible for reporting and disseminating the outcomes in peer-reviewed journals and conferences. LS recruited at the University Hospital NICU in Bern and supported the doctoral student in the data collection process. RB supported the doctoral student in the data analyses. All authors read and approved the final manuscript.

Ethics approval and consent to participate
The study was approved by the Ethics Committee Bern (2015–238), the Ethics Committee northwest/central Switzerland EKNZ (2015–385) and the Ethics Committee Zurich (2015–563).
Written informed consent was obtained from parents according to the protocol approved by the ethics committees. We did not expose infants to additional painful situations. No heel sticks were performed solely for research purposes. We upheld the current standard of care in pain prevention by giving oral sucrose to all infants before the heel stick procedure.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Division of Midwifery, Department of Health Professions, Bern University of Applied Sciences, Murtenstrasse 10, 3008 Bern, Switzerland. [2]Neonatalogy, Children's Hospital, University Hospital of Bern, Bern, Switzerland. [3]Lawrence S. Bloomberg Faculty of Nursing and Faculties of Medicine and Dentistry, University of Toronto, Toronto, ON, Canada. [4]Department of Neonatology, University Hospital Zurich and University of Zurich, Zurich, Switzerland. [5]Department of Neonatology, University of Basel Children's Hospital (UKBB), Basel, Switzerland.

Received: 16 April 2018 Accepted: 20 December 2018
Published online: 15 January 2019

References

1. Sellam G, Cignacco EL, Craig KD, Engberg S. Contextual factors influencing pain response to heelstick procedures in preterm infants: What do we know? A systematic review. European Journal of Pain. 2011;15:661.e661–15.
2. Sellam G, Engberg S, Denhaerynck K, Craig KD, Cignacco EL. Contextual factors associated with pain response of preterm infants to heel-stick procedures. Eur J Pain. 2013;17:255–63.
3. Anand KJ. International evidence-based Group for Neonatal Pain. Consensus statement for the prevention and management of pain in the newborn. Arch Pediatr Adolesc Med. 2001;155(2):173–80.
4. Cong X, McGrath JM, Cusson RM, Zhang D. Pain assessment and measurement in neonates: an updated review. Advances in Neonatal Care. 2013;13(6):379–95.
5. Anand KJS. Pain assessment in preterm neonates. Pediatrics. 2007;119(3):605–7.
6. Lee GY, Stevens BJ. Neonatal and infant pain assessment. In: McGrath P, Stevens B, Walker SM, Zempsky WT, editors. Oxford textbook of Paediatric pain. Oxford: Oxford University Press; 2014. p. 353–69.
7. Pillai Riddell R, Fitzgerald M, Slater R, Stevens B, Johnston C, Campbell-Yeo M. Using only behaviours to assess infant pain: a painful compromise? Pain. 2016;157(8):1579–80.
8. Ranger M, Johnston CC, Anand K. Current controversies regarding pain assessment in neonates. Semin Perinatol. 2007;31:283–8.
9. Holsti L, Grunau RE. Initial validation of the behavioral indicators of infant pain (BIIP). Pain. 2007;132(3):264–72.
10. Holsti L, Grunau RE, Oberlander TF, Osiovich H. Is it painful or not? Discriminant validity of the behavioral indicators of infant pain (BIIP) scale. Clin J Pain. 2008;24(1):83–8.
11. Lucas-Thompson R, Townsend EL, Gunnar MR, Georgieff MK, Guiang SF, Ciffuentes RF, et al. Developmental changes in the responses of preterm infants to a painful stressor. Infant Behav Dev. 2008;31(4):614–23.
12. Morison SJ, Grunau RE, Oberlander TF, Whitfield MF. Relations between behavioral and cardiac autonomic reactivity to acute pain in preterm neonates. Clin J Pain. 2001;17(4):350–8.
13. Hummel P, van Dijk M. Pain assessment: current status and challenges. Semin Fetal Neonatal Med. 2006;11(4):237–45.
14. Johnston CC, Fernandes AM, Campbell-Yeo M. Pain in neonates is different. Pain. 2011;152(3):S65–73.
15. American Academy of Pediatrics, Committee on Fetus and Newborn, Section on Anesthesiology and Pain Medicine. Prevention and Management of Procedural Pain in the Neonate: An Update. Pediatrics. 2016;137(2):e20154271.
16. Stevens B, Johnston C, Petryshen P, Taddio A. Premature infant pain profile: development and initial validation. Clin J Pain. 1996;12(1):13–22.
17. Johnston CC, Stevens B, Craig KD, Grunau RVE. Developmental changes in pain expression in premature, full-term, two- and four-month-old infants. Pain. 1993;52:201–8.
18. Johnston CC, Stevens BJ. Experience in a neonatal intensive care unit affects pain response. Pediatrics. 1996;98(5):925–30.
19. Grunau RE, Oberlander TF, Whitfield MF, Fitzgerald C, Lee SK. Demographic and therapeutic determinants of pain reactivity in very low birth weight neonates at 32 Weeks' postconceptional age. Pediatrics. 2001;107(1):105–12.
20. Gibbins S, Stevens B, Beyene J, Chan PC, Bagg M, Asztalos E. Pain behaviours in extremely low gestational age infants. Early Hum Dev. 2008;84:451–8.
21. Gibbins S, Stevens B, McGrath PJ, Yamada J, Beyene J, Breau L, et al. Comparison of pain responses in infants of different gestational ages. Neonatology. 2008;93:10–8.
22. Johnston CC, Stevens BJ, Yang F, Horton L. Differential response to pain by very premature neonates. Pain. 1995;61(3):471–9.
23. American Academy of Pediatrics, Committee on Fetus and Newborn, Section on Surgery, and Section on Anesthesiology and Pain Medicine, Canadian Paediatric Society, Fetus and Newborn Committee. Prevention and Management of Pain in the neonate: an update. Pediatrics. 2006;118(5):2231–41.
24. Cignacco E, Mueller R, Hamers JPH, Gessler P. Pain assessment in the neonates using the Bernese pain scale for neonates. Early Hum Dev. 2004; 78:125–31.
25. Boettcher M, Göttler S, Stoffel L, Schwab K, Berger S, Mérat M. Schmerzmanagement bei Kindern in der Schweiz. Monatsschrift Kinderheilkunde. 2012;160(9):887–94.
26. Stevens BJ, Gibbins S, Yamada RN, Dionne K, Lee G, Johnston C, et al. The premature infant pain profile-revised (PIPP-R) initial validation and feasibility. Clin J Pain. 2014;30(3):238–43.
27. Cignacco E, Schenk K, Stevens B, Stoffel L, Bassler D, Schulzke S, et al. Individual contextual factors in the validation of the Bernese pain scale for neonates: protocol for a prospective observational study. BMC Pediatr. 2017;17(1):171.
28. Stevens B, Yamada J, Ohlsson A, Haliburton S, Shorkey A. Sucrose for analgesia in newborn infants undergoing painful procedures (review). Cochrane Database of Systematic Review. 2016;7:CD001069.
29. Final Cut Pro X. Version 10.2.3. Cupertino: Apple Inc.; 2001-2016.
30. Gibbins S, Stevens BJ, Yamada J, Dionne K, Campbell-Yeo M, Lee G, et al. Validation of the premature infant pain profile-revised (PIPP-R). Early Hum Dev. 2014;90:189–93.
31. Stevens B, Johnston C, Taddio A, Gibbins S, Yamada J. The premature infant pain profile: evaluation 13 years after development. Clin J Pain. 2010;26(9):813–30.
32. Cignacco EL, Sellam G, Stoffel L, Gerull R, Nelle M, Anand KJS, et al. Oral sucrose and "facilitated tucking" for repeated pain relief in Preterms: a randomized controlled trial. Pediatrics. 2012;129(2):299–308.

33. SPSS Inc. Released 2015. IBM SPSS statistics for windows, version 23.0. Armonk, NY: IBM Corp.
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. https://www.R-project.org/. Accessed 16 Apr 2018.
35. Bland JB, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995;310:170.
36. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 2008.
37. Hothorn T, Zeileis A. partykit: A modular toolkit for recursive partytioning in R. Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, University of Innsbruck. 2014;10.
38. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat. 2006;15(3):651–74.
39. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were propsed. Int J Nurs Stud. 2011;48:661–71.
40. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol. 2012;8(1):23–34.
41. Streiner DL, Norman GR, Cairney J. Health mesasurement scales, a practical guide to their development and use. 5th ed. Oxford: University Press; 2015.
42. Little TD. Longitudinal structural equation modeling. New York: The Guilford Press; 2013.
43. Rosseel Y. Lavaan: an R package for structural equation modeling. J Stat Softw. 2012;46(2):1–36.
44. Bentler PM. Comparative fit indexes in structural models. Psychological bulletin. 1990;107:2.
45. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973;38(1):1–10.
46. Brown TA. Confirmatory factor analysis for applied research. 1th ed. New York: The Guilford Press; 2006.
47. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J. 1999;6(1):1–55.
48. Jackson DL, Gillaspy JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: an overview and some recommendations. Psychol Methods. 2009;14(1):6–23.
49. Satorra A, Bentler PM. Ensuring positiveness of the scaled difference chi-square test statistic. Psychometrika. 2010;75(2):243–148.
50. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model. 2002;9(2):233–55.
51. Gliem JA, Gliem RR. Calculating, Iterpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. 2003 Midwest Research to Practice Conference in Adult, Continiuing, and Community Education. https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem%20%26%20Gliem.pdf?sequence=1&isAllowed=y. Accessed 27 Aug 2018.
52. Diedenhofen B, Musch J. Cocron: a web Interface and R package for the statistical comparison of Cronbach's alpha coefficients. Int J Internet Sci. 2016;11(1):51–60.
53. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull. 1980;87(2):245–51.
54. Bates D, Machler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48.
55. Winter B. Linear models and linear mixed effects models in R with linguistic applications. University of California, Merced, Cognitive and Information Sciences. 2013. http://arxiv.org/pdf/1308.5499.pdf. Accessed 16 Apr 2018.
56. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561–77.
57. World Health Organization. Preterm birth. 2017. https://www.who.int/en/news-room/fact-sheets/detail/preterm-birth. Accessed Dec 2018.
58. Kaplan RM, Saccuzzo DP. Psychological testing: principles, applications, and issues. 4th ed. Brooks/Cole: Pacific grove, CA; 1997.
59. Hatfield LA, Ely EA. Measurement of acute pain in infants: a review of behavioral and physiological variables. Biological Research for Nursing. 2015;17(1):100–11.
60. Gibbins S, Stevens B. The influence of gestational age on the efficacy and short-term safety of sucrose for procedural pain relief. Advances in Neonatal Care. 2003;3(5):241–9.
61. Hudson-Barr D, Capper-Michel B, Lambert S, Palermo TM, Morbeto K, Lombardo S. Validation of the pain assessment in neonates (PAIN) scale with the neonatal infant pain scale (NIPS). Neonatal Netw. 2002;21(6):15–21.
62. Hummel P, Puchalski M, Creech SD, Weiss MG. Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. J Perinatol. 2008;28(1):55–60.
63. Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. The FLACC: a behavioral scale for scoring postoperative pain in young children. Pediatr Nurs. 1997;23(3):293–7.
64. Craig KD, Korol CT, Pillai RR. Challenges of judging pain in vulnerable infants. Clin Perinatol. 2002;29(3):445–57.
65. Carbajal R, Paupe A, Hoenn E, Lenclen R, Olivier-Martin M. DAN: une échelle comportementale d'évaluation de la douleur aiguë du nouveau-né. Arch Pédiatr. 1997;4:623–8.
66. Lawrence J, Alcock D, McGrath P, Kay J, MacMurray SB, Dulberg C. The development of a tool to assess neonatal pain. Neonatal Netw. 1993;12(6):59–66.
67. Holsti L, Grunau RE, Oberlander TF, Whitfield MF. Specific newborn individualized developmental care and assessment program movements are associated with acute pain in preterm infants in the neonatal intensive care unit. Pediatrics. 2004;114(1):65–72.
68. Morison SJ, Holsti L, Grunau RE, Whitfield MF, Oberlander TF, Chan HW, et al. Are there developmentally distinct motor indicators of pain in preterm infants? Early Hum Dev. 2003;72:131–46.
69. Välitalo PAJ, Van Dijk M, Krekels EHJ, Gibbins S, Simons SHP, Tibboel D, et al. Pain and distress caused by endotracheal suctioning in neonates is better quantified by behavioural than physiological items: a comparison based on item response theory modelling. Pain. 2016;157(8):1611–7.
70. van Dijk M, de Boer JB, Koot HM, Duivenvoorden HJ, Passchier J, Bouwmeester N, et al. The association between physiological and behavioral pain measures in 0- to 3-year-old infants after major surgery. J Pain Symptom Manag. 2001;22:600–9.
71. Vederhus BJ, Eide GE, Natvig GK. Psychometric testing of a Norwegian version of the premature infant pain profile: an acute pain assessment tool. A clinical validation study. Int J Nurs Pract. 2006;12(6):334–44.
72. Craig KD, Whitfield MF, Grunau RVE, Linton J, Hadjistavropoulos D. Pain in the preterm neonate: behavioural and physiological indices. Pain. 1993;52:287–99.
73. Gicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284–90.
74. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas. 2007;1(1):77–89.
75. Cignacco E, Denhaerynck K, Nelle M, Bührer C, Engberg S. Variability in pain response to a non-pharmacological intervention across repeated routine pain exposure in preterm infants: a feasibility study. Acta Paediatr. 2009;98(5):842–6.
76. Holsti L, Grunau RE, Whifield MF, Oberlander TF, Lindh V. Behavioral responses to pain are heightened after clustered care in preterm infants born between 30 and 32 weeks gestational age. Clin J Pain. 2006;22(9):757.
77. Williams AL, Khattak AZ, Garza CN, Lasky RE. The behavioral pain response to heelstick in preterm neonates studied longitudinally: description, development, determinants, and components. Early Hum Dev. 2009;85(6):369–74.
78. Johnston CC, Stevens BJ, Franck LS, Jack A, Stremler R, Platt R. Factors explaining lack of response to heel stick in preterm newborns. JOGNN. 1999;28(7):587–94.
79. Agresti A. Analysis of ordinal categorical data. 2nd ed. Hoboken, NJ: Wiley; 2010.

