University of Alberta

Instruments for Scoring Pain, Non-Pain Related Distress, Analgesia, and Sedation in Pediatric Mechanically Ventilated Patients and their Efficacy and Effectiveness in Practice: A Systematic Review

by

Tamara Lynn Dorfman

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Master of Nursing

Faculty of Nursing

Fall 2012

Edmonton, Alberta

Dedication

This thesis is dedicated to my parents.

~For their endless love, support, and encouragement~

Abstract

Objective pain and sedation assessment scales to standardize measures of distress in mechanically ventilated pediatric patients are increasingly available, but few have been evaluated for efficacy and effectiveness in this population. The purpose of paper one was to identify and evaluate available instruments appropriate for measuring physiological and behavioural cues of pain, non-pain related distress, and adequacy of sedation and analgesia in mechanically ventilated PICU or NICU patients. Twenty-eight articles were included in this systematic review, identifying 17 instruments. Three, additional, articles were found examining the efficacy and effectiveness of two of these scales in practice. In paper two, para-clinical tests used in the assessment of sedation (i.e. Bispectral Index Scale, auditory-evoked potential index, and skin conductance) were compared with three of the instruments for measuring sedation from paper one. Nine articles were identified in this systematic review and a correlation was found between the two methods for assessing sedation.

Acknowledgements

There are a number of people without whom this thesis might not have been written, and to whom I am greatly indebted.

First, I would like to thank Dr. Gwen Rempel, who was both my supervisor for this thesis and my advisor for the Master's program. Dr. Rempel worked tirelessly to ensure I completed my thesis and was instrumental in the writing of both the thesis and the thesis proposal. She fulfilled several roles in the completion of this thesis including supervisor, second reviewer, editor, and cheerleader. I am extremely glad to have had her as both a supervisor and advisor, and will never forget all that she has done for me.

I would like to recognize the contributions of Dr. Christine Newburn-Cook to this thesis. Dr. Newburn-Cook was a co-supervisor on my thesis, but passed away suddenly in the summer of 2011. She was instrumental in the writing of my proposal for this thesis and made tremendous contributions to both the theory and the methodology behind this thesis. She is greatly missed and I know she would be proud of what was accomplished in writing this thesis.

Thank you to Dr. Shannon Scott and Dr. Lisa Hartling, for their help with the theory and methodology behind my thesis. They were both very valuable committee members.

Thank you to Elizabeth Sumamo for her role as a second reviewer on this thesis. Elizabeth stepped in as second reviewer in April of 2012. Without her hard

work and expertise in data abstraction I would not have been able to successfully complete this knowledge synthesis project.

Thank you to Sana Ishaque for her role as a second reviewer on this thesis. Sana was the primary second reviewer during the first phase of screening for this review. Her hard work in screening the initial 2275 articles was essential to the completion of this review.

Thank you to Louisa Fricker for all her help with the formatting of my thesis and instruction in the various computer programs used in the completion of my thesis. Without, Louisa I would not be as adept at REDCap and Microsoft word as I am today.

Thank you to Andrew Lejeune for helping edit my thesis. Your love of correcting my grammar definitely came in handy.

To my friends, thank you for your understanding and encouragement during my many moments of crisis. Your support and encouragement never wavered. Your friendship means the world to me. I cannot list all your names here, but your sacrifices are not forgotten.

Lastly, thank you to my parents, who are my greatest supporters and have made several sacrifices to allow me accomplish my goals. I would not be where I am today if it were not for your love and encouragement.

Pain and non-pain related distress is experienced every day by children cared for in Pediatric and Neonatal Intensive Care. I hope that the findings of this

review will help improve the management of pain and non-pain related distress and the experiences of the children cared for in these setting,

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Pain and Non-Pain Related Distress in Pediatric Critical Care

Pain is an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage. It is a subjective experience and differs among individuals. Characteristics of pain include intensity, location, sensory qualities, cognitive appraisal, and affective reactions (von Baeyer, 2009). Pain, itself, may lead to negative emotions such as anxiety, fear, depression, anger, distress, dysphoria, unhappiness, frustration, or to efforts to cope with fear and pain (McGrath et al., 2008; von Baeyer & Spagrud, 2007). Researchers tend to explore distress in children as related to their pain responses. However, the same negative emotions can exist without pain. These emotions occur when the child is exposed to non-pain related external stimuli which elicit fear, anxiety, anger, frustration, depression, distress, dysphoria, or unhappiness (von Baeyer & Spagrud, 2007). Behaviours observed in both pain and non-pain related distress are similar. These behaviours include: crying, whining, physical tension, clinging, restlessness, seeking or avoiding touch, withdrawal and resistance (McGrath et al., 2008; von Baeyer & Spagrud, 2007).

In the Pediatric Intensive Care Unit (PICU) and Neonatal Intensive Care Unit (NICU), patients are exposed to both aversive internal and external stimuli which may lead to both pain and non-pain related distress. The bedside of a ventilated child in the PICU and NICU is overwhelming and variable with bright lights, multiple alarms, frequently changing caregivers, several types of mechanical

equipment, medical emergencies, inconsistent sleep/wake cycles, and separation from family and the familiar. In fact, PICU noise levels have been shown to be eight times higher than suggested by the American Environmental Protection Agency and the World Health Organization (Milette & Carnevale, 2003). Furthermore, children on critical care units (PICU and NICU) undergo five times more painful procedures than children on general pediatric wards (median of 10 per day in PICUs and NICUs; median of 2 per day on pediatric medical and surgical units) (Stevens et al., 2011). Therefore, it is not surprising that the percentage of children prescribed medication to treat pain and non-pain-related distress is higher in the PICU than on general pediatric units due to the complexity and invasiveness of the treatments they require in the PICU (Rennick, Johnston, Dougherty, Platt, & Ritchie, 2002). Consequently, it is increasingly important that these children receive adequate analgesia, to relieve the pain associated with invasive procedures, and sedation to minimize non-pain related distress they may experience as a result of the life-saving treatments they may require during their stay in the PICU or NICU.

Use of Mechanical Ventilation in Pediatric Critical Care

One life-saving treatment that can cause both pain and non-pain related distress in pediatric patients is mechanical ventilation. Mechanical ventilation is a central treatment modality in PICUs and NICUs and its use is increasing. PICUs worldwide have reported 20% to 64% of the patients they admit require ventilation (Farias et al., 2004; Khemani, Markovitz, & Curley, 2009). Indications for mechanical ventilation of children admitted to critical care include: acute

respiratory failure, acute pulmonary disease, pneumonia, acute respiratory disease syndrome, bronchiolitis, aspiration, cyanotic heart disease, congestive heart failure, chronic lung disease (i.e., chronic obstructive lung disease or asthma), upper airway obstruction (i.e., postoperatively, or due to sepsis/ trauma), reactive airway disease, spinal instability, neuromuscular disease, abdominal wounds, support on extracorporeal membrane oxygenation, bone marrow/lung transplantation, cerebral hypertension, and altered mental state (Farias et al., 2004; Khemani et al., 2009). The most common indications for mechanical ventilation are acute respiratory failure/ acute lung injury, ranging from 26% to 72%, and congenital heart disease, 22% (Farias et al., 2004; Khemani et. al., 2009). The numbers of patients requiring mechanical ventilation due to congenital heart disease may have been under represented in the Khemani study as not all the diagnoses of patients with non-acute lung injuries were available to the researchers; thus, this percentage could be significantly higher (Khemani et al., 2009).

Reasons for variations in rates of mechanical ventilation depend not only on disease state but also PICU or NICU size, location, season/time of year, patient population served, political admissions and discharges of PICUs and NICUs, and severity of illness (Farias et al., 2004; Khemani et al., 2009). For instance, a United States study of 16 PICUs found that mechanical ventilation was required by 30% of patients admitted to the PICU. There was, however, significant centre variability, with ventilation rates varying between 20 and 64% (Khemani et al., 2009).

Ventilation rates also vary across countries. A study of six PICUs in Mexico and Ecuador indicated that 64% of their patients required mechanical ventilation; however, a study of PICUs in Spain reported a similar value to the United States study, of 32% of patients, admitted to the PICU requiring mechanical ventilation (Farias et al., 2004). Nevertheless, a study examining 36 volunteer PICUs in seven countries (Argentina (n=10); Costa Rica (n=1); United Kingdom (n=2); Greece (n=2); Panama (n= 1); Spain (n=18); and the United States (n=2) found that 35% of infants and children admitted to the PICUs required mechanical ventilation for more than 12 hours (Farias et al., 2004). However, the study missed the respiratory syncytial virus season for most of the participating PICUs; therefore, it is possible the percentage of patients receiving mechanical ventilation may have been higher (Farias et al., 2004). Despite variation in reported ventilation rates in PICUs and NICUs worldwide, the prevalence of mechanical ventilation in these settings necessitates that clinicians and researchers review the adequacy and effectiveness of current available methods for evaluating and addressing both pain and non-pain related distress experienced by mechanically ventilated pediatric patients.

Statement of the Problem

The "gold standard" for assessment of pain and non-pain related distress in patients is the person's own self-report (Ambuel, Hamlett, Marx, & Blumer, 1992; Ista, van Dijk, Tibboel, de Hoog, & Aneja, 2005; Marx et al., 1994; von Baeyer & Spagrud, 2007). Although this "gold standard" may be achievable in adults and older children, mechanically ventilated patients are sedated and often unable to

speak; making self-report difficult. Self-report in young children is more complicated as these children may lack the verbal and cognitive ability to effectively articulate/verbalize their pain or emotions related to their pain, and other non-pain related negative emotions. This problem is further complicated in certain groups of pediatric patients. For example, infants born with congenital heart disease often lack the physiological proficiency to exhibit nonverbal signs of pain or non-pain related distress such as tachycardia or hypertension. Furthermore, during their postoperative stay in the PICU or NICU, these infants are more likely to experience conduction abnormalities, require use of pacemakers and cardiovascular medications, and experience catecholamine abnormalities which may alter the physiologic responses that caregivers commonly look for as cues to intervene (Connolly, McClowry, Hayman, Mahony, & Artman, 2004). Therefore, self-report and observational approaches are difficult, if not impossible, in mechanically ventilated children due to age, intubation, sedation, physiology, and the potential inability to manifest some of the predetermined non-verbal signs which caregivers may recognize in assessing pain and non-pain related distress.

Significance of Study

The objective of analgesia and sedation in the PICU and NICU is to minimize pain and non-pain related distress in the child and to optimize the delivery of care and the child's recovery. Functions of analgesia and sedation include: 1) facilitation of mechanical ventilation; 2) induction of sleep and control of

agitation; 3) induction of amnesia during paralysis and painful procedures; and, 4) decrease of cellular metabolism (Bavdekar, Mahajan, & Chandu, 1999).

Unrelieved pain can have adverse physical and psychological consequences leading to a stress response which includes tachycardia, hypercoagulability, immunosuppression, and a persistent catabolic state (Playfor et al., 2006). Pain can lead to pulmonary complications in postoperative patients due to decreased movement of the chest wall and diaphragm (Playfor et al., 2006). Conversely, over-sedation, to control non-pain related distress, can lead to prolonged mechanical ventilation, ventilator-associated pneumonias, lung injury, or neuromuscular disorders. Conversely, under-sedation can lead to interference with effective mechanical ventilation, myocardial and cerebral ischemia, and dangerous outcomes such as self-extubation or the removal of other mechanical devices including intravenous lines or chest tubes (Jin et al., 2007).Therefore, well titrated analgesia and sedation are essential components of intensive care.

Inconsistencies in Clinical Judgement

When the "gold standard" for the assessment of pain and non-pain distress is not available, the clinical judgement of the attending nurse or physician is the next logical choice. This measure, however, can result in many interpretations and lead to disagreements and variation within the team of nurses and clinicians. Disagreements -- about the patient's actual condition, what level of analgesia and sedation is optimal, and how this optimal level should be achieved – can result in significant fluctuations in the administration and discontinuation of analgesia and sedation (Carnevale & Ducharme, 1997). This can predispose the patient to the

development of adverse reactions and potentially result in over- or under-sedation. Other factors such as length of ventilation, length of stay, and long-term psychological and neurodevelopmental factors may also be affected by fluctuations in analgesia and sedation (Carnevale & Ducharme, 1997).

To address problems arising from the subjective assessments of pain and non-pain distress, assessment scales have been developed to standardize or objectively measure the effectiveness of analgesia and sedation in treating pain and non-pain related distress in mechanically ventilated and non-verbal patients (Marx et al., 1994). A ventilated patient's response to pain and non-pain related distress consists of two elements: 1) behavioural (e.g., crying, avoidance, agitation, and grimacing); and, 2) physiological (e.g., increases in heart rate, blood pressure, muscle tension, and hormonal response) (Ambuel et al., 1992). Systematic assessment scales for children have been developed with these elements as their foundation (Ambuel et al., 1992; Carnevale & Razack, 2002; Ista et al., 2005; Merkel, Voepel-Lewis, Shayevitz, & Malviya, 1997; van Dijk et al., 2000). Theoretically, these scales provide a more consistent measure of the adequacy of analgesia and sedation in controlling patient's pain and non-pain related distress than do a nurse or physician descriptive analog or visual analog scale (Marx et al., 1994). These scales allow for fewer discrepancies between individual assessments of pain and non-pain related distress and the patient's response to the analgesia and sedation.

Study Objectives

Objective assessment scales to standardize evaluation of pain and non-pain distress in mechanically ventilated, non-verbal pediatric patients are increasingly available but few have been identified and evaluated for efficacy/effectiveness in the subpopulation of PICU and NICU mechanically ventilated patients. The objectives of this systematic review were to: 1) identify available scales appropriate for measuring physiological and behavioural cues of pain, non-pain related distress, and adequacy of analgesia and sedation in PICU and NICU patients who are invasively mechanically ventilated; 2) describe the instruments in terms of how they were developed, the elements of behavioural and physiological cues of pain, non-pain related distress, analgesia, and sedation they assess, and the results of any validity or reliability testing completed on these scales; and, 3) determine the efficacy/effectiveness of these instruments on patient care outcomes, including total use of analgesics and sedatives, fluctuations in analgesia and sedation between nursing shifts, length of ventilation and PICU/NICU stay, and any adverse withdrawal effects.

Definitions of Terms

Systematic assessment of pain and non-pain related distress is an objective standardized measure for assessing behavioural and physiological cues of pain, non-pain related distress, sedation, and/or analgesia. **Pain** is an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage. Pain may lead to negative emotions such as anxiety, fear, depression, anger, distress, dysphoria,

unhappiness, frustration; or efforts to cope with fear and pain (McGrath et al., 2008; von Baeyer & Spagrud, 2007). Pain related behaviours include crying, whining, physical tension, clinging, restlessness, seeking or avoiding touch, withdrawal and resistance (McGrath et al., 2008; von Baeyer & Spagrud, 2007). **Distress** is an individual's response to aversive intrinsic and extrinsic stimuli. It may include discomfort, anxiety, or fear, and can exist with or without the presence of pain (Ambuel et al., 1992). **Non-pain related distress** is an organism's response to non-pain related aversive stimuli which elicit negative emotions and behaviours. These aversive stimuli elicit emotions such as fear, anxiety, anger, frustration, depression, distress, dysphoria, or unhappiness and behaviours such as crying, whining, physical tension, clinging, restlessness, seeking or avoiding touch, withdrawal and resistance (McGrath et al., 2008; von Baeyer & Spagrud, 2007).

An **analgesic** is medication used to relieve pain. **Analgesia** is the relief of pain. A **sedative** is a substance that induces sedation by reducing irritability or excitement. **Sedation** is the reduction of irritability or agitation by administration of sedative drugs, and generally to facilitate a medical procedure or diagnostic procedure.

For the purposes of this review, a **pediatric patient** is defined as any patient under the age of 18 years who requires care in a PICU or NICU setting. An **invasively mechanically ventilated patient** is any patient who requires assistance for ventilation via a nasally or orally inserted endotracheal tube and a ventilator.

Outcomes of interest in relation to scales that measure pain and sedation are as follows: **Total usage of sedatives** includes any medication that is used for the purpose of sedating the patient (i.e., continuous infusions, intermittent and as needed (*prn*) or break through doses). **Total usage of analgesics** includes any medication that is used for the purpose of treating the patient's pain (i.e., continuous infusions, intermittent and *prn* or break through doses). **Nursing shifts in PICU or NICU** will be defined as a period of time, typically eight to 12 hours, where the majority of care is provided by one nurse. When a new nurse comes on to provide the majority of care for a patient, a new shift starts. **Fluctuations in analgesia and sedation between nursing shifts** will be defined as the changes or variations in analgesics and sedatives given between nursing shifts. These fluctuations can include the use of *prn* medications or changes in infusion rates of continuous morphine or midazolam. **Length of mechanical ventilation** is the time from intubation to the time of extubation, measured in hours. Types of mechanical ventilation will include endotracheal intubation (both nasally and orally). **Length of stay in PICU/ NICU** will be defined as the time from which the child is admitted to PICU/NICU to the time of discharge from PICU/NICU or transfer to another unit. **Adverse withdrawal effects** are any changes in a patient that is unwanted or undesirable following the removal or weaning of the analgesic or sedative medications the patient is receiving.

