CHAPTER 2

Literature Review

The literature review is presented in six sections. The first section is an overview of the pain situation in preterm neonates. The second section is related to pain indicators and pain assessment for preterm neonates. The third section reviews the instrument development process. The fourth and fifth sections are the psychometric and clinical utility evaluation, respectively. The last section is the conceptual framework.

2.1 An overview of pain situation in preterm neonates
  2.1.1 Epidemiology of burden of pain in preterm neonates
  2.1.2 Adverse effects of pain in preterm neonates
  2.1.3 Pain pathways in preterm neonates
    1) Development of a preterm neonates' pain circuit
    2) Nociception and a preterm neonate's reactivity
2.2 Pain indicators and assessment for preterm neonates
  2.2.1 Pain indicators in preterm neonates
    1) Behavioral indicators of pain
    2) Physiological indicators of pain
    3) Factors affecting pain reactivity
  2.2.2 Existing pain assessment tools for preterm neonates
2.3 Instrument development process
2.4 Psychometric testing
  2.4.1 Validity
  2.4.2 Reliability
2.5 Clinical utility evaluation
2.6 Conceptual framework

2.1 An Overview of Pain Situation in Preterm Neonates

The knowledge on pain in preterm neonates is mainly based on two sources including evidences on rat pups and neonatal humans. First, the current knowledge regarding neonatal pain has relied on experimental research involving rodent models because of the ethical considerations for such a vulnerable preterm neonate (Eiland, 2012). Although precise correlations between the ages of animals and humans are limited in their validity, it is possible to use animal models as a means of studying changes occurring in humans. It is generally believed that the first 12 days of a rat's life corresponds to gestational weeks of 24 to 40 in humans (Eiland, 2012; Pattinson & Fitzgerald, 2004). Although these developmental parallels are not exact, rodent research has greatly increased the understanding of pain in preterm neonates, especially the study regarding adverse impacts of their neurodevelopment. Second, the human neonates' pain studies mostly involved observational studies comparing within or between groups of preterm infants during procedures, especially in a heel stick (Warnock & Lander, 2004). Most preterm infants must stay in NICU for a long period of time depending on their illness condition. In a recent study, for instance, the median hospital stay of preterm neonates less than 32 weeks of the gestational age was 58 days (Korvenranta et al., 2007). The care in NICU involves a high number of routine procedures which are associated with pain for diagnostic and therapeutic purpose in the preterm neonates, especially in ELGA and VLGA infants. The most frequent painful procedures performed in preterm neonates were heel stick, intravenous cannulation, tracheal or nasal aspiration, removal of intravenous lines, adhesive removal, and gastric tube insertion (Carbajal et al., 2008; Chen et al., 2012). Therefore, early and repeated exposures to painful procedures are common for neonates in intensive care, particularly for preterm neonates.

2.1.1 Epidemiology of burden of pain in preterm neonates

Painful procedures, such as heel sticks, insertion of intravenous or arterial lines, and tracheal or nasogastric tubes, are essential for monitoring and intensive care management. These invasive procedures produce measurable cortical pain responses in even the youngest preterm infants. The number of such procedures to which a neonate is exposed varies from study to study. The nature and number of invasive procedures was studied in 54 consecutive infants admitted to NICU in England (Barker & Rutter, 1995). Of the 3,283 procedures recorded, 74% of procedures were performed in infants below 31 weeks of gestation. Heel stick

blood sampling was the most common procedure (56%), followed by endotracheal suction (26%) and intravenous cannula insertion (8%). Heel stick is the most common painful procedures in preterm neonates in NICU.

Preterm neonates in NICU tend to have early and repeated exposures to painful procedures, and the younger preterm neonates are likely to have more procedures than the older group. One epidemiological study of procedural pain in neonates was designed as a prospective observational study to collect around-the-clock bedside data on all painful or stressful procedures performed in neonates admitted to 13 NICUs and pediatric intensive care units (PICUs) in the Paris Region (Carbajal et al., 2008). Preterm neonates (mean age = 33.00 weeks of gestation) experienced 60,969 first-attempt procedures, with 42,413 (69.60%) painful procedures. Of painful procedures, 79.20% were performed without specific analgesia. Each neonate experienced a mean of 16 procedures per day of hospitalization (range, 0 to 62). Another study in Switzerland (Cignacco et al., 2009) also found the similar mean number of painful procedures per preterm neonates. All 120 ventilated preterm neonates (mean age = 29.70 weeks of gestation) underwent a total of 38,626 procedures during the first 14 days of life, 75.60% of them were painful procedures. When the same dataset was reanalyzed by Sellam and colleagues (2013), the mean number of painful procedures was 14.38 procedures per day (*SD* = 7.43). There was significant difference in the mean of painful procedures between the two gestational age groups. The younger group (≥ 24 to 28 weeks) had twice as many procedures (mean = 21 procedures per day, *SD* = 9.00) as the older group (29 to 32 weeks) (mean = 11.00 procedures per day, *SD* = 7.00) ($p < .0001$). The mean number of painful procedures per preterm neonates was approximately 14 procedures per day and it increased in the younger preterm neonates. Similarly, a study in China recorded all painful procedures performed on preterm infants from admission to discharge (Chen et al., 2012). Most of the painful procedures were performed within the first three days of life. Preterm neonates, especially those born at 28 and 29 weeks' gestation, experienced more pain than those born at 30 weeks' gestation or later ($p < .001$). Notably, none of the painful procedures was accompanied by analgesia. It can be concluded that the wide variation in number of procedures from studies depends on gestational age of infants and duration of NICU stay (Badr, 2013).

2.1.2 Adverse effects of pain in preterm neonates

Pain during the early days of life can have short-term and long-term impacts on children born preterm in many developing systems (Anand, 1998; Fitzgerald & Walker, 2003; Fitzgerald, 2005). Those short-term impacts included alterations in physiological stability, the caution of early intraventricular hemorrhage, and a prolonged cutaneous sensitization or hyperalgesia for days or weeks or an increased sensitivity to pain. Invasive procedures in the NICU incite inflammatory responses and tissue damage that can last from hours to several days. Long-term impacts included pain in early life influencing not only long-term neurodevelopmental, cognitive outcomes, behavioral and social-emotional outcomes, but also pain reactivity in later life. Preterm neonatal pain may also indirectly affect postnatal growth beyond this early neonatal window. The prospective study in 78 preterm infants found that greater neonatal pain predicted lower body weight and head circumference percentiles, after adjusting for birth weight percentile and postnatal risk factors of illness severity, duration of mechanical ventilation, infection, and morphine and corticosteroid exposure (Vinall et al., 2012). Taken together, adverse neonatal experience alters the brain development and subsequent behavior of preterm neonates (Anand & Scalzo, 2000). The early and prolonged exposures to pain in the neonate cause early neurologic injury in peripheral, spinal cord and supraspinal processing neuroendocrine functions, and neurological development (Bouza 2009; Hall & Anand, 2005; Vinall & Grunau, 2014). Therefore, early and repeated nociceptive input in preterm neonates can be harmful to the nervous system, especially in the neonatal period of life and can lead to long-term undesired changes and adverse outcomes observed in later childhood and adulthood (Brummelte et al., 2012; Grunau, 2013; Grunau et al., 2006; Vinall & Grunau, 2014).

Besides neonatal development, preterm neonates who are exposed to cumulative early procedural pain or exposed to painful procedures during the plasticity of the developing peripheral and central nervous system without analgesia demonstrated higher sensitivity and lower pain thresholds compared to healthy full term controls. Brummelte et al. (2012) found that higher numbers of skin breaking procedures were significantly associated with early brain development in the very preterm infants brains (n = 86, 24 to 32 weeks' gestation) by reducing white matter (β = -.0002, p = .028) and subcortical gray matter maturation (β = -.0006, p = .004). Interestingly, early but not later pain exposure was a significant predictor of abnormal

white matter microstructure. Additionally, preterm neonates who received repeated heel sticks in intensive care had a lowering of cutaneous flexion reflex thresholds in the long-term (Andrews & Fitzgerald, 1999). The data from a systematic review of the effects of repeated painful procedures in infants supported that preterm infants exposed to cumulative early procedural pain demonstrated higher sensitivity and lower pain thresholds compared to healthy full term controls (Hatfield, Meyers, & Messing, 2013). These alterations may place preterm infants at greater risk for developing chronic pain conditions. Thus, valid interpretation of pain response may lead to appropriate management, decrease adverse outcomes and ultimately promote brain development.

2.1.3 Pain pathways in preterm neonates

The preterm neonate has an anatomical and functional pain system which can perceive and respond to tissue injury even if it is immature (Eiland, 2012; Fitzgerald, 2005, 2015; Lundeberg & Lundeberg, 2013). Pain pathways or nociceptive circuits are the detection and transmission of information regarding the presence and quality of a painful stimulus from the site of stimulation to the higher brain centers (Eiland, 2012; Simon & Tibboel, 2006; Verriotis et al., 2016). A better understanding of the anatomical and functional of nervous system of developing pain pathways in preterm neonates is described in terms of development of nociceptive circuits and related to a preterm neonate's reactivity (see Table 2.1) as follows:

Table 2.1

Anatomic pain pathways of preterm neonates

Anatomic pathways	Pain circuit	Pain reactivity
Peripheral and ascending pathways		
Peripheral nervous system	Greater density of nociceptor per area of skin	Increased sensitivity (decreased threshold)
- *Nociceptors*	Lack of myelination in the A delta or C-fibers	Slower conduction rate
- *Afferent fibers*		
- *Neurotransmitter*	Biochemical mediators released and the repeated stimuli	Primary & secondary hyperalgesia, allodynia
Spinal cord	Weak linkage between afferent fibers and dorsal horn (*Rexed's laminae*)	Latency response, diffused and exaggerate behaviors
-*Spinothalamic*		
-*Spinomesencephalic*		
-*Spinorecticular*		Cutaneous withdrawal reflex & muscle spasm
	Anterior horn (*laminar VII*) stimulate autonomic nervous system	Vasoconstriction in the skin and splanchnic region
Supraspinal area	Thalamic synaptic connections	Interpretation of pain
- *Thalamus*	Periaqueductal gray area	NIRS responses (HbO_2)
- *Mid brain*	Connected with bilateral somatosensory and motor cortex areas	Facial responses & arousal
- *Cortex*		Avoidance behaviors
	Stimulation of NMDARs releasing excitatory neurotransmitters	Activate descending pathways
		Central sensitization
		Windup phenomena

Table 2.1 (continues)

Anatomic pathways	Pain circuit	Pain reactivity
Descending pathway		
Lateral corticospinal tract	Sensory input is attenuated before being transmitted to the ascending pathway	Releasing of inhibitory neurotransmitters Endogenous analgesia
Reticulospinal tract	Connections associated with periaqueductal gray area	Dopamine and norepinephrine released
Raphe spinal tract	Inhibit impulse at spinal level	Serotonin (analgesia at the spinal level) released Inhibit motor responses

1) Development of a preterm neonate's nociceptive circuits

The anatomical and functional of nervous system support the preterm neonates' capacity for nociception develop early in gestation, thus pain perception of preterm neonates begins before birth. The CNS development related to components of pain pathways is described in four, five or six depending on knowledge available at that time and the purpose of information sources. Four processes of brain development include neuronal proliferation, migration, organization, and myelination (Brab & Lemons, 1989). Holditch-Davis and Blackburn (2014) stated that the development of CNS can be divided into six overlapping stages (1) neurulation, (2) prosencephalic development, (3) neuronal proliferation, (4) neuronal migration, (5) organization, and (6) myelination. These stages support the understanding about congenital birth defect. Neurulation composes two stages including formation of the brain and spinal cord and the caudal neural tube which occurs in three to four weeks of gestation. Prosencephalic development occurs in two to three months gestation involving development of brain and ventricular system. During this period the three primary brain bulges (or vesicles) and cavities are formed, after fusion of the neural folds in the cranial area. Development of the face is associated with prosencephalic development of the CNS. Alterations in brain development in this period result in facial malformations. For the rest of stages, from proliferation to myelination is slightly the same content with the five stages of brain development including

proliferation, migration, synaptogenesis, organization, and myelination which described elsewhere (Linderkamp, Janus, Linder, & Skoruppa, 2009) as follows:

1.1) Neuronal proliferation

Neuronal proliferation involves with the development of neurons in the germinal matrix which lasts from two to four months gestation (Holditch-Davis & Blackburn, 2014) and glial cells proliferation at approximately five months (Brab & Lemons, 1989). Neurons from peripheral neuron system up to spinal cord are produced and migrate to their final locations in the supraspinal area and the brain during week seven to 24 of gestation, while the tactile system fully appears in early as weeks eight of gestation and is functional by week 12 of gestation (Eiland, 2012; Fitzgerald, 2005; Linderkamp et al., 2009; Slater, Cantarella, et al., 2006). Cutaneous receptors and sensory fiber development begin to appear in the area of the mouth at the week seven of gestational age, then spreads to the palms of the hands, the soles of the feet, the trunk and proximal parts of limbs and then spreading to all cutaneous and mucous surface completely by week 15 to 20 of gestation (Bouza, 2009; Eiland, 2012; Simons & Tibboel, 2006). Transmission of pain signals involves the activation of those nociceptors in the periphery through afferent fibers to the dorsal root ganglion and project to dorsal horn of the spinal cord.