Appendix IV: Manuscript 3

Schenk, K, Stoffel, L., Bürgin, R., Stevens, B., Bassler, D. Schulzke, S., Nelle, M., & Cignacco, E. The Bernese Pain Scale for Neonates-Revised accounts for individual contextual factors. *Manuscript submitted for publication*.

Draft March 20, 2019

The Bernese Pain Scale for Neonates-Revised accounts for individual contextual factors

K. Schenk[1], L. Stoffel[2], R. Bürgin[1], B. Stevens[3], D. Bassler[4], S. Schulzke[5], M. Nelle[4], E. Cignacco[1]

[1]Division of Midwifery, Department of Health Professions, Bern University of Applied Sciences, Switzerland; [2]Neonatalogy, Children's Hospital, University Hospital of Bern, Switzerland; [3]Lawrence S. Bloomberg Faculty of Nursing and Faculties of Medicine and Dentistry, University of Toronto, Canada; [4]Department of Neonatology, University Hospital Zurich and University of Zurich, Switzerland; [5]Department of Neonatology, University Children's Hospital Basel UKBB, Switzerland

Corresponding author:

Karin Schenk, Bern University of Applied Sciences, Department of Health Professions, Murtenstrasse 10, 3008 Bern, Switzerland

Phone: +41 31 848 47 25, E-Mail: karin.schenk@bfh.ch

Category for which the manuscript is being submitted: Original article

Funding: This research was funded by the Swiss National Science Foundation (SNF 320030_159573).

Conflicts of interest disclosures: The authors declare that they have no conflicts of interest disclosures.

Significance: We identified individual contextual factors associated with dampened pain response in neonates and incorporated them into the revised version of the Bernese Pain Scale for Neonates-Revised (BPSN-R), providing clinicians with a tool they can use to more accurately assess and then manage pain in this vulnerable population.

Running head: The Bernese Pain Scale for Neonates-Revised 3

Abstract

Background: Individual factors like gestational age (GA) or a history of painful experiences create a context for neonates' pain responses. If pain assessment instruments do not account for these individual contextual factors, they may not adequately capture a neonate's pain level.

Objectives: We set out to determine the influence of individual contextual factors on variability in pain response in neonates, and, if necessary, to revise the BPSN to reflect our findings. This is part of a comprehensive validation study designed to improve the Bernese Pain Scale for Neonates' (BPSN) by incorporating relevant individual factors that may influence a neonate's pain response.

Methods: We videotaped 154 full-term and preterm neonates of different GAs during 1-5 capillary heel sticks in their first 14 days of life. For each heel stick, we produced three video sequences: baseline, heel stick, and recovery. The randomized sequences were rated on the BPSN by five blinded nurses. Individual contextual factors were retrospectively extracted from patient charts and from the video recordings. We analysed the data in single and multiple linear mixed models.

Results: Premature or term birth ($p < 0.001$), behavioural state ($p < 0.01$), and caffeine ($p < 0.001$) were associated with changes in behavioural pain scores. Premature or term birth ($p = 0.004$), mechanical ventilation ($p = 0.002$), and duration of the heel stick procedure ($p < 0.001$) were associated with changes in physiological pain scores.

Conclusions: Individual contextual factors change behavioural or physiological pain patterns, so we incorporated relevant factors in the revised BPSN.

1. Introduction

Even though our understanding of pain in neonates has advanced significantly over the last 30 years, and spurred the development of myriad pain assessment tools for neonates, it is still clinically challenging to ascertain a painful status (AAP, Committee on Fetus and Newborn, and Section of Anesthesiology and Pain Medicine, 2016; Anand et al., 2017). A neonate's pain response may be influenced by more than the invasive procedure itself (Sellam et al., 2011). Individual contextual factors play a major role in the ability of a child to express his or her experienced pain. These include demographic factors (e.g., age, sex) and medical factors (e.g., health status, medication). For example, more premature neonates tend to demonstrate lower behavioural pain responses than preterm neonates of more mature gestational age (GA) or than full-term neonates (Johnston et al., 1995; Grunau et al., 2001; Gibbins et al., 2008a; Gibbins et al., 2008b; Williams et al., 2009). Ignoring individual factors that create a context in which neonates feel and express pain could reduce the accuracy of a clinician's estimate of how much pain the neonate is in, and make it harder to successfully manage their pain (Sellam et al., 2011; Valeri and Linhares, 2012; Hatfield and Ely, 2015). But considering these individual contextual factors could make pain assessment more accurate (Craig et al., 1993; Johnston et al., 1999; Sellam et al., 2013; AAP et al., 2016).

The consequences of painful experiences in very early life are well known. Early life painful experiences are associated with negative short- and long-term neurological consequences (Brummelte et al., 2012; Ranger et al., 2013; Ranger and Grunau, 2014; Vinall and Grunau, 2014; Vinall et al., 2014), like changes in structural and functional brain development in preterm neonates (Schneider et al., 2018). Since we need to effectively manage pain in neonates based on accurate pain assessments, we need to determine which factors clinicians should include in their clinical assessments of pain (Sellam et al., 2011).

There are already a few pain assessment scales for neonates that account for individual contextual factors like GA and behavioural state (e.g., the Premature Infants Pain Profile-Revised [PIPP-R; Gibbins et al., 2014]; the Neonatal Pain, Agitation and Sedation Scale [N-PASS; Hummel et al., 2008; Hummel, 2017]). Until today, none of the existing scales is considered as gold standard for clinical pain assessment in preterm and full-term neonates (Lee and Stevens, 2014).

The Bernese Pain Scale for Neonates (BPSN) is a multidimensional pain assessment tool that is already widely used in the clinical setting of Swiss neonatal intensive care units (NICU; Boettcher et al., 2012). Feedback from clinicians related to challenges in pain assessment with the BPSN in preterm neonates with low GA. Therefore, we decided to validate and revise the BPSN to better assess pain in neonates. We conducted a comprehensive validation study of the BPSN with a sample of 154 neonates of different GAs. The results of this study indicated a significant reduction of the scale from nine to four items: facial expression, crying, posture, and heart rate (Schenk et al., 2019). This modified BPSN showed good reliability and validity, but our analyses showed that the younger the GA, the lower the cut-off point should be that discriminates between no or low pain and moderate to severe pain (Schenk et al., 2019). At the same time, we set out to identify individual contextual factors that could explain variability in neonates' pain response across repeated measurement points. We set out to improve our assessment of acute procedural pain in preterm and full-term neonates of different GAs by exploring individual contextual factors, and to revise the BPSN accordingly to our study results.

2. Methods

2.1 Design

This study is an integral part of a prospective multicentre validation study with repeated measurement design (Cignacco et al., 2017). The main goal of the comprehensive validation study was to develop a revised version of the BPSN for clinical pain assessment that accounts for relevant individual contextual factors that influence neonates' pain response. In the first sub-study, we analysed the psychometric properties of the BPSN. The results of these analyses suggested that the original 9-item scale should be limited to four items (Schenk et al., 2019). In this sub-study, we explored the influence of individual contextual factors on the variability in neonates' pain response across repeated measurement points.

2.2 Ethical considerations

The study was approved by the Ethics Committee Bern (2015-238), the Ethics Committee Northwest/Central Switzerland (2015-385), and the Ethics Committee Zurich (2015-563). All parents gave written informed consent before participation, as described in the protocol approved by the ethics committees. No infant received a heel stick only for research purposes; infants were not exposed to any other type of pain. To reduce pain, all infants received oral sucrose before each heel stick.

2.3 Sample and Setting

Preterm and full-term neonates were recruited in the neonatal intensive care units (NICU) of three Swiss university hospitals (Basel, Bern and Zurich). Recruitment and data collection were ongoing, from January 1 to December 31, 2016. We extended data collection in Bern until January 31, 2017, to recruit more extremely premature neonates. Premature neonates born between 24 0/7 and 36 6/7 weeks of gestation were included if they were

expected to undergo 2-5 routine capillary heel sticks; full-term neonates born between 37 0/7 and 42 0/7 weeks of gestation were included if they were expected to have at least two capillary heel sticks during their first 14 days of life. Exclusion criteria were 1) high-grade intraventricular haemorrhage (grades III and IV); 2) severe life-threating malformation or any condition that caused partial or total loss of sensitivity; 3) arterial cord pH < 7.15 at birth; 4) previous surgery for any reason; or, 5) congenital malformation that impaired brain circulation and/or cardiovascular function.