Description of Papers

Objectives of Paper 1

The objectives of paper one are to :1) identify available instruments appropriate for measuring physiological and behavioural cues of pain, non-pain related distress, and adequacy of analgesia and sedation in PICU and NICU patients who are invasively mechanically ventilated; 2) describe the instruments in terms of how they were developed, 3) the physiological and behavioral variables they assess, 4) evaluate the instruments in terms of their psychometric properties, and 5) determine the efficacy/effectiveness of these instruments on patient care outcomes, including total use of analgesics and sedatives, fluctuations in analgesia and sedation between nursing shifts, length of ventilation and PICU/NICU stay, and any adverse withdrawal effects.

Objective of Paper 2

The objective of paper two is to compare two methods for assessing sedation: validated systematic observational scales and para-clinical tests. Para-clinical tests are investigations that assess the underlying biochemical and morphological clinical manifestations of sedation (e.g. Bispectral Index Scale).

References

Ambuel, B., Hamlett, K. W., Marx, C. M., & Blumer, J. L. (1992). Assessing distress in pediatric intensive care environments: The COMFORT scale. *Journal of Pediatric Psychology, 17*(1), 95-109. doi: 10.1093/jpepsy/17.1.95

Bavdekar, S. B., Mahajan, M. D., & Chandu, K. V. (1999). Analgesia and sedation in paediatric intensive care unit. *Journal of Postgraduate Medicine, 45*(3), 95-102. Retrieved from: http://www.jpgmonline.com/text.asp?1999/45/3/95/338

Carnevale, F. A., & Ducharme, C. (1997). Adverse reactions to the withdrawal of opioids and benzodiazepines in paediatric intensive care. *Intensive & Critical Care Nursing, 13*(4), 181-188. doi: 10.1016/S0964-3397(97)80012-2

Carnevale, F. A., & Razack, S. (2002). An item analysis of the COMFORT scale in a pediatric intensive care unit. *Pediatric Critical Care Medicine, 3*(2), 177-180. Retrieved from: http://ovidsp.tx.ovid.com/sp-3.4.0b/ovidweb.cgi?&S=FGPLFPCOGEDDMDKFNCCLLGIBEIMCAA00

Connolly, D., McClowry, S., Hayman, L., Mahony, L., & Artman, M. (2004). Posttraumatic stress disorder in children after cardiac surgery. *Journal of Pediatrics, 144*(4), 480-484. doi: 10.1016/j.jpeds.2003.12.048

Farias, ,J., Frutos, ,F., Esteban, ,A., Flores, ,J., Retta, ,A., Baltodano, ,A., . . . Johnson, ,M. (2004). What is the daily practice of mechanical ventilation in

pediatric intensive care units? A multicenter study. *Intensive Care Medicine*, (5), 918-925. doi: 10.1007/s00134-004-2225-5

Ista, E., van Dijk, M., Tibboel, D., de Hoog, M., & Aneja, R. (2005). Assessment of sedation levels in pediatric intensive care patients can be improved by using the COMFORT "behavior" scale. *Pediatric Critical Care Medicine*, *6*(1), 58. doi: 10.1097/01.PCC.0000149318.40279.1A

Jin, H. S., Yum, M. S., Kim, S. L., Shin, H. Y., Lee, E. H., Ha, E. J., . . . Park, S. J. (2007). The efficacy of the COMFORT scale in assessing optimal sedation in critically ill children requiring mechanical ventilation. *Journal of Korean Medical Science*, *22*(4), 693-697. doi: 10.3346/jkms.2007.22.4.693.

Khemani, R. G., Markovitz, B. P., & Curley, M. A. (2009). Characteristics of children intubated and mechanically ventilated in 16 PICUs. *Chest*, *136*(3), 765-771. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=19542258

Marx, C. M., Smith, P. G., Lowrie, L. H., Hamlett, K. W., Ambuel, B., Yamashita, T. S., & Blumer, J. L. (1994). Optimal sedation of mechanically ventilated pediatric critical care patients. *Critical Care Medicine*, *22*(1), 163-170.

McGrath, P. J., Walco, G. A., Turk, D. C., Dworkin, R. H., Brown, M. T., Davidson, K., . . . PedIMMPACT. (2008). Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. *Journal of Pain, 9*(9), 771-783. doi: 10.1016/j.jpain.2008.04.007

Merkel, S. I., Voepel-Lewis, T., Shayevitz, J. R., & Malviya, S. (1997). The FLACC: A behavioral scale for scoring postoperative pain in young children. *Pediatric Nursing, 23*(3), 293-297. Retrieved from: http://find.galegroup.com/gtx/retrieve.do?contentSet=IAC-Documents&resultListType=RESULT_LIST&qrySerId=Locale%28en%2CUS%2C%29%3AHQE%3D%28__HR__%2CNone%2C43%29sn+0097-9805+and+iu+3+and+sp+293+and+vo+23+%24&sgHitCountType=None&inPS=true&sort=

Milette, I. H., & Carnevale, F. A. (2003). I'm trying to heal...noise levels in a pediatric intensive care unit. *Dynamics (Pembroke, Ont.), 14*(4), 14-21.

Playfor, S., Jenkins, I., Boyles, C., Choonara, I., Davies, G., Haywood, T., . . . Analgesia and Neuromuscular Blockade Working,Group. (2006). Consensus guidelines on sedation and analgesia in critically ill children. *Intensive Care Medicine, 32*(8), 1125-1136.

Rennick, J. E., Johnston, C. C., Dougherty, G., Platt, R., & Ritchie, J. A. (2002). Children's psychological responses after critical illness and exposure to

invasive technology. *Journal of Developmental & Behavioral Pediatrics, 23*(3), 133-144. doi: Retrieved from: http://ovidsp.tx.ovid.com/sp-3.4.0b/ovidweb.cgi?WebLinkFrameset=1&S=FGPLFPCOGEDDMDKFNCCLLGIBEIMCAA00&returnUrl=ovidweb.cgi%3f%26Full%2bText%3dL%257cS.sh.70.71%257c0%257c00004703-200206000-00002%26S%3dFGPLFPCOGEDDMDKFNCCLLGIBEIMCAA00&directl

Stevens, B. J., Abbott, L. K., Yamada, J., Harrison, D., Stinson, J., Taddio, A., . . . CIHR Team in Children's, P. (2011). Epidemiology and management of painful procedures in children in canadian hospitals. *CMAJ Canadian Medical Association Journal, 183*(7), E403-10. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=21464171; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:21464171&id=doi:&issn=0820-3946&isbn=&volume=183&issue=7&spage=E403&pages=E403-10&date=2011&title=CMAJ+Canadian+Medical+Association+Journal&atitle=Epidemiology+and+management+of+painful+procedures+in+children+in+Canadian+hospitals.&aulast=Stevens&pid=%3Cauthor%3EStevens+BJ%2CAbbott+LK%2CYamada+J%2CHarrison+D%2CStinson+J%2CTaddio+A%2CBarwick+M%2CLatimer+M%2CScott+SD%2CRashotte+J%2CCampbell+F%2CFinley+GA%2CCIHR+Team+in+Children%27s+Pain%3C%2Fauth

or%3E%3CAN%3E21464171%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

van Dijk, M., de Boer, J. B., Koot, H. M., Tibboel, D., Passchier, J., & Duivenvoorden, H. J. (2000). The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain, 84*(2-3), 367-377. doi: 10.1016/S0304-3959(99)00239-0

von Baeyer, C. L. (2009). Children's self-report of pain intensity: What we know, where we are headed. *Pain Research & Management, 14*(1), 39-45. Retrieved from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2706563/pdf/prm14039.pdf

von Baeyer, C. L., & Spagrud, L. J. (2007). Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. *Pain, 127*(1-2), 140-150. doi: 10.1016/j.pain.2006.08.014

Chapter 2: A Systematic Review of Instruments for Scoring Physiological and Behavioural Cues of Pain, Non-Pain Related Distress, and Adequacy of Analgesia and Sedation in Pediatric Mechanically Ventilated Patients

Tamara Dorfman, RN, BScN, BSc Honors Neuroscience

Dr. Gwen Rempel, PhD, RN

University of Alberta, Edmonton, Alberta, Canada

Potential Venue: Pediatric Critical Care Medicine

Mechanical ventilation is a central and increasingly used treatment modality in Pediatric Intensive Care Units (PICUs) and Neonatal Intensive Care units (NICUs). PICUs worldwide have reported 20% to 64% of the patients they admit require ventilation (Farias et al., 2004; Khemani, Markovitz, & Curley, 2009). Mechanical ventilation complicates optimal pain control and sedation to manage pain and non-pain related distress, that is, responses that occur when the child is exposed to non-pain related external stimuli that elicit fear, anxiety, anger, frustration, depression, distress, dysphoria, or unhappiness (von Baeyer & Spagrud, 2007). Ventilated patients are sedated and often unable to speak, which makes self-report, the "gold standard" of pain and non-pain related distress difficult (Ambuel, Hamlett, Marx, & Blumer, 1992; Marx et al., 1994; M. Van Dijk et al., 2000; von Baeyer & Spagrud, 2007). This difficulty is further complicated as young children lack the verbal and cognitive ability to effectively express their emotions related to pain and other negative non-pain related emotions. Physiological proficiency to display nonverbal signs of pain or non-pain related distress is also lacking. For example, infants born with congenital heart disease cannot exhibit tachycardia or hypertension. Postoperatively, these infants are likely to experience conduction abnormalities, require cardiac pacing and cardiovascular medications, and experience catecholamine abnormalities that alter the physiologic responses that caregivers commonly look for as cues to intervene (Connolly, McClowry, Hayman, Mahony, & Artman, 2004). Therefore, self-report is difficult in mechanically ventilated children due to age, intubation,

sedation, and the potential inability to manifest the predetermined non-verbal signs which caregivers require to recognize pain and non-pain related distress.

When the "gold standard" for the assessment of pain and non-pain distress is not available, the clinical judgement of the attending nurse or physician is the next logical choice. This measure, however, can result in many interpretations and lead to disagreements within the clinical team. Disagreements can result in significant fluctuations in the administration and discontinuation of analgesia and sedation (Carnevale & Ducharme, 1997). This predisposes the patient to adverse reactions and over- or under-sedation. Length of ventilation and ICU stay, and long-term psychological and neurodevelopmental factors may also be affected by fluctuations in analgesia and sedation (Carnevale & Ducharme, 1997).

To address these problems, systematic assessment instruments have been developed to objectively measure the effectiveness of analgesia and sedation in treating pain and non-pain related distress in mechanically ventilated and non-verbal patients (Marx et al., 1994). Theoretically, these scales provide a more consistent measure of the adequacy of analgesia and sedation in controlling patient's pain and non-pain related distress than do a nurse or physician descriptive analog or visual analog scale (Marx et al., 1994). Consequently, these scales allow for fewer discrepancies between individual assessments of pain and non-pain related distress and the patient's response to the analgesia and sedation. These scales are increasingly available but few have been identified and evaluated for efficacy and/or effectiveness in PICU and NICU mechanically ventilated patients.

Two previous systematic reviews have examined instruments for assessing either pain or sedation in critically ill patients. De Jonghe et al. (2000) summarized available systematic assessment instruments for the evaluation of sedation in both adult ICU and PICU patients. The authors identified 25 scales for assessing sedation in critically ill patients of which only five had been evaluated in a PICU setting. They concluded that the Comfort scale was the most appropriate measure of sedation in PICU patients at the time of the review, but there not enough was known about the instrument's ability to detect change in a patient's condition over time (De Jonghe et al., 2000). The authors limited their assessment of the Comfort scale's psychometric properties to the evaluation completed in the original report of this study and did not comment on the psychometric properties of the other four scales identified. Since the time of this review additional sedation scales have been developed and published. Additionally, the concept of non-pain related distress was not assessed as a measurement property of the instruments.

A more recent systematic review identifed all published observational measures for pain in children age 3 to 18 regardless of patient setting. The authors did not limit the measures they reviewed to systematic assessment scales. They included global rating scales such as visual analog scales and numerical rating scales. They recommended that the Comfort scale be used to assess pain in PICU patients based their assessment of its level of evidence (von Baeyer & Spagrud, 2007). The authors did not identify any other instruments that may be suitable in this population and they did not comment on the modified form of the Comfort

scale, the Comfort-Behavioural Scale. Limited information was provided on the psychometric properties of the measures other than a statement of whether it was determined to have good validity and reliability.

Neither of these two reviews considered mechanically ventilated pediatric patients as a subgroup within the critical care setting, and the concept of delirium as a component of non-pain related distress was not considered. No instruments for the assessment of pain or sedation in mechanically ventilated patients who were muscle relaxed were identified. None of the instruments evaluated in these two reviews commented on the efficacy (i.e., the ability of instrument to cause an effect) and effectiveness (i.e., the capacity of an instrument to produce the desired effect in the real world) of these instruments on patients on patient care outcomes.

The objectives, therefore, of part one of this systematic review were to: 1) identify available instruments appropriate for measuring physiological and behavioural cues of pain, non-pain related distress, and adequacy of analgesia and sedation in mechanically ventilated PICU and NICU patients; 2) describe the instruments in terms of how they were developed, 3) describe the physiological and behavioral variables they assess, and, 4) evaluate the instruments in terms of their psychometric properties. The objective of the second part of this systematic review was to determine the efficacy and effectiveness of these instruments on patient care outcomes, including total use of analgesics and sedatives, fluctuations in analgesia and sedation between nursing shifts, length of ventilation and PICU/NICU stay, and any adverse withdrawal effects.

Methods

Search Strategy

A comprehensive search strategy for published articles, theses, and dissertations was used to search MEDLINE, CINAHL, EMBASE, Web of Science (WOS), BIOSIS Previews, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Database of Abstracts of Review and Effects (DARE), and Scopus, and Proquest Dissertations from January 1970 to June 2011. The search was limited to literature published during this timeframe because previous systematic reviews of pain and sedation observational assessment measures did not reveal published scales for children or adults before 1970 (De Jonghe et al., 2000; von Baeyer & Spagrud, 2007). Keywords for searching were used in free text form as well as mapped to MeSH headings (Table 2- 1). The reference lists of relevant articles were reviewed to identify additional articles. Identified sources of information were entered into a citation database and duplicates were removed (Refworks-COS, 2011).

Selection Criteria for Part One

Studies appropriate for analysis were identified using the following criteria: 1) study population of inpatient critical care, invasively mechanically intubated patients, age 0 days to 18 years; 2) study describes the development and evaluation of a systematic assessment instrument to measure physiological and behavioral cues of pain and/or non-pain related distress and adequacy of analgesia and/or sedation, or the evaluation of a previously published instrument; 3) the instrument consists of physiological and/or behavioral cues of pain, non-pain

related distress, analgesia, or sedation; 4) at least one psychometric property of the instrument is evaluated (Table 2-2); 5) instrument items have a number of response options, which may be measured as categorical variables, either numerical (e.g., with a 5 or 7 point scale) or non-numerical; 6) the instrument does not require self-report from patient; and, 7) the instrument is not a global rating scale, meaning it does not require the observer's global impression of the patient's pain, non-pain related distress, analgesia, or sedation.

Selection Criteria for Part Two.

Studies appropriate for analysis were identified using the following criteria: 1) the study is Randomized Control Trial (RCT), Controlled Clinical Trial (CCT), Controlled before and after study (CBA), Cohort Study, Case control study, retrospective chart review, crossover study, or Quasi-experimental design; 2) the study population includes inpatient, critical care, invasively mechanically intubated patients, age 0 days to 18 years; 3) the intervention, in at least one group of mechanically ventilated patients, is the use of an objective, systematic assessment tool to measure physiological and behavioural cues of pain and/or non-pain related distress and/or adequacy of analgesia and/or sedation is the intervention in at least one group of mechanically ventilated patients in the study (note: the tool can be part of a sedation protocol); 4) each item on the instrument has a number of response options, which may be measured as categorical variables, either numerical (e.g. with a 5 or 7 point scale) or non-numerical; 5) the instrument does not require self-report from patient; 6) the instrument is not a global rating scale, meaning it does not require the observer's global impression

of the patient's pain, non-pain related distress, analgesia, or sedation; and, 7) numeric data is reported on at least one of the five outcomes: total use of analgesics and/or sedatives, fluctuations in analgesia and sedation between nursing shifts, length of ventilation, length of PICU stay, or adverse withdrawal effects.