1.2) Neuronal migration

The peak period of neuronal migration stage is three to five months gestation (Brab & Lemons, 1989; Holditch-Davis & Blackburn, 2014). This stage is characterized by movement of millions of cell from their point of origin to their eventual loci in the cerebral cortex and cerebellum. The cerebral cortex has essentially archived its full complement of neurons by week 20 to 24 of gestation. In terms of ascending pathway, signals from nociceptors are transmitted through three types of primary afferent fibers (i.e., A-beta, A-delta, and C-fibers), carried across and ascend in the contralateral spinal cord. During development, the thickly myelinated A-beta fibers, which transmit light touch and proprioception in the adult, also appear to transmit noxious information to pain processing areas of the spinal cord. However, A-delta fibers and C nociceptors mediate “first” and “second” pain, respectively, namely the myelinated, rapid, acute, sharp pain and the unmyelinated, delayed, more diffuse, dull pain evoked by noxious stimuli (Julius & Basbaum, 2001). In the

rat pup, C-fiber polymodal nociceptors have mature thresholds and firing pattern at birth, A-delta mechanoreceptors are less mature, and A-beta mechanoreceptors are the most immature at birth (Fitzgerald, 2005). Afferent fibers and synapses between sensory fibers are completed by the week 30 of gestation (Bouza, 2009; Ponder, 2002; Stevens, Johnston, & Grunau, 1995). According to information from rodent models, the central terminals of low-threshold tactile and nociceptive inputs are relatively diffuse and initially overlap with those of nociceptive C-fibers in lamina II of the neonatal dorsal horn, slowly becoming pruned over the postnatal period, thus discrimination between noxious and non-noxious stimuli is less efficient (Fitzgerald, 2005; Fitzgerald & Walker, 2009). The specific neural circuits necessary for discrimination between touch and nociception emerge from 35 to 37 weeks of gestation in the human brain (Fabrizi et al., 2011).

1.3) Synaptogenesis

Nociceptive sensory input reaches the thalamus through second-order neurons in the spinothalamic, spinorecticular, and spinomesencephalic tracts and is then distributed throughout the brain (Walden, 2013). Excitatory synapse transmission is modulated by three receptor types (α-amino-3-hydroxy-5-mythyl-4-isoxazolepropinic acid receptors [AMPARs], kainate receptors [KARs], and *N*-methyl-D-aspartate receptors [NMDARs] (Eiland, 2012; Patterson & Fitzgerald, 2004). Synaptogenesis is mediated by excitatory neurotransmitter such as glutamate and glutamate acts on NMDARs to enhance neuronal proliferation, migration, and synaptic plasticity. Overall, excitatory of these three receptors is more than those existing in adults, whereas inhibitory circuits have delayed functionality and depolarizing rather than hyperpolarizing (Eiland, 2012). As mentioned overlapping stages, development of synaptic connections is explained under the organization stage (Holditch-Davis & Blackcurn, 2014).

1.4) Organization

The organization stage is a period for hardwiring specialized function and action of neuronal cells (Linderkamp et al., 2009). The knowledge of fine-tune developing pain circuits to the established and pruning of synaptic connection is derived from rodent models. In human, the peak period of organization is about the fifth month of gestation to a few years after birth (Holditch-Davis & Blackcurn, 2014). Organization processes allow the

nervous system to act as an integrated whole. These processes include establishment of subplate neuron, attainment of cortical neuron, differentiation and branching of axons and dendrites, differentiation of the glial cells, development of synaptic connections, balancing of excitatory and inhibitory, and cell death and selective elimination of neuronal process. Lack of connection between neurons can result in hypersensitivity, poorly modulation behaviors, and all –or-nothing responses, which can often be observed in preterm infant in NICU.

1.5) Myelination

Myelinization occurs between eight months gestation and two years of age, with the peak period thought to be occurring within the first eight postnatal months of life (Holditch-Davis & Blackcurn, 2014). The neuronal cells are covered with a lipoprotein shell that helps facilitate conduction of neuronal impulses. Myelinization of ascending pathways in the spinal cord, brainstem, and thalamus is completed by about 30 weeks of gestation, and myelinization from thalamus to cortex is completed by 37 weeks of gestation (Holditch-Davis & Blackcurn, 2014). Thus, the underdeveloped myelination and slower synaptic transmission from immature ion channel kinetics cause sensory and motor neurons in the CNS to respond at longer and more-variable latencies, producing less-synchronized responses to peripheral stimulation (Linderkamp et al., 2009).

2. Nociception and a preterm neonate's reactivity

Nociception is composed of the processes of transduction, conduction, transmission, perception and modulation and described elsewhere (Marko & Dickerson, 2017). Ascending pathways begin from the transduction to perception phase and descending pathways are the modulation and inhibition phase. These five basic processes of pain perception in infant and children are similar to those of adults, but some unique differences exist (Fitzgerald, 2005, 2015; Hall & Anand, 2005; Walden, 2013) (see Table 2.1). The areas of the nervous system are indicated where developmental changes and plasticity impact pain detection and treatment in preterm neonates (Fitzgerald & Walker, 2009). Repetitive painful stimulus as early adverse experiences leads to abnormal development and behaviors. For example, repetitive painful stimulus causes excessive NMDA/ecitatory amino acid activition resulting in excitotoxic damage to developing neurons in nervous system (Anand & Scalzo, 2000; Fitzgerald, 2005). Preterm neonates respond to painful stimulus associated with changes in their arousal, specific

behaviors, cerebral blood flow, physiology, and cellular and molecular in pain pathways (Hall & Anand, 2005). To understand the underlying mechanisms, the pain pathway, and a preterm neonate response, the key sites of developmental transition in infant pain pathways are described in parallel with these five basic phases of nociception.

Transduction phase is the conversion of a noxious thermal, mechanical, or chemical stimulus into electrical activity in the peripheral terminals of nociceptor sensory fibers. Activation of nociception occurs when sensory information or noxious stimuli is detected by specific receptors of free nerve endings in the dermis of the skin. This process is mediated by converting noxious stimuli to electrical current. Nociceptors, one of the five main categories of receptors, are the pain receptors or free-nerve endings that detect and transmit general physical and chemical tissue damage (Jorgensen, 1999). The other four categories of receptors are mechanoreceptors, thermoreceptors, chemoreceptors, and electromagnetic receptors. In young rats, the cutaneous receptive fields of individual spinal sensory neurons are larger and gradually become smaller and more organized with over the first two postnatal weeks (Fitzgerald, 2005; Fitzgerald & Jennings 1999). In the human fetus, the number and types of peripheral nociceptors are similar to adult numbers by 20 to 24 weeks' gestation, implying a greater density per area of skin (Hall & Anand, 2005). Thus, peripheral innervation is vulnerable and sensitive to tissue injury (Fitzgerald & Walker, 2009).

The conduction phase is the passage of action potentials from the peripheral terminal along axons to the cell body located in the dorsal root ganglion in the spinal cord. Then, the current goes to their central terminal in the dorsal horn of the spinal cord where the electrical impulses initiate neurotransmitter release and are the very first synaptic processing station for pain. In the adult, lamina II is exclusively occupied by C fibers, but in the neonatal period it is also occupied by transient A-fiber terminals (Fitzgerald, 2005). The A-delta fibers synapse with neuron in the dorsal horn of the spinal cord known as Rexed's laminae, especially I, II, and V (Hall & Anand, 2005) at early postnatal age, with subsequent gradual withdrawal down to laminae III or below over the first three postnatal weeks leading to sensory processing and may underlie the increase cutaneous thresholds (Fitzgerald, 2005; Fitzgerald & Walker, 2003). Unmyelinated C-fiber synapse with laminae II and III. Signal spread to the anterior horn cell producing muscle spasm, resulting in reflex withdrawal and to the preganglionic sympathetic neurons of the autonomic nervous system to produce vasoconstriction of the skin. All infants

display a robust and long-duration flexion reflex (> four seconds) to a single noxious skin lance, the magnitude of which decreases significantly with gestational age (Cornelissen et al., 2013). In contrast between the preterm and term infant period, a single heel lance evokes biceps femoris activity decreasing significantly for at least two to four seconds, but remaining longer in duration than adult reflexes. It is now known that noxious stimuli are carried by both myelinated and unmyelinated fibers because the velocity of transmission is influenced by the size of the nerve fiber and the presence of myelin along the nerve fiber sheath. Incomplete myelination, before the 30th week of gestation, implies only a slower conduction speed in the nerves or central nerve tracts of infants.

Transmission phase is the synaptic transfer and modulation of input from one neuron to another. A delta fiber pain is conducted by spinothalamic fibers which are responsible for transmission of most nociceptive signals to somatosensory areas (Eiland, 2012; Ponder, 2002). The spinomesencephalic tracts terminate in the nuclei of the midbrain such as the periaqueductal gray area which contributes to avoidance behavior and activates the structure's descending analgesia system. The C-fiber pain passes through spinorecticular tract projects to the limbic system and mediates pain-associated arousal, activates the endogenous analgesia system, and triggers the affective, neuroendocrine, and autonomic responses to pain. Thus, this nociceptive circuit results in reflex muscle contraction and withdrawal from noxious stimuli and this circuit does not involve conscious perception of pain (Eiland, 2012). The ability to sense and transmit painful stimuli to the brain present by about 20 weeks of gestation, and a response to pain can be observed from 22 weeks of gestation. However, less than 10% of dorsal horn neurons are projection neurons that send information up to higher centers in the brain (Pattinson & Fitzgerald, 2004). There is a weak connection between the afferent nerve fibers and dorsal horn cells before the second postnatal week of preterm neonates. Immediately after tissue damage, there will be local releasing of mediators including peptides and neurotransmitters, for example substance P (Fitzgerald & Beggs, 2001). These substances stimulate the responses of the peripheral nociceptors to painful stimuli. The great number of substance P receptors diminish in the first two postnatal weeks, and the immaturity of the central inhibitory control system of the newborn, are probably responsible for this exaggerated reaction (Fitzgerald, 1993). This immaturity of sensory processing leads to lower thresholds for excitation and sensitization, thus after a noxious stimulating such as heel stick, they show the same long-lasing excitation called the "windup phenomena" (Fitzgerald & Beggs, 2001; Hall

& Anand, 2005; Koch & Fitzgerald, 2013). While the nociceptors were still immature, they are triggered by activation tactile afferents of sensory neurons that underline this hypersensitivity (Woodbury, 2008). Primary hyperalgesia in sites of surgical injury or inflammation is evident from birth in both rat pups and human infants, but secondary hyperalgesia that spreads outside the area of injury is slower to mature and may be less obvious in infants. The capacity for referred hyperalgesia, whereby visceral pathology evokes abdominal reflex hypersensitivity, also increases with postnatal age in human infants.

The perception phase is pain processing in the brain by signals via the ascending nociceptive pathway to the brain. The supraspinal processing of noxious stimuli imply that activation of cortical neuronal circuits is the first step toward formation of primary consciousness (Elderman 2004 as cited in Bartocci, Anand, & Lagercrantz, 2006). The consciousness of pain in human neonate has been an issue for critical commentary since 1996 (Anand & Craig, 1996). Currently comparing between the two studies by Anand et al. (2006) and Slater, Cantarella, et al. (2006), the indistinct picture of the cortical processing in preterm neonates appears apparently. Both studies found changes in cerebral blood volume reflected oxygenated hemoglobin (HbO_2) and somatosensory cortical activation responses tested by near-infrared spectroscopy (NIRS) (Bartocci, Bergqvist, Lagercrantz, & Anand, 2006; Slater, Cantarella et al., 2006). The conclusion is robust cortical activation but not in the occipital cortex after acute pain in preterm neonates. In addition, periaqueductal gray area at midbrain receives signals from all types of afferent fibers, thus it is extremely important in the modulation, interpretation, and response to pain. Facial expressions reflecting pain, arousal effects of pain, and autonomic response to pain because of the stimulation of periaqueductal gray area. This area is also connected with the reticulospinal tract of the descending inhibitory pathways.