2.4 Data collection

We consecutively sampled for the recruitment and stratified neonates into six different groups based on their GA at birth (Cignacco et al., 2017). Trained study assistants collected data from 1-5 routine heel stick procedures that were performed on each neonate. Each heel stick procedure was captured in three video sequences: baseline, heel stick, and recovery phases. We filmed with an HC-V757 high-definition camcorder (Panasonic, Osaka, Japan). Since causing neonates additional stress that could affect their pain reaction, bedside nurses were asked not to handle the neonates before we recorded the baseline phase. After the baseline phase was recorded, the bedside nurse warmed the neonate's heel; every neonate received a dose of 24% oral sucrose (0.2 ml/kg bodyweight). We started recording the heel stick phase when the nurse disinfected the neonate's heel; we recorded the recovery phase immediately after the heel stick phase. For each phase, the camera focused on the neonate's face for at least a minute. Afterwards, we recorded the infant's body for at least one more minute. Trained study assistants digitally elaborated each video sequence with Final Cut Pro X (Apple Inc., Cupertino, CA, USA) video editing software. To ensure rater were blinded, we removed any information that might have shown the heel stick phase (e.g., the arm of the nurse while she was performing the heel stick).

We produced a total of 1'941 video sequences, but excluded 130 because video quality was poor (e.g., neonate's face was not visible, poor light quality, reflections on the incubator). The remaining1'811 video sequences were uploaded into a web-based rating tool and randomized by sequence number, phase, and presentation order. Five nurses experienced in using the BPSN (*Mean* = 8.3 years, *Range* = 3.5-15 years) and blinded to the phase of the heel stick procedure and measurement points, independently rated neonates' behavioural pain responses on both the BPSN and the PIPP-R. Therefore, the nurses were also trained to use and score the PIPP-R. Our study assistants recorded the neonate's highest heart rate and lowest oxygen saturation from the neonate's monitors while the three phases of video were recorded. Individual contextual factors were retrospectively retrieved from patient charts. We captured a few of these contextual factors during the video recording (e.g., duration of the heel stick procedure).

2.5 Measures

2.5.1 Pain assessment

Neonates' pain response was measured with the BPSN (Cignacco et al., 2004) and the PIPP-R (Stevens et al., 2014). The BPSN is a multidimensional pain assessment tool that includes seven subjective items (sleeping, crying, consolation, skin colour, facial expression, posture, and breathing) and two physiological items (changes in heart rate and oxygen saturation). Each item is rated on a 4-point Likert scale (0, 1, 2, and 3); the higher the score, the more intense the pain response. Based on the results of our first sub-study, we significantly reduced the scale from nine to four items. The modified BPSN includes facial expression, crying, posture, and heart rate (Schenk et al., 2019). We then used the modified BPSN to analyse the effect individual contextual factors might have had on a neonate's pain level, and to see if they accounted for variation in pain reaction among infants. The modified

BPSN is reliable and valid for preterm and full-term neonates (Schenk et al., 2019). We calculated the interrater reliability for the behavioural subscale of the modified BPSN with two-way random-effects, average measure, absolute agreement models (Streiner et al., 2015); interrater reliability ranged from 0.874 to 0.931 (*Mdn* = 0.906) during the heel stick phases of the five measurement points.

In the first sub-study, we used the PIPP-R, a well-validated multidimensional pain assessment tool widely used in the clinical setting of North America and for research purposes (Gibbins et al., 2014; Stevens et al., 2014), to test concurrent validity and sensitivity and specificity of the modified BPSN (Schenk et al., 2019). The PIPP-R includes three behavioural items (brow bulge, eye squeeze, and naso-labial furrow) and two physiological items (changes in heart rate and oxygen saturation). The PIPP-R also accounts for GA and baseline behavioural state as contextual factors. Extremely preterm neonates and neonates in a quiet sleep state at baseline score the highest if their behavioural and/or physiological sub score is ≥ 1 (Stevens et al., 2014). Each of the seven items of the PIPP-R is rated on a 4-point Likert scale (0, 1, 2, and 3). In this sub-study, we considered the PIPP-R item baseline behavioural state as a contextual factor. We calculated interrater reliability for the item behavioural state with two-way random-effects, average measures, absolute agreement models; interrater reliability ranged from 0.874 to 0.915 (*Mdn* = 0.896) during the baseline phase of the five measurement points.

2.5.2 Individual contextual factors

We determined individual contextual factors based on the findings of a systematic review (Sellam et al., 2011) and a study that analysed the effect of individual contextual factors on variability in preterm neonates' pain response (Sellam et al., 2013). We collected contextual factors retrospectively from patient charts and during video recordings. Individual

contextual factors had three dimensions. The first is comprised of demographic contextual factors including GA at birth; sex; birth weight; nationality; parity; and, mode of delivery. The second are medical contextual factors including primary diagnosis (premature or term birth) and common comorbidities in preterm neonates like necrotizing enterocolitis, respiratory distress syndrome, patent ductus arteriosus, septic events, apnoea bradycardia syndrome. Other medical factors are the number of comorbidities since birth, neonate's health status during the first 12 hours since birth (measured by the Clinical Risk Index for Babies [CRIB; Bührer et al., 2000]), CPAP or mechanical ventilation at the time of the heel stick procedure, and medication administered from birth to the first recorded heel stick procedure and between recorded heel stick procedures, including sedatives (e.g., Propofol, Chloralhydrate), opioids (e.g., Morphine, Fentanyl), non-opioids (e.g., Paracetamol, Indomethacin), steroids, caffeine, antibiotics, catecholamines, medication for lung maturation during pregnancy, surfactant administered after birth, or muscle relaxant. The third dimension includes a neonate's history of painful and non-painful interventions, including the number of previous painful and non-painful interventions from birth and in the past 24 hours, the time since the last painful and non-painful interventions (0-15 minutes, 16-30 minutes, 31-45 minutes, 46-60 minutes, and, > 60 minutes), and the type of the last painful and non-painful intervention. We defined painful (e.g., heel stick, intubation) and non-painful (e.g., changing diaper, removal of nasogastric tube) interventions based on a previous study (Cignacco et al., 2008). At each measurement point, we retrieved the following contextual factors from the patient chart or from observations made while the video was recorded: postnatal age (PNA); post-menstrual age (PMA; GA at birth combined with PNA); weight; the duration of each heel stick procedure; and, the number of additional sucrose doses given during the heel stick procedures. The behavioural state at baseline was rated by the five nurses with the PIPP-R item behavioural state (Stevens et al., 2014).

2.6 Statistical methods

The five raters could rate single items of the BPSN and PIPP-R as "non-evaluable". For BPSN items crying and posture, 1.0% or less of the data were missing, and for facial expression 0.1 to 4.0% (*Mdn* = 0.8%) of the data were missing. For PIPP-R item behavioural state, 0.1 to 0.9% (*Mdn* = 0.4%) were missing and for heart rate, less than 1% of data. Since we could not compute BPSN sum scores when an item was unrated, we used multiple imputation (Rubin, 2008) and the R-package *partykit* (Hothorn and Zeileis, 2014) to replace the values of unrated items with random substitutes generated from conditional inference regression trees (Hothorn et al., 2006). We generated five data sets this way, performed statistical analyses with each data set separately, and then pooled the results.

We analysed the influence of individual contextual factors on the variability in neonates' pain reaction by conducting linear mixed effect analyses using the statistical program R (R Core Team, 2017) and specifically the R-package *lme4* (Bates et al., 2015). Because the behavioural BPSN sub-score and the BPSN item heart rate had low correlation *(r* = 0.102-0.379) during the heel stick procedures (Schenk et al., 2019), we analysed the modified BPSN total scale and the behavioural and physiological subscales separately.

We divided the analysis into two stages, based on the methodology of a study that analysed the influence of individual contextual factors on pain reaction (Sellam et al., 2013). First, we tested the effect of each contextual factor on the level of pain scores separately in simple linear mixed effect models. The simple linear mixed effect models included the contextual factor and considered the three phases and five measurement points as fixed effects; neonates and raters were random intercepts. Raters were not included in our analyses of the physiological subscale, since we captured heart rate from the patient's monitor during the video recording. We determined the effect of each contextual factor on pain scores with

likelihood ratio tests (e.g., Winter, 2013). We included all contextual factors into the second variable selection stage that reached a *p*-value below 0.20. Because we considered a high number of contextual factors, this method was useful in the first round of excluding irrelevant predictors.

At the end of this first stage, we used Pearson product moment correlations for continuous variables and Spearman's Rho for ordinal variables to screen individual contextual factors for multicollinearity. If we identified two or more highly correlated variables ($r \geq 0.90$), the researchers designated one of those variables as eligible for the second selection stage. We excluded contextual factors in this second selection stage if they occurred in less than 1% of our study sample. Since data was collected in three different NICUs, we also tested the effect of the study centre on the pain scores. If this effect was smaller than $p < 0.20$, we included study centre as an organisational contextual factor in the multiple linear mixed effect model to control this possible influence.

In the second stage, we included all the contextual factors we selected in the first stage and ran large multiple linear mixed effect models. As in the single linear mixed model analyses, we included phases and measurement points as fixed effects and neonates and raters as random intercepts. For all three outcome variables (modified BPSN total score, behavioural and physiological BPSN sub-scores), we used likelihood ratio tests to see if the multiple models that included a random slope for neonate and measurement point fit the data better than the same model without a random slope (Twisk, 2006). Adding a random slope for neonate and measurement point allows the model to return different slopes for different neonates. After we decided on the specification for the random effect, we used the R-package *LmerTest* (Kuznetsova et al., 2017) to perform backward elimination of non-significant contextual factors. This package uses Satterthwaite's method for approximating degrees of freedom for the F-Test and a fixed alpha of 0.05. We also analysed the remaining contextual

factors for multicollinearity by calculating variance inflation factors (VIFs) for each predictor with the R-package *car* (Fox and Weisberg, 2011). Multicollinearity may exist when VIF values ≥ 10 (Myers, 1990; O'Brien, 2007). In the last model, we considered a level of $p < 0.01$ to be statistically significant. We performed statistical analyses on each of the five imputed data sets separately, and then pooled the results of the five analyses.

3. Results

3.1 Characteristics of the sample

Between January 2016 and January 2017, we assessed a total of 1'673 consecutive neonates for eligibility and finally enrolled 162. A flow chart with reasons for exclusion appears in our previous study (Schenk et al., 2019). Eight neonates were excluded from data analysis for missing or poor-quality video sequences. Mean GA at birth of the total sample was 30.85 (± 4.5) weeks, ranging from 24.29 to 41.57 weeks. Characteristics of the sample are summarized in Table 1.