Studies published in any language were eligible for the review; however, when unable to locate translation assistance, a Bulgarian article was excluded. Two reviewers (TD and SI) independently screened all abstracts to ensure they met the criteria for inclusion. When no abstract was available, subject headings and extracts from the text were used. The full texts of all studies potentially meeting our inclusion criteria, as determined by either reviewer, were retrieved. The retrieved articles were independently screened by two reviewers (TD and either SI or GR) using the same criteria to confirm inclusion. Disagreements between the reviewers were resolved through discussion between the two initial reviewers, and when needed the involvement of the third reviewer (SI or GR depending on who did the initial review).

Assessment of Quality for Studies Meeting Inclusion Criteria for Part One

Studies that met the inclusion criteria were assessed for quality by two independent reviewers (TD and GR or ES). Disagreements between the reviewers were resolved through discussion and, when necessary, the involvement of the third reviewer. Methodological quality was assessed using the COSMIN checklist with four point scale (Appendix A, B) (Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010b; Terwee et al., 2012). The COSMIN

checklist was developed through an international Delphi study in which consensus was reached on the definitions of domains, measurement properties, aspects of measurement properties to be used to evaluate health-- related patient- reported outcomes. This checklist can also be used to evaluate the quality of studies on the measurement properties of other instruments (Mokkink et al., 2009; Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010a).

The COSMIN checklist consists of 12 boxes. Two boxes are used to evaluate the general requirements of a study on measurement properties, in terms of generalizability and whether Classical test theory (CTT) or Item Response theory (IRT) methodology was used to evaluate the instrument(s)'s psychometric properties. Nine boxes are used to evaluate internal consistency, reliability, measurement error, content validity (including face validity), construct validity (i.e., structural validity, hypotheses testing, and cross-cultural validity), criterion validity, and responsiveness. The remaining box is used to evaluate the interpretability of the instrument (Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010b). The four point scale for the COSMIN checklist allows the reviewer to give the measurement property a score of poor, fair, good, or excellent. There is no scoring system for interpretability and generalizability (Terwee et al., 2012). The two independent reviewers, therefore, used the COSMIN checklist questions for interpretability and generalizability as a guide and assigned scores of limited generalizability/ interpretability, somewhat generalizable/ interpretable, acceptable generalizability/ interpretability, and superior generalizability/ interpretability.

Assessment of Quality for Studies Meeting Inclusion for Part Two

Methodological quality for studies that met the inclusion criteria for part two was assessed using a before and after quality assessment tool (BAQA) that was adapted from the New Castle Ottawa Assessment Scale (NOS) by Alberta Research Centre for Health Evidence (Wells et al., 2011) (See appendix C). The NOS was developed to provide an easy and convenient tool for quality assessment of nonrandomised studies to be used in a systematic review. Each study is judged on four broad perspectives: the selection of the study groups; the comparability of the groups; assessment of outcome, and intervention. Using the four perspectives, each study received a score of 0 to 13. Higher scores indicate better quality.

Data Abstraction.

Study data was collected and managed using REDCap (Research Electronic Data Capture) hosted at University of Alberta's Women and Children's Health Research Institute (Harris et al., 2009). Separate data abstraction forms were designed in Microsoft Word for abstraction of data for part one and part two of the review. The forms were initially piloted on four studies (three studies for part one and one study for part two), independently, by reviewers TD and ES. Revisions and modifications were then made, based on the opinions and consensus of the two reviewers, and the forms were created in REDCap (Appendix D). Data from all eligible studies were abstracted by the first reviewer (TD) and then independently verified by a second reviewer (ES). Data abstraction included coder information, study characteristics, dimensions measured on the scale, characteristics of the study population, study design, and psychometric

properties of the scales for part one. Data abstraction for part two included coder information, study characteristics, a description of the intervention and outcome measured, and results of the intervention.

Synthesis of Data

Following data abstraction for part one, the studies were grouped by instrument. Data was then synthesized into tables to provide the characteristics of the instruments, compare the behavioral and physiological variables assessed, and evaluate the psychometric properties of each scale. Definitions of the psychometric properties used to assess the reliability, validity, responsiveness, and interpretability of the instruments identified in this review are presented in Table 2-2. Data from studies which met criteria for part two was synthesized into a table which illustrated the intervention and resulting patient care outcomes.

Results

The search strategy identified 2269 articles. An additional six articles were found from the reference list of the articles identified in the search. Evaluation of the articles by abstract, title, and keywords, during the first screening, left 312 articles to be evaluated as full texts. The full texts of the eligible studies were screened for inclusion and exclusion criteria, leading to a total of 28 articles being included in part one of the review and three studies being included in part two of the review. None of the studies met inclusion for both part one and part two of the review. The study selection process and reasons for exclusion are summarized in Figure 2-1.

Among the 28 articles, 17 scales for measuring physiological and behavioural cues of pain, non-pain related distress, and adequacy of sedation or analgesia in mechanically ventilated PICU or NICU patients were identified (Table 2-3; Appendix E). (Ambuel et al., 1992; Curley, Harris, Fraser, Johnson, & Arnold, 2006; Hartwig, Roth, & Theisohn, 1991; Hughes et al., 1994; Hummel, Puchalski, Creech, & Weiss, 2008; Johansson & Kokinsky, 2009; Macnab, Levine, Glick, Susak, & Baker-Brown, 1991; Parkinson et al., 1997; Popernack, Thomas, & Lucking, 2004; Ramelet, Rees, McDonald, Bulsara, & Abu-Saad, 2007; Razmus, Clarke, & Naufel, 2003; Smith et al., 2011; Spence, Gillies, Harrison, Johnston, & Nagy, 2005; Suominen et al., 2004; M. Van Dijk et al., 2000; Voepel-Lewis, Zanotti, Dammeyer, & Merkel, 2010). The instrument structure including number of items per scale, parameters assessed by each scale and scoring of the instrument are presented in Table 2-3.

The behavioral and physiological variables assessed by each instrument are compared in Table 2-4. A total of six physiological variables were assessed by at least one of the 17 different scales. The most common physiological variables assessed are heart rate and blood pressure. Muscle tone was originally identified as a physiological variable, but fits more congruently with behavioral variables (Ambuel et al., 1992; Carnevale & Razack, 2002). A total of 19 behavioral variables were assessed in at least one of the 17 instruments. Common behavioral variables are alertness/ level of consciousness, calmness/agitation, posture/ muscle tone, and facial tension or expression. Only five instruments (Hartwig Scale, Comfort and Comfort-B scale, State Behavioral Scale, and the the Penn

State Children's Sedation Algorithm), assess the patient's response to the ventilator (Ambuel et al., 1992; Curley et al., 2006; Hartwig et al., 1991; Popernack et al., 2004; M. Van Dijk et al., 2000). Only six instruments (Hartwig Scale, State Behavioral Scale, Hughes unnamed Scale, Parkinson unnamed scale, Modified Comfort Scale for PICU for Muscle Relaxed Patients, and the Penn State Children's Sedation Algorithm), assess the patient's response to endotracheal suctioning (Curley et al., 2006; Hartwig et al., 1991; Hughes et al., 1994; Parkinson et al., 1997; Popernack et al., 2004; Razmus et al., 2003). Table 2-5 describes study design, population, number of assessments, and assessment process for both original and validation studies for each scale. The internal consistency and reliability of each of the scales is shown in Table 2-6. The validity, responsiveness, interpretability of each of the scales is shown in Table 2-7. The quality of each of the studies using the COSMIN checklist is described in Table 2-8. Only one of the studies used IRT to evaluate the methodology of the studies. The remaining studies used CTT.

Instruments that Assess Pain, Non-pain Related Distress, Analgeisa, and Sedation

Comfort and Comfort-Behavioral scale. The Comfort scale was developed to assess pain and non-pain related distress in PICU patients. The research team identified potential scale variables by reviewing behavioral science and medical literature on the assessment of pediatric distress and pain and surveying 20 experienced PICU nurses. They selected eight physiologic and behavioral variables for the pilot version of the scale (see Table 2-3). These variables were

chosen as they were commonly used by clinicians, responded rapidly to changing levels of distress, could be assessed quickly and non-intrusively, reflected distress in children of all ages, and were likely to remain variable in the face of changes in disease state and drug treatment (Ambuel et al., 1992).

Table 2-6 and 2-7 indicate the psychometric properties of the Comfort scale as determined by the original study on the scale and several validation studies. The quality assessment of all these studies (Table 2-8) ranges from poor to excellent depending on the psychometric property evaluated. The internal consistency of the scale is strong. In the original study, all items correlated with the adjusted total score, although, muscle tone stood out as having a lower correlation than other items (Ambuel et al., 1992). Three studies have demonstrated high inter-rater agreement for all items on the Comfort Score including total score (Ambuel et al., 1992; Marx et al., 1994; M. Van Dijk et al., 2000). The two most objective variables, heart rate and MAP tend to have the lowest inter-rater agreement while variables that require global judgments (alertness, calmness, respiratory response, and movement) have the highest levels of agreement (Ambuel et al., 1992). A comparison of inter-rater reliability between paired Comfort Score ratings and paired intensivist's ratings for adequacy of sedation demonstrated significantly less variability in the Comfort score ratings (Marx et al., 1994). Strong correlations between the Comfort Score and a visual analog scale (VAS; a horizontal continuous 100 mm line; graded from 0 (no distress) to 100 (worst imaginable)) of distress and intensivist's descriptive sedation adequacy ratings indicate that the scale is a valid measure of

both distress due to pain, non-pain related distress, and sedation (Ambuel et al., 1992; Marx et al., 1994). The scale itself demonstrates good interpretability as the rating scale for each variable has been constructed to allow scoring of extreme values and avoid ceiling or floor effects. The eight individual dimensions and total scale have central tendencies near the midpoint with distributions that should accommodate extremes of high and low distress (Ambuel et al., 1992).

The Comfort Scale's validity has been questioned by some studies, including concerns regarding structural validity. Grap et al. (2006) found Comfort Scale scores did not significantly correlate with a dominant behavioral state or individualized activities of leg movement, head movement, twitching, and coughing. These authors suggested that overall behavioral state should be taken into account when assessing sedation in addition to the Comfort Score (Grap, Pickler, & Munro, 2006). No further studies were found investigating this suggestion.

Additionally, analyses of item-to-item correlations of the Comfort Scale and structural validity have continued to call into question three variables within the scale: muscle tone, heart rate, and MAP. Muscle tone stood out as having relatively low correlations with other variables, and MAP and heart rate stood out as having a higher correlation with each other than with other variables in the original study (Ambuel et al., 1992). Principal component analysis suggested that the scale had two related dimensions that accounted for 84% of the variance. The first dimension, behavioral comfort, accounted for 58% of the variance and included alertness, calmness, respiration, movement, facial tension, and muscle

tone. The second dimension, physiologic comfort, accounted for 26% of the variance and included heart rate, mean arterial pressure, and muscle tone. Behavioral and physiologic comforts were correlated at 0.69 (Ambuel et al., 1992).

These findings lead to the question of whether both dimensions should be used. Comparison of the Comfort-- Behavior dimension with physiological pain measures of MAP, MAP variability, heart rate, and heart rate correlations were significant indicating an association between the behavioral dimension of the Comfort Scale and physiological variables of pain (M. van Dijk et al., 2001). In a separate study, a stepwise multiple regression analysis for predictors of total comfort score found that 97% of the total score variance was explained by six of the eight items (all the items except for heart rate and mean arterial blood pressure). These authors proposed a modified Comfort Scale based on these six items, which has since been referred to the Comfort-Behavioral (Comfort-B Scale) (Carnevale & Razack, 2002). The psychometric properties of the Comfort-B scale are provided in Table 2- 6 and 2-7. The quality of the studies assessing these properties is shown in Table 2-8. The scale has high internal consistency and inter-rater reliability. It also correlates well with VAS scales for pain and Nurse Interpretation Score of Sedation (NISS) indicating that it is a valid scale for assessing both pain and sedation (Ista, Van Dijk, Tibboel, & De Hoog, 2005; M. Van Dijk et al., 2000).

Advantages of the the Comfort Scale and its modified form the Comfort-B Scale are that both scales have been extensively researched giving them a strong

evidence base (Alexander, Carnevale, & Razack, 2002; Ambuel et al., 1992; Boerlage, Ista, de Jong, Tibboel, & van Dijk, 2011; Brunow de Carvalho, Lucas da Silva, Paulo, Fonseca, & Belli, 1999; Carnevale & Razack, 2002; Grap et al., 2006; Ista et al., 2005; Ista, De Hoog, Tibboel, & Van Dijk, 2009; Jin et al., 2007; Johansson & Kokinsky, 2009; Kusahara, Rego, Pedreira, Peterlini, & Carvalho, 2005; Marx et al., 1994; Valkenburg et al., 2011; M. van Dijk et al., 2001; M. Van Dijk et al., 2000). Both scales are age independent and relatively non-intrusive. They can be adminstered in 2 to 3 minutes so time sampling is possible (Ambuel et al., 1992; Brunow de Carvalho et al., 1999; Ista et al., 2005). The scales offer a further advantage of not being based on responsiveness criteria (meaning the child does not have to be disturbed from rest) (Marx et al., 1994; M. Van Dijk et al., 2000). They can also be easily integregated into routine nursing care and health care professionals can be trained to use the scale in approximately 2 hours (Ambuel et al., 1992; Ista et al., 2005; M. Van Dijk et al., 2000). The Comfort-B scale has additional benefits as the respiratory dimension has subcategories (which had been validated) that allow for its assessment in non-ventilated PICU patients and it has been validated in a subgroup of PICU patients, children three and under with Down Sydrome (Johansson & Kokinsky, 2009; Valkenburg et al., 2011; M. Van Dijk et al., 2000).

Although there are several advantages to the use of Comfort and Comfort-B Scale, there are some disadvantages. Both scales are not intended to be used to evaluate pain during intermittent painful procedures and consequently are not useful in assessing procedural pain (Marx et al., 1994; M. Van Dijk et al., 2000).

Furthermore, critiques of the Comfort Scale include that it is time consuming given there are eight variables (Brunow de Carvalho et al., 1999). Similarly, the Comfort-B scale requires a 2-minute assessment. Shorter observation periods (30 seconds) have been shown to increase the risk of underscoring the patient's pain (Boerlage et al., 2011). The Comfort Scale is measured at a single point in time, rather than continuously and, therefore, may not accurately assess behavior states. It has been suggested that its assessment of sedation may be more accurate if combined with an assessment of the dominant behavior state (Grap et al., 2006). The Comfort Scale is also unable to accurately score pain or sedation when patients are chemically paralyzed (Lamas et al., 2008; Razmus et al., 2003). Finally, the area in the Comfort-B score between 11 and 22 does not adequately predict under- or over-sedation. Consequently, a NISS score needs to be completed in combination with the Comfort-B score to adequately predict the level of sedation.

Instruments that Assess Sedation Only

The Ramsay Scale. The Ramsay Scale was developed to assess the level of sedation of adult ICU patients. Its reliability and validity has not been assessed in PICU mechanically ventilated patients. However, one published abstract was found comparing the Ramsay Scale to the Comfort Scale. An attempt was made to contact the author by e-mail to determine if there was a published full text with no response. The abstract reported the percentage of agreement between Comfort and Ramsay scales was 80.0%, with a kappa of 0.59 expressing a tendency to concordance. The percentage of agreement between professional caregivers'

assessment of sedation (rating sedation as adequate or inadequate) was 47.0% and 50.0% for the Comfort and Ramsay scale, respectively. But, the kappa statistics showed a higher concordance level between the caregivers' clinical assessment and the Ramsay Scale than with COMFORT scale (Kusahara et al., 2005).

Hartwig Scale. Hartwig et al. (1991) developed a scale to assess level of sedation at different fentanyl and midazolam doses. The scale assesses behavioral variables only (See Table 2-3). In contrast, to the Comfort Scale, the Hartwig scale requires tracheal suction for assessment. Advantages of the Hartwig Scale are: it only measures behavioral variables, and therefore, does not require monitors for assessment of physiological variables, it is age independent, and it can be used at any time to assess adequacy of sedation (Brunow de Carvalho et al., 1999). One disadvantage of the scale is its need for endotracheal aspiration. This assessment variable is questionable due to the painfulness of the procedure and the possibility it may affect the degree of sedation of the child. However, the fact that this is a routine procedure in PICU patients minimizes this disadvantage (Brunow de Carvalho et al., 1999).

There is limited research on the psychometric properties of the Hartwig Scale (Table 2-6, 2-7) and quality of these studies is poor to fair (Table 2-8). The original report offers no information on how the scale was developed and only asseses inter- rater reliability, reported as moderate to high for the different variables (Table 2-6). A comparison of the Hartwig scale with the Comfort Scale showed statistically significant agreement between the two scales suggesting convergent validity (Brunow de Carvalho et al., 1999).