The modulation phase can be regulated by local inhibitory circuits in the spinal cord and descending inhibitory pathways from the brainstem. The three descending modulatory tracts through the spinal cord including corticospinal tract, reticulospinal tract, and raphe spinal tract (Hall & Anand, 2005). For examples, the corticospinal tract inhibits spinal processing of pain through the release of inhibitory neurotransmitters (serotonin, dopamine, and norepinephrine) that inhibit substance P and other excitatory neurotransmitters. Thus, endogenous analgesia is mediated by inhibiting these pathways that might dampen the pain. In

adults, the descending brainstem pathways provide a prominent mechanism in controlling pain transmission. In the neonates, the modulation and inhibition of pain incoming impulse is weak, poorly targeted and does not develop until late gestation (36 to 40 weeks of the gestation) or even six to eight weeks after birth (Anand, 2004). Thus, the exaggerated spinal responses contrast with the facial expression response, which is weaker in younger infants and increases with postnatal age.

In summary, the nervous system requirements for nociceptive pathway of preterm neonates are present before birth. The functional signaling pathways involved in the pain experience of preterm neonates are different from those of adults. It does not involve evaluating or attributing meaning to the subjective experience called pain but it involves peripheral sensory receptors whose afferent fibers synapse in the spinal dorsal horn and supraspinal areas (ascending pathways). While much of the system described in adults is also functional in newborns, the postnatal period is a time of great structural and functional change in pain pathways (descending pathway). As a result, noxious stimulation does not evoke the same pattern of activity as it does in the adult CNS. Dampened behavioral and physiological responses to pain do not necessarily represent an absence of nociceptive processing in the CNS.

2.2 Pain Indicators and Assessment for Preterm Neonates

2.2.1 Pain indicators in preterm neonates

According to a review of pain pathways and pain reactivity in preterm neonates, the surrogate measures of pain in preterm neonates must be viewed in the context of CNS sensory circuits and motor pathways that are undergoing considerable structural and functional maturation over the postnatal period. A combination of behavioral indicators of pain, physiological indicators of pain, and factors affecting pain reactivity should be considered for assessing pain in preterm neonates.

1) Behavioral indicators of pain

Regarding the preterm neonates undergoing painful procedures, there are generally five phases of continuous event on each occasion. First, before the procedure called the baseline phase, the preterm neonates are mainly in an active sleep state, with no perceptible signs of pain and no infant crying. Second, in the handling or preparing phase, immediately

before the puncture, they remain primarily drowsy or in active sleep but some infants cry and increase facial expressions. There is a gradual increase in the awake-state from the baseline to handling phase, and from the handling phase to the puncture phase. Third, in the puncture phase, most infants show more facial expressions and a behavioral state of arousal, indicating a significant increase in reactivity to pain during the puncture procedure in relation to the baseline pattern. Fourth, in the dressing phase, the neonates are predominantly in the quiet alert state. Finally, in the recovery phase, most infants are in the same arousal state as in the puncture. The reactivity to pain is unquestionable, and available behavior indicators indicated for assessment of pain responses and stress reactions to events prior to procedures involve exposure to acute pain. Currently, studies demonstrated the variation in response patterns across preterm infants (Chimello, Gaspardo, Cugler, Martinez, & Linhares, 2009). The latency, magnitude, and duration of behavioral responses in preterm neonates are observed to understand the individual patterns of reactivity and recovery after painful procedures. The results of observational study found that in the recovery-resting phase, only half of the sample returned to the baseline pattern of behavioral responses. However, ten minutes after the puncture, 38% of the neonates kept showing signs of a pain response and 31% of the neonates were behaviorally activated. Behaviors are the first indicators of pain that alert caregivers to provide assistance (McGrath, 1998). Even though they are indirect measures, behavior deserves usage as a measure of pain in children, especially in young children. Behavioral measures of pain or behavioral indicators are intended to indicate the existence of pain through reactions to pain which are observed by caregivers, especially a nurse. Based on observational studies in each phase and evidence of pain pathway in preterm neonates, four behavioral indicators of pain including facial expressions, body movement, change of sleep-wake states, and cry are described as follows:

Facial expression is the most effective and most specific indicator of pain in preterm neonates; in the meanwhile it is one of the subjective indicators (Mitchell et al., 2000). Most of facial expression indicators in most scales for infants are based on Neonatal Facial Coding System (NFCS). Using the NFCS needs to be trained and training of NFCS takes eight hours for one individual (Hudson-Barr et al., 2002). According to NFCS, brow bulge is defined as "*bulging, creasing and vertical furrows above and between brows occurring as result of the lowering and drawing together of the eyebrows*" (Grunau & Craig, 1987, p.399). Brow bulge, the contraction of the corrugator muscle, is the most sensitive marker of pain in preterm

neonates (Holsti, Grunau, Oberlander, Whitfield, & Weinberg, 2005) including ELGA neoantes (Gibbins, Stevens, Beyene et al., 2008). Changes in four facial actions (brow bulge, eye squeeze, nasolabial furrow, and vertical mouth stretch) were the most sensitive indicators of pain in neonates and ELGA infants (< 28 weeks gestation) measured by NFCS (Gibbins, Stevens, McGrath et al., 2008; Xia, Yang, Zhao, & Zhang, 2002). In Bernese Pain Scale for Neonates (BPSN), the result of using this scales suggested that eyebrow bulge with eye squeeze as a proxy for upper facial actions might be suitable for preterm neonates with mechanical ventilation (Cignacco, Mueller, Hamers, & Gessler, 2004). Additionally, brow bulge, eye squeeze, and nasolabial furrow indicators of the PIPP scores were well correlated with cortical hemodynamic activity using NIRS (Slater et al., 2007; Slater, Cantarella, Franck, Meek, & Fitzgerald, 2008). Interestingly, the latency to the facial expression response ranged from one to 17 seconds which was dependent on the infants' postmenstrual age at the time of the heel stick. Infants below 32 weeks' postmenstrual age had a significantly longer latency for change in facial expression than older infants (Slater et al., 2009). Even though facial expressions seem to be a subjective indicator and appear to have different patterns of response in younger and older infants, they need to be considered when choosing only sensitive indicators for the future pain assessment tool of preterm neonates (Grunau, Oberlander, Holsti, & Whitfield, 1998).

Sleep-wake organization which involves coordination of physiological, biochemical, and neurobehavioral system is the way of communication an infant uses to express internal needs in response to external events. Transition among sleep-wake states, called transitional sleep or indeterminate sleep, is commonly found in a preterm neonate due to immature neurological function. Sleep states are categorized as quiet or deep sleep and active or light sleep. Awake state is composed of drowsy, quiet alert, active alert, and crying. Preterm infants spend more than 85% of sleep time in active sleep and more defined and less diffuse states with increasing postmenstrual age (Foreman, Thomas, & Blackburn, 2008). Neonates respond to painful stimuli with complex behavior, such as alterations in sleep-wake cycles, sleeplessness and increased irritability (Bouza, 2009). The results from a systematic review by Sellam and colleagues (2011) found that sleep-wake states can be used as a predictor of pain response because the odds of a sleeping infant demonstrating behavioral and physiological indicators were 17.87% of those for an infant who was awake (Johnston et al., 1999). Infant behavioral state appears to be a useful additional global indicator of biobehavioral shifts (Grunau et al., 1998). The descriptive, cross-sectional study was used to compare the pain

responses of 72 preterm infants to a heel stick procedure taking into consideration a variety of factors, including several background and contextual variables. Four predictor variables including behavioral state, gestational age, type of needle, and severity of illness were correlated with pain responses assessed on the PIPP scale (Badr et al., 2010).

Body movements as another indicator of pain refers to the movement of whole body or extremities, except the face. Lack of body movement, a suppression of the grasp reflex, decreased muscle tone, and decreased arousal to stimuli are observable in sedated infants. ELGA infants demonstrated greater movements of the body during diaper changes and no specific changes in the body's movement following the heel stick, which may reflect immature motor coordination (Gibbins, Stevens, Beyene et al., 2008). Therefore, it is necessary to consider the condition of infants with careful interpretation and to differentiate pain from stress.

Crying is an indicator that is commonly used for assessing pain in a full-term infant. The presence of a cry might be helpful in distinguishing between the intensity and duration of pain, but the absence of an audible cry provides no information in those who exhibit silent cries. Thus, an audible cry was not sensitive pain indicators in using with the ELGA infants, due to inhibiting drugs or the presences of endotracheal tubes in this high risk population (Gibbins, Stevens, Beyene et al., 2008). The immature of motor system and illness of the ELGA infants also influence their crying. The inclusion of only crying with voice as a pain indicator for the ELGA infants may not be justified. The study of using the Behavioral Indicators of Infant Pain (BIIP) scale found that almost half of the infants did not cry during the skin-breaking phase (Holsti & Grunau, 2007). The preterm neonates' audible crying is a possible indicator during on continuous positive airway pressure (CPAP) or cannula. For preterm neonates with endotracheal tube, the silent cry is a possible pain indicator. The ventilation infants who can make no sound, but is seen to gape around endotracheal tube, which is called "silent cry" (Sparshott, 1997).

2) Physiological indicators of pain

Physiological responses to pain result from activation of sympathetic and parasympathetic nervous systems, and responses to noxious stimuli are limited to the immediate post noxious period (Gibbins & Stevens, 2001). Thus, these responses are

inconsistent and unsustainable over time. Physiological measures of pain are intended to indicate the existence of pain through changes of physiological variables. Physiological indicators commonly used in the assessment of acute procedural pain in preterm neonates include heart rate, oxygen saturation, and respiratory rate or breathing pattern. Even though there are many physiological measures in research settings, heart rate appears to be the most promising physiological measure in clinical settings (Raeside, 2011; Sweet & McGrath, 1998).

Heart rate, which is respond to stimulus, can be bi-directional (Sparshott, 1997), increased as response to stress and fear but decreased as response to visual and auditory stimulation. Morison and colleagues (2001) suggested that the relationship of association and concordance among and between behavioral and autonomic reactivity to pain are complex and mediated by multiple sources. The best approach to pain measure may be a multi-dimensional one. The low to moderate correlations between the behavioral and physiologic response systems to pain may be due to several reasons such as maturation and development, response specificity and stimulus specificity, and individuals response (Sellam et al., 2011). Since heart rate is inversely correlated with gestational age, a sample with older gestational age may have a lower resting heart rate. The mean resting heart rate of neonates for three to five days of life was 152 beats/minute (*SD* = 7) for 27.00 to 28.90 weeks, 153 beats/minute (*SD* = 9) for 29.00 to 31.90 weeks, and 149 beats/minute (*SD* = 8) for 32 to 34 weeks. The mean resting heart rate was higher for all three groups on day eight to ten (Ruth, Kennedy, Rehm, Ariagno, & Lee, 2010). The consistent results from two studies found that a higher minimum, maximum and mean heart rate during heel stick in younger infants (< $31^{6/7}$ weeks of the gestational age) compared with the older infants (> 32 weeks of the gestational age) (Gibbins, Stevens, McGrath et al., 2008; Lucas-Thompson et al., 2008). Even though, the specific changes in heart rate variability were found to relate more to the gestational age of the infant than to the noxious stimulus, most of infants had heart rate changes during noxious stimulus. Heart rate change is the most specific indicator of painful procedures of preterm neonates because the increase in heart rate is greater during painful procedure compared to clustered care (Holsti et al., 2005).

Oxygen saturation is the other objective parameter of existing pain assessment in preterm neonates. However, the drops in oxygen saturations were greater during the non-painful handling than painful events (Holsti et al., 2005). During each phase of the heel stick procedure, the ELGA infants reached a lower minimum, maximum and the mean oxygen

saturation was compared to all other gestational age strata and might reflected the physiological immaturity or severity of illness (Gibbins, Stevens, McGrath et al., 2008). Oxygen saturation, blood pressure and respiratory rate lacked sensitivity and specificity and cannot be used independently (Raeside, 2011).