<< INSERT TABLE 1 – CHARACTERISTICS OF THE TOTAL SAMPLE AND THE FOUR GESTATIONAL AGE GROUPS – HERE>>

3.2 Individual contextual factors associated with the modified BPSN scores

We present the results of the single linear mixed model analyses we conducted for each individual contextual factor with the modified behavioural and physiological BPSN sub scores and with the modified BPSN total score as outcome variables, in AppendixS1. There were high correlations between GA at birth and birth weight ($r = 0.901$), between birth weight and PMA at heel stick ($r \geq 0.875$), between GA at birth and weight at heel stick ($r \geq 0.903$), between PMA and weight at heel stick ($r \geq 0.899$), and between GA at birth and PMA at heel

stick ($r \geq 0.990$), so we decided to include only the variable PMA in our multiple mixed model analyses. We did not include individual medications in our multiple mixed model analyses if less than 1% of our study sample had taken them.

3.2.1 Behavioural sub score

Based on the results of the single mixed model analyses, where the modified behavioural BPSN sub score was the outcome, we included all individual contextual factors in the multiple linear mixed model that had a p-value < 0.20 (see AppendixS1). Study centre ($p = 0.163$) was included in this model. A model that included a random slope for neonate and measurement point fit the data better than a model without random slope ($p < 0.001$), so we included a random slope for infant and measurement point into the multiple linear mixed model. After backward elimination, the factors premature or term birth, CRIB, PNA, caffeine, indomethacin, time since the last painful procedure, and baseline behavioural state were included in the multiple mixed model analysis (Table 2). The behavioural pain scores of preterm neonates were 0.72 points lower than the scores of full-term neonates ($p < 0.001$). The behavioural pain scores of neonates who received caffeine were 0.30 points lower than the scores of neonates who did not receive caffeine ($p < 0.001$). Neonates who were in an active and awake state before the heel stick procedure always had the highest behavioural pain scores; their pain scores were 0.28 points higher than those of neonates in a quiet and awake state ($p < 0.001$), 0.16 higher than neonates in an active and asleep state ($p = 0.006$), and 0.50 points higher than neonates in a quiet and asleep state ($p < 0.001$). CRIB ($p = 0.024$), PNA ($p = 0.013$), indomethacin ($p = 0.053$) and time since the last painful intervention ($p \geq 0.045$) were not significantly associated with the modified behavioural sub-score. None of the contextual factors had a VIF ≥ 2.7.

3.2.2 Physiological sub score

Based on the results of our single mixed model analyses, where the modified physiological BPSN sub score was the outcome, we included all individual contextual factors in the multiple linear mixed model that had a p-value < 0.20 (see AppendixS1) and considered study centre ($p = 0.007$). For the physiological pain scores, a model that included a random slope for neonate and measurement point was not a better fit than a model without random slope ($p = 0.752$). After backward elimination, we were left with the factors premature or term birth, apnoea bradycardia syndrome, ventilation status, duration of the heel stick procedure, and administration of additional sucrose for the multiple mixed model analysis (Table 2). The physiological pain scores of preterm neonates were 0.23 points lower than those of full-term neonates ($p = 0.004$). The physiological pain scores of neonates who received mechanical ventilation were 0.20 lower than those of neonates who had not been ventilated ($p = 0.002$). The length of the heel stick procedure and physiological pain scores were positively associated; pain scores increased 0.02 points ($p < 0.001$) with each additional minute. Apnoea bradycardia syndrome ($p = 0.050$), CPAP ($p = 0.062$), and additional sucrose ($p = 0.014$) were not significantly associated with the modified physiological sub-score. None of the contextual factors reached a VIF ≥ 1.4.

3.2.3 Total score

Based on the results of our single mixed model analyses, for which the modified BPSN total score was the outcome, we included in the multiple linear mixed model all individual contextual factors that had a p-value < 0.20 (see AppendixS1) and study centre ($p = 0.015$). A model that included a random slope for neonate and measurement point fit the data better than a model without random slope ($p < 0.001$), so we included the random slope in the model.

After backward elimination, we included the factors premature or term birth, PNA, caffeine, time since the last painful intervention, baseline behavioural state, and duration of the heel stick procedure in the multiple mixed model analysis (Table 2). Overall scores for preterm neonates were 0.91 points lower than the scores of full-term neonates ($p < 0.001$). The pain scores of neonates who received caffeine were 0.40 points lower than the pain scores of neonates who had no caffeine ($p < 0.001$). The pain scores of neonates in an active and awake state were 0.27 points higher than the scores of neonates in a quiet and awake state ($p < 0.001$) and 0.46 points higher than the scores of neonates in a quiet and asleep state ($p < 0.001$). Every minute added to the heel stick procedure increased pain scores 0.06 points ($p < 0.001$). PNA ($p = 0.029$) and time since the last painful intervention ($p \geq 0.055$) were not significantly associated with the modified BPSN total scale. None of the contextual factors reached a VIF ≥ 1.8.

<< INSERT TABLE 2 – THE INDIVIDUAL CONTEXTUAL FACTORS THAT INFLUENCE PAIN SCORES – RESULTS OF THE MULTIPLE LINEAR MIXED MODEL ANALYSES AFTER BACKWARD ELIMINATION – HERE>>

4. Discussion

Our results indicate that individual contextual factors account for variability in neonates' pain response. Behavioural pain scores were mainly influenced by premature or term birth, caffeine, and baseline behavioural state. Physiological pain scores were mainly influenced by premature or term birth, ventilation status and the duration of the heel stick procedure.

4.1 Postmenstrual and postnatal age

Preterm neonates had lower behavioural and physiological pain scores than full-term neonates. Our single mixed model analysis revealed a positive association between PMA and behavioural and physiological pain scores confirm the results of earlier studies. Because preterm neonates with younger GA have a more immature nervous system and less muscular strength, posture, and body movements, their behavioural pain responses are less obvious and more inconsistent than those of more mature preterm or full-term neonates (Craig et al., 1993; Gibbins and Stevens, 2003; Morison et al., 2003; Ahn, 2006b; Ranger et al., 2007; Gibbins et al., 2008a; Gibbins et al., 2008b; Johnston et al., 2011). Heart rate may be more stable in preterm infants because they are in a state of constant autonomic arousal caused by repeated painful and stressful procedures they often experience during their NICU stay (Grunau et al., 2001). Gibbins et al. (2008b) did not find the increase in heart rate differed between different GA-groups, but extremely and very preterm neonates had a higher heart rate before and during the heel stick procedures than older preterm and full-term neonates. In our study, the categorization of the contextual factor primary diagnosis in premature or term birth may have reduced the effect of PMA in the multiple mixed model.

Postnatal age tended to be positively associated with behavioural pain scores. As preterm neonates develop and mature with increasing PNA, neonates may express a more robust and sustained behavioural pain response with increasing PNA (Craig et al., 1993; Johnston et al., 1996).

4.2 Baseline behavioural state

Neonates in a quiet (asleep or awake) behavioural state before the heel stick procedure had lower behavioural pain scores than neonates in an active (asleep or awake) behavioural state. Several studies indicated that behavioural state before a procedure influences preterm or

full-term neonates' pain response (Grunau and Craig, 1987; Stevens and Johnston, 1994; Stevens et al., 1994; Johnston et al., 1999; Ahn, 2006a; Ranger et al., 2007; Badr et al., 2010). Neonates in a more active and awake state have higher pain responses than neonates in a quiet sleeping state. Since preterm neonates spend up to 70% of their time in a sleep state (active sleep is the main behavioural state), their response to a painful procedure may often be dampened (Foreman et al., 2008; Werth et al., 2017), which is why it is appropriate to consider a neonate's behavioural state when assessing acute pain.

4.3 Medical contextual factors

Neonates treated with mechanical ventilation had lower physiological pain scores than neonates who were not ventilated during the recorded heel stick. Neonates treated with CPAP tended to have lower pain scores than neonates who were not ventilated. These findings align with the results of a study that analysed the effect of different contextual factors on pain responses in neonates who were treated with non-pharmacological, pain-relieving interventions (Sellam et al., 2013). Other studies found no association between ventilation status and behavioural or physiological pain responses (Johnston et al., 1996; Grunau et al., 2001; Grunau et al., 2006). When we examined the effect of ventilation in a single mixed model we identified a significant association between ventilation status and behavioural pain scores, which accords with the findings of Williams et al. (2009). Neonates who received CPAP or mechanical ventilation had lower behavioural pain scores than neonates who did not receive any ventilation. Ventilation status may lower behavioural pain scores because CPAP or tapes that fix the tube hide the neonate's face and make it harder to assess their pain level. Mechanically ventilated neonates may show dampened physiological and behavioural pain responses due to administered medications.

Neonates who received caffeine before or during the recorded heel stick procedure had lower behavioural pain scores than neonates who did not receive caffeine. Caffeine prevents preterm neonates from apnoea of prematurity (Schmidt et al., 2006), so caffeine administration is associated with a neonates' need for ventilation and PMA. A study that examined the association between caffeine and PIPP scores found these two variables were not correlated (Johnston et al., 1999).

We found no association between the number of earlier painful procedures and a neonate's behavioural or physiological pain scores. There was a tendency that neonates who had had experienced a painful procedure immediately before the heel stick tended to have lower behavioural pain scores than neonates who had not had a painful intervention within an hour of the heel stick. This aligns with Johnston et al. (1999)'s findings, which suggested that neonates who were younger, sleeping, and had had a recent painful procedure were less likely to show behavioural and physiological pain responses.

4.1 Bernese Pain Scale for Neonates-Revised

We revised the BPSN based on these results and the results of our psychometric tests of the BPSN (Schenk et al., 2019). The BPSN-Revised (AppendixS2) includes three behavioural items (crying, facial expression, and posture), one physiological item (heart rate), and three individual contextual factors (PMA, baseline behavioural state, and ventilation status). The results of our first sub-study showed the importance of considering a neonate's GA when assessing their pain. The cut-off that discriminates between pain and non-pain should be lowered in proportion to neonate's GA (Schenk et al., 2019).

4.2 Limitations and strengths

This study had several limitations. First, the blinded raters may have had trouble assessing pain response in lower quality videos (e.g., poor lighting conditions) even though we tried to exclude videos of very poor quality. Second, the way individual contextual factors are operationalized may vary across studies (e.g., health status or kind of painful procedures) and this heterogeneity needs to be considered if researchers want to compare our study results to others. Third, individual contextual factors are often correlated strongly (e.g., younger and sicker neonates endure greater number of painful procedures), which can affect study results. Fourth, the BPSN was validated and revised to assess acute procedural pain, but the BPSN-Revised needs to be validated for different painful and stressful procedures and for different types of pain (Anand, 2017) to increase its validity and applicability in the clinical settings. It may not be appropriate to consider baseline behavioural state when assessing pain that has no clear beginning and end (e.g., prolonged or chronic pain). Future studies should also test the feasibility and clinical utility of the BPSN-Revised.