Vancouver Sedative Recovery Scale (VSRS). The VSRS was developed to address prolonged recovery times from a sedated state for children receiving continuous infusions of analgesic and sedation. Visible characteristics that differentiated a fully alert child were initially examined in critically ill children at various stages of recovery from sedation. Based on face validity and applicability, a total of 12 items were selected for the VSRS. These items fell into three broad categories: response; eye appearance and function; and body movement. A pilot study revealed high inter-rater reliability. However, item analysis indicated that some items were not measuring the same concept as the other scale items (Macnab et al., 1991). Scoring modifications were made but there are no published validation studies on the amended scale. The quality of the initial study was excellent (Table 2-8).

State Behavioral Scale (SBS). *SBS* was derived from the authors' previous work describing PICU nurses' descriptions of agitation, literature on sedation tools used in the adult population, and expert opinion from a pediatric anesthesiologist and PICU clinical nurse specialist. Two adult ICU sedation scales served as templates for the scale (Curley et al., 2006). The tool includes eight dimensions. Each dimension contains three to six levels that incrementally describe the sedation–agitation continuum (Table 2-3) (Curley et al., 2006).

The reliability and construct validity of the scale was assessed in the original report (Table 2-6, 2-7). The quality of this study is excellent (Table 2-8).The weighted kappa coefficients for each of the eight dimensions ranged from 0.44 (consolability) to 0.76 (respiratory drive/response to ventilation), indicating

moderate to good inter-rater reliability. The level of inter-rater agreement did not differ by patient age group, except for the coughing dimension. In analyses of all 198 paired ratings, the weighted kappa for this dimension was significantly lower in the youngest age group of 6 weeks to 1 year (0.55), compared with the weighted kappa in the oldest age group of 3-6 years (0.81). The middle age group, ages 1-3 years, had a weighted kappa of 0.73. Cluster analysis revealed five distinct state profiles: unresponsive (-3), responsive to noxious stimuli (-2), responsive to gentle touch or voice (-1), awake and able to calm (0), and restless and difficult to calm (+1), all of which differed significantly from each other ($p <$ .001). A sixth profile (agitated) was added to the scale, although not observed in the study (Curley et al., 2006).

The main disadvantage of the SBS is assessment of sedation with this scale requires progressive stimulation. The scale itself has only been tested in patients aged 6 weeks to 6 years who were physiologically stable and not rated in pain (Curley et al., 2006). It is, therefore age dependent and can be only used in a subgroup of PICU patients.

Penn State Children's Hospital Sedation Algorithm (PSCHSA). The PSCHSA was developed to reduce unplanned extubations in PICU patients. An open-ended survey indicated that the communication of sedation goals for each patient needed to be more objective. Based on this finding, the PSCHSA was developed to provide six levels of sedation with target goals for objective patient behaviors (Table 2-3) (Popernack et al., 2004). A disadvantage of the scale is patient must meet all the behavioral descriptors of that level in order for their

level of sedation to be classified. There is limited published data on the validity of this algorithm, but the authors indicate a significant decrease in the number of unplanned extubations post institution of the PSCHSA (Table 2-7), supporting construct validity (Popernack et al., 2004).

Hughes Parkinson Unnamed Sedation Scales. Hughes et al. (1994) and Parkinson et al. (1997) developed sedation scales with the goal of testing the efficacy of specific medications for sedation. The structure and scoring of these instruments is described in Table 2-3. The development of these scales is not well described and only the inter-reliability of the scales was evaluated (Table 2-6) (Hughes et al., 1994; Parkinson et al., 1997). The quality of the assessment of this property is good for the Parkinson study and fair for the Hughes study (Table 2-8). Both of these scales require the patient to be suctioned via the endotracheal tube meaning the patient must be disturbed from rest (Hughes et al., 1994; Parkinson et al., 1997).

Instruments that Assess Pain and Adequacy of Analgesia

FLACC and Modified FLACC Scale. The Faces, Legs, Activity, Cry, and Consolability (FLACC) scale was developed to measure postoperative pain in young children (2 months to 7 years) and has shown a high degree of clinical usefulness (Merkel, Voepel-Lewis, Shayevitz, & Malviya, 1997; Voepel-Lewis et al., 2010). The FLACC scale has been studied in multiple acute hospital settings and has been shown as effective in assessing acute pain. The FLACC is widely recognized and used internationally. It has been translated into several languages, including French, Chinese, Portuguese, Swedish, and Italian (Voepel-Lewis et al.,

2010). There are three studies on the scale or its modified form in pediatric critical care (Johansson & Kokinsky, 2009; Manworren & Hynan, 2003; Voepel-Lewis et al., 2010). The quality of these studies as assessed by COSMIN range from fair to good (Table 2-8) and one of the studies consists mainly of adult ICU patients, thereby limiting its generalizability to the pediatric population (Voepel-Lewis et al., 2010). The advantages are that FLACC is simple to use, is not time consuming, and does not require the assessment of any physiological variables. Additionally, the behavioral components of the scale have excellent internal consistency (Voepel-Lewis et al., 2010).

Its reliability and clinical validity was tested in adult and pediatric ventilated and non-ventilated critical care patients. The internal consistency and inter-rater reliability of the scale was excellent. Exploratory factor analysis showed that one component accounted for 68.9% of the variance in the FLACC scores and that four categories of FLACC (face, legs, activity, and consolability) reflected the pain expression factor. The FLACC correlated highly with the Comfort Scale, supporting criterion validity in critically ill children. Furthermore, FLACC pain scores decreased significantly after the adminstration of analgesia for painful to non-painful situations, supporting construct validity in PICU patients (Manworren & Hynan, 2003; Voepel-Lewis et al., 2010).

The usefulness of FLACC, however, does not extend beyond the assessment of acute pain. The descriptors included in the FLACC tool were meant to indicate some of the differences observed from patient to patient. They cannot be used to assess chronic or long term pain as it does not include other observations such as

activity, quality of sleep, and expressions of depression (Voepel-Lewis et al., 2010).

A modified version of FLACC has been compared with the Comfort-B scale. In the modified version of FLACC, the cry component of the scale was altered offering alternative behavioral signs of pain to describe 'Cry' in the ventilated patient through the 'Cry face' - facial expressions of moaning or crying. Inter-rater reliability was high for both scales (Table 2-6). Comfort-B scores were significantly different for all 3 NISS categories (over-sedated, sufficiently sedated, and under-sedated) and for VAS scores of pain >3 and no pain <3. Modified FLACC scores, however, were not significantly different for all 3 NISS categories, but were significantly different for VAS scores of pain >3 and no pain <3. Modified FLACC showed good construct validity as scores decreased following the administration of morphine (Table 2-7). These results indicated that the Comfort-B scale is a more reliable scale than modified FLACC for the assessment of sedation. Modified FLACC, however, showed good construct and criterion validity for the measurement of pain (Johansson & Kokinsky, 2009).

Multidimensional Assessment of Pain Scale (MAPS). The Multidimensional Assessment of Pain Scale (MAPS) was generated from observations of critically ill infants experiencing postoperative pain and postoperative pain exacerbated by painful procedures (e.g., tracheal suctioning, positioning and chest drains removal). It is, therefore, useful in assessing both of these types of pain. The scale consists of 5 items. Its structure and scoring is described in Table 2-3 (Ramelet et al., 2007; Ramelet, Rees, McDonald, Bulsara, & Huijer Abu-Saad, 2007). The

main categories of the MAPS scale are not different from those of previous measures, but the components of the categories at the extreme end of the scale (extreme pain) are novel. It is the first scale to use a decrease in heart rate and/ or blood pressure, severe respiratory distress compromising oxygenation, sustained grimacing associated with a silent or weak cry, rigid or limited movements and decreased physical activity with no eye contact to define worst pain. All these components remained in the revised version of MAPS except for a decrease in heart rate and blood pressure which was changed to a 20% increase from baseline in those variables (Ramelet et al., 2007). The MAPS cannot be used to assess pain in patients who are chemically paralyzed, and it cannot assess chronic or long-term pain (Ramelet et al., 2007). It is also not applicable to chemically paralyzed children, as it includes behavioral responses as indicators of pain (Ramelet et al., 2007).

The reliability and validity of the MAPS scale has been evaluated in two different studies (Table 2-6, 2-7) (Ramelet et al., 2007; Ramelet, Rees et al., 2007). The quality of these studies is good to excellent (Table 2-8). Face and content validity was established by a panel of experts in pain assessment and pediatric critical care (Ramelet et al., 2007). The inter-rater reliability was good for all MAPS categories with the exception of breathing pattern. This may be explained by individual differences in assessing the quality of breathing, especially in intubated patients. Agreement measurements between MAPS and FLACC, and MAPS and VAS pain showed that the risk of measurement error was small indicating strong convergent and criterion validity, respectively (Table 2-6)

(Ramelet et al., 2007; Ramelet, Rees et al., 2007). Internal consistency of the MAPS was good as demonstrated by a Cronbach's alpha coefficient of 0.68 in the initial study and 0.62 in the follow-up study (Ramelet et al., 2007; Ramelet, Rees et al., 2007). The homogeneity of the scale would improve considerably if the vital signs category was deleted. However, compared with other behavioral items, vital signs achieved the lowest item total correlation. Therefore, based on the theory behind item analysis, it would be inappropriate to remove this item relying on one value only. Furthermore, the heart rate, systolic, mean, and diastolic arterial pressure significantly decreased after a bolus of morphine indicating a response in vital signs to pain relief and likewise the scale was responsive as the scores decreased significantly by four points (40% of total score) after the administration of the morphine ($P < 0.001$) (Ramelet, Rees et al., 2007). Based on these two validation studies and the supporting literature, a revised version of the MAPS has been published (Ramelet, Rees et al., 2007). No further studies have been published evaluating this revised version.

Cardiac Analgesic Assessment Scale (CAAS). There is only one report on this scale and the quality of this study is poor (Table 2-8) (Suominen et al., 2004). The CAAS was developed for an intrathecal morphine study (unpublished data) to standardize the protocol for rescue pain management of mechanically ventilated children after cardiac surgery and to avoid bias caused by different PICU nurses. The authors identified potential variables by reviewing the literature on pain assessment and distress in children and by surveying pain assessment variables used by PICU nurses. Four physiological variables (Table 2-3) were selected for a

pilot version. Inter-rater reliability was high for all of the CAAS items (Suominen et al., 2004). The CAAS significantly reflected changes in pain status over time. Specifically, the mean of the CAAS scores obtained after the administration of analgesic medication was significantly lower than the previous score measured before the intervention. The CAAS was not found to correlate well with VAS pain. The authors, however, found poor inter-reliability of VAS pain, which was completed at the same time as the CAAS. They suggested this was due to the more subjective nature of the VAS score (Suominen et al., 2004).

Instruments that Assess Sedation in Muscle Relaxed Patients

PICU Modified Comfort Sedation Scale for Muscle Relaxed Patients. The PICU Modified Comfort Sedation Scale for muscle-relaxed patients is the only scale identified that systematically assesses the level of sedation of patients who have received muscle relaxants. The scale was developed as a quality improvement project. The developers modified the Comfort Scale to exclude parameters that were not applicable to the muscle relaxed patient (Ambuel et al., 1992; Razmus et al., 2003). A total of six items are assessed by the scale (see table 2-3). The scale exhibited poor concurrent validity (low correlation) when assessed against the adequacy of sedation score (Like Scale) although the distributions of both scales were similar (Table 2-7) (Razmus et al., 2003). The quality of this study is fair because of its small sample size (Table 2-8).

Instruments for the Assessment of Delirium

The Pediatric Confusion Assessment Method for Intensive Care Unit (pCAM-ICU).The pCAM-ICU was the only instrument identified in this review

that assesses delirium (a component of non-pain related distress). Two adult-oriented instruments were adapted for age-appropriate cognitive assessment based on comprehensive literature review and consultation with child development and delirium experts (Smith et al., 2011). The p-CAM-ICU can be used in both verbal and nonverbal children with at least the cognition expected of a developmentally appropriate 5-yr-old child (Smith et al., 2011). The instrument requires the completion of four steps and may be time consuming. The instrument has excellent inter-rater reliability ($\kappa = 0.96$).The predictive and concurrent validity of the pCAM-ICU was also high. Non--psychiatric trained nurses and physicians using the pCAM-ICU demonstrated a high sensitivity of 83%, specificity of 99%, in the diagnosis of delirium among critically ill children (Smith et al., 2011).

Instruments for the Assessment of Neonates

Neonatal Pain, Agitation, and Sedation Scale (N-PASS).The N-Pass was developed for infants in the NICU. Indicators for the scale were chosen from literature review and expert opinion based on clinical applicability, ease of assessment and established validity (Hummel et al., 2008). The structure and scoring of the instrument is described in Table 2-3. The scale has been validated in assessing prolonged pain, acute pain, agitation, and sedation in both ventilated and non-ventilated NICU patients (Hummel et al., 2008; Hummel, Lawlor-Klean, & Weiss, 2010). The original report indicated higher internal consistency on the sedation scale than pain scale of N-PASS. Overall, both scores are high. High inter-rater reliability was found for both the pain and the sedation scale. The correlation between the Premature Infant Pain Profile (PIPP) and the N-PASS was

strong, particularly at high pain scores indicating convergent validity. The correlation was acceptable, but lower for low pain scores, possibly because the PIPP tool assigns a higher score for a quiet infant, while the N-PASS gives a lower score for a quiet infant (Hummel et al., 2008). The same was true when N-PASS and PIPP scores were compared for actual and sham heel stick procedure (Hummel et al., 2008). The scores on N-PASS were significantly different when assessed during actual heel sticks and sham heel sticks, indicating that N-PASS scores can discriminate between actual painful procedure and non-painful procedures. This study also found strong internal consistency of the pain scale of N-PASS and high inter-rater reliability (0.874, $p<0.0001$). Test–retest reliability was demonstrated by repeat scoring of videotaped heel sticks, measured by Spearman's rho correlation (Table2- 6) (Hummel et al., 2010).

Pain Assessment Tool (PAT). The PAT is reliable and valid for assessing pain in mechanically ventilated neonates in NICU. The PAT measures both behavioral and physiological parameters, as well as the bedside nurse's perception of the infant's pain (Table 2-3). The tool was developed for use in term infants following neonatal surgery (Spence et al., 2005). The inter-rater reliability of the PAT is high (0.85). It has high convergent validity and moderate concurrent validity based on a strong correlation between the PAT and CRIES pain scores ($r = 0.76$) and a moderate correlation ($r=0.38$) between the PAT score and the VAS scores for pain as rated by the infant's mother. The correlation coefficient between the PAT score and CRIES score was significant for all groups ($p < .01$).

Both scales are limited as they are only applicable in neonates (although they show reliability and validity in both pre-term and term neonates) (Hummel et al., 2008; Hummel et al., 2010; Spence et al., 2005). They have only been tested in an NICU setting so their usefulness for neonates being care for in a PICU setting has not been determined.

Efficacy and Effectivness of Objective Systematic Scales in Practice

There is limited research on the efficacy and effectiveness of these instruments in practice. Only three studies included in this review examined the efficacy of using the scales identified in part one of review on patient care outcomes (Alexander et al., 2002; Ista et al., 2009; Jin et al., 2007; Jin et al., 2007). The study designs, population, and interventions are identified in Table 2-9. Each study utilized a sedation protocol which required scoring of either the Comfort Scale or the Comfort-B Scale to determine if sedation or analgesia was required (Alexander et al., 2002; Ista et al., 2009; Jin et al., 2007).

The quality of all three studies is very good to excellent and indicates efficacy of these scales in practice (Table 2-10). Two studies showed an increase in the amount of pain and sedation medication administered following the implemention of the sedation directed protocol guided by these scales and one showed a decrease (Alexander et al., 2002; Ista et al., 2009; Jin et al., 2007). This diffference may be related to the sedation protocol used with scale. The use of the sedation directed protocols, however, positively influence patient outcomes as the intervention was shown to decrease length of ventilation, PICU stay, and development of adverse withdrawal effects (Jin et al., 2007). It also appeared to

improve the assessment of sedation as there were fewer incidents of undersedation in one of the studies and level of sedation was rated as adequate in another study (Alexander et al., 2002; Ista et al., 2009). None of the studies assessed the patient care outcomes of fluctuations in analgesia and sedation between nursing shifts. The limited number of studies assessing the effect of using these scales on patient outcomes makes it difficult to determine their effectiveness in the real world as there is simply not enough data.