3) Factors affecting pain reactivity

Factors affecting pain reactivity, called contextual factors (Stevens, Johnston, Petryshen, & Taddio 1996), are variables that impact pain responses of preterm neonates which are categorized into six categories including age, previous pain exposure, health status, therapeutic interventions, behavioral status, and demographic factors (Sellam et al., 2011). Evidence shows many factors affecting pain reactivity in preterm neonates (Badr et al., 2010; Johnston et al., 1999; Stevens et al., 1996) which corresponds with the results of two systematic reviews (Sellam et al., 2011; Valeri & Linhares, 2012). The first systematic review examined 23 studies covering the period from 1990 to 2009 and found varied strength of the association between contextual factors and pain response of preterm neonates (Sellam et al., 2011). The second systematic review examined 18 studies covering the period from 2004 to 2009 and focused on the effects of sex, gestational age, and neonatal illness severity on pain response of preterm neonates (Valeri & Linhares, 2012). The following factors that influence pain responses are summarized from the results of those two systematic review studies and adding literature review from beyond year 2009. These following factors do not influence pain in isolation but are listed separately for clarity and understanding.

3.1) Gestational age

Age is commonly measured by using the gestational age at birth. The results of the systematic review found that most studies reported a statistically significant gestational age effect on behavioral response to pain with greater behavioral response as gestational age increased (Sellam et al., 2011). Consistent with fourteen of eighteen studies which analyzed the effect of gestational age on pain responses studies found significant associations between gestational age and pain responses (Valeri & Linhares, 2012). It can be concluded that gestational age influences in both behavioral and physiological responses in infants born preterm. The facial expression response to pain tended to increase with age because of the development of the nervous system. The cephalocaudal development of facial

musculature may influence the magnitude of facial activity as evidenced by the fact that preterm infants have less muscular strength, posture, tone, and body movement compared to term infants (Sellam et al., 2011). ELGA infants, especially those with less than 32 weeks of gestation, demonstrated the least amount of change in facial activity compared to more mature infants (Gibbins, Stevens, Beyene et al., 2008).

3.2) Previous pain exposures

Previous pain exposures differs from pain experience. Pain experiences must involve the cerebral cortex and cognitive brain function. Anand and Scalzo (2000) explained that how infants remembered painful experiences which were biologically memories, not conscious recall. Therefore, pain-produced reflex behavior cannot be equated as pain experience. Previous pain exposures were measured in many ways including the cumulative previous number of painful procedures, the time since the last painful procedures and the number of painful procedures during the 24 hours prior to the heel stick (Sellam et al., 2011). However, the cumulative previous number of painful procedures was commonly reported significantly in related with pain reactivity as follows:

Johnston and Stevens (1996) compared pain reactivity between two groups of infants. They were the newly born group (n = 53) and the earlier-born group (n = 36). The mean of total number of pain exposures of the newly born group and the earlier-born group were 12.20 (SD = 6.70) and 71.50 (SD = 27.60) procedures, respectively. The preterm neonates who had a higher number of pain exposures and were exposed to care in NICU for four weeks had diminished behavioral responses to heel sticks as compared to age-matched controls. The maturation of facial expressions in response to heel sticks was delayed by frequent invasive procedures. Within the groups, the correlation between facial scores and invasive procedures was higher for the earlier-born infants (r > .34) than the newly born infants (r > .18). The preterm infants of 32 week postconceptional age who were born at the gestational age of 28 weeks showed depressed behavioral responsiveness associated with the number of painful procedures that they had undergone. The other cohort study examined the relationship of factors and biobehavioral reactivity to acute pain in 136 preterm infants at 32 weeks' postconceptional age (Grunau, Oberlander, Whitfield, Fitzgerald, & Lee, 2001). lnvasive procedures were defined as those involving tissue damage, including but not limited to heel lance, venipuncture, insertion of arterial and venous lines, and injection. Intrusive procedures,

such as endotracheal tube suctioning, were not included. A number of previous invasive procedures since birth (mean = 86.10 [*SD* = 55.20]; range, 5 to 283) was a significant factor associated with altered behavioral (r = .29) and autonomic pain reactivity (r = .26). Interestingly, those preterm neonates who received more than 20 procedures were associated with dampened behavioral response measured by NFCS; therefore it might be a cutoff point number of procedures to shift the subsequent response (Grunau et al., 2001). For hierarchical linear regression analysis, higher number of previous pain procedures was associated with dampened pain response at 32 weeks' postconceptional age (β = -.51; p < .0001). With regard to the results of both studies, the idea that higher numbers of skin breaking procedures predicted dampened facial responses is supported. Even though the numbers of skin breaking procedures have not been included in any existing scales, the evidence shows that the number of previous pain procedures influences pain reactivity and pain scores.

3.3) Length of NICU stay and respiratory suppport

The length of stay in the NICU and receiving respiratory support such as mechanical ventilation were factors related to frequency of painful procedures influencing pain reactivity of preterm infants (Cruz, Fernandes, & Oliveira, 2015; Williams, Khattak, Garza, & Lasky, 2009). They often experience multiple painful procedures during their stay in NICU. The systematic review of 18 observational studies in frequency of painful procedures summarized that preterm infants at the lowest gestational age with very low birth weight and/or neonates receiving nasal oxygen, continuous positive airway pressure and ventilation support were exposed to a high number of painful procedures (Cruz et al., 2015). They also reported that six studies reported a relationship between the number of painful procedures and day of admission or postnatal age. In addition, the other study aimed to evaluate the behavioral responses of 35 preterm infants to multiple heel stick procedures during their stay in the NICU (Williams et al., 2009). Sixty-one video recordings of blood collection by heel lance were evaluated for behavioral pain response using the NIPS. Mean of length of NICU stay was 81 days (range, 62 to 126) and 63% of infants used mechanical ventilation during hospital stay. The results found that length of stay was negatively associated with pain scores. They had lower pain scores by 0.05 points (range, 0.02 to 0.09) for each additional week they stayed in the NICU. Infants receiving mechanical ventilation at the time of testing had lower pain scores by an average of 0.69 points (range, 0.10 to 1.27) (p = .02). It can be concluded that the longer the

length of NICU stay, the less robust the preterm neonates are in their behavioral responses to painful procedures, especially in infants receiving mechanical ventilation.

2.2.2 Existing pain assessment tools for preterm neonates

Recently, thirteen pain assessment tools for preterm infants exist, yet a gold standard pain assessment in clinical setting has not yet emerged (Anand, 2007; Fitzgerald & Walker, 2009; Franck & Miaskowski, 1997; Lake, 2013; Gibbins, Stevens, Beyene et al., 2008; American Academy of Pediatrics and Canadian Paediatric Society, 2000). The measurement of preterm neonate response to painful stimuli remains a significant clinical problem. The feasibility and practicality of using most of the pain measures in clinical practice are questionable because the previous scales have been developed for research purposes and have not been used routinely at the bedside. The following contents present overview of 13 published pain assessment tools for preterm neonates in terms of development and utilization. Concerning the components of these exiting tools, the pain assessment tools can be categorized into two groups; uni-dimension and multi-dimension scales. The uni-dimensional tools use a single indicator of pain assessment (physiological or behavioral aspect), whereas multi-dimensional tools include both the physiological and behavioral aspects that provide a more comprehensive assessment of pain (de Melo et al., 2014).

1) Uni-dimensional scale

Three scales are categorized in the uni-dimensional scale measuring only behavioral indicators. They are Neonatal Facial Coding System (NFCS), Behavioral Indicators of Infant Pain (BIIP), and Pain Assessment and Care for the Extremely Low Gestational Age Infant Focused Instrument (PACEFI).

1.1) Neonatal Facial Coding System (NFCS)

The nine facial expressions indicators of NFCS was initially developed by Grunau and Craig (1987). The facial expressions are brow bulge, eyes squeeze, naso-labial furrow, open lips, vertical mouth stretch, horizontal mouth stretch, lip purse, taut tongue, and chin quiver. They observed facial expressions of the 140 full-term newborn infants (38 to 42 week's gestational age) from a well-baby unit. The vocalization and sleep states were also observed using videotaping during heel stick to draw blood samples. Inter-observer reliability

was .88, computed for randomly selected 20% of the subjects. A three-way univariate ANOVA (4x2x2) of summed facial movement across state, sex, and the heel-rub/heel-lance condition with repeated measure on the last factor was carried out to conduct the hypothesis testing. Following heel-lance, neonates showed a cluster of facial changes together with vocalization (cry) that differed substantially from the amount and types of facial action provoked by heel-rub (F [1,132] = 295.35, $p < .0001$) (Grunau & Craig, 1987). Remarkably, by using video recordings during heel lance to heel squeezing, over 96% of term neonates responded consistently to a painful stimulus with a cluster of facial actions, namely brow bulge, eye squeeze, nasolabial furrow, and open mouth. The later study by Grunau, Johnston, and Craig (1990) added one facial action relating to the mouth area, tongue protrusion, which was an out ward thrusting movement of the tongue between the lips. The result found that the tongue protrusion was associated with non-nociceptive tactile stimulation. It appeared to be a 'no pain' response in full-term neonates, and thus considered useful as a fairly conclusive counter indicator to pain.

The NFCS also has been used to study pain behavior of preterm infants (Craig, Whitfield, Grunau, Linton, & Hadjistavropoulos, 1993; Stevens, Johnston, & Horton, 1994; Johnston et al., 1995, 1996; Johnston & Stevens, 1996). In 1998, Grunau and colleagues tested the ten facial expressions of the NFCS with 40 preterm infants at 32 weeks to 32 weeks 6 days post conceptional age (mean of gestational age at birth = 28.09 weeks) in real time at bedside. Each face action was coded as 1/0 (occurred/did not occur). Repeated measures analysis of variance (ANOVA) showed statistically significant differences across events for facial activity ($p < .0001$). The reliability coefficients were .83 for the total NFCS and .86 for the NFCS excluding tongue protrusion. Interestingly, the NFCS reliability of four face expression indicators (brow lower, eye squeeze, nasolabial furrow, and open mouth) was .91. This item-reduced NFCS was used by Rushforth and Levene (1994) to assess 36 term infants and 31 preterm infants. Thirty-five term infants (97%) and 26 preterm infants (84%) showed an increase in the number of behaviors in response to heel lance.

The NFCS has been extensively applied to combine with physiological parameters. Even though it was the starting point of many pain scales' development such as PIPP, NIPS, and the Crying, Requires oxygen for saturation, Increased vital signs, Expression, Sleepless (CRIES) (Chen, Chang, Hsiao, Chen, & Lin, 2005), it still has some limitations. The

limitations of this scale are subjective individual variations in the vigor of infant's response, inability to use in sedated infant, and the requirement of training to recognize cues. The training to use and interpret the infant's facial activity on NFCS has been reported to require extensive training for eight hours per one individual as mentioned and extensive time to code by experienced coders for scoring (Abu-Saad, Bours, Stevens, & Hamers, 1998; Hudson-Barr et al., 2002; Steven et al., 1996). The full NFCS was used primarily in research and information about its clinical use is limited. However, item-reduced NFCS (i.e., brow bulge, eye squeeze, nasolabial furrow, and open mouth) is claimed that it is a reliable, feasible, and valid bedside facial coding for neonates (Rushforth & Levene, 1994) and infants (0-18 months) in postoperative pain (Peters et al., 2003).

1.2) Behavioral Indicators of Infant Pain (BIIP)

The seven-indicators of BIIP was developed by Holsti and Grunau (2007) for assessing acute pain behaviors in preterm infants by combining reliably pain-associated five facial actions from the NFCS with sleep-wake indicators and two hand movements. Each indicator is evaluated on different point scales such as zero to two for sleep-wake states, zero to one for face actions and two hand actions. The presence of pain is defined as a BIIP score ≥ 5. To establish construct validity, the BIIP scale was measured across phases of blood collection with videotape recording in 92 preterm infants (23 to 32 week gestational age) in NICU (Holsti & Grunau, 2007). The BIIP scale discriminates between procedures phases with low scores occurring before and after the procedure and the highest scores occurring during the heel stick or squeeze phase. The alpha coefficient for the BIIP during the heel stick/squeeze phase was .82. When the analysis was separated by gestational age, it was .89 for < 29 weeks and .76 for 29 to 32 weeks. The inter-rater reliability of NICU examined among trained nurses ranged from .80 to .92 depending on the phase. When the analysis was separated by gestational age, it ranged from .64 to .92 for < 29 weeks and .83 to .94 for 29 to 32 weeks.