This study also has several strengths. Our study sample included full-term and preterm neonates, covering a wide range of GAs, when other studies mainly focused on preterm or full-term infants. We repeatedly measured neonates' pain responses across the first 14 days of life, improving on the cross-sectional design of existing studies that evaluated neonates' pain responses (Williams et al., 2009). Since cohort or history effects, rather than systematic individual changes, could account for differences in pain responses between different GA-groups, longitudinal data are more suitable for investigating change over time (Singer and Willett, 2003). Finally, we analysed the influence of contextual factors on physiological and behavioural pain scores separately.

5. Conclusions

Individual contextual factors like premature birth, quiet behavioural state and mechanical ventilation dampen neonates' response to acute procedural pain. The presented BPSN-Revised accounts for these relevant contextual factors. However, while the behavioural and physiological items of the BPSN-Revised were subject to rigorous testing and proved to have good psychometric properties (Schenk et al., 2019), the adding of the three contextual factors based on the results of this sub-study demands further testing of the validity, feasibility and clinical utility of the BPSN-Revised.

Acknowledgements

We are grateful to all parents who gave us permission to include their infant in the study. We thank the Swiss National Science Foundation for their financial support. We also thank the nurses, head nurses, and medical directors of the University Hospital NICUs in Basel, Bern, and Zurich, and everyone else who supported us and helped to make this study possible. We thank Dr. Kali Tal for her editorial work.

Author Contributions

EC, the primary investigator, conceived of and designed the study. DB, SS, MN, and BS co-authored the study proposal and supported EC in designing the study. DB, SS, and MN provided access to the research field. KS, the doctoral student, handled all tasks related to data collection and entry, management, and analysis. She was also responsible for reporting and disseminating outcomes in peer-reviewed journals and at conferences. LS recruited at the University Hospital NICU in Bern and supported the doctoral student in the data collection process. RB supported the doctoral student in the data analyses. All authors discussed the results and commented on the manuscript.

References

AAP, Committee on Fetus and Newborn, & Section on Anesthesiology and Pain Medicine (2016). Prevention and Management of Procedural Pain in the Neonate: An Update. *Pediatrics, 137*, e20154271.

Ahn, Y. (2006a). The relationship between behavioral states and pain responses to various NICU procedures in premature infants. *Journal of Tropical Pediatrics, 52*, 201-205.

Ahn, Y., & Jun, Y. (2006b). Measurement of pain-like response to various NICU stimulants for high-risk infants. *Early Human Development, 83*, 255-262.

Anand, K. J. S. (2017). Defining pain in newborns: need for a uniform taxonomy? *Acta Paediatr, 106*, 1438-1444.

Anand, K. J. S., Eriksson, M., Boyle, E. M., Avila-Alvarez, A., Andersen, R. D., Sarafidis, K., et al. (2017). Assessment of continuous pain in newborns admitted to NICUs in 18 European countries. *Acta Paediatr, 106*, 1248-1259.

Badr, L.K., Abdallah, B., Hawari, M., Sidani, S., Kassar, M., Nakad, P., et al. (2010). Determinants of Preamture Infant Pain Responses to Heel Sticks. *Pediatric Nursing, 36*, 129-136.

Bates, D., Machler, M., Bolker, B.M., & Walker, S.C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48.

Boettcher, M., Göttler, S., Stoffel, L., Schwab, K., Berger, S., & Mérat, M. (2012). Schmerzmanagement bei Kindern in der Schweiz. *Monatsschrift Kinderheilkunde, 160*, 887-894.

Brummelte, S., Grunau, R. E., Chau, V., Poskitt, K. J., Brant, R., Vinall, J, et al. (2012). Procedural pain and brain development in premature newborns. *Ann Neurol, 71*, 385-396.

Bührer, C., Grimmer, I., Metze, B., & Obladen, M. (2000). The CRIB (Clinical Risk Index for Babies) score and neurodevelopmental impairment at one year corrected age in very low birth weight infants. *Intensive Care Medicine, 26*, 325-329.

Cignacco, E., Hamers, J., Stoffel, L., Van Lingen, R. A., Schütz, N., Müller, R., et al. (2008). Routine procedures in NICUs: factors influencing pain assessment and ranking by pain intensity. *Swiss Med Weekly, 138*, 484-491.

Cignacco, E., Mueller, R., Hamers, J. P. H., & Gessler, P. (2004). Pain assessment in the neonates using the Bernese Pain Scale for Neonates. *Early Human Development, 78*, 125-131.

Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., et al. (2017). Individual contextual factors in the validation of the Bernese pain scale for neonates: protocol for a prospective observational study. *BMC Pediatr, 17*(171). DOI 10.1186/s12887-017-0914-9

Craig, K. D., Whitfield, M. F., Grunau, R. V. E., Linton, J., & Hadjistavropoulos, D. (1993). Pain in the preterm neonate: behavioural and physiological indices. *Pain, 52*, 287-299.

Foreman, S. W., Thomas, K. A., & Blackburn, S. T. (2008). Individual and gender differences matter in preterm infant state development. *J Obstet Gynecol Neonatal Nurs, 37*, 657-665.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression*, Second Edition. Thousand Oakds CA: Sage.

Gibbins, S., & Stevens, B. (2003). The influence of gestational age on the efficacy and short-term safety of sucrose for procedural pain relief. *Advances in Neonatal Care, 3*, 241-249.

Gibbins, S., Stevens, B., Beyene, J., Chan, P. C., Bagg, M., & Asztalos, E. (2008a). Pain behaviours in Extremely Low Gestational Age infants. *Early Human Development, 84*, 451-458.

Gibbins, S., Stevens, B., McGrath, P. J., Yamada, J., Beyene, J., Breau, L., et al. (2008b). Comparison of Pain Responses in Infants of Different Gestational Ages. *Neonatology, 93*, 10-18.

Gibbins, S., Stevens, B. J., Yamada, J., Dionne, K., Campbell-Yeo, M., Lee, G., et al. (2014). Validation of the Premature Infant Pain Profile-Revised (PIPP-R). *Early Human Development, 90*, 189-193.

Grunau, R. E., Holsti, L., & Peters, J. W. B. (2006). Long-term consequences of pain in human neonates. *Seminars in Fetal & Neonatal Medicine, 11*, 268-275.

Grunau, R. E., Oberlander, T. F., Whitfield, M. F., Fitzgerald, C., & Lee, S. K. (2001). Demographic and Therapeutic Determinants of Pain Reactivity in Very Low Birth Weight Neonates at 32 Weeks' Postconceptional Age. *Pediatrics, 107*, 105-112.

Grunau, R. V. E., & Craig, K. D. (1987). Pain expression in neonates: Facial action and cry. *Pain, 28*, 395-410.

Hatfield, L. A., & Ely, E. A. (2015). Measurement of Acute Pain in Infants: A Review of Behavioral and Physiological Variables. *Biological Research for Nursing, 17*, 100-111.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics, 15*, 651-674.

Hothorn, T., & Zeileis, A. (2014). partykit: A Modular Toolkit for Recursive Partytioning in R. *Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics*, Universität Innsbruck, 2014;10.

Hummel, P. (2017). Psychometric Evaluation of the Neonatal Pain, Agitation, and Sedation (N-PASS) Scale in Infants and Children up to Age 36 Months. *Pediatr Nurs*, *43*, 175-184.

Hummel, P., Puchalski, M., Creech, S. D., & Weiss, M.G. (2008). Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. *J Perinatol*, *28*, 55-60.

Johnston, C. C., Fernandes, A. M., & Campbell-Yeo, M. (2011). Pain in neonates is different. *Pain*, *152*, S65-S73.

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

Johnston, C. C., Stevens, B. J., Franck, L. S., Jack, A., Stremler, R., & Platt, R. (1999). Factors Explaining Lack of Response to Heel Stick in Preterm Newborns. *JOGNN*, *28*, 587-594.

Johnston, C. C., Stevens, B. J., Yang, F., & Horton, L. (1995). Differential response to pain by very premature neonates. *Pain*, *61*, 471-479.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1-26.

Lee, G. Y., & Stevens, B. J. (2014). *Neonatal and infant pain assessment*. In P. McGrath, B. Stevens, S. M. Walker, & W. T. Zempsky (Eds.), Oxford Textbook of Paediatric Pain (pp. 353-369). Oxford: Oxford University Press.

Morison, S. J., Holsti, L., Grunau, R. E., Whitfield, M. F., Oberlander, T. F., Chan, H. W. P., et al. (2003). Are there developmentally distinct motor indicators of pain in preterm infants? *Early Human Development*, *72*, 131-146.

Myers R. (1990). *Classical and modern regression with applications*. Boston: Duxbury.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*, 673-690.

Ranger, M., Chau, C. M. Y., Garg, A., Woodward, T. S., Beg, M. F., Bjornson, B., et al. (2013). Neonatal pain-related stress predicts cortical thickness at age 7 years in children born very preterm. *PLoS ONE*, *8*, e76702.

Ranger, M., & Grunau, R. E. (2014). Early repetitive pain in preterm infants in relation to the developing brain. *Pain management*, *4*, 57-67.

Ranger, M., Johnston, C. C., & Anand, K. (2007). Current controversies regarding pain assessment in neonates. *Seminars in perinatology*, *31*, 283-288.

R Core Team (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org

Rubin, D. B. (2008). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., et al. (2019). The influence of gestational age in the psychometric testing of the Bernese Pain Scale for Neonatse. *BMC Pediatrics*, *19*(20). https://doi.org/10.1186/s12887-018-1380-8

Schmidt, B., Roberts, R. S., Davis, P., Doyle, L. W., Barrington, K. J., Ohlsson, A., Solimano, A., & Tin, W. (2006). Caffeine therapy for apnea of prematurity. *N Engl J Med*, *354*, 2112-2121.

Schneider, J., Duerden, E. G., Guo, T., Ng, K., Hagmann, P., Bickle Graz, M., et al. (2018). Procedural pain and oral glucose in preterm neonates: brain development and sex-specific effects. *Pain*, *159*, 515-525.

Sellam, G., Cignacco, E. L., Craig, K. D., & Engberg, S. (2011). Contextual factors influencing pain response to heelstick procedures in preterm infants: What do we know? A systematic review. *European Journal of Pain*, *15*, 661.e1-661.e15.

Sellam, G., Engberg, S., Denhaerynck, K., Craig, K. D., & Cignacco, E. L. (2013). Contextual factors associated with pain response of preterm infants to heel-stick procedures. *European Journal of Pain*, *17*, 255-263.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford: University Press.

Stevens, B. J., Gibbins, S., Yamada, R. N., Dionne, K., Lee, G., Johnston, C., & Taddio, A. (2014). The Premature Infant Pain Profile-Revised (PIPP-R) Initial Validation and Feasibility. *Clinical Journal of Pain*, *30*, 238-243.

Stevens, B. J., & Johnston, C. C. (1994). Physiological responses of premature infants to a painful stimulus. *Nursing Research*, *43*, 226-231.

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

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health mesasurement scales, a practical guide to their development and use*. Oxford: University Press.

Twisk, J. W. R. (2006). *Applied Multilevel Analysis A Practical Guide*. New York: Cambridge University Press.

Valeri, B. O., & Linhares, M. B. M. (2012). Pain in preterm infants: Effects of sex, gestational age, and neonatal illness severity. *Psychology & Neuroscience*, *5*, 11-19.