Discussion

The Debate about Physiological Variables

There has been debate about the relevance of physiological variables in assessing pain and sedation. Both the Comfort Scale and the MAPS include heart rate and blood pressure (BP). However,these items, in both scales, have been shown to have the lowest item total correlations indicating that the internal consistency of both these scales would improve if these variables were excluded (Table 2-6) (Ambuel et al., 1992; Carnevale & Razack, 2002; Ramelet et al., 2007; Ramelet, Rees et al., 2007). This discrepancy may be related to the variability of the parameters during the 2 minute rating period. Ambuel et al. (1992) suggested this problem could be rectified if raters were instructed to make a continuous printed record of both parameters during the 2 minute observation period and then rate each parameter after the observation period by reviewing the record. van Dijk et al. (2000) verified this suggestion by having nurses (the raters) note the heart rate and BP from the monitor six times every 20 seconds (with the aid of a stopwatch) during the 2-minute scoring period achieving an excellent

Kappa score (see Table 2-6). Similarly, the heart rate and BP variables on the MAPS were also found to have high inter-reliability (See Table 2-6). To ensure the accuracy of these variables, baseline measures were taken when the patient was stable. In instances where the patient could not be hemo-dynamically stabilized, normal range for age, and the condition of the patient was used for baseline measures. Reliability of the vital signs scoring was excellent (Table 2-6) (Ramelet et al., 2007).

An additional investigation into the relationship between physiologic indicators of pain (heart rate, heart rate variability, mean arterial pressure (MAP), and MAP variability (MAPV)) and the Comfort-behavior dimension of the Comfort scale showed significant correlations between physiological and behavioral indicators of pain. In particular, the correlation between the behavioral measure, Comfort-B and the MAP and MAPV were fairly high. Furthermore, the physiology– behavior correlation improved with increasing pain intensity, suggesting that the combination of both dimensions may confirm the diagnosis of a higher intensity of postoperative pain, but not moderate pain (M. van Dijk et al., 2001). These results indicate there is relevance to the physiological variables in the assessment of pain.

Other scales that utilize physiological variables as a part of their assessment are the CAAS, N-PASS, PAT, Hughes (unnamed sedation scale), and the PICU Modified Comfort Scale for muscle relaxed patient (Hummel et al., 2008; Razmus et al., 2003; Spence et al., 2005; Suominen et al., 2004). Both the N-PASS and PAT are designed for assessment of neonates so it is possible that these

physiological variables are necessary for the pain assessment in neonates because they are too young to display certain behavioral variables of pain and non-pain related distress. Similarly, the Modified Comfort Scale for muscle relaxed patients requires more physiological variables in its assessment because these patients are paralyzed and not able to express many of the behavioral signs of pain and non-pain related distress outlined in the original Comfort Scale (Razmus et al., 2003). Therefore, there appears to validity in the use of physiological variables for the assessment of pain, non-pain related distress, and sedation.

Recommendations for Selection of Scales

This review identified 17 instruments as appropriate measures of physiological and behavioural cues of pain, non-pain related distress, sedation, and/or analgesia in mechanically ventilated PICU and NICU patients and involved rigorous evaluation of study quality. The author's recommendations follow and are summarized in Table 2-11.

Of these 17 instruments only three of the instruments asssess all four of these components: the Comfort Scale, the Comfort-B Scale, and the N-PASS. All three scales are easy to use, non-invasive, and exhibit strong psychometric properties (Ambuel et al., 1992; Hummel et al., 2008; Hummel et al., 2010; Ista et al., 2005; M. Van Dijk et al., 2000). However, the Comfort Scale emerges as the most clinically useful of all 17 instruments in the assessment of post-operative pain, non-pain related distress, and adequacy of analgesia and sedation in mechanically ventilated pediatric patients. Specifically, the Comfort Scale is not age dependent, like the N-PASS , which can can only be used in assessment of neonates, and it

does not exclude physiological variables which, as stated above, have definite value in the assessment of postoperative pain, non-pain related distress, analgesia, and sedation. Additionally, the Comfort Scale does not have the grey area in scoring that the Comfort-B scale has, and consequently does not require the use of a global rating scale to compensate for this grey area (see Table 2-3) (Ista et al., 2005).

The Comfort scale, however, does have it limitations. It intended to be used to evaluate pain during intermittent painful procedures and consequently is not useful in assessing procedural pain (Marx et al., 1994; M. Van Dijk et al., 2000). It is also not appropriate for the assessment of chronic or long-term pain, and does not adequately assess pain, non-pain related distress, analgesia, and sedation in mechanically ventilated patients who are muscle relaxed due to use chemical paralysis (Marx et al., 1994).

For the assessment of sedation only, the Comfort Scale is recommended for the same reasons mentioned above.The State Behavioral Scale (SBS) does appear to have some clinical value in the assessment of sedation in patients aged 6 weeks to 6 years who were physiologically stable and not rated in pain, given its high reliability and strong construct validity. It, therefore, is an appropriate second choice for the assessment of sedation. It does have one main disadvantage; the assessment of sedation with this scale requires progressive stimulation (Curley et al., 2006). Furthermore, it is not appropriate for the assessment of sedation in mechanically ventilated patients who are muscle relaxed due to use chemical paralysis. More research on the SBS, is recommended in order to substantiate its

use in the assessment of sedation. The other six scales identified in part one of this review are not recommended because there is limited research supporting their reliability and validity.

Although the FLACC Scale and the MAPS have no clinical usefulness is the assesment of non-pain related distress or sedation, they are recommended for the assessment of procedural pain and other brief acute pain as well as postoperative pain based on their high reliability and validity in these settings (Johansson & Kokinsky, 2009; Manworren & Hynan, 2003; Ramelet et al., 2007; Ramelet, Rees et al., 2007; Voepel-Lewis et al., 2010). The main difference between the FLACC scale and the MAPS scale is that the FLACC scale includes only behavioral variables and the MAPS scale employs both behavioral and physiological variables. Modified FLACC should be used in the assessment of mechanically ventilated pediatric patients, instead of FLACC in its original form, because it breaks down the parameter of 'cry' into separate descriptions for ventilated and non-ventilated children (Table 2-3) allowing for more effective assessment of the ventilated children (Johansson & Kokinsky, 2009). Both of these scales are not appropriate for the assessment of chronic or long term pain or the assessment of pain in mechanically ventilated patients which are muscle relaxed. Research is, therefore, needed on instruments that asssess chronic or long term pain in this population.

Part one of this review identified only one scale that assessed sedation in mechanically ventilated pediatric patients who are muscle relaxed, the PICU Modified Scale for Muscle Relaxed Patients. The scale, however, has only been

assessed in one study and did not exhibit strong concurrent validity (possibly due to small sample size) (Razmus et al., 2003). More research is needed on this scale before any recommendations for or against its use can be determined.

Similarly, only one instrument for the assesment of delirium (a component of non- pain related distress), was identified in part one of this review, the PCAM-ICU. This instrument was found to have high inter-rater reliability and good criterion validity (Table 2-6, 2-5) in its initial evaluation, indicating it has promise for assessment of delirium in pediatric mechanically ventilated patients (Smith et al., 2011). More research on this scale is required, however, before it can be recommended as the first choice for assessement of delirium in this population. It should be noted, however, that delirium is a relatively new concept in PICU settings, so it is likely that more scales will be developed and published.

Two scales were identified as specific to the assessment of neonates, the N-PASS and the PAT. Both exhibited strong psychometric properties in the initial assessments (Table 2-6, 2- 7) but the N-PASS appears to be more clinically useful than the PAT as it can assess pain (procedural and non-procedural pain), agitation, and sedation (Hummel et al., 2008; Hummel et al., 2010; Spence et al., 2005).

Efficacy and Effectivness of Objective Systematic Scales in Practice

There is limited research on the efficacy of these instruments in practice. Only three studies examined the effect of systematic scales for the assessment of pain, non-pain distress, sedation, and analgesia on patient outcomes. In all three studies the scales were used to guide the rater in following a sedation protocol (Alexander et al., 2002; Ista et al., 2009; Jin et al., 2007). Therefore, differences in

patient outcomes may be related to the sedation protocol used with scale rather than the scale itself. Further research, however, is needed to assess the effect of these scales on patient care outcomes, and where possible these scales should be assessed without the co-intervention of a sedation protocol. None of the studies assessed fluctuations of analgesia and sedation between nursing shifts. Fluctuations in analgesia and sedation predispose the patient to adverse reactions and over- or under-sedation. Length of ventilation and ICU stay, and long-term psychological and neurodevelopmental factors may also be affected by fluctuations in analgesia and sedation (Carnevale & Ducharme, 1997).This may be a more conclusive method of assessing how these scales affect the amount of analgesia and sedation patients receive when they are assesed using these scales as they are meant to standardize and increase the objectively of the assessment. The limited number of studies assessing the effect of using these scales on patient makes it difficult to determine their effectiveness in the real world. Furthermore, none of these studies looked the long term effect of using these scales. More research is therefore needed on the efficacy of using the recommended systematic assessment scales in practice.

The Knowledge Translation Problem

It has been argued that the core challenge in improving management of pain in pediatrics is a knowledge translation challenge or failure to put what we already know into use rather than a knowledge deficit or lack of research (Scott-Findlay & Estabrooks, 2006). The same can be said for the management of non-pain related distres, and the assessment of adequacy of analgesia and sedation. These

instruments cannot be shown to have efficacy or effectiveness in practice if they are used consistently by healthcare team in the management of pain and non-pain related distress, and in the assessment of adequacy of analgesia and sedation. Simply giving people the information will not cause them to change their practice. Instead individuals must be given this information about the use of these instruments and how to use them, decide to adopt them in their practice, implement them in their practice, and then see confirmation that this adoption improved patient care and patient outcomes (Scott-Findlay & Estabrooks, 2006). Based on this information the challenge of knowledge translation needs to be addressed to ensure not only that these scales are used, but that they are used properly. Methods for improving knowledge translation, therefore, is also an important component of future research

Conclusion

In conclusion, of the 17 instruments evaluated, the Comfort Scale has the greatest clinical utility in the assessment of pain, non-pain related distress, and sedation in mechanically ventilated pediatric patients. Modified FLACC and the MAPS are more appropriate, however, for the assessment of procedural pain and other brief painful event, and N-PASS is the most appropriate tool for the assessment of neonates. When choosing an instrument to use in PICU or NICU, clinicians should choose a scale or multiple scales that are easy to use in assessing the condition(s) of concern whether it be acute pain, procedural pain, agitation, sedation, delirium, analgeisa, or a another form of non-pain related distress. Further research is required on the use of SBS in the assessment of sedation,

instruments for the assessment of sedation and/or pain in pediatric patients who are mechanically ventilated and muscle relaxed, and in the assessment of delirium in mechanically ventilated pediatric patients. Lastly, instruments that assess chronic and long term pain in mechanically ventilated pediatric patients may be needed for patients who require long term mechanical ventilation. Effective knowledge translation is essential in the implementation, adoption, and successful use of these instruments.

Figure 2-1. Flow Diagram of Review #1

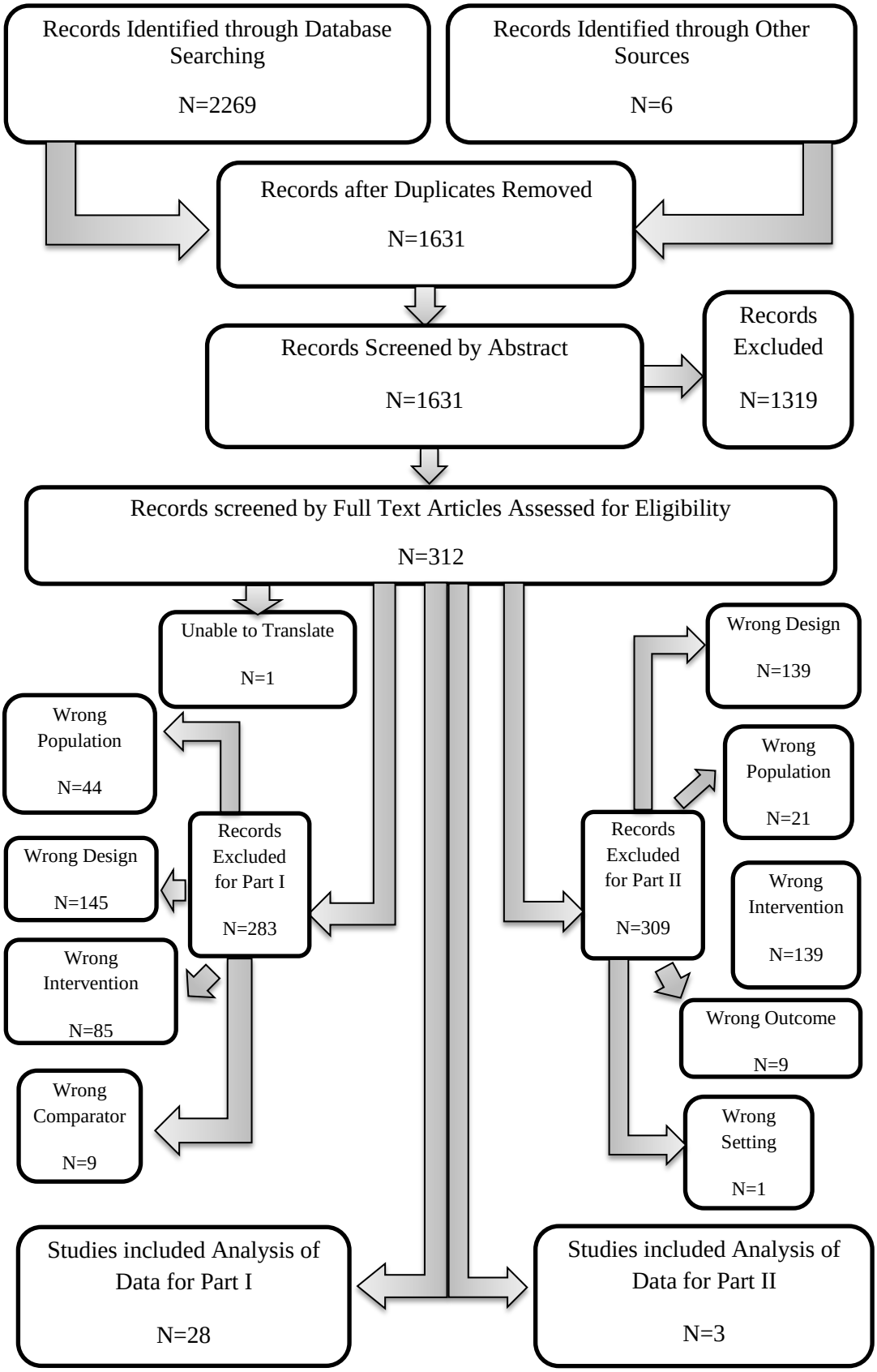


Table 2-1

Search Strategy and Key Word

	Population: Mechanically Ventilated Pediatric Patients	**Environment: Pediatric Intensive Care**	**Intervention: Systematic assessment scale**	**Pain/analgesia**	**Sedation**	**Pain and Non-Pain related Behaviours**	**Pain and Non-Pain related Emotions**
	mechanically ventilate OR ventil* AND	AND	AND	AND	OR	OR	OR
OR	infant*	pediatric critical care	diagnos*	pain	sedat*	crying	anxiety
OR	neonate*	paediatric critical care	measur*,	agony	tranquility		fear
OR	newborn*	paediatric intensive care	evaluat*	aching	distress	physical tension	dysporia
OR	pediatric	pediatric intensive care	tool*	throbbing	torpor	restless*	anger
OR	paediatric	PICU	battery	hurt*	restfulness	seeking or avoiding touch	depression
OR	child		instrument*	sting	peaceful*	avoidance	unhappiness
OR	children		inventor*	soreness	calm*	withdrawal	frustration
OR	teen*		checklist	tenderness	hypnotic	resistance	nervousness
OR	adolescent*		indicator	discomfort		agitation	apprehension
OR	youth		score*	comfort			worry
OR	juvenile		scoring	analgesic*			concern
OR			questionnaire				unease
OR			series				aggravate
OR			scale*				irritat*
OR			protocol*				terror
OR			appraisal				dread
OR			assessment				* panic
OR							alarm
OR							apprehension
OR							anxious*
OR							delirium

Table 2-2

Definition of psychometric properties by domain

Term			***Definition***
DOMAIN	**Measurement property**	Aspect of measurement property	**Sources:** (Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010b; Waltz, Strickland, & Lentz, 2010)
RELIABILITY	The degree to which the measurement is free from error. The extent to which scores for patients that have not changed are the same for repeated measurement		
	Internal consistency		Consistency of performance of one group of individuals across items of a single measure. The extent to which scores for patients have not changed using different sets of items from a single measure
	Test-retest reliability		Is used to determine the quality of measures and other methods designed to assess characteristics known to be relatively stable over time. Administer to same group of subjects under same conditions at different times
	Inter-rater reliability		Refers to the degree of agreement among different raters in assigning scores to the same objects or responses
	Intra-rater reliability		The consistency in which one rater assigns scores to a single set of responses on two occasions
	Measurement Error		The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured. Random error would be misreading a measurement. Systematic error there is an error in the measuring device itself
VALIDITY	The degree to which an instrument measures what it is supposed to measure.		
	Content validity		The degree to which the content of an instrument is an adequate reflection of the construct to be measured. The focus is on determining whether the items sampled for inclusion on the tool adequately represent the domain of content addressed by the instrument and relevance of the content domain to the proposed interpretation of scored obtained when the measure is employed.. When a domain is adequately defined, objectives that clearly that represent that domain are clearly explicated, an exhaustive set of items to measure to measure each objective is constructed, and then random sampling is employed to select a subset of items from this