In a subsequent study, the discriminant validity was evaluated in 69 preterm infants (24 to 32 weeks) by comparing changes in heart rate during blood collection and diaper changes (Holsti, Grunau, Oberlander, & Osiovich, 2008). Significant main effects were found for changes in the BIPP score and mean heart rate across the procedure phase. It

effectively differentiated painful from non-painful phases of the procedure. The development of the BIIP by combining theoretically derived indicators, developmentally relevant hand movements and sleep-wake states with anatomically derived facial actions can improve the accuracy of the scale. However, the initial validation of the BIIP was scored from videotape rather than at the bedside. This method of scoring is practical for research purposes, but is not feasible for clinical use of the scale.

1.3) Pain Assessment and Care for the Extremely Low Gestational Age Infant Focused Instrument (PACEFI)

The PACEFI scale was developed by Francis (2012) for her dissertation. It was constructed from the synactive theory of development model and literature review of behaviors related to acute pain. The PACEFI is a 20-item dichotomous (yes/no) procedural pain assessment instrument designed from observation of 16 ELGA infants ($24^{0/7}$ to $29^{6/7}$ weeks' gestational age infants) and 15 VLGA infants ($30^{0/7}$ to $33^{6/7}$ weeks' gestational age infants) during the first 14 days of postnatal age. Twenty behavioral indicators include eye squeeze, hands on face, crying, furrow, grimace, nasolabial bulge, lips pursed, stretch mouth, taut tongue, push away arms, push away legs, pull away arms, pull away legs, full body pull away, sit on air, fisting, mouthing, finger splay, pull extremities midline, and curling toes.

The PACEFI instrument was developed using the information from a preliminary study to ascertain the registered nurses perceptions of the most accurate pain cues in extremely and very low gestational age infants, a literature review, views of expert consultants, review of preexisting pain instruments, and the investigator's own experience. Due to zero variance (i.e., infants did not demonstrate behaviors), the excluded items for baseline, during invasive procedure, and recovery phase were 17, 6, and 2 items, respectively. When those items were excluded, the scale internal consistency estimated with KR-20 was .14 for baseline phase, .88 during invasive procedure, and .79 for recovery phase. For construct validity testing, the scores of ELGA infants and VLGA infants across three phases were compared using repeated measures ANOVA. The score of during had significantly higher scores than baseline and recovery score in both groups. Information about clinical utility and feasibility of PACEFI is not available in any published study.

2) Multi-dimensional scale

A multi-dimensional assessment perspective is to combine a variety of subjective and objective measurement approaches or to utilize composite measures that include a variety of self-report, physiologic, behavioral, and contextual factors within one instrument (Stevens, 1998). This type of scale is particular appropriate when the accepted proxy for self-report (a gold standard of pain) is not possible. However, Lake (2013) analyzed multi-dimensional scales used to assess pain in preterm and critically ill neonates and found that scales to assess every neonate across continuum of prematurity, severity of illness, and sedation remain elusive. Presently, there are ten scales that are categorized in this group including Neonatal Infant Pain Scale (NIPS), Scale for Use in Newborns (SUN), Premature Infant Pain Profile (PIPP), Pain Assessment in Neonates (PAIN), Bernese Pain Scale for Neonates (BPSN), COMFORTneo scale, COVERS, Faceless Acute Neonatal Pain Scale (FANS), Pain Assessment Scale for Preterm Infants (PASPI), and Neonatal Infant Acute Pain Assessment Scale (NIAPAS).

2.1) Neonatal Infant Pain Scale (NIPS)

The NIPS was developed by Lawrence et al. (1993) for assessment of pain in neonates by adapting the Children's Hospital of Eastern Ontario Pain Scale (CHEOP) (McGrath et al., 1985). It was developed from a survey results of 43 experienced neonatal nurses. They identified six indicators associated with varying levels of pain and distress in preterm and full term infants responding to painful procedures. The six indicators of NIPS scale are heavily weighted with behavioral parameters including five behavioral parameters (i.e., facial expression, cry, arms, legs, and state of arousal) and one physiological parameter (i.e., breathing patterns). Each indicator is scored as 0 or 1 with the exception of crying, which is scored zero to two, resulting in a total possible score of zero to seven (Lawrence et al., 1993).

In the initial validation and reliability evaluation, the NIPS was used to assess pain during 90 needle-insertion procedures (capillary, venous, or arterial punctures) with videotape recording in 38 infants (28 to 38 weeks gestational age) (Lawrence et al., 1993). To establish construct and concurrent validity, changes of scores over time and scores obtained from NIPS and visual analogue scale (VAS) were compared. The changes in NIPS scores over time were statistically significant with the main effect of time (p <.001) and the correlations

between the NIPS and the VAS at each minute of observation were .53 to .84. Inter-rater reliability of the NIPS was high, ranging from .92 to .97 across successive minutes of observation. Internal consistency was measured with Cronbach's alpha before (.95), during (.87), and after (.88) the procedures.

Comparisons of the NIPS and other tools were subsequently studied. The study comparing among NIPS, COMFORT, and Scale for Use in Newborns (SUN) evaluated four common procedures performing in NICU (intubation, intravenous catheter insertion, endotracheal suctioning, and diaper changes) indicated that using the NIPS might detect change but it might not accurately describe the patient state. In addition, parts of this scale involve observing behaviors that are visible only when the infants are totally exposed. The NIPS scale has only breathing pattern but does not include any other physiological parameters such as heart rate which are often early indicators of pain and/or distress in premature, sedated, or paralyzed infants (Lake, 2013). Most of preterm neonates in the NICU are too ill and need mechanical ventilation, therefore the utility of scale which has cry indicator or heavily weighted with only behavioral parameter might be limited.

2.2) Scale for Use in Newborns (SUN)

The SUN scale was initially developed by Blauer and Gerstmann (1998). This scale contains seven indicators (three behavioral and four physiologic indicators) including tone, face, movement, central nervous system state, breathing, heart rate changes, and mean blood pressure changes. Each scoring category had five possible outcomes with symmetric scale levels for scoring (zero to four) and scoring with the third gradation or score of two level was the normal or baseline value. Thus, SUN scale yields a baseline score of 14 and the possible maximum score was 28.

The scale was used to assess pain in 33 infants who had a gestational age ranging from 24 to 40 weeks during intubation, intravenous catheter insertion, endotracheal tube suctioning, and diaper change. To evaluate for consistency of scoring using a coefficient of variation, the SUN had smaller coefficient of variation value than those of NIPS (33% ± 8% and 188% ± 99%, respectively) indicating consistency in score more than NIPS. The very high coefficient of variation makes this less a desirable tool because for any one patient it is difficult to assess what the value of a single score represent. For intravenous catheter insertion

procedure, the SUN scale had a baseline score of 10 to 12, rose to a level of 17, and returned to a baseline of 12 to 13. The mean score of during procedure was significantly different from mean score of baseline and recovery phases ($p < .01$). Because Blauer was the only observer in the initial study, inter-rater reliability was not tested.

2.3) Premature Infant Pain Profile (PIPP)

The PIPP was initially developed by Steven and colleagues (1996) to assess acute pain in preterm and term neonates. The PIPP was developed and subjected to initial validation through an interactive process that involved seven steps: (1) identification of the indicators, (2) pilot testing of the indicators, (3) evaluation of the specificity and sensitivity of the indicators, (4) determination of the factor structure of the indicators, (5) developing indicator scales, (6) establishing internal consistency of the indicators, and (7) establishing construct validity of the measure. The seven indicators were brow bulge, eye squeeze, nasolabial furrow, heart rate change, decrease in oxygen saturation, baseline behavioral state, and gestational age. Each of seven indicators was evaluated on a four-point scale (zero, one, two, and three). For the maximum possible scoring from all seven indicators depended on the gestational age of the infant. The infants in the youngest gestational age category (< 28 weeks) and those in quiet sleep states received the highest score of "3" (on the zero to three scale) for these variables; thus could obtain a total score of "six" (six out of a possible 21) before any response to pain. For all age groups, a total score of six or less generally indicated minimal or no pain and scores greater than 12 reflected moderate to severe pain.

The alpha coefficients of the indicators calculated after deleting each individual indicator ranged from .76 for eye squeeze (.74 brow bulge, .72 nasolabial furrow, .66 oxygen saturation, and .64 heart rate) to .59 for behavioral state. The standardized item alpha for the six items was.71. Four different methods were used to establish construct validity including; (1) comparing a pain scores of preterm infants during handing (a non-pain situation) and heel stick (a pain situation), (2) comparing pain scores of two different groups of infants with the same gestational age during heel stick and handing, (3) comparing a pain scores during heel stick and sham heel stick, and (4) comparing a pain scores of two groups of term neonates during circumcision. The results were supported construct validity of the PIPP scale.

In terms of clinical utility, Schiller (1999) reported that PIPP was found to be clinically useful in all domains assessed but especially in relation to acceptability, with respect to its potential for future use in applicability for pain situations in the NICU. Using the PIPP required a set of instructions for users and individual practice sessions. On the contrary, Bellieni and colleagues (2007) conclude that the instructions for PIPP scoring are quite difficult to follow. The difficulty of thorough PIPP scoring may be since it involves complex monitoring and timing of all seven parameters (except gestational age and arousal state are not subjective). For example, duration of facial movements must be estimated a very precise time in seconds because a few seconds mean a different score (zero for 0% to 9% of time, one for 10% to 39% of time, two for 40% to 69% of time, three for ≥ 70% of time). Moreover, changes in heart rate and saturation cannot be measured without the help of a chronometer and a calculator.

With regard to its limitations, Stevens et al. (2014) revised the PIPP, named PIPP-R, to enhance validity and feasibility. The measure was revised in three points. First, decimal points to the oxygen saturation variable were rounded to the nearest whole number. For example, 2.50% to 4.90% was changed to 3% to 5%. Second, the percentage of time facial expression was changed to a 30-second block to the duration of time. For example, 10% to 39% of the time was changed to three to ten seconds. Finally, the instruction of behavioral state was made more specific and clarifications of definitions. For example, "active/awake, eyes open, facial movements" was changed to "active and awake." The PIPP-R scale was examined in infants of varying gestational ages, diagnoses, and procedures. Interestingly, the PIPP-R scores of non-painful events (diaper change, temperature taking, and repositioning) varied from 4.30 to 8.30 (Gibbins et al., 2014; Stevens et al., 2014). As mentioned in using of PIPP scale, a total score of six or less generally indicates minimal or no pain for all age groups. Therefore, further delineation of mild, moderate or severe pain based on PIPP-R scores is required and reevaluation of "cut-off" scores denoting no pain needs to occur.

In terms of feasibility of PIPP-R, 195 nurses indicated that additional training on the scoring of the PIPP-R and establishing a baseline behavioral state prior to handling the infant were required and often a step that had been overlooked (Gibbins et al., 2014). Although, the results reported that the majority of nurses completed the PIPP-R in less than one minute, its feasibility should be carefully interpreted because all nurses were clinicians

who had experience and were responsible for managing pain in high-risk infants. Even though the PIPP was revised, NICU nurses who used the PIPP-R suggested additional clarifications for how to use the measure and calculate the total score (Stevens et al., 2014). Gibbins and colleagues (2014) also suggested that further studies examining the clinical utility of the PIPP-R are required.

2.4) Pain Assessment in Neonates (PAIN)

The PAIN was developed by Hudson-Barr and colleagues (2002) by combing the components of the NIPS scale and the Crying, Requires oxygen for saturation, Increased vital signs, Expression, Sleepless (CRIES) scale that is a postoperative pain measurement developed by Krechel and Bildner (1995). Seven indicators of PAIN scale are facial expression, crying, breathing pattern, movement of extremities, state of alertness, oxygen saturation and heart rate. The PAIN scale was validated with a convenience sample from NICU (n = 106) and from the step-down unit (n = 90) with gestational ages of 26 to 47 weeks and most of the neonates were greater than 32 weeks gestational age (Hudson-Barr et al., 2002). Validation was assessed by bedside nurses observed the neonates for two minutes and then scored their responses on NIPS and PAIN scales. Inter-rater reliability was not assessed. Evidence of concurrent validity was supported by high correlation between pain scores on NIPS and PAIN scales (.93). It is not surprising that a high correlation was found due to sharing common items because four indicators of PAIN were obtained from NIPS.