Vinall, J., & Grunau, R. E. (2014). Impact of repeated procedural pain-related stress in infants born very preterm. *Pediatric Research*, *75*, 584-587.

Vinall, J., Miller, S. P., Bjornson, B. H., Fitzpatrick, K. P. V., Poskitt, K. J., Brant, R., et al. (2014). Invasive Procedures in Preterm Children: Brain and Cognitive Development at School Age. *Pediatrics*, *133*, 412-421.

Werth, J., Atallah, L., Andriessen, P., Long, X., Zwartkruis-Pelgrim, E., & Aarts, R. M. (2017). Unobtrusive sleep state measurements in preterm infants - A review. *Sleep Med Rev*, *32*, 109-122.

Williams, A. L., Khattak, A. Z., Garza, C. N., Lasky, R. E. (2009). The behavioral pain response to heelstick in preterm neonates studied longitudinally: Description, development, determinants, and components. *Early human development*, *85*, 369-374.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. University of California, Merced, Cognitive and Information Sciences. Retrieved from http://arxiv.org/pdf/1308.5499.pdf

Tables

Table 1 Characteristics of the total sample and the four gestational age groups

Table 2 The individual contextual factors that influence pain scores – results of the multiple linear mixed model analyses after backward elimination

Appendixes

AppendixS1 Results of the single linear mixed regression models of the modified behavioural and physiological BPSN sub scores and the modified BPSN total score.

AppendixS2 The Bernese Pain Scale for Neonates-Revised for acute procedural pain assessment

Table 1 Characteristics of the total sample and the four gestational age groups

Characteristic	Total Sample	Gestational age groups			
Mean ± SD or n (%)		Extremely preterm neonates	Very preterm neonates	Moderate to late preterm neonates	Full-term neonates
Sample	154 (100%)	50 (32.5%)	45 (29.2%)	38 (24.7%)	21 (13.6%)
At time of birth					
Male	87 (56.5%)	31 (62.0%)	23 (51.1%)	20 (52.6%)	13 (61.9%)
GA (weeks)	30.85 ± 4.5	26.23 ± 1.2	29.44 ± 1.0	34.21 ± 1.0	38.81 ± 1.3
Weight (g)	1630.1 ± 934.3	851.4 ± 196.4	1285.1 ± 328.2	2093.7 ± 377.5	3384.5 ± 811.6
Way of delivery:					
- Vaginal-spontan	36 (23.4%)	10 (20.0%)	4 (8.9%)	13 (34.2%)	9 (42.9%)
- Vaginal-operativ	4 (2.6%)	0 (0%)	1 (2.2%)	2 (5.3%)	1 (4.8%)
- Planned c-section	23 (14.9%)	3 (6.0%)	8 (17.8%)	7 (18.4%)	5 (23.8%)
- Emerg. c-section	91 (59.1%)	37 (74.0%)	32 (71.1%)	16 (42.1%)	6 (28.6%)
Number of birth:					
- Single	104 (67.5%)	43 (86.0%)	20 (44.4%)	21 (55.3%)	20 (95.2%)
- One of twins	44 (28.6%)	4 (8.0%)	22 (48.9%)	17 (44.7%)	1 (4.8%)
- One of triplets	6 (3.9%)	3 (6.0%)	3 (6.7%)	0 (0%)	0 (0%)
CRIB	3.8 ± 3.9	7.5 ± 3.7	3.2 ± 2.8	1.1 ± 1.7	0.9 ± 1.6
At HS 1					
PNA (days)	3.95 ± 2.0	4.80 ± 2.2	3.56 ± 1.9	3.18 ± 0.9	4.19 ± 2.6
PMA (weeks)	31.42 ± 4.5	26.91 ± 1.2	29.94 ± 1.0	34.67 ± 1.0	39.41 ± 1.4
Weight (g)	1573.8 ± 916.6	812.3 ± 183.1	1220.1 ± 322.0	2032.0 ± 354.1	3315.5 ± 755.1
CPAP	77 (50.0%)	40 (80.0%)	30 (66.7%)	7 (18.4%)	0 (0%)
Mech. ventilation	14 (9.1%)	10 (20.0%)	1 (2.2%)	2 (5.3%)	1 (4.8%)
Total comorbidities	5.7 ± 4.4	10.1 ± 4.2	5.4 ± 2.4	2.7 ± 1.4	1.4 ± 1.1
NEC	2 (1.3%)	1 (2%)	1 (2.2%)	0 (0%)	0 (0%)
RDS	113 (73.4%)	50 (100%)	41 (91.1%)	19 (50.0%)	3 (14.3%)
PDA	44 (28.6%)	40 (80.0%)	4 (8.9%)	0 (0%)	0 (0%)
ABS	91 (59.1%)	46 (92.0%)	32 (71.1%)	13 (34.2%)	0 (0%)
Neonatal infection	82 (53.2%)	35 (70.0%)	27 (60.0%)	13 (34.2%	7 (33.3%)
Sedatives	21 (13.6%)	11 (22.0%)	6 (13.3%)	2 (5.3%)	2 (9.5%)
Opioid	49 (31.8%)	31 (62.0%)	12 (26.7%)	4 (10.5%)	2 (9.5%)
Non-opioid	24 (15.6%)	21 (42.0%)	0 (0%)	1 (2.6%)	2 (9.5%)
Steroid	5 (3.2%)	1 (2.0%)	2 (4.4%)	2 (5.3%)	0 (0%)
Caffeine	80 (51.9%)	46 (92.0%)	31 (68.9%)	3 (7.9%)	0 (0%)
Antibiotic	95 (61.7%)	46 (92.0%)	29 (64.4%)	10 (26.3%)	10 (47.6%)
Catecholamine	8 (5.2%)	7 (14.0%)	1 (2.2%)	0 (0%)	0 (0%)
Medication for lung maturation during pregnancy	82 (53.2%)	38 (76.0%)	33 (73.3%)	11 (28.9%)	0 (0%)
Surfactant after birth	59 (38.3%)	38 (76.0%)	16 (35.6%)	3 (7.9%)	2 (9.5%)
Muscle relaxant	29 (18.8%)	19 (38.0%)	8 (17.8%)	1 (2.6%)	1 (4.8%)
Total number PI	43.49 ± 32.1	69.46 ± 34.0	40.29 ± 22.5	22.32 ± 16.5	26.86 ± 24.0
PI past 24 hours	9.62 ± 5.6	12.96 ± 5.6	9.29 ± 4.2	6.76 ± 5.0	7.52 ± 5.4
Total number nPI	43.95 ± 25.2	57.86 ± 27.9	39.22 ± 23.7	30.55 ± 10.2	45.19 ± 25.1
nPI past 24 hours	12.61 ± 3.3	12.92 ± 2.6	12.07 ± 3.0	12.08 ± 2.9	14.00 ± 5.2
At HS 5					
PNA (days)	9.01 ± 3.0	10.04 ± 3.1	8.60 ± 2.7	7.61 ± 2.7	
PMA (weeks)	30.19 ± 3.1	27.54 ± 1.2	30.45 ± 0.9	34.99 ± 1.0	
Weight (g)	1289.2 ± 526.4	895.0 ± 204.2	1257.6 ± 247.3	2108.7 ± 291.9	
CPAP	52 (50.5%)	33 (73.3%)	17 (48.6%)	2 (8.7%)	
Mech. ventilation	10 (9.7%)	10 (22.2%)	0 (0%)	0 (0%)	

Note. SD, standard deviation; extremely preterm neonates, < 28 weeks of gestation; very preterm neonates, 28 0/7-31 6/7 weeks of gestation; moderate to late preterm neonates, 32 0/7-36 6/7 weeks of gestation; full-term neonates, ≥ 37 0/7 weeks of gestation; GA, gestational age; emerg. c-section, emergency cesarean-section; CRIB, Clinical Risk Index for Babies; HS, heel stick; PNA, postnatal age; PMA, postmenstrual age; Mech. ventilation, mechanical ventilation; CPAP, continuous positive airway pressure; NEC, necrotizing enterocolitis; RDS, respiratory distress syndrome; PDA, patent ductus arteriosus; ABS, apnoea bradycardia syndrome; PI, painful intervention; nPI, non-painful intervention.

Table 2 The individual contextual factors that influence pain scores – results of the multiple linear mixed model analyses after backward elimination

Outcome variable	Contextual factor	Fixed effect	SE	df	*t*-value	*p*-value
Modified behavioural BPSN sub score	Term birth[a]	0.721	0.199	172.916	3.620	0.000**
	CRIB	0.024	0.010	95.535	2.300	0.024
	PNA	0.033	0.013	186.919	2.503	0.013
	Caffeine	-0.302	0.084	212.593	-3.580	0.000**
	Indomethacin	-0.205	0.106	399.963	-1.939	0.053
	Last PI: 16-30 min	-0.029	0.197	426.866	-0.148	0.883
	Last PI: 31-45 min	-0.077	0.239	426.321	-0.322	0.748
	Last PI: 46-60 min	0.033	0.208	363.230	0.160	0.873
	Last PI: > 60 min	0.275	0.137	371.246	2.011	0.045
	Quiet & awake state	-0.283	0.058	8027.899	-4.848	0.000**
	Active & asleep state	-0.158	0.057	5980.223	-2.764	0.006*
	Quiet & asleep state	-0.498	0.061	4272.785	-8.226	0.000**
Modified physiological BPSN sub score	Term birth[a]	0.232	0.080	460.300	2.909	0.004*
	ABS	-0.083	0.042	144.280	-1.985	0.050
	CPAP	-0.071	0.038	228.620	-1.880	0.062
	Mechanical ventilation	-0.196	0.063	178.080	-3.123	0.002*
	Duration HS (sec)	0.0004	0.000	934.160	3.337	0.000**
	Additional sucrose	0.144	0.058	902.800	2.492	0.014
Modified BPSN total scale	Term birth[a]	0.907	0.234	184.280	3.870	0.000**
	PNA	0.036	0.017	193.740	2.206	0.029
	Caffeine	-0.402	0.093	205.280	-4.331	0.000**
	Last PI: 16-30 min	-0.007	0.235	350.731	-0.029	0.971
	Last PI: 31-45 min	-0.152	0.290	441.660	-0.526	0.599
	Last PI: 46-60 min	-0.054	0.251	390.400	-0.216	0.829
	Last PI: > 60 min	0.317	0.167	397.760	1.922	0.055
	Quiet & awake state	-0.274	0.069	8139.200	-3.956	0.000**
	Active & asleep state	-0.127	0.068	6192.200	-1.118	0.062
	Quiet & asleep state	-0.459	0.072	4482.200	-6.395	0.000**
	Duration HS (sec)	0.001	0.000	378.200	3.698	0.000**

Note. BPSN, Bernese Pain Scale for Neonates; SE, standard error; df, degree of freedom; CRIB, Clinical Risk Index for Babies; PNA, postnatal age; PI, painful intervention; ABS, apnoea bradycardia syndrome; CPAP, continuous positive airway pressure; HS, heel stick.

[a]=Primary diagnosis was categorized into premature or term birth.; *$p < 0.01$; **$p < 0.001$.

AppendixS1: Results of the single linear mixed regression models of the modified behavioural and physiological BPSN sub scores and the modified BPSN total score.