Term			*Definition*
DOMAIN	**Measurement property**	Aspect of measurement property	**Sources:** (Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010b; Waltz, Strickland, & Lentz, 2010)
			larger pool for inclusion on the instrument, the probability that the instrument will have adequate content validity is high.
		Face validity	The degree to which the items of an instrument indeed look as though they are an adequate reflection of the construct to be measured Face validity is not validity in the true sense and refers only to the appearance of the instrument to the lay person; that is if upon cursory inspection, and instrument appears to measure what the test constructor claims to measure it is said to have face validity
	Criterion Validity		The degree to which the scores of an instrument are an adequate reflection of a gold standard. How well each item measures its respective objective (item objective congruence) and helps classify persons or objects into their appropriate category (item discrimination). Note: for the purpose of this review acceptable gold standards are clinician team opinions using visual analog or global rating scales (e.g., NISS, VAS pain).
		Concurrent	the extent to which a measure may be used to estimate an individual's present standing on the criterion
		Predictive	the extent to which future level of performance on a criterion can be predicted knowledge of performance on a prior measure
	Construct Validity		The degree to which scores of an instrument are consistent with hypotheses (e.g., with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the instrument validly measures the construct to be measured. Is the extent to which relationships among items included in the measure are consistent with the theory and concepts operationally defined
		Convergent	Two basic principles underlying the multi-trait approach are 1. That different measures of the same construct should correlate highly with each other (the convergent validity principle), and 2. Those measures of different constructs should have a low correlation with each other (the discriminant validity hypothesis)

Term			*Definition*
DOMAIN	**Measurement property**	Aspect of measurement property	**Sources:** (Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010b; Waltz, Strickland, & Lentz, 2010)
		Structural	The degree to which the scores of an instrument are an adequate reflection of the dimensionally of the measured construct
		Hypotheses-Testing	An approach to construct validity. The investigator uses theory or conceptual framework underlying the measure's design to state hypotheses regarding the behavior of individuals with varying scores on the measure, gathers data to test the hypotheses, and makes inferences on the basis of the findings regarding whether the rationale underlying the instrument's construction is adequate to explain the data collected.
		Cross-cultural	The degree to which the performance of items on a translated or culturally adapted instrument are an adequate reflection of the performance of items of the original version of the instrument
RESPONSIVENESS	The ability of an instrument to detect change over time in the construct to be measured		
INTERPRETABILITY	Is the degree to which one can assign quantitative meaning – that is clinical or commonly understood connotations- to an instrument's quantitative scores or change in scores. It is not considered a measurement property, but an important characteristic of an instrument		

Table 2-3

Description of Instruments

		Instrument Structure						
Name of Instrument **N=17**	**Study** **(Original first author of scale)**	**Number of items**	**Parameters Assessed**	**Sub-parameters dependent on population**	**Range of response items**	**Scoring**	**Highest / lowest possible score**	**Meaning of score**
Instruments which assess Pain, Non-Pain, Related Distress, and Sedation								
COMFORT scale	Ambuel (Ambuel et al., 1992) (same)	8 items	Alertness; Calmness; Respiratory response; Movement; Mean arterial pressure; Heart rate; Muscle tone; Facial expression	none	8 numerical (scored from 1 to 5)	summed	8 to 40	Scores of: 8-16 =over-sedated 17-26= optimally sedated 27-40= inadequately sedated *
COMFORT -B Scale	van Dijk, (M. Van Dijk et al., 2000) (Ambuel) (Ambuel et al., 1992)authored this dimension originally as	6 items	Alertness; Calmness; Muscle tone; Movement; Facial tension; Respiratory response or cry	Cry assessed in non-ventilated children, respiratory response assessed in ventilated children	6 numerical (scored from 1 to 5)	summed	8 to 40	Under-sedation: COMFORT-B scores of ≥23 or 11–22 (also called grey area) in combination with NISS of 1. Adequate sedation: scores of 11–22 (grey area) in combination with NISS of 2.. Over-sedation- scores of 6–10; T-B scores 11–22 in combination with

		Instrument Structure						
Name of Instrument **N=17**	**Study** **(Original first author of scale)**	**Number of items**	**Parameters Assessed**	**Sub-parameters dependent on population**	**Range of response items**	**Scoring**	**Highest / lowest possible score**	**Meaning of score**
	part of the Comfort Scale)							NISS of 3.** 12 for no pain; 17 for pain (median scores)***
Neonatal Pain, Agitation and Sedation Scale (N-PASS)	Hummel (Hummel et al., 2008) (same)	5 items	crying/irritability, behavior state, facial expression, extremities/tone and vital signs (heart rate, respiratory rate, blood pressure and/ oxygen saturation)	none	5 numerical (scored from -2 to 2)	summed	-10 to +10	A high pain/ agitation score indicates more frequent or intense behaviors, and a low sedation score indicates a decreased response to stimulation, or a deeper level of sedation. From N-pass website: 'Deep sedation' goal score of -10 to -5 'Light sedation' goal score of -5 to -2
Instruments which assess Sedation Only								
Ramsay Sedation Scale / Modified Ramsay Scale	Kushahara (Kusahara et al., 2005) (Ramsay(Ramsay, Savege,	1 item	Calmness/ agitation, physical movement	Modified Ramsay takes into account acceptance of mechanical ventilation	6 levels	Choose the level that matches the patient	1 to 6	Level 1 is awake and Level 6 is Asleep and unresponsive

		Instrument Structure						
Name of Instrument **N=17**	**Study** **(Original first author of scale)**	**Number of items**	**Parameters Assessed**	**Sub-parameters dependent on population**	**Range of response items**	**Scoring**	**Highest / lowest possible score**	**Meaning of score**
	Simpson, & Goodwin, 1974))							
The Vancouver Sedative Recovery Scale (VSRS)	Macnab (Macnab et al., 1991) (same)	12 items	Response; Eye appearance and function; Body movement	none	12 numerical (scored from 0 to 4 or 0 to 2 or 0 to 1)	summed	0 to 22	higher scores indicate more alertness
Hartwig scale	Hartwig (Hartwig et al., 1991) (same)	5 items	motor function, mimic ability, eye-opening, toleration of ventilation and reactions to painful measures, i.e. tracheal aspiration	none	5 numerical (scored from 1 to 5)	Summed	8 to 25.	8 points indicated deepest sedation, whereas 25 points indicated insufficient sedation. . Reliable sedation was achieved when a total of 15-18 points was reached.
Sedation Scale	Parkinson (same) (Parkinson et al., 1997)	1 item	Response to suction, agitation, movement	none	5 possible choices	Choose the description that matches the pt.	1 to 5	Description given for each of the 5 possible ratings.
No name-Assessment	Hughes (Hughes et	8 items	Level of consciousness; Pattern	Verbal response and motor	4 (scored from 1 to 3, 1	summed	Highest children-	Patients were considered fully conscious if they had

		Instrument Structure						
Name of Instrument **N=17**	**Study** **(Original first author of scale)**	**Number of items**	**Parameters Assessed**	**Sub-parameters dependent on population**	**Range of response items**	**Scoring**	**Highest / lowest possible score**	**Meaning of score**
of Level of Sedation	al., 1994) (same)		of respiration; Spontaneous respiratory rate; Presence of cough in response to suction; Eye opening, and verbal and motor responses to stimuli	response have subscales for assessing infants or children	to 4, 1 to 5 , or 1 to 6), 1 (yes/no), 2 (dichotomous)		19 lowest-5 highest infants-15 lowest-5	a score of 17 out of 19 (children) or 13 out of 15 (infants), with a normal respiratory pattern and a respiratory rate of at least 20 breaths/min., and the presence of a cough in response to suction
Penn State Children's Hospital Sedation Algorithm (PSCHSA)	Popernack (Popernack et al., 2004) (same)	1 item	Alertness/ Level of consciousness, physical movement/motor response, respiratory response, response to suctioning, tolerance to care	none	6 levels (choice the level for which the description matches the pt.)	Patient fits into a specific sedation level	Level 1 to 6	Each level has a description of sedation
State Behavioral Scale (SBS)	Curley (Curley et al., 2006) (same)	6 items	respiratory drive, response to ventilation, coughing, best response to stimulation, attentiveness to care provider, tolerance to care, consolability, and movement after	none	6 descriptive	matching pt.'s response to appropriate score	-3 to +2	The higher the score the more agitated the patient is, the lower the score the more unresponsive the patient is.

		Instrument Structure						
Name of Instrument **N=17**	**Study** **(Original first author of scale)**	**Number of items**	**Parameters Assessed**	**Sub-parameters dependent on population**	**Range of response items**	**Scoring**	**Highest / lowest possible score**	**Meaning of score**
			consoled.					
Instruments which assess Pain and Analgesia Only								
FLACC	Voepel-Lewis (Voepel-Lewis et al., 2010) (Merkel (Merkel et al., 1997))	5 items	facial expression, leg movement, bodily activity, cry or verbalization, and consolability	none	5 numerical (scored from 0 to 2)	summed	0 to 10	0- no Pain 10- max pain
Modified FLACC	Johansson (Johansson & Kokinsky, 2009) (same)	5 items	facial expression, leg movement, bodily activity, cry or verbalization, consolability	Cry – ventilated or non-ventilated	5 numerical (scored from 0 to 2)	summed	0 to 10	0 no pain; 10 worst pain, 0 for over-sedated and adequately sedated; 4 for under-sedated; 0.5 no pain; 3.5 for pain
Multidimensional Assessment Pain Scale (MAPS)	Ramelet (Ramelet et al., 2007) (same)	5 items	Vital signs HR and/or BP, Breathing pattern, Facial expressions, Body movements, State of arousal	none	5 numerical (scored from 0 to 2)	summed	0 to 10	with a minimum score of 0 (no pain) and a maximum score of 10 (extreme pain) **

		Instrument Structure						
Name of Instrument **N=17**	**Study** **(Original first author of scale)**	**Number of items**	**Parameters Assessed**	**Sub-parameters dependent on population**	**Range of response items**	**Scoring**	**Highest / lowest possible score**	**Meaning of score**
Pain Assessment Tool (PAT)	Spence (Spence et al., 2005) (Hodgkinson (Hodgkinson, Bear, Thorn, & Van Blaricum, 1994))	10 items	Posture/tone , Cry , Sleep pattern , Expression , Color , Respirations , Heart rate, Oxygen saturation, Blood pressure, Nurse's perception	none	10 numerical (scored from 1 to 2)	summed	0 to 20.	Scores greater than 5 indicate that comfort measures such as tactile soothing, use of pacifier, and repositioning should be instituted, and scores greater than 10 require adjustment of the analgesia dose.
Cardiac Analgesic Assessment Scale (CAAS)	Suominen (Suominen et al., 2004) (same)	4 items	pupil size, heart rate, blood pressure and respiration, gross motor movement)	none	4 numerical (scored from 0 to 2)	summed	0 to 8	The highest aggregate score possible for any patient is 8, and when a patient scored 4 or more, pain was assumed.
Instruments which assess Sedation in Muscle Relaxed Patients Only								
PICU Modified Comfort Sedation Scale for the muscle	Razmus (Razmus et al., 2003) (same)	6 items	heart rate, blood pressure changes, skin perfusion, pupil size, and response to auditory and tactile	none	6 numerical (scored from 1 to 5)	summed	6 to 30	target range 13-20 for optimal sedation

		Instrument Structure						
Name of Instrument **N=17**	**Study** **(Original first author of scale)**	**Number of items**	**Parameters Assessed**	**Sub-parameters dependent on population**	**Range of response items**	**Scoring**	**Highest / lowest possible score**	**Meaning of score**
relaxed patient			stimuli					
Instruments which assess Delirium (a component of non-pain related distress) Only								
Pediatric Confusion Assessment Method for Intensive Care Unit (pCAM-ICU)	Smith (Smith et al., 2011) (same)	4 Steps rather than items	Acute change or fluctuating course of mental status; Inattention; Altered level of consciousness; Disorganized thinking	none	3 numerical (scored from 0 to 2)	Findings of each step indicates delirium or no delirium	N/A	Positive delirium diagnosis requires both feature 1 (acute change or fluctuating course of mental states) and feature 2 (inattention) with either positive feature 3 (altered LOC) or feature 4 (Disorganized thinking)

*Scoring of Comfort Scale determine by Mar x et al. (Marx et al., 1994)** Scoring of Comfort-B for Sedation determined by Ista et al.(Ista et al., 2005);
***Scoring of Comfort-B for pain determined by Johansson et al. (Johansson & Kokinsky, 2009)

Note: N-PASS and PAT are met to be used for the assessment of neonates

Table 2-4

Comparison of Behavioral and Physiological Variables assessed by the Instruments for Measuring Pain, Non-Pain Related Distress, Sedation, and Analgesia in Mechanically Ventilated PICU Patients

	Behavioral Variables																		Physiological Variables					
Name of Scale	Alertness/ Level of Consciousness	Calmness/Agitation	Respiratory Response	Physical Movement/ Motor Response	Posture/ Muscle Tone	Facial Tension/ Mimic	Response to Auditory/visual/ noxious stimuli	Response to Suctioning	Eye Opening	Cry	Verbal response	Spontaneous cough	CONSOL ability	Attentiveness to care provider	Acute Change or Fluctuating Course of Mental Status	Inattention	Disorganized Thinking	Tolerance to care	Nurse's Perception of Pain	Mean Arterial Blood Pressure	Heart Rate	Respiratory rate	Oxygen Saturation	Pupil size /Reactivity	Color/ Skin Perfusion
Hartwig			X[1]	X		X		X[2]	X																	
N-PASS	X[3]	X[3]		X[3]	X	X				X[4]											X	X	X	X		
PAT		X[5]	X[6]		X	X				X										X	X		X	X		X
State Behavioral Scale			X[7]	X[8]			X	X[9]				X	X	X				X								
CAAS				X								X			X	X	X			X	X			X	
pCAM-ICU	X	X																								
COMFORT Scale	X	X	X[1]	X	X	X															X	X				
Modified COMFORT scale for Muscle Relaxed pts.							X[10]	X[11]													X	X			X	X
COMFORT-B	X	X	X[1]	X	X	X				X[12]																
FLACC				X[13]		X				X			X													
Modified FLACC				X[13]		X				X[14]			X													
Penn State Children's Sedation	X	X	X[1]	X				X										X								

	Behavioral Variables																			Physiological Variables					
Name of Scale	Alertness/ Level of Consciousness	Calmness/Agitation	Respiratory Response	Physical Movement/ Motor Response	Posture/ Muscle Tone	Facial Tension/ Mimic	Response to Auditory/visual/ noxious stimuli	Response to Suctioning	Eye Opening	Cry	Verbal response	Spontaneous cough	CONSOL ability	Attentiveness to care provider	Acute Change or Fluctuating Course of Mental Status	Inattention	Disorganized Thinking	Tolerance to care	Nurse's Perception of Pain	Mean Arterial Blood Pressure	Heart Rate	Respiratory rate	Oxygen Saturation	Pupil size /Reactivity	Color/ Skin Perfusion
Algorithm																									
Vancouver Sedative Recovery Scale	X			X[15]		X	X[16]		X[17]																
Hughes (unnamed)			X[18]	X[19]				X	X	X[20]	X[21]											X			
Parkinson (unnamed)		X		X				X[22]																	
MAP	X		X[23]	X		X														X	X				
Revised MAP	X		X[23]	X		X														X	X				
Ramsay		X		X[24]																					
Modified Ramsay		X	X[25]	X[6]																					

Footnotes

FN1	Includes response to ventilator	FN14	For ventilated children cry is assessed by facial expression
FN2	Facial expression, movement, opposition, coughing	FN15	Breaks into 4 components of movement
FN3	Part of behavior state	FN16	Non-specific stimuli
FN4	Includes irritability	FN17	Actual description of eyes and eye movement
FN5	Called sleep pattern	FN18	Shallow/ normal respirations, pattern of breathing
FN6	Presence apnea	FN19	In response to stimuli
FN7	Respiratory drive and response to ventilation	FN20	In infants
FN8	Movement after consoled	FN21	Assessed instead of cry in older children
FN9	Coughing	FN22	Movement, agitation, cough
FN10	Auditory	FN23	Development of respiratory distress
FN11	Vital signs and tears	FN24	Response to galbellar tap
FN12	Assessed instead of respiratory response in non-ventilated children	FN25	Accepts Mechanical Ventilation
FN13	Legs movement and activity	FN26	Response to galbellar tap or loud Noise