2.5) Bernese Pain Scale for Neonates (BPSN)

The BPSN scale was developed by the nurses of the NICU of the University Hospital of Berne to assess acute pain in preterm and term neonates with or without ventilation on CPAP (Cignacco et al., 2004). The BPSN consists of nine indicators including changes in heart rate, breathing pattern, transcutaneous measured oxygen saturation, skin color, alertness, eye bulge with eye squeeze, duration of crying, time to clam, and posture. Each indicator was rated on a defined 4-point scale (0 to 3) with a total score of 0 to 27.

The BPSN scale was tested in 12 term and preterm neonates (27 to 41 weeks gestational age) for validity and reliability in four given situations (10 minutes after feeding or baseline, warming of the foot with a warm towel or touching, heel stick, and 15

minutes after heel stick or recovery) (Cignacco et al., 2004). Assessment of the situations was done by two nurses at the bedside. In addition, video sequences were made (n = 48 different situations), which were assessed by four different nurses. Thus, a total of 288 events analyses was performed. Construct validity was calculated using data of the video analyses. The inter-rater and intra-rater reliability for each of four situations was .86 to .97 and .98 to .99, respectively. The difference between the situations was highly significant (F = 41.27, p < .0001). The concurrent validity of the BPSN was supported by high correlations with the VAS and PIPP scales (.86 and .91, respectively). However, the BPSN was tested with a small sample and did not include critically ill neonates requiring mechanical ventilation (only four infants required continuous positive airway pressure); therefore this scale might not be appropriate for use in all preterm neonates in NICU.

2.6) COMFORTneo Scale

The original COMFORT tool is used to assess psychological distress of critically ill children under the age of 18 months in the Pediatric Intensive Care Unit (Ambuel, Hamlett, Marx, & Blumer, 1992). The COMFORT indicators include alertness, calmness, respiratory response, movement, mean arterial blood pressure, heart rate, muscle tone, and facial expression. The CONFORTneo scale was modified from the COMFORT scale to assess pain in neonates. It composes of alertness, calmness/agitation, respiratory response, crying, body movement, muscle tone and facial tension. Each indicator is evaluated on a one to five Likert scale, total scores range from six to 30, with higher scores indicating more pain because the rater will actually rate six indicators (Dijk et al., 2009). The respiratory response applies to ventilated neonates only, and crying to spontaneously breathing applies to neonates only (including those requiring continuous positive airway pressure).

The CONFORTneo scale was used in 286 neonates (24.60 to 42.60 week gestational age) (Dijk et al., 2009). In total 141 nurses and three researchers, inter-rater reliability ranged from .65 to .97. The concurrent validity of the CONFORTneo scale was supported by correlations with the Numeric Rating Scale score for pain (NRS-pain) of .51. The researcher claimed that the scale has adequate level of reliability even the alpha coefficients was quiet low. Using the COMFORTneo measuring acute pain in the NICU is questionable because the original COMFORT and COMFORTneo scale intended to be used as an

assessment tool for pain or distress during routine nursing care (Marx et al., 1994; Dijk et al., 2009), not to evaluate response to procedural pain.

2.7) COVERS

The COVERS scale, a measure which incorporates six physiological and behavioral indicators, was developed for clinical assessment of pain in newborns and infants regardless of their gestational ages and disease states. Two physiological indicators are changes in vital signs and oxygen requirement. The behavioral indicators include crying, expression, resting, and signaling distress. In order to establish construct and concurrent validity, the COVERS scales as well as three previously established scales (CRIES, PIPS, and NIPS) were used to assess pain during two procedures, a heel stick and diapering in 21 infants (27 to 40 weeks, mean 34.90 weeks) admitted to the NICU (Hand, Noble, Geiss, Wozniak, & Hall, 2010). Comparing the pain scores during pain and non-pain procedures demonstrated a significant difference between values supporting that the scale could discriminate pain from non-pain situation. Scores on the COVERS scale correlated well with scores on the PIPP (r = .84) and NIPS (r = .95) reflecting evidence of concurrent validity. Currently, the COVERS and Pain Assessment Tool (PAT) were compared to assess pain in a neonatal unit (O'Sullivan, Rowley, Ellis, Faasse, & Petrie, 2016). The internal consistency and intra-class correlation coefficients the COVERS during heel lance was .78 and .80, respectively. The concurrent validity of the COVERS was supported by high correlations with the PAT scale (.91). However, the samples of the study included both preterm and term infant with gestational age at birth ranged from 23.60 to 41.10 weeks (mean = 31.60, SD = 4.70).

The unique feature of the COVERS scale is that the criteria used for scoring are applicable to a wider range of infants. The visible crying as a behavioral response in COVERS scale is the limitation to assess paralyzed infants since they cannot perform behavioral responses such as crying, grimacing, or signaling distress. High-pitched crying is one of the behavioral responses to pain, but an intubated infant physically cannot make such a cry. The COVERS scale also addresses oxygen requirements from a new perspective. Rather than recording the infant's oxygen requirement, which is not always indicative of pain, it looks at a change in the need for oxygen.

2.8) Faceless Acute Neonatal Pain Scale (FANS)

The FANS was especially developed to assess acute pain in preterm and term neonates (Milesi et al., 2010). The FANS is a four-indicator pain measure that includes heart rate variation, acute discomfort, limb movements, and vocal expression. These four indicators were inspired by other validated scales including CRIES, NIPS, PIPP, BIIP, NFCS, and Douleur aiguë du Nouveau-né (DAN). The DAN scale is a three-indicator behavioral scale developed to rate acute pain in term and preterm neonates (Carbajal, Veerapen, Couderc, Jugie, & Ville, 2003). The initial validation and reliability of DAN has not been published in English. Each indicator is evaluated on a different weighting point scale: heart rate variation (zero, one, and two) acute discomfort (zero and one), limb movements (zero, one, two, three, and four), and vocal expression (zero, one, two, and three) for a possible score of zero to ten.

The FANS scale was assessed in 20 preterm neonates who were born at 32 (30 to 35) weeks gestational age (Milesi et al., 2010). Inter-rater agreement was .92 (.90 to .98); the internal consistency of the scale evaluated by Cronbach's alpha was .72. The researcher assessed the validity of the FANS by comparing the FANS results with those obtained using the reference scale (DAN). The researcher claimed that the intraclass correlation coefficient of .88 between the two scales indicating well correlated with the reference scale (DAN), as there are no objective markers of pain, and in particular no validated biological marker to evaluate transitory nociceptive stimulation in the newborn. The FANS is the first scale that permits the acute pain of non-intubated, preterm neonates to be scored when the face is not visible. However, the FANS scale is appropriate only in infants who are not intubated because the vocal expressions indicator needs to be scored.

2.9) Pain Assessment Scale for Preterm Infants (PASPI)

The PASPI was developed in Taiwan to measure procedural pain in full term and preterm infants over 27 weeks of gestational age (Liaw et al., 2012). The development process of the PASPI scale was composed of three steps including identifying pain indicators by neonatal experts, reviewing the indicators by content experts, and psychometric testing. The initial ten pain indicators was identified by a panel of six neonatal experts (two researchers in neonatal pain, two neonatologists and two neonatal nurses). Another ten neonatal clinicians (five nurses and five neonatologists) answered two questions about how clear and how effective

each PASPI item to establish content validity. The CVIs of those two questions of the original 10 PASPI items were .82 and .80, respectively. Moreover, clinical feasibility of the PASPI was tested by two trained nurses at bedside to measure pain with ten preterm infants across the four phases of heel stick procedures. The inter-rater reliabilities between the two observers for Phase I (three minutes baseline), Phase II (during the procedure), Phase III (three minutes after procedure) and Phase IV (the tenth minutes after procedure) were acceptable of .91, .85, .90, and .86, respectively (Liaw et al., 2012). Content validity of the 10-items of PASPI was evaluated to develop the final six-item PASPI. The six-item PASPI covers the transition between states, facial expression, heart rate, oxygen saturation, limb and body movement and, hand behavior. Of the original ten items, four items were deleted (gestational age, respiration, nurse consolation, and skin color change).

The six-item PASPI was tested by assessing 60 preterm infants for 240 observations comparing with the PIPP and VAS via videotape recording and direct observation. Internal consistency of the six-item PASPI was .84 for during procedure. The inter-rater reliability of the researcher and three research nurses for the six-item PASPI was ranged from .88 to .93. Scores on the six-item PASPI scale across the four phases correlated well with scores on the PIPP (.74 to .83) and VAS (.72 to .81) (Liaw et al., 2012) supporting concurrent validity of the scale.

The PASPI instrument provided clinicians' involvement in the development process to establish a clear, feasible and valid pain scale in the direct observation at bedside. However, Liaw et al. (2012) demonstrates six issues of the PASPI limitations which are very useful for further study. First, the gestational age of infants was limited to older than 27 weeks. Thus it is not known whether the PASPI is valid and reliable for assessing pain in infants with the gestational age less than 27 weeks. Second, although the PASPI measured infants' state changes during procedures, such changes might not be observed in infants with ELGA because the ability of those changes is influenced by their gestational age. Third, the sample for testing the feasibility was small and the PIPP and VAS were not tested in practice but coded from video recording. Fourth, the observational context in this study focused on heel stick procedures. Fifth, observers were only partially blind to the phase of the study, as they could easily see the heel stick occurring. Finally, further study is needed to measure concurrent

validity of the PASPI with the level of cortical activity and continuously modify PASPI to be a more sensitive pain measurement.

2.10) Neonatal Infant Acute Pain Assessment Scale (NIAPAS)

The development of NIAPAS was based on the results of national and international studies in NICU indicating that pain assessment of neonates was unsystematic and that few nurses used the pain assessment tools in clinical practice. Thus, the items relevant to NICU setting of previously available pain scales were determined by nurses. The NIAPAS consists of eight pain indicators (five behavioral: alertness, facial expressions, crying, muscle tension, and reacting to handling; and three physiological indicators: breathing, heart rate, and oxygen saturation) and includes the gestational age of neonates as a contextual factor (PÖlkki et al., 2014). The indicators are rated on a two, three, or four-point scale (zero and one; or zero, one, and two; or zero, one, two, and three) for a possible total score of 18.

The NIAPAS and NIPS scales were assessed 34 neonates (23 and 42 weeks' gestational age) who were undergoing 60 painful procedures. The internal consistency of NIAPAS was .72 and the coefficients for inter-rater and intra-rater reliability were high across three phases. The very good with I-CVI and S-CVI which were reviewed by five content experts reflecting evidence of content validity. The high correlations between two scales across three phase ranged from .75 to .87 reflecting evidence of concurrent validity. The NIAPAS score during the procedure increased significantly from the baseline and recovery phases indicating a good construct validity.

Summary of existing pain assessment tools for preterm neonates.

The interesting pain indicators including facial expressions, behavioral states, and heart rate from 13 existing pain assessment tools for preterm neonates were analyzed and summarized (see Tables 2.2-2.3). Facial expressions were called in variety terms such as face, expression, and facial tension. Facial expression and scoring of ten out of 13 scales (except NFCS, PACEFI, & FANS) were compared in Table 2.4. The change of each part of the face, the change of whole face, and timing of facial expressions are used as pain indicators. Those ten scales also consist of behavioral state as a pain indicator which was called in variety terms such as alertness, state of arousal, CNS state, and transition between states (see Table 2.5).

However, none of the scale describes facial expressions or behavioral states incorporating with infants' maturation.

Regarding heart rate, there were various criteria for scoring such as increased or decreased heart rate, the change of heart rate in beat per minute comparing with baseline, and heart rate variation in percent comparing with mean baseline (see Table 2.6). Complicated scoring of each score level such as calculating percentage heart rate increase makes pain assessment more difficult. Existing pain assessment tools mostly use similar or overlapping parameters, adding other indicators such as factors affecting pain reactivity and determining pain with the spectrum of gestational age need additional development. The major issues concerning existing pain assessment scales leave room for the improvement of clinical pain scales. Even though 13 pain assessment scales exit, a clinical pain scale that is appropriate for preterm neonates is not available.