	BPSN behavioural subscale					BPSN physiological subscale					BPSN total scale				
Individual contextual factors	Fixed effect	SE	X2	df	*p*-value	Fixed effect	SE	X2	df	*p*-value	Fixed effect	SE	X2	df	*p*-value
Sex (female)	-0.015	0.11	0.02	1	0.894	0.013	0.04	0.13	1	0.726	-0.013	0.14	0.01	1	0.924
Nationality			1.54	2	0.464			0.15	2	0.926			0.87	2	0.648
- other than Swiss	-0.148	0.12				-0.015	0.04				-0.116	0.15			
- unknown	0.063	0.40				-0.022	0.13				0.217	0.49			
Weight birth (g)	0.230	0.06	15.02	1	0.000	0.087	0.02	17.94	1	0.000	0.341	0.07	22.76	1	0.000
GA birth (weeks)	0.063	0.01	27.65	1	0.000	0.022	0.00	27.79	1	0.000	0.091	0.01	40.90	1	0.000
Parity			1.72	2	0.423			1.70	2	0.430			1.32	2	0.516
- one of twins	-0.149	0.12				0.020	0.04				-0.142	0.15			
- one of triplets	0.097	0.28				0.114	0.09				0.180	0.35			
Birth method			0.35	3	0.951			3.79	3	0.289			0.61	3	0.894
- vaginal-operativ	-0.094	0.36				0.222	0.12				0.291	0.44			
- planned c-section	0.071	0.18				0.044	0.06				0.095	0.22			
- emerg. c-section	0.052	0.14				0.041	0.04				0.092	0.16			
Term birth[a]	0.927	0.15	33.81	1	0.000	0.178	0.06	8.95	1	0.003	1.244	0.18	41.93	1	0.000
CRIB Total	-0.023	0.01	2.68	1	0.102	-0.018	0.00	17.44	1	0.000	-0.045	0.02	7.03	1	0.008
Number of comorbidities	-0.042	0.01	11.63	1	0.001	-0.017	0.00	15.90	1	0.000	-0.061	0.02	17.07	1	0.000
NEC	0.003	0.27	0.00	1	0.992	-0.101	0.14	0.50	1	0.478	0.021	0.32	0.00	1	0.949
RDS	-0.412	0.12	11.03	1	0.001	-0.131	0.05	7.76	1	0.006	-0.575	0.15	14.60	1	0.000
PDA	-0.331	0.12	7.68	1	0.006	-0.150	0.04	13.72	1	0.000	-0.518	0.14	12.79	1	0.000
ABS	-0.406	0.10	16.81	1	0.000	-0.173	0.04	18.88	1	0.000	-0.486	0.12	16.04	1	0.000
Infect	0.099	0.12	0.75	1	0.387	-0.041	0.04	0.95	1	0.331	0.048	0.14	0.12	1	0.733
At time of HS															
PNA	0.030	0.01	7.15	1	0.008	0.000	0.01	0.01	1	0.935	0.041	0.01	9.18	1	0.002
PMA	0.068	0.01	32.37	1	0.000	0.028	0.01	33.71	1	0.000	0.098	0.01	47.44	1	0.000
Weight	0.233	0.06	14.95	1	0.000	0.124	0.02	25.24	1	0.000	0.363	0.07	25.24	1	0.000
Ventilation			41.09	2	0.000			28.93	2	0.000			62.77	2	0.000
- CPAP	-0.127	0.06				-0.156	0.04				-0.234	0.07			
- Mechanical ventilation	-0.687	0.11				-0.304	0.07				-1.020	0.13			
Duration of HS (sec)	0.000	0.00	10.90	1	0.001	0.001	0.00	23.24	1	0.000	0.001	0.00	36.85	1	0.000
Additional sucrose	0.016	0.06	0.08	1	0.795	0.150	0.05	8.45	1	0.004	0.135	0.07	3.46	1	0.071
Baseline behavioural state (PIPP-R)			333.18	3	0.000			2.63	3	0.455			183.41	3	0.000
- quiet and awake	-0.551	0.05				0.000	0.07				-0.525	0.07			
- active and asleep	-0.220	0.05				0.043	0.06				-0.172	0.06			
- quiet and asleep	-0.701	0.05				0.066	0.06				-0.641	0.06			
Medication (no)															
Sedatives	0.279	0.08	11.29	1	0.001	0.128	0.07	2.96	1	0.086	0.315	0.10	10.26	1	0.001
Opioids	0.247	0.06	15.31	1	0.000	0.121	0.05	5.46	1	0.020	0.303	0.08	16.35	1	0.000
Non-opioids	0.192	0.06	11.00	1	0.001	0.148	0.05	8.57	1	0.003	0.322	0.07	22.08	1	0.000
Steroids	0.339	0.16	4.32	1	0.039	0.083	0.16	0.28	1	0.594	0.325	0.19	2.84	1	0.093
Caffeine	0.263	0.06	17.82	1	0.000	0.146	0.04	15.43	1	0.000	0.379	0.07	25.98	1	0.000
Antibiotics	0.127	0.05	6.55	1	0.011	0.028	0.04	0.57	1	0.458	0.164	0.06	7.76	1	0.005
Catecholamines	0.113	0.13	0.71	1	0.399	0.143	0.12	1.35	1	0.246	0.166	0.16	1.09	1	0.298
Propofol	0.367	0.09	17.01	1	0.000	0.107	0.08	1.61	1	0.205	0.400	0.11	14.47	1	0.000
Midazolam	0.045	0.20	0.05	1	0.822	0.186	0.15	1.49	1	0.223	0.060	0.24	0.07	1	0.802

Chloralhydrate	-0.420	0.29	2.16	1	0.146	0.229	0.27	0.72	1	0.397	-0.338	0.34	1.00	1	0.322
Morphine	0.312	0.11	7.79	1	0.005	0.159	0.07	5.04	1	0.025	0.588	0.13	19.50	1	0.000
Fentanyl	0.021	0.08	0.07	1	0.789	0.074	0.08	0.94	1	0.332	0.017	0.09	0.04	1	0.853
Pethidin	0.305	0.40	0.63	1	0.464	0.337	0.38	0.78	1	0.378	0.778	0.48	2.70	1	0.110
Remifentanyl	0.365	0.09	16.24	1	0.000	0.084	0.09	0.93	1	0.336	0.376	0.11	12.30	1	0.000
Paracetamol	0.139	0.10	1.99	1	0.160	0.011	0.09	0.01	1	0.909	0.114	0.12	0.96	1	0.329
Tylenol	0.645	0.16	15.97	1	0.000	0.128	0.16	0.67	1	0.412	0.710	0.19	13.86	1	0.000
Becetamol	0.325	0.15	4.78	1	0.030	0.151	0.14	1.09	1	0.296	0.419	0.18	5.70	1	0.017
Indomethacin	0.307	0.06	26.84	1	0.000	0.174	0.05	11.21	1	0.001	0.467	0.07	44.48	1	0.000
Ibuprofen	0.645	0.16	15.97	1	0.000	0.128	0.16	0.67	1	0.412	0.710	0.19	13.86	1	0.000
Lung maturation during pregnancy	0.360	0.07	27.80	1	0.000	0.159	0.06	7.07	1	0.008	0.442	0.08	29.84	1	0.000
Surfactant	0.131	0.07	4.02	1	0.046	0.141	0.06	5.35	1	0.021	0.151	0.08	3.82	1	0.051
Muscle relaxation	0.005	0.08	0.01	1	0.951	0.067	0.07	0.83	1	0.364	-0.010	0.09	0.02	1	0.914
Painful and non-painful interventions															
PI total	0.001	0.00	0.65	1	0.421	-0.001	0.00	4.91	1	0.028	0.000	0.00	0.01	1	0.933
PI past 24 hours	-0.011	0.00	6.49	1	0.011	-0.009	0.00	9.18	1	0.003	-0.015	0.01	7.80	1	0.005
Type of last PI			18.98	10	0.041			17.53	11	0.096			24.63	10	0.006
- CPAP prongs insertion/removal	0.060	0.05				-0.089	0.04				0.054	0.06			
- Heel stick	0.034	0.07				0.048	0.06				0.122	0.08			
- Tape removal	0.007	0.11				0.155	0.11				0.176	0.13			
- Endotracheal suctioning	0.223	0.12				-0.125	0.08				0.252	0.14			
- Insertion nasogastric tube	-0.333	0.12				-0.083	0.31				-0.359	0.14			
- Naso-pharyngeal suctioning	-0.043	0.10				-0.100	0.09				-0.028	0.12			
- Removal thoracal drain	-0.468	0.45				-0.271	0.39				-0.482	0.54			
- Extubation	-0.004	0.22				-0.262	0.22				-0.178	0.26			
- Insertion intravenous canula	-0.252	0.19				-0.299	0.20				-0.546	0.23			
- Venipuncture	-0.093	0.38				-0.122	0.39				-0.059	0.45			
Time since last PI			54.90	4	0.000			3.84	4	0.429			50.14	4	0.000
- 16-30 minutes	-0.058	0.11				-0.126	0.11				-0.215	0.13			
- 31-45 minutes	-0.294	0.13				-0.113	0.13				-0.441	0.16			
- 46-60 minutes	-0.271	0.12				-0.132	0.12				-0.356	0.14			
- > 60 minutes	0.206	0.08				-0.019	0.08				0.150	0.09			
nPI total	0.003	0.00	9.20	1	0.003	0.000	0.00	0.01	1	0.948	0.004	0.00	11.83	1	0.001
nPI past 24 hours	0.014	0.00	4.14	1	0.042	0.008	0.01	2.35	1	0.131	0.022	0.01	7.85	1	0.005
Type of last nPI			41.99	8	0.000			10.37	8	0.242			27.28	8	0.001
- Position	-0.014	0.04				-0.038	0.03				-0.059	0.04			
- Kangaroo	0.092	0.08				-0.002	0.08				0.086	0.09			
- Cranial ultrasound	-0.084	0.19				0.168	0.20				0.172	0.22			
- X-ray	0.326	0.40				-0.192	0.39				0.233	0.48			
- Removal nasogastric tub	0.631	0.19				0.121	0.20				0.795	0.23			
- Removal intravenous cannula	-0.136	0.27				0.661	0.27				0.377	0.32			
- Removal EKG/EEG stickers	0.527	0.38				-0.325	0.39				0.181	0.45			
- Other	0.569	0.11				-0.026	0.21				0.364	0.13			
Time since last nPI			28.50	4	0.000			1.19	4	0.878			41.43	4	0.000
- 16-30 minutes	-0.141	0.09				-0.046	0.08				-0.266	0.11			
- 31-45 minutes	-0.215	0.12				-0.069	0.12				-0.247	0.15			

- 46-60 minutes	0.018	0.09				-0.083	0.09				-0.044	0.11			
- > 60 minutes	0.148	0.05				-0.029	0.05				0.195	0.06			
Organizational contextual factor															
Study centre			3.63	2	0.163			10.05	2	0.007			8.39	2	0.015
- B	0.265	0.14				0.141	0.05				0.488	0.17			
- C	0.069	0.13				0.026	0.04				0.117	0.15			

Note. BPSN, Bernese Pain Scale for Neonates; SE, standard error; df, degree of freedom; GA, gestational age; emerg. c-section, emergency cesarean-section; CRIB, Clinical Risk Index for Babies; NEC, Necrotizing enterocolitis; RDS, Respiratory distress syndrome; PDA, Patent ductus arteriosus; ABS, Apnoea bradycardia syndrome; HS, heel stick; PNA, postnatal age; PMA, postmenstrual age; CPAP, continuous positive airway pressure; Mech. ventilation, mechanical ventilation; PI, painful intervention; nPI, non-painful intervention; PIPP-R, Premature Infant Pain Profile-Revised.