Table 2-5

Description of the Population Assessed, Study Design, the number, and Method of Assessment involved in the Validation of the Identified Instruments for Assessing Pain, Non-Pain Related Distress, Sedation, and Analgesia in Mechanically Ventilated PICU Patient

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
Hartwig Scale	Hartwig, et al. (Hartwig et al., 1991)	-	24 ventilated PICU pts. Diagnoses- cardiac surgery, ARDS, brain damage with cerebral edema, acute renal failure with pulmonary edema, aortoventroprexia, submandibular abscess Age: 26 days- 5 yrs.	Prospective cohort study	24 assessments for to determine Inter-rater reliability	6 nurses assessing, 4 patients at the same time for reliability
Comfort Scale	Ambuel et al. (Ambuel et al., 1992)	-	37 PICU pts. receiving intermittent mandatory ventilation or continuous positive airway pressure Principal diagnoses: cardiac disease, respiratory, infectious disease, and post trauma Age Range: Newborn-204 mos., mean age -37.1 mos. $\pm$ 52.7 mos.	Prospective convenience sample	37 patients, each observed on one to two occasions, generating 50 observations	Each participant was observed concurrently by three raters, a principal investigator (PI), and a research assistant (RA) trained in use of the COMFORT scale, and an experienced intensive care nurse who was unfamiliar with the COMFORT scale and had not been caring for the pt. The PI and RA completed the Comfort score and the PICU nurse completed the VAS independently.
Comfort Scale	-	Marx et al. (Marx et al., 1994)	85 ventilated PICU pts. (Study 1 n=34, Study 2 n=30, Study 3 n=21) Diagnoses: cardiac, respiratory, other	Serial Prospective agreement cohort studies	316 observations: study 1 - 100 observations study 2- 96 observations study 3- 120	Each study consisted of a set of simultaneous, independent observations by multiple observers over a 2 min. period. The intensivist rated adequacy of sedation and an independent rater rated the Comfort Score

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
			Age: Study 1: 15.9 ± 30.9 mos. Study 2-20.1 ± 30.9 mos. Study 3- 27.5 ± 30.9 mos.,		observations	
Comfort Scale and Hartwig Scale	-	Brunow de Carvalho, et al (Brunow de Carvalho et al., 1999)	18 ventilated PICU pts. Diagnoses: cardiac disease, neurologic disease, infectious disease, respiratory disease Mean age 16.45 mos., SD 17.27 mos. Range- 16 days-5 yrs.	Prospective cohort study	30 observations	Simultaneous independent ratings conducted by a specialist PICU physicians using the Comfort scale and Hartwig scale
Comfort Scale/Comfort-B Scale	-	van Dijk, et al. (M. Van Dijk et al., 2000)	158 PICU pts. 19 received short term ventilation (<36 hrs.) 43 received prolonged ventilation (> 36 hrs.) 96 non-ventilated Diagnoses: Congenital and acquired anomalies requiring surgery Age 0-3 yrs.	Prospective convenience sample	158 cases - 13 assessments each.	Bedside nurses observed each child for 2 min at bedside.
Comfort Scale	-	van Dijk et al. (M. van Dijk et al., 2001)	204 post-operative PICU pts. Neonates (>35 weeks) n=66, 77 % ventilated Young infants (1 mo. to 6 mo.) n=67, 33% ventilated	Randomized double blind clinical trial	Pain assessment was performed at baseline, after return to the PSICU, and every 3 hours during the first 24 hours postoperative.	During the 2 min interval period needed to assess the COMFORT scale, 6 HR and 6 MAP values are registered from the monitor and compared with the baseline range of the child, assessed prior to surgery. VAS pain was also assessed. -. Infants were rated by at least three different nurses during a 24- hour period. -. Pain assessment was performed prior to handling of the child and morphine or placebo bolus

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
			Infants (6 mos. to 12 mos.) n=31, 13 % ventilated Toddlers (1-3 yrs.) n= 40, 25 % ventilated		(~ 8 observations per child?)	administration (when children were considered to be in pain; VAS > or equal to 4) at any time after surgery)
Comfort Scale	-	Carnevale et al. (Carnevale & Razack, 2002)	18 ventilated PICU pts. Diagnoses: Cardiovascular surgery, respiratory dysfunction, liver dysfunction Age < 1mo. -7 yrs.	Observational Study	514 individual completed ratings	All scores were recorded by the nurses routinely assigned to care for the children in this study. The scale was used for variable intervals with each child.
Comfort-Scale/ Comfort-B Scale	-	Ista et al. (Ista et al., 2005)	78 PICU pts. 66 ventilated, 12 non-ventilated Diagnoses: congenital, cardiac-other, respiratory failure, sepsis/septic shock, other Age 0-223 mos. Mean age 17 mos.	Prospective observational study	A total of 843 paired observations , for inter-rater reliability-40 observations	The care-giving nurse assessed the patient every 8-hr shift at set times (2, 10, and 18 hrs.) determining the NISS score before COMFORT score. Paired scores were obtained when the patient was uncomfortable (NISS =1) or when sedation medication was administered or changed. score
Comfort Scale and Ramsay Scale	-	Kusahara, et al. (Kusahara et al., 2005)	77 ventilated PICU pts. Diagnoses not stated Mean age 5.08 yr. Median age 5.6 yrs.	A descriptive and correlative study	77 children-assume at least one per pt.	The sedation scales were applied by trained observers, and at the same time the health care professionals responsible by the child care expressed their clinical opinion about the sedation level.

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
Comfort Scale	-	Grap, et al. (Grap et al., 2006)	20 ventilated PICU pts. Diagnoses not stated Mean age 3 yrs. Range 1 mo. to 14 yrs. and 7 mos.	Observational, non-experimental correlational study	40 total assessments	Each participant is observed continuously for 2 hrs. The Comfort Scale was administered at the beginning and end of the 2-hour observation period along with heart rate, respiratory rate, oxygen saturation, and blood pressure.
Comfort-B and Modified FLACC	-	Johansson, et al.(Johansson & Kokinsky, 2009)	40 ventilated PICU pts. Diagnoses: Cardiac malformations, other malformations, miscellaneous Age 0 to180 mos. median age 4 mos.	prospective observational design	119 paired assessments In 40 pts. for both modified FLACC and Comfort-B, for 7pts. there was only one bedside nurse's assessment of pain and sedation available for assessment. In these pts. there were 98 paired, 21 single assessments 20 extra pts. modified FLACC before and after boluses was scored	Two study nurses simultaneously and independently observed the patient over a full 2 min confirmed by an alarm clock. The observations took place randomly during the day, e.g. after cares, before and after analgesics. Concurrently, the two care giving intensive care nurses assigned to the patient independently scored pain using an observational visual analogue scale (VASobs). They also scored the sedation using NISS observations were performed on two to three different occasions during 1 day.
Comfort-	-	Valkenburg, et al.	466 PICU patients (218 ventilated & 248 non-	Prospective cohort study	Down syndrome	All Comfort-B, NRSOBS, and NISS scores for the included patients were retrieved from our

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
B Scale		(Valkenburg et al., 2011)	ventilated) & 76 PICU pts. with Down Syndrome (56 ventilated & 20 non-ventilated) Diagnoses: Cardiothoracic surgery, GI surgery, ENT surgery, craniofacial surgery, cardiorespiratory failure, GI/GU illness, Metabolic, trauma, infection/sepsis, other Age median (IQR) Down's group- 81days [42-273] Control group- 119 days [22-355]	comparing two groups	(1163 scores) Control (6276 scores)	PDMS, used all scores of the first 7 days of a patient's admission. If patients had been admitted more than once, the data of the longest admission were used.
Comfort-B Scale	-	Boerlage et al. (Boerlage et al., 2011)	80 ventilated PICU pts., 60 of which were surgical pts.	Observational Study	The 236 observations	In the first of the two observations, either the pain specialist or the nurse singly started a 2-min observation; after 90 secs had elapsed, the other observer started a 30-sec observation. Thus they stopped simultaneously. Both observers then independently completed the Comfort-B scoring form and assigned an additional pain rating on the 11-point Numerical Rating Scale (NRS-11) (. This was always done directly after the Comfort-B assessment, and thus in this study, it was done concurrently with the 30-sec and 2-min assessments. Roles were reversed for a second observation in another patient.
Faces, Legs, Activity, Cry, Consolability Scale (FLACC)	_	Manworren et al. (Manworren & Hynan, 2003)	22 PICU pts., 78 PACU pts., 7 surgical/trauma pts., 7-hematology/oncology pts., 33 infant unit Mean age 1 yr. 40 days, range 1 days to 34 mos.	Prospective convenience sample	147 patients	Each patient assessed and treated by a standardized method used in the hospital Time 1- nurse would calculate the pain score using FLACC -analgesic was given based on what was ordered, if more than one ordered nurse chose what to give based on her judgement

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
Faces, Legs, Activity, Cry, Consolability Scale (FLACC)	-	Voepel-Lewis et al. (Voepel-Lewis et al., 2010)	29 Adult ICU pts., (23 ventilated, 6 non-ventilated); 8 Pediatric ICU pts. (3 ventilated, 5 non-ventilated	29 Adult ICU pts., (23 ventilated, 6 non-ventilated); 8 Pediatric ICU pts. (3 ventilated, 5 non-ventilated	A total of 73 observations were obtained in 29 critically ill adults and 8 children -only 13 of these observations were in children	Observations were made by 3 intensive care unit nurses during the routine care of each patient as follows: Before administration of an analgesic, or during a painful procedure such as turning or suctioning, nurses observed the patient and simultaneously, but independently, scored pain behaviors during a 1- to 2- minute period. Two of the nurses used the FLACC tool to score pain behaviors; the third nurse used the CNPI for adults and the Comfort Scale for children. Each patient was observed again by the same nurses approximately 15 to 30 minutes after the first observation
Cardiac Analgesic Assessment Scale (CAAS)	Suominen et al. (Suominen et al., 2004)	-	Study 1 -32 ventilated PICU pts., Study 2- 37 ventilated PICU pts. All had cardiac surgery with a steronomy incision Study 1- mean age- 2.5 yrs. (0.02-19 yrs.) Study 2- mean age 2.6 yrs. (0.08-12.6 yrs.)	two prospective studies	Part 1: Four concurrent observers performed paired observations with the CAAS. Total = 64 assessments Part 2: 37 assessments	First part of the study: Four concurrent observers performed paired observations with the CAAS or the VAS once for every patient stimulus, two nursing observers independently scored the patient using the CAAS, and another two independent nursing observers (the bedside nurse and an adjacent bedside nurse) simultaneously scored the patient using the VAS. The two nursing observers who used the CAAS evaluated all the patients. Second part of the study: Patients were randomly allocated to one of two treatment groups: group A (intrathecal morphine) and group B (control) by a research team member p. After induction of anaesthesia and placement of central venous and arterial catheters, group A patients received 20 mg/kg) of intrathecal morphine for postoperative analgesia. After the stimulus the bedside nurse scored the patient using the CAAS. Three different CAAS values were analysed to study the changes in the CAAS values and response to analgesia over time: (i) CAAS value before the administration of i.v. morphine; (ii) the highest CAAS score indicating the need for i.v. morphine; and (iii) the CAAS score following the intervention. Scoring was performed at least every 2 h.

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
Multidimensional Assessment of Pain Scale (MAPS)	Ramelet et al. (Ramelet et al., 2007)	-	43 PICU pts. 30 ventilated, 13 non-ventilated All postoperative Types of surgeries: cardiac, cranial vault re-modelling, thoracotomy, Laparotomy, other Age 0 to 31 months, median age 9.5 mos.	prospective observational study	43 children up to 8 assessments by 2 independent observers (>100)	Up to eight measurements of pain using MAPS, FLACC, and VAS were performed by two independent raters - the bedside nurse and the clinical nurse on duty - over a maximum period of 48 h post-surgery. VASobs was administered first by a trained clinical nurse and senior medical registrar on duty, both with expertise in pain assessment. Immediately after that, the same clinical nurse and senior medical registrar administered the MAPS and the FLACC scores respectively. Simultaneously and independently, the bedside nurse attributed another MAPS score.
Multidimensional Assessment of Pain Scale (MAPS)	-	Ramelet, et al. (Ramelet, Rees et al., 2007)	20 post-operative PICU pts., 18 ventilated, 2 non-ventilated Diagnoses: Cardiac condition, cranial synostosis, respiratory condition, abdominal condition Age 4 days to 31 mos. Median age 7.5 mos.	Prospective convenience sample	Assessments done for up to 3 rescue doses by 2 observers- on each child	Before rescue intravenous morphine bolus was administered, pain was measured simultaneously and independently by the bedside nurse using VAS score (VASobs), the trained clinical nurse on duty using MAPS, and the nurse researcher using FLACC. Pain was measured again at 15, 30, and 60 min after the administration of the rescue bolus of Morphine. This was repeated for up to three rescue Doses for each of the participants. The nurse caring for the child attributed a VASobs score; the

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
						trained clinical nurse attributed a MAPS score; the Nurse Researcher attributed a FLACC score.
Neonatal Pain, Agitation and Sedation Scale (N-PASS)	Hummel et al. (Hummel et al., 2008)	-	46 ventilated and/or post-operative NICU pts. Surgeries- bowel resection, exploratory laparotomy for NEC, Esophageal Atresia with trachoesophageal fistula repair, PDA ligation Age 0- 100 days old	Prospective psychometric evaluation	72 data sets (before and after intervention assessments)	The N-PASS tool was independently and concurrently administered when two data collection nurses were present, for 5 to 10 min before and 1 hour after analgesic and/or sedative administration on an infant with a pain score over 3. One nurse also administered the PIPP tool concurrently with the N-PASS.
Neonatal Pain, Agitation and Sedation Scale (N-PASS)	-	Hummel, et al. (Hummel et al., 2010)	42 NICU pts. 11 ventilated, 21 non-ventilated Diagnoses not given Age 1 to 30 days	Prospective psychometric evaluation, randomized crossover design	Fifty-nine observations	The N-PASS was studied with routine heel stick procedures in the NICU, compared with a sham heelstick procedure. The bedside nurse randomly determined the order of events through coin toss. A researcher and a trained staff nurse scored the infant with both the sham and real heelstick procedures using the N-PASS tool. One observer also scored the infant using the PIPP tool. The infant was assessed before and during each procedure.
No name	Hughes et al. (Hughes et al., 1994)	-	53 ventilated PICU pts. Diagnoses: croup, bronchiolitis, epiglottis, surgical, pneumonia, asthma, burns, other Age 6 days to 11 yrs. Mean & median age of pts. <12 mos.-5.3 mos.	Prospective cohort study	38 joint assessments	Assessments carried out by the research nurse and the individual nurse caring for each patient. Assessment were carried out at intervals of 3 hours as midazolam was withdraw; every hour for 6 h when midazolam was stopped; and subsequently every 3 h until the patient was full conscious

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
			and 6 mos., respectively Mean and median age of pts.> 12 mos. 3.2 yrs. and 2 yrs. respectively			
Pain Assessment Tool (PAT)	Spence et al. (Spence et al., 2005)	-	144 NICU pts. 48 ventilated (17 term, 31 pre-term) 96 non-ventilated 65 pre-term, 79 term, 84 had undergone neonatal surgery Median age at assessment 22 days, range0-182 days	Prospective reliability/ validity study	144 assessments	Twelve clinical nurses collected data throughout the study. A third observer assessed the infants' CRIES scores. The mother's perception of her infant's pain or discomfort was collected using the VAS same time if the mother was present.
Pediatric Confusion Assessment Method for Intensive Care Unit (pCAM-ICU)	Smith et al. (Smith et al., 2011)		146 PICU pts., 17 ventilated, 129 non-ventilated Diagnoses: Acute lung injury, Avascular malformation, Brain mass, Cardiogenic shock. Congenital heart repair. Craniotomy, Encephalitis/meningitis, Endocrine/metabolic. Hydrocephalus. Orthopedic trauma, Overdose/poisoning, Septic shock, Status asthmaticus, Status epilepticus, Stroke , Traumatic brain injury, Vascular trauma Mean age 12.3 yrs. (SD-	prospective observational cohort study	146 blinded, paired assessments	The blinded paired assessments included patients' evaluation with the pCAM-ICU either before or after full diagnostic psychiatric evaluation for delirium. pCAM-ICU assessed by a member of the pCAM-ICU team (two intensivists, one NP, and 3 RNs), psychiatric assessment done by psychiatrist using DSM criteria

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
			3.4 yrs.)			
Penn State Children's Hospital Sedation Algorithm (PSCHSA)	Popernack et al. (Popernack et al., 2004)	-	PICU ventilated pts. (N not stated; pre-PSCHA % -39.8 ± 5.4, post- PSCHA 60.8 ± 2.4) Median age Pre-PSCHSA-31.4 mos. Median age Post-PSCHSA-35 mos.	Prospective, observational study with historical controls.	Not stated	After intubation, the sedation level is established by the team to create an individualized patient behavior goal. With appropriate medications prescribed, the nurse uses clinical assessment skills to administer pharmacological and age-appropriate psychological support to achieve the established goal.
PICU Modified Comfort Sedation Scale for the muscle relaxed patient	Razmus, et al. (Razmus et al., 2003)	-	20 ventilated PICU pts. Age and Diagnosis not stated. All received muscle relaxants.	Prospective cohort study	40 rating for modified Comfort and 40 for adequacy of sedation scale	each child was evaluated simultaneously by 4 raters, 2 using the Likert like adequacy of sedation scale and 2 using the modified Comfort scale
Sedation Scale (no name)	Parkinson et al. (Parkinson et al., 1997)	-	44 ventilated PICU pts. Diagnoses: Croup, pneumonia, Bronchiolitis, epiglottis, whooping cough, convulsions, sepsis, burns, asthma Age- 1 day to 15 yrs.	Pilot study and then a randomized controlled trial	154 assessments	The level of sedation was recorded as part of the routine nursing observations of the patients. Independent assessment was also carried out by the Research Nurse. Assessments were carried out when neuromuscular blocking agents were not used or when the effect of boluses had worn off.
State Behavioral Scale (SBS)	Curley et al. (Curley et al., 2006)	-	91 ventilated PICU pts. Diagnoses: pulmonary, cardiac, neurologic, infectious disease/sepsis, congenital anomaly,	Prospective psychometric evaluation	198 paired assessments	A pair of trained pediatric critical care nurse evaluators simultaneously and independently conducted state behavioral assessments of each intubated, mechanically ventilated pediatric patient in the sample. Data collection was conducted at a time when the bedside nurse was

Scale	Original Report	Validation Study	Population	Study Design	Number of Assessments	Assessment
			gastrointestinal Median age- 18 mos. Interquartile range 4.4-34.5 mos.			completing planned care when two evaluators were available. If endotracheal extubation was planned, the patient was assessed just before the procedure. Observed undisturbed, then with stimulus provided, and then after repositioning. NRS rating was also recorded.
The Vancouver Sedative Recovery Scale (VSRS)	Macnab et al. . . .(Macnab et al., 1991)	-	91 pediatric patients in the post anesthetic recovery (PAR) and ICU patients included postoperative cardiac patients and trauma at different stages of recovery Mean age 6.2 yrs., range 8 mos.- 17.7 yr.	Prospective Pilot Study	91 paired assessments	The VSRS was then administered simultaneously by pairs of experienced observers (either two pediatric ICU nurses, or one pediatric ICU nurse and one research assistant).