Table 2.2

Uni-dimension pain assessment scales for preterm neonates

Pain scale (Number of indicators)	Country	Development method	Psychometric testing
1. NFCS (9 indicators)	USA.	Slow motion video coding 140 infants (38-42 weeks)	Construct validity Inter-rater reliability .88
2. BIIP (7 indicators)	Canada	Combined existing scales 92 infants (23-32 weeks)	Construct validity Internal consistency .82 Inter-rater reliability .80
3. PACEFI (20 indicators)	USA.	Theoretically derived 31 infants (24-$33^{6/7}$ weeks)	Construct validity Internal consistency .88

Table 2.3

Multi-dimension pain assessment scales for preterm neonates

Pain scale (Number of indicators)	Country	Development method	Psychometric testing
Multi-dimension			
1. NIPS (6 indicators)	Canada	Redefined previous scales 38 infants (32-40 weeks)	Concurrent validity Internal consistency .87 Inter-rater reliability .92
2. SUN (7 indicators)	USA.	N/A 33 infants (24-40 weeks)	Coefficient of variation
3. PIPP (6 indicators)	Canada	Development process 238 infants (28-40 weeks)	Construct validity Internal consistency .71
4. PAIN (7 indicators)	USA.	Combined existing scales 196 infants (26-47 weeks)	Concurrent validity
5. BPSN (9 indicators)	Switzerland	NICU nurses review 12 infants (27-41 weeks)	Concurrent validity Inter-rater reliability .86-.97 Intra-rater reliability .98-.99
6. COMFORTneo (7 indicators)	Netherland	Redefined previous scales 286 infants (24.6-42.6 weeks)	Concurrent validity Inter-rater reliability .65-.97
7. COVERS (6 indicators)	USA.	Redefined previous scales 21 infants (27-40 weeks)	Construct & concurrent validity Internal consistency .78 Inter-rater reliability .80
8. FANS (4 indicators)	France	Combined existing scales 53 infants (30-35 weeks)	Concurrent validity Internal consistency .72 Inter-rater reliability .92
9. PASPI (7 indicators)	Taiwan	Development process 60 infants (27.6-36.3 weeks)	Construct & concurrent validity Internal consistency .84 Inter-rater reliability .88-.93
10. NIAPAS (8 indicators)	Finland	NICU nurses review 34 infants (23-42 weeks)	Construct validity Internal consistency .72 Inter-rater reliability .99

Table 2.4

Facial expressions and scoring of existing pain assessment scales for preterm neonates

Facial expressions	Score zero	Score ≥1
NIPS (facial expression)	0= relaxed muscles	1= grimace
BIIP (face)	N/A	1= brow bulge 1= eye squeeze 1= nasolabial furrow 1= horizontal mouth stretch 1= taut tongue
COMFORTneo (facial tension)	N/A	1= facial muscles fully relaxed, relaxed open mouth 2= normal facial tension 3= intermittent eye squeeze and brow furrow 4= continuous eye squeeze and brow furrow 5= facial muscles contorted and grimacing
SUN (face)	0= totally relaxed, no tone or expression	1= reduced facial tone or expression 2= normal, neutral, no tension 3= increased tension, furrowed brow 4= contortion, grimace, vigorous cry
PIPP (brow bulge, eye squeeze, & nasolabial furrow)	brow bulge 0= none (< 3 seconds) eye squeeze 0= none (< 3 seconds) nasolabial furrow 0= none (< 3 seconds)	brow bulge 1= Minimal (3-10 seconds) 2= Moderate (11-20 seconds) 3= Maximal (> 20 seconds) eye squeeze (scoring same as brow bulge) nasolabial furrow (scoring same as brow bulge)

Table 2.4 (continue)

Facial expressions	Score zero	Score ≥1
PAIN (facial expression)	0= relaxed muscles	1= grimace
BPSN	0= relaxed	1-3 (eyebrow bulge with eye squeeze as one indicator but could not find original tool)
COVERS (expression)	0= none/facial muscles relaxed	1= grimace, min-mod brow bulge, eye squeeze, nasolabial furrow 2= grimace/grunt, mod-max row bulge eye squeeze, nasolabial furrow
PASPI (facial expression)	0= eyebrows and face relaxed	1= slight brow bulge and eye squeeze 2= moderate brow bulge and eye squeeze, nasolabial furrow and mouth stretch 3= severe brow bulge and eye squeeze, nasolabial furrow and mouth stretch
NIAPAS (facial expression)	0 = relaxed (relaxed face, natural expression)	1 = dissatisfied (knitted brows/dissatisfied expression, frown, grin) 2 = grimace (taut facial muscles, tense brows, cheeks and chin, grimaces)

Table 2.5

Sleep-wake states and scoring of existing pain assessment scales for preterm neonates

States	Score zero	Score ≥1
NIPS	0= sleeping/awake	1= fussy
(state of arousal)		
BIIP	0= deep sleep	1= active awake
(state)	0= active sleep	2= agitated/crying
	0= drowsy	
	0= quiet awake	
COMFORTneo	N/A	1= quiet sleep
(alertness)		2= active sleep
		3= quietly awake
		4= actively awake
		5= awake and hyper-alert
SUN	0= deeply asleep	1= drowsy, light sleep
(CNS state)		2= awake, quiet alert, calm
		3= anxious, fussy
		4= hyper-alert, panicked
PIPP	0= active and awake	1= quiet and awake
(behavioral state)		2= active and asleep
		3= quiet and asleep
PAIN	0= sleeping/awake	1= fussy
(state of arousal)		
BPSN	0= sleep or quiet awake,	1= drowsy
(alertness)		2= active awake
		3= highly aroused and agitated
COVERS	0= sleeping most of time	1= wakes at frequent intervals—fussy
(resting)		2= constantly awake (even when not disturbed)
PASPI (transition between states)	0= maintains the same state (asleep or awake or drowsy)	1= transit from sleep to drowsy or from drowsy to awake with more movement

Table 2.5 (continue)

States	Score zero	Score ≥1
	with face relaxed; no change	2= transit from sleep or drowsy to awake with fussy and much more movement 3= cannot go to sleep; remains crying or fussing; transit from sleep, drowsy or awake to loudly crying
NIAPAS (alertness)	0= calm/quiet (calm, quiet, asleep/awake)	1= restless (restless and flailing a little/at times, can be calmed down) 2= remarkably restless (restless and flailing nearly continuously, intermittent sleep)

Table 2.6

Heart rate and scoring of existing pain assessment scales for preterm neonates

Heart rate	Score zero	Score ≥1
BIIP (heart rate change)	(just notes) no change increase decrease	N/A
SUN (heart rate change)	0= depressions >15% below baseline	1= depressions to 15% below baseline 2= baseline 3= elevations to 15% above baseline 4= elevations >15% above baseline
PIPP (heart rate change)	0= 0-4 beats/minute	1= 5-14 beats/minute 2= 15-24 beats/minute 3= >24 beats/minute
PAIN (increased vital signs)	0= HR and BP no increase	1= HR or BP increased < 20% 2= HR or BP increased ≥ 20%
COVERS (vital signs)	0= HR &/or BP within normal limits for age or at baseline	1= HR &/or BP↑ < 20% of baseline 2= HR &/or BP↑> 20% of baseline
FANS (heart rate variation)	0= < 10%	1= > 10% 2= > 50%
PASPI (heart rate change)	0= normal and regular (baseline mean)	1= increased or decreased 12–16 beats from baseline mean 2= increased or decreased 17–24 beats from baseline mean 3= increased or decreased ≥ 25 beats from baseline mean

Table 2.6 (Continue)

Heart rate	Score zero	Score ≥1
NIAPAS (heart rate change)	0= normal (heart rate decreases/increases 0–5 beats from baseline)	1= slight change (heart rate decreases/increases 6–20 beats from baseline or heart rate is 170–189 beats/minute) 2= clear change (heart rate decreases/increases > 20 beats from baseline or heart rate ≥ 190 beats/minute)

2.3 Instrument Development Process

According to literature review in relation to development of the existing pain assessment scales, there is a variety of strategies for identifying pain indicators including bedside observation (Grunau & Craig, 1987), identification by NICU nurses (Lawrence et al., 1993), and combination of indicators from previous scales (Hudson-Barr et al., 2002). However, there is no specific guideline for development process of the pain assessment scales. Generally, several guidelines and approaches for instrument development process are available. For example, Mishel (1998) provides the process of instrument development from the perspective of enhancing control measurement error. While Waltz, Strickland, and Lenz (2005) suggest four essential steps in the instrument development design if a norm-referenced measure and eight stages in the design if criterion-referenced measure.

Burns and Grove (2005) provide eleven steps describing how to construct a scale.

1) Defining the concept is to clarify the definition which is achieved through concept analysis. According to Walker and Avant (2005), eight steps of concept analysis by include: (1) select a concept, (2) determine the aims or purposes of analysis, (3) identify all uses of the concept, (4) determine the defining attributes, (5) identify a model case, (6) identify borderline, relate, contrary, invented, and illegitimate cases, (7) identify antecedents and consequences, and (8) define empirical referents.

2) Designing the scale relates to number of items or indicators which can reflect the concept as fully as possible. The indicators previously can be retrieved if they have been shown empirically to be good indicators of the concept. Clarity and conciseness of the indicators should be considered in this step.

3) Seeking item review refers to revising the indicators by asking qualified individuals to review them.

4) Conducting preliminary items tryout should be performed on a limited number of subjects (15 to 30) representatives of the target population. The results of descriptive and exploratory statistical analyses as well as the comments from respondents are used to revise items.

5) Performing a field test requires administration of the scale to a large sample of subjects representative of the target population.

6) Conducting item analyses is performed to identify items and exam the extent of inter-correlation among items.

7) Selecting item with the highest coefficient for retention has to be performed with concern. This step is balancing between statistical approach and concern about achievement of the concept measurement.

8) Conducting validity studies requires the collection of additional data from a large sample. As part of this process, scale scores need to be correlated with scores on other variables proposed to be related to the concept being put into operation.

9) Evaluating the reliability of the scale can be performed on the data collected to evaluate validity.

10) Determination of norms requires administration of the scale to a large sample and as many diverse groups as possible. The alternative way to obtain the large sample is giving permission to others to use the scale with the condition that data from these studies is provided for compiling norms.

11) Publishing the results of the development of the scale is recommended.

More recently, DeVellis (2012) provided eight steps in developing measurement scales which are similar to those identified in other research texts (Polit & Hungler, 1999). Even though this instrument development process mostly emphasizes psychosocial measurement, Richmond and Wright (2006) used eight steps of scale development described by Devellis

(1991) as the core essential steps for developing a constipation risk assessment scale. They include: (1) clearly articulating what the phenomenon of interest is, (2) generating a pool of suitable items, (3) selecting a response format for those items, (4) having a panel of experts review the items pool, (5) considering inclusion of validation items, (6) testing a measure with a development sample, (7) evaluating the items, and (8) optimizing the scale length.

According to a set of specific guidelines for developing scale by Devellis (2012), *the first step* is to determine clearly what it is scale developers want to measure. Due to the fact that most phenomena cannot be observed directly, including pain in preterm neonates, a tentative theoretical model is needed to be specified to offer a guide to scale development. In addition, having a clear frame of reference that determines what level of specificity is appropriate, given the intended function of the scale. Concerning about the indicators that might cross over into the related construct should be done in this step. *The second step,* generating an item pool involves choosing all relevant items from a rich source that reflect the scale's purpose. Determining the format for measurement including selecting a response format, *the third step,* occurs simultaneously with the second step. *The fourth step,* a group of people who are knowledgeable in the content area review the item pool for maximizing the content validity. The questions for a panel of experts will relate to relevance, clarity, and the other ways to capture phenomenon. *The fifth step,* inclusion consideration of validation items involves at least two types of items including measures of social desirability and measures of relevant constructs. *The sixth step,* scale developers administer to an appropriate large and representative sample. *The seventh step,* evaluating the performance of the individual items provides relationships to true scores from correlations among items. *The final step,* the scale developers have to give some thought to the optimal scale length because a shorter scale places less of a burden on respondents while a longer scale tends to be more reliable. Since preterm neonates cannot verbalize their pain, pain measures for them must rely on physiological and behavioral observation by caregivers. All eight steps by Devellis (2012) provide a framework for development clinical pain assessment scale.

2.4 Psychometric Testing

Psychometric properties of biophysiologic measure is composed of reliability and validity which more often refer to precision and accuracy. Measurement validity is assessed with three different operations including content, criterion-related, and construct validity

(DeVellis, 2012; Polit & Beck, 2008). In accordance with DeVellis (2012), methods of reliability testing based on the analysis of variance including continuous versus dichotomous items and internal consistency. Methods of reliability testing based on correlations between scale scores including alternative-forms, split-haft, inter-rater agreement, temporal stability reliability.

2.4.1 Validity

With regard to the step of conducting validity study, this review focuses on only content and construct validity. As mentioned, no gold standard tools for clinical practice exist, thus criterion-related validity is not applicable.