[a]=Primary diagnosis was categorized into premature or term birth.

AppendixS2: The Bernese Pain Scale for Neonates-Revised for acute procedural pain assessment

Pain indicators	0	1	2	3	Score
Crying	Not crying	Brief period of crying (less than 2 minutes)	Increased crying (more than 2 minutes)	Increased, shrill crying (more than 2 minutes)	
Facial expression	Face relaxed	Brief grimace	Increase grimace and trembling of chin	Permanent grimace of face and trembling of chin	
Posture	Body relaxed	Mainly relaxed, short bouts of tension	Frequent bouts of tension but relaxation possible	Permanently tense	
Heart rate (bpm) Baseline score: _____	Normal (Baseline)	Increase of 20 bpm or more over the baseline **with** return to baseline within 2 minutes	Increase of 20 bpm or more over baseline **without** return to baseline within 2 minutes	Increase of 30 bpm or more over baseline or more frequent episodes of bradycardia within 2 minutes	
				Subtotal →	
				Subtotal ≥ 1: Contextual factors need to be added!	
Contextual factors	**0**	**1**	**2**	**3**	**Score**
Postmenstrual age (GA + number of days since birth)	Full-term neonates (≥ 37 0/7 weeks)	Moderate to late preterm neonates (32 0/7 – 36 6/7 weeks)	Very preterm neonates (28 0/7 – 31 6/7 weeks)	Extremely preterm neonates (< 28 weeks)	
Behavioural state (baseline)	Active (awake or asleep)	Quiet (awake or asleep)			
Ventilation status	CPAP or no ventilation	Mechanical ventilation			
				Overall Total →	

Overall total = Subtotal of pain indicators + score of contextual factors (if subtotal of pain indicators ≥ 1).

0-4 points = **no pain** or no observable pain reaction
≥ 5 points = **pain**

Procedure:

1. Observation of the neonate during the **baseline phase** for 15 seconds: Assessment of the highest heart rate and behavioural state
2. Observation of the neonate **during the procedure** for **2 minutes**: Assessment of the three behavioural pain indicators (crying, facial expression, posture) and the highest heart rate.
3. Calculation of the sub-total based on the 4 pain indicators.
4. If **subtotal ≥ 1 point**, evaluation of the three contextual factors and calculation of the overall total. **Overall total = subtotal + contextual factors**

Appendix V: Curriculum Vitae

Name:	Karin Schenk
Date of Birth:	December 22th, 1980
Nationality:	Swiss
Marital Status:	Single, no children
Business Address:	Bern University of Applied Sciences Department of Health Professions Murtenstrasse 10 CH – 3008 Bern T +41 31 848 47 25 karin.schenk@bfh.ch

Education

Since October 2015	**PhD**, University of Basel, Faculty of Psychology PhD student in the SNF project "Validation of the Bernese Pain Scale for Neonates with Consideration for Individual Contextual Factors" (SNF 320030_159573)
2011 – 2015	**Master of Science in Psychology**, University of Bern Master thesis "Zur Wirkung von Mindfulness-Based Stress Reduction (MBSR) in einem nicht-klinischen Kontext"
2008 – 2011	**Bachelor of Science in Psychology**, University of Bern Bachelor thesis "Menschen ohne Rückenschmerzen – was charakterisiert sie?"
2008 – 2011	**Bachelor Minor in General Ecology,** University of Bern Project thesis "Erreichung der Aichi-Ziele auf kantonaler Ebene"
2005 – 2006	**General qualification for university entrance** "Passerellenkurs der TSME", Frauenfeld
2004 – 2005	**Vocational school** Gesundheitliche und soziale Berufsmaturitätsschule, St. Gallen
1996 – 1999	**Diploma of retail assistant (apprenticeship)** Berufsschule des Detailhandels, Bern & St. Gallen

Employment History

Since October 2017	**Research assistant** Project "Perinatal mental health care in Switzerland: Unraveling the perspective of affected women and health professionals (MADRE)" Bern University of Applied Sciences
Since January 2017	**Research assistant** Coordination and organization of the 12th International Symposium on Pediatric Pain (ISPP) in Basel, 2019 Bern University of Applied Sciences
10.2015 – 12.2018	**PhD student** SNF project "Validation of the Bernese Pain Scale for Neonates with Consideration for Individual Contextual Factors" Bern University of Applied Sciences, Department of Health Professions
07.2015 – 09.2015	**Trainee in clinical psychology** Privatklinik Aadorf
05.2015 – 06.2015	**Research assistant** Bern University of Applied Sciences, Department of Health Professions
07.2013 – 06.2015	**Clerk, pension fund** PANVICA, Münchenbuchsee
07.2013 – 10.2013	**Shop assistant** Eiselin Sport, Bern
01.2013 – 05.2013	**Clerk, old age and survivors insurance** PANVICA, Müchenbuchsee
12.2012 – 12.2012	**Shop assistant** Thalia Bücher AG, Bern
07.2012 – 11.2012	**Research internship in psychosomatic medicine** University hospital of Bern
12.2011 – 07.2012	**Research internship in health psychology** University of Bern
11.2011 – 07.2012	**Clerk, engineering office** hpb consulting ag, Bern
06.2011 – 09. 2011	**Trainee in the field of environment/sustainability** Praktikum Gruner AG, Ingenieure und Planer, Ittigen

02.2010 – 05.2011	**Therapy assistant, working with a child with autism spectrum disorder** Proinfirmis, Bern
10.2009 – 04.2011	**Shop assistant** Globus AG, Bern
07.2009 – 08.2009	**Kitchen help** Hotel Gasterntal-Selden, Kandersteg
02.2007 – 08.2008	**Clerk, internal sales office** Ditzler AG, Möhlin
10.2006 – 01.2007	**Shop assistant** Globus AG, Basel
08.2004 – 09.2004	**Temporary job, photographic laboratory** Photocolor Kreuzlingen AG
12.2003 – 04.2004	**Shop assistant** Koller Mode-Sport AG, Arosa
05.2003 – 11.2003	**Temporary job, photographic laboratory** Photocolor Kreuzlingen AG
12.2002 – 04.2003	**Shop assistant** Koller Mode-Sport AG, Arosa
05.2002 – 11.2002	**Temporary job, photographic laboratory** Photocolor Kreuzlingen AG
12.2001 – 04.2002	**Shop assistant** Koller Mode-Sport AG, Arosa
05.2001 – 11.2001	**Trainee in a home for the handicapped** Heim für Seh- und Mehrfachbehinderte, Homburg
12.2000 – 04.2001	**Temporary job, restaurant** Bergrestaurant Alp Lavoz, Lenzerheide
08.1999 – 08.2000	**Shop assistant** Charles Vögele Mode AG, St. Gallen
08.1996 – 07.1999	**Apprenticeship as a retail assistant** Charles Vögele Mode AG, Bern/St. Gallen

Supervision of students

Since 2019	Supervision of two master's thesis
Since 2016	Supervision of trainees

Membership in Scientific Societies & Congress Organization Committees

Since 2019	Member of the International Association for the Study of Pain, Special Interest Group on Pain in Childhood
Since 2017	Member of the Local Organizing Committee for the International Symposium on Pediatric Pain 2019 in Basel
Since 2016	Member of the Canadian trainee program "Pain In Child Health (PICH)", a strategic training initiative in health research of the Canadian institutes of health research.

Awards

March 2018	PICH2GO Award (travel grant) for attending PICH2GO Copenhagen, 25-27 March 2018.
November 2016	PICH2GO Award (travel grant) for attending PICH2GO Toronto, 9-10 November 2016.

Publications and presentations

Publications in peer-reviewed journals

Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., Nelle, M., & Cignacco, E. The Bernese Pain Scale for Neonates-Revised accounts for individual contextual factors. (*Manuscript submitted for publication, 20 March 2019*).

Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., Nelle, M., & Cignacco, E. (2019). The influence of gestational age in the psychometric testing of the Bernese Pain Scale for Neonates. *BMC Pediatrics, 19*:20.

Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. (2017). Individual contextual factors in the validation of the Bernese pain scale for neonates: protocol for a prospective observational study. *BMC Pediatrics, 17*:171.

Publications in non peer-reviewed journals

Cignacco, E., Schenk, K., & Stoffel, L. (2018). Schmerzen bei Neugeborenen: die Geschichte einer Vernachlässigung. Hebamme.ch, 7/8, 15-18.

Schenk, K., & Cignacco, E. (2016). Schmerzzustände von Neugeborenen zuverlässiger beurteilen. *Hebamme.ch*, 7/8, 11-13.

Presentations

Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., & Nelle, M. Gestational age in the validation of the Bernese Pain Scale for Neonates. *12th International Symposium on Pediatric Pain (ISPP) 2019*, Basel, 16-20 June 2019. (*Accepted abstract for presentation*)

Schenk, K., Stoffel, L., Bürgin, R., Stevens, B., Bassler, D., Schulzke, S., Nelle, M., & Cignacco, E. Psychometric testing of the Bernese Pain Scale for Neonates considering the full range of gestational ages. *PICH2GO Copenhagen*, 25-26 March 2018.

Schenk, K., Cignacco, E., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. (2017). Individuelle Kontextfaktoren in der Validierung des Berner Schmerzscores für Neugeborene: Ergebnisse der Validierungsstudie. *28. Deutscher Kongress für Perinatale Medizin*, Berlin, 30 November to 2 December 2017.

Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. Validation of the Bernese Pain Scale for Neonates with consideration for Individual Contextual Factors. Short oral presentation at the Forschertag der Kinderkliniken Bern, 8 November 2017.

Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. Individual contextual factors in the validation of the Bernese Pain Scale for Neonates: preliminary results of the psychometric testing. *11th International Symposium on Pediatric Pain (ISPP) 2017*, Kuala Lumpur, 6-9 July 2017.

Cignacco, E., Schenk, K., Stevens, B., Stoffel, L., Bassler, D., Schulzke, S., & Nelle, M. Validierung des Berner Schmerzscores für Neugeborene unter Berücksichtigung von individuellen Kontextfaktoren. *Swiss Congress for Health Professions (SCHP),* Lugano, 1-2 September 2016.

Personal Skills

Languages	German (native), English (very good), French (basic knowledge)
Technical Skills	Microsoft, Mac OS, SPSS, EndNote, secuTrial, R
2016 – 2018	Mediation teacher training, Lharampa Tenzin Kalden
2011 – 2016	Training in “Buddhistic philosophy and psychology” and “Lojong and Buddhistic philosophy”, Lharampa Tenzin Kalden