Table 2-6

Internal Consistency and Reliability of the Identified Instruments for Assessing Pain, Non-Pain Related Distress, Sedation, and Analgesia in Mechanically Ventilated PICU Patients

Scale	Original Study	Validation Study	Internal Consistency	Reliability
Hartwig Scale	Hartwig, et al. (Hartwig et al., 1991)	-	NA	***Inter-rater reliability*** : Sum of Total Score of Sedation: Mean difference of 2.5 points (SD ± 1.3 points). Evaluation of the parameters: motor response, mimic and respiration: mean difference of 1 point (SD ± 0 point) Eye Opening and aspiration: mean difference of 0.75 points (SD ± 0.5 points)
Comfort Scale	Ambuel et al. (Ambuel et al., 1992)	-	Cranach's alpha for whole scale=0.90 Evaluation by Parameter: Item-Item total correlation (Cronbach's alpha): Alertness r= .70 α= 0.88 Calmness r = .90 α= 0.86 Respiratory response r= .73 α= 0.88 Movement r= .80 α=0.87 Mean arterial pressure r= .66 α=0.88 Heart rate r=.60 α=0.89 Muscle tone= .30 α= 0.91 Facial Expression r= .81 α=0.87	***Inter-rater Reliability (expressed as spearman correlations):*** Total Comfort score, r = .84; n = 50; p < .01 Individual scale dimensions, r = .51 to .75; n = 50; p < .01 Inter-rater Reliability by parameter: Alertness r= .73 Calmness r= .69 Respiratory response r=.70 Movement r=.75 Mean arterial pressure r=.51 Heart rate r= .66 Muscle tone r=.52 Facial expression r=.51
Comfort Scale	-	Marx et al. (Marx et al., 1994)	NA	*__Inter-rater reliability:__ Study 3: Reliability of paired intensivists ratings using adequacy of sedation- r^2= 0.632 was lower than paired Comfort score ratings- r^2= 0.822
Comfort Scale and Hartwig Scale	-	Brunow de Carvalho, et al. (Brunow de Carvalho et al., 1999)	NA	NA
Comfort Scale/Co	-	van Dijk, et al. (M. Van	Cronbach α = 0.90, 0.92 and 0.92 for Comfort	***Inter-rater reliability:*** Linearly weighted Kappa (number of paired assessments) Alertness 0.74 n=302 Calmness 0.69 n=302 Respiratory

Scale	Original Study	Validation Study	Internal Consistency	Reliability
mfort-B Scale		Dijk et al., 2000)	behavioural' for the three assessments was. The loadings of the behavioural items were consistent across time and were significant. With Calmness fixed at 1.00, the other items had high loadings (0.76-0.85) with the exception of Muscle tone, with a loading of 0.51. The items were well represented by the model. Associations between the latent variables HR and VAS pain at 3 and 9 h postoperative assessments were non-significant (0.13 and 0.15, respectively) and moderate but significant at 6 h (0.24).	response 0.54 n=131 Crying 0.70 n=170 Physical movement 0.70 n=302 Muscle tone 0.66 n=302 Facial tension 0.63 n=296 Blood pressure baseline 0.93 n=232 Heart rate baseline 0.93 n=290
Comfort Scale	-	van Dijk, et al. (M. van Dijk et al., 2001)	NA	NA
Comfort Scale	-	Carnevale et al. (Carnevale & Razack, 2002)	* A factor analysis of scale items identified two principal components. Most scale items were highly loaded on the first component, except for heart rate and blood pressure. Total score correlations with : movement r= .807 calmness r= .791 facial tension r =.759 alertness r=.732 respiratory r=.719 muscle tone r= .701 HR r= .460 BP r=.347 Movement correlations with: calmness r= 709 facial tension r= .631 alertness r=.605 respiratory r= .479 muscle tone r=.567 HR r = .223 BP r=.088 Calmness correlations with facial tension r= .635 alertness r=.591 respirator r= .518 muscle tone r= .461 HR r=.264 BP r= .104 Facial tension correlations with alertness r=.543 respiratory r= .477 muscle tone r=.673 HR r= .224 BP r=.056	NA

Scale	Original Study	Validation Study	Internal Consistency	Reliability
			Alertness correlations with respiratory r= .427 muscle tone r =.496 HR r =.152 BP r= .165 Respiratory correlations with : muscle tone r= .424 HR r= .276 BP r= .193 Muscle Tone correlations with : HR r= .220 BP r= .058 Heart rate correlations with :BP= .06 - Heart rate and blood pressure were the items with the smallest correlations with the total score.	
Comfort-Scale/ Comfort-B Scale	-	Ista et al. (Ista et al., 2005)	Cronbach's α=.78, including all Item-total correlation of all Comfort items: MAP r=0.27 HR r=0 .31 Alertness .r= 0.60 Calmness r=0.65 Respiratory response or crying 0.48 Physical movement r= 0.71 Muscle tone .r=0.43 Facial tension r=0.54 The internal consistency, presented by the alpha if item deleted, increased to .80 (if MAP deleted) or .79 (if HR deleted). The Spearman's rank order correlation coefficient of HR with the behavioral items ranged from 0.18 to 0.30 and for the MAP items with the other items ranged from 0.05 to 0.20. Cronbach's alpha increased to 0.84 when both MAP and HR were excluded. In this analysis, all corrected item total correlations were ≥ 50.	***Inter-rater reliability:*** The intra-class correlation coefficient of 40 paired observations was .99 for the Comfort scale The inter-observer reliability (linearly weighted Cohen's kappa) for the COMFORT items ranged from .77 to 1.00.

Scale	Original Study	Validation Study	Internal Consistency	Reliability
Comfort Scale and Ramsay Scale	-	Kusahara, et al. (Kusahara et al., 2005)	NA	NA
Comfort Scale	-	Grap et al. (Grap et al., 2006)	NA	NA
Comfort-B and Modified FLACC	-	Johansson et al. (Johansson & Kokinsky, 2009)	NA	***Inter-rater Reliability:*** Comfort-B total score for Cohen's Kappa- 0.71 (95 %Confidence interval 0.65-0.77); modified FLACC total score= 0.63 (95 % CI 0.53- 0.72). Comfort-B item Weighted kappa (95% CI): Alertness 0.69 (0.60-0.77) Calmness 0.54 (0.42-0.66) Respiratory response 0.78 (0.64-0.92) Physical movement 0.69 (0.60-0.77) Muscle tone 0.43 (0.29-0.57) Facial tension 0.54 (0.41-0.67) FLACC item Weighted kappa (95 % CI) Face 0.61 (0.44-0.78) Legs 0.56 (0.38-0.72) Activity 0.51 (0.35-0.67) Cry (signs of crying) 0.52 (0.34-0.71) Consolability 0.60 (0.46-0.74)
Comfort-B Scale	-	Valkenburg, et al. (Valkenburg et al., 2011)	*Comfort-B item scores and internal consistency measures, by group. Down syndrome (DG) (1163 scores) Control (CG) (6276 scores) P value Standard Mean Difference (SMD) Items, mean (SD) Alertness- DG- 2.2 (1.1) CG- 2.1 (1.1) P=.10 SMD= 0.05 Calmness-DG- 1.3 (0.7) CG- 1.4 (.7) p=.50 SMD= 0.02 Respiratory response DG=1.8 (0.8) CG=1.8 (.8) p= .61 SMD= 0.02 Crying- DG- 1.3 (0.7) CG- 1.5 (1.0) p <.001 SMD= 0.23 Physical movements DG= 2.3 (1.0) CB= 2.2 (.9) p <.001 SMD= 0.12 Facial tension DG- 2.0 (0.6) CG- 2.0 (.6) P <.001 SMD= 0.12 Muscle tone DG- 2.8 (0.6) CG- 2.9 (.5) P <.001 SMD= 0.21 Corrected item--total correlation 0.54-0.72 0.57-0.76 Cronbach's a, unstandardized 0.84 0.87 Cronbach's a, standardized 0.86 0.88 Confirmatory factor analysis was applied on 347 scores in the Down syndrome group and 2067 scores in the control group. The most plausible model included equal factor loadings, equal residual variances, unequal error variances, and unequal factor means. I n this model, the item "Calmness" seemed to be correlated with "Respiratory response/crying." "Facial expression" was correlated with "Muscle tone." The fit indices were satisfactory (x2 of 101	***Inter-rater reliability***: Median linearly weighted kappa values were 0.81 (interquartile range [IQR] 0.77-0.87) for 103 nurses

Scale	Original Study	Validation Study	Internal Consistency	Reliability
			with 19 degrees of freedom, x2 /df of 5.3, SRMR of 0.03, RMSEA of 0.06). The unstandardized factor loadings varied from 0.36 for muscle tone to 0.86 for body movements in both groups.	
Comfort-B Scale	-	Boerlage et al. (Boerlage et al., 2011)	NA	NA
Faces, Legs, Activity, Cry, Consolability Scale (FLACC)	_	Manworren et al. (Manworren & Hynan, 2003)	NA	Nurses were dropped if Kappa < 0.60 initial 25 nurses >0.54 kappa score (see discussion) average weighted Kappa for 14 remaining collectors was 0.61
Faces, Legs, Activity, Cry, Consolability Scale (FLACC)	-	Voepel-Lewis et al. (Voepel-Lewis et al., 2010)	Internal consistency of the FLACC was excellent, as indicated by Cronbach α= 0.882, when all items were included. Each category correlated highly with the others (Spearman ρ=0.69-0.92; P < .001) except for the cry category (ρ = 0.18-0.36). The Cronbach α improved to 0.934 when the cry category was removed, but decreased slightly with removal of other items. In the exploratory factor analysis, 1 component accounted for 68.9% of the variance in FLACC scores; 4 items contributed to this component: face (0.86), legs (0.94), activity (0.90), and consolability (0.95).	***Inter-rater reliability***: give ICC, % and Kappa Total observations: Face: 0.90 [0.83-0.94] 80% (0.68) Legs: 0.956 [0.93-0.98] 89% (0.82) Activity:0.91 [0.85-0.95] 82% (0.70) Cry:0.67 [0.44-0.81] 89% (0.69) Consolability:0.95 [0.91-0.97] 86% (0.78) Total Score:0.96 [0.93-0.97] Adults Face: 0.93 [0.88-0.96] 84% (0.76) Legs:0.97 [0.94-0.98] 91% (0.85) Activity:0.93 [0.87-0.96] 91% (0.74) Cry: 0.72 [0.53-0.86] 93% (NA)b Consolability: 0.96 [0.93-0.98] 89% (0.82) Total Score: 0.98 [0.97-0.99] Children Face:0.74 [0.10 -0.93] 58% (0.33) Legs:0.92 [0.73-0.98] 83% (0.71) Activity: 0.76 [0.17-0.93] 75% (NA)b Cry: 0.43 [-0.97 to 0.84] 75% (NA)b Consolability: 0.87 [0.56-0.96] 75% (0.60) Total Score: 0.85 [0.52-0.96] Mechanically Ventilated Face:0.90 [0.80-0.95] 76% (0.64) Legs: 0.94 [0.88-0.97] 79% (0.74) Activity: 0.95 [0.89-0.97] 85% (0.76) Cry:0.60 [0.20-0.80] 91% (0.72) Consolability: 0.96 [0.93-0.98] 88% (0.81) Total Score: 0.96 [0.92-0.98] b- unable to calculate k values in unequal contingency tables.
Cardiac Analgesic Assessment Scale (CAAS)	Suominen, et al. (Suominen et al., 2004)	-	NA	**Inter-rater reliability**, represented by Lin's concordance correlation coefficient was very high for all the CAAS items ranging from 0.86 to 1.0. Lin's concordance correlation coefficient for total CAAS scores was 0.97 (95% CI: 0.95, 0.99) indicating that one rater's CAAS scores were in almost perfect agreement with the CAAS scores from the other rater. Of the 32 patients who were CAAS scored, 91% received the same total score from two observers. The average difference of 0.03 between the two CAAS observations for each patient was close to zero indicating no substantial

Scale	Original Study	Validation Study	Internal Consistency	Reliability
				bias between the two raters. The Bland and Altman limits of agreement provided a 95% CI for this average difference of (0.64, 0.58) indicated that the total CAAS scores were generally no more than 0.6 different. VAS had low inter-rater reliability. Lin's concordance correlation coefficient (10) of 0.61 (95% CI: 0.38, 0.83) indicates that the VAS scores by one observer do not correlate well with the VAS scores from the second observer. Twenty-six of 32 patients (81%) were assessed as having VAS scores that differed by 5 mm or more by two observers. About 28% had VAS scores that differed by more than 20 mm. As the VAS scale is a continuous scale from 0 to 100, a difference of 20 indicates a substantial variation. The average difference between the two VAS scores (VAS2) VAS1) is 3.1 mm. The Bland and Altman limits of agreement for this difference are wide indicating that a range of () 30.2, 36.3) is needed to contain 95% of the differences between the two VAS scores. Inter-rater reliability of CAAS components: Lin's Concordance Correlation Coefficient (CCC). (95% Confidence Interval)- Pupillary size 1.0 (1.0, 1.0) Heart rate 0.86 (0.76, 0.95) Blood pressure 0.97 (0.95, 0.99) Respiratory and motor response 1.0 (1.0, 1.0) The CAAS scores obtained by the two independent nurses were dichotomized to determine whether each score indicated that analgesic medication was required or not. For the 32 patients, the dichotomized CAAS scores of the two nurses indicated that for 26 (81%) patients the nurses agreed that the patient did not need treatment (CAAS < 4), and five (16%) patients needed treatment (CAAS > or equal to 4). The two nurses scoring with CAAS disagreed about whether analgesic medication was needed for only one patient. The VAS scores above 30 were interpreted to indicate that the patient required analgesic medication .Applying this cut-off to the two VAS scores obtained for each patient by two nurses indicated that the nurses disagreed about the need for treatment in seven (22%) patients. For all four comparisons of dichotomized VAS scores with dichotomized CAAS scores, kappa statistics were used to calculate the measure of agreement about the need for treatment of pain. The average of the four kappas was 0.33, which can be interpreted as fair agreement.
Multidimensional Assessment of Pain Scale (MAPS)	Ramelet et al. (Ramelet et al., 2007)	-	Examination of the correlation matrix showed no redundant items. The item total correlations ranged from 0.10 to 0.65. Overall Cronbach's alpha coefficient for the five-item MAPS was 0.68. The internal consistency