1) Content validity

Content validity refers to an adequate and specific set of indicators that reflect a content domain (DeVellis, 2012). Due to being concerned about the adequacy of indicators on the scale, content validity is obtained from three sources including the literature, representativeness of relevant populations, and content experts (Burn & Grove, 2005). In order to establish content validity, judgments of clinical experts can be calculated and indicators that have high endorsement will remain. A numerical value reflecting the level of content-related validity evidence can be obtained by using the content validity index. Clinical experts should be people who work closely with preterm neonates and are able to identify indicators that they have observed to be related to pain. Experts have an intuitive grasp of whole situations and are able to accurately diagnose and respond without wasteful consideration of ineffective possibilities, therefore they are often consulted by other nurses and relied upon to be preceptors (McHugh & Lake, 2010). Selection of at least five experts in various fields, such as instrument development, clinical area, and another discipline relevant to the content area, is recommended (Burn & Grove, 2005). These experts might respond individually or in the format of a focus group but they must be given specific guidelines for judging the appropriateness, accuracy, representativeness of the specifications.

2) Construct validity

Construct validity is defined as the extent to which a test measures what it is intended to measure. Several approaches can be used for supporting the construct validity of

the measure of pain in preterm neonates. Hypothesis testing approach is one of activities undertaken to obtain evidence for construct validity. The researcher uses the theory or conceptual framework underlying the measure's design to states hypotheses regarding the behavior of individuals with varying scores on the measure (Waltz, Strickland, & Lenz, 2005). Comparison of scores in situations with and without pain has been used by many researchers to establish evidence to support construct validity. Development of many pain assessment instruments such as BIPP and PASPI also used comparing among phases of painful procedures to establish construct validity (Holsti & Grunau, 2007; Liaw et al., 2012).

Examining evidence of convergent validity is the other way to test the construct validity. As stated by Burns and Grove (2005), the newly developed instrument and the existing instrument are administered to a sample concurrently to determine how closely they measure the same construct. The highly positive correlation is expected for evidence of validity of each instrument. DeVellis (2012) also mentioned about convergent validity that it is evidence of similarity between measures of theoretically related constructs.

Systematic errors that lie at the very heart of validity are predictable errors, occurring in one direction only, constant and biased. In this study the most important systematic error is improper indicators for pain in preterm neonates. Choosing the indicator that could not provide a picture as possible of pain is the systematic error. For example, some indicators reflect the other concept such as stress rather than the specific construct of pain. If an indicator contains error, the resultant correlations will be an underestimate of the actual relationship between the constructs. In addition, improper calibration of monitor and timer are also the systematic errors that should be concerned. Therefore, to minimize the systematic errors and maximize the validity of the scale being developed, researcher should clearly identify pain indicators from several sources such as literature review and clinical observation.

2.4.2 Reliability

Scale reliability is the proportion of variance attributable to the true score of the latent variable (DeVellis, 2012). There are several methods for computing reliability. With regard to the step of evaluating the reliability of the scale, this review focuses on inter-rater reliability (equivalence) and internal consistency (homogeneity). Since pain is a state that can vary rapidly, there has been little point in estimating association of scores over time. The

stability testing makes no sense in this situation due to the stability requires an assumption that the factors to be measured remain the same at the two testing times and that any change in the values or score is a consequence of random error.

1) Internal consistency

Internal consistency refers to the homogeneity of items within a scale (DeVellis, 2003). It is assessed by analyzing simultaneous response to multiple indicator items. A scale cannot be homogeneous unless all of its items are interrelated. The greater the dispersion of the scores, the larger the variance; the more homogeneous of the scores, the smaller the variance. The correlation coefficient give information about the degree of association between two sets of data. High internal consistency implies that all the items within a scale are interrelated and measure the same characteristics or share a common cause.

Because all measurement techniques contain some random error, reliability exists in degrees and is usually expressed as a form of correlation coefficient with one indicating perfect reliability. In this study, random errors are ambiguous instruction of scoring and fatigue of the scale users. The amount of random error is inversely related to the degree of reliability of the measuring instrument. Thus, alpha is that it equals one minus error variance (DeVellis,, 2012). Alpha was developed by Lee Cronbach in 1951 to provide a measure of the internal consistency of a test or scale; it is expressed as a number between 0 and 1. When reporting the internal consistency of scales, the usual statistics is Cronbach's alpha coefficient. The generally agreed upon lower limit for Cronbach's alpha is .70 (DeVellis, 2003). To evaluate whether each item was contributing properly to the scale, the item-to-total correlations and inter-item correlations should be computed. Items are often discarded (or revised) if the items-total correlation coefficient is less than .50 or if an inter-item coefficient is less than .30 (Polit, 2010).

Streiner (2003b) reviewed the principles of internal consistency and discussed four good points about internal consistency. First, alpha is not a fixed property of a scale. Because the reliability depends as much on the sample being tested as on the test, it is a characteristic of the test scores, not of the test itself. Second, the high value of alpha means that the high correlations among items of a scale, but not always high internal consistency because it depends on the length of the scale and multi-dimension of a scale. Third, the high level of alpha doesn't mean only the homogeneity of the items because it also means the homogeneity

of what is being assessed. It can also reflect redundancy than to homogeneity. Finally, theoretically alpha should be a number of between 0 and 1, but negative alpha could exit in case of the items are tapping a variety of different constructs or in case of the scoring of the revered items.

2) Inter-rater reliability

Inter-rater reliability refers to the degree of agreement among raters and the nature of that agreement can vary depending on the goals of the researcher and the approach taken to estimate inter-rater reliability (DeVellis, 2012). Inter-rater reliability or inter-rater agreement is assessed by comparing of multiple observers to determine the degree of agreement in the scoring of the similar instrument by the observers must be of equal status. If two data collectors are observing the same event and recording their observations on a carefully designed data collection instrument, the measurement would be reliable if the recordings from the two data collectors are comparable (Burn & Grove, 2005). High inter-rater reliability is especially important in observational studies.

Inter-rater reliability can be determined in a number of ways, but the two most common reliability indices are percentage agreement and the intraclass correlation coefficient. The intraclass correlation coefficient is an attempt to overcome some of the limitations of the classic correlation coefficients (Bruton, Conway, & Holgate, 2000). It is a single index calculated using variance estimates obtained through the partitioning of total variance into between and within subject variance (known as one way analysis of variance [ANOVA]). It thus reflects both degree of consistency and agreement among ratings. Theoretically, the intraclass correlation coefficient can range from 0 (no agreement) to 1.0 (perfect agreement) with values closer to one representing the higher reliability. There is no absolute value below which inter-rater reliability is unacceptable. Chinn (1991) recommends that any measure should have an intraclass correlation coefficient of at least .60 to be useful. Thus, a value larger than .80 is regarded as satisfactory and found for widely used scales as a general rule (Carmines & Zeller, 1979), conversely, any value below .80 should generate serious concern (Burn & Grove, 2009) .

2.5 Clinical Utility Evaluation

The representative of pain in the cerebral cortex of humans has unique featured and may reflect adaptive developments allowing for language and the social complexities of human lifestyles (Craig, Korol, & Pillai, 2002). In particular vulnerable preterm neonates who have a minimal capacity to react or escape pain totally depend on judgments of the caregiver. Therefore, pain management requires not only of a good psychometric properties of assessment scale, but also how observers or nurses can interpret the infant's reactions and clinical decision to choose appropriate intervention. Regarding the sociocommunication model of pain, the process of pain whereby caregivers arrive at judgments of infant's needs and make decision concerning interventions (Craig et al., 2002). This model composes of pain experience and expression of infants, assessment of caregivers, and action dispositions. If this process is completed, the preterm neonates will received effectiveness of pain assessment and relieving interventions. However, there are still report underestimate and undertreat of pain in this vulnerable populations. The least attention being paid to clinical utility might be causes of this problem.

Clinical utility in the title of academic papers has increased tremendously over the past two decades (Smart, 2006). Within the context of clinical practice or NICU setting, clinical utility of pain measurement instrument must evolve toward high levels of acceptability and convenience for health care professions who use it. Nevertheless, the characteristics of clinical utility are not consistently defined among measurement experts or those researchers who have attempted to establish this measurement property for a particular instrument. In 1999, Schiller conducted her doctoral dissertation by evaluate the clinical utility of the PIPP and the CRIES through questionnaire responses provided by a group of NICU nurses. The clinical utility was evaluated by the questionnaire based on the definition of clinical utility by Law and Letts (1989). Questions include (1) time (e.g., training time, time for pain assessment, scoring time); (2) cost (e.g., time, training and personnel costs for clarification, and assessment time, as well as practical costs such a printing the measures); (3) instructions (e.g., written and verbal directions); (4) acceptability (e.g., perceived usefulness of the measure); and (5) format (overall layout of the measure and scoring method). A Cronbach's alpha of the questionnaire including nine items which derived from those five concepts of clinical utility conceptual model (Schiller, 1999) was only .0385. Cronbach's alphas of each concept which measured multiple items range

from .03 to.73. Currently, Gibbins et al. (2014) measured feasibility of PIPP-R with five-item questionnaire, but they referred to the study of Schiller (1999) leading even more confusions.

The literature review by Stevens and Gibbins (2002) summarized the results of clinical utility testing in five infant pain measures including fuller infant pain assessment tool, PIPP, CRIES, SUN, and COMFORT scales with difference properties of clinical utility evaluated. They also concluded that there is no consistent conceptualization of this measurement construct. Concerning the lack of an agreed formal definition or conceptualization, Smart (2006) conducted a review on clinical utility in common usage and summary of clinical utility dimensions. Clinical utility can be conceptualized as a multi-dimensional judgment about the usefulness, benefits, and drawbacks of intervention. The four dimensions of clinical utility can be measured from practitioners' perspectives about an innovation (or a new scale in this circumstance) for their working practice including appropriateness, accessibility, practicability, and acceptability. Questions about a new scale being appropriate encompass crucial information its effectiveness and relevance. Under the accessible component, formal evidence of an intervention's resource implications is likely to be important in a practitioner's judgment about its clinical utility. Questions such as the availability of a new scale or managing relationships with suppliers' relation to organizational contexts as well as appropriate and available training or technical support. The component of practicability is focused on working practice, particularly the relationship between a new scale and the practitioners' specific needs and capabilities. Questions relate to the functionality of a new scale and also its suitability in a particular working context. Finally, an acceptable component considers whether practitioners have moral objections to an innovation because a practitioners' willingness may be affected by their moral concerns. The clearly definitions and questions guidance of four components in this model might be the optimal way in which to establish clinical utility for a clinical pain assessment scale.

2.6 Conceptual Framework

A developmental neurobiological approach was used to guide construction of an assessment scale for preterm neonates in NICU. It explains the pain pathway, from injury site to behavioral and physiological consequences, and pain reactivity of preterm neonates in a quantitative manner (Anand & Scalzo, 2000; Fitzgerald & Walker, 2009). The anatomical and functional requirements for nociceptive pathways are established by 24 weeks' gestational age

(Hall & Anand, 2005) and develop continuously through their postnatal age. Immediately after a needle invades through the skin, noxious stimulus is converted to electrical activity and transmitted into the CNS through unmyelinated C-fiber. The lack of myelination and lower conduction velocities contribute to low speed of CNS processing. Therefore, the latency of pain response in preterm neonates varies depending on their neurobiology of axonal and synaptic development. The perception of pain occurs when action potential reaches to the thalamus and cortex. Facial expressions, arousal, and increased heart rate are reflexes mediated at the level of spinal cord through supraspinal area (Fitzgerald, 2005; Slater et al., 2007). These behavioral and physiological responses of preterm neonates could be derived as pain indicators of the scale.

The perception and meaning of pain of preterm neonates are complex and not determined by structural and functional maturation alone, but they are influenced by multiple factors. Gestational age impacts on the ability of facial expressions and transition of sleep-wake states (Sellam et al., 2011; Valeri & Linhares, 2012). The length of NICU stay, the number of invasive procedure exposures since after birth, and the mode of respiratory support influence pain reactivity of preterm neonates (Cruz et al., 2015; Grunau et al., 2001; Sellam et al., 2011). Therefore, these three factors affecting pain reactivity need to be considered and included in the pain scale for preterm neonates.

With respect to clinical practice and research, the development of a pain scale for preterm neonates consists of three phases including the construction of initial scale, psychometric testing and clinical utility evaluation. The construction of the initial scale followed the steps of constructing a scale by DeVellis (2012) and the evaluation of clinical utility of the initial scale used the multi-dimensional model of clinical utility as stated by Smart (2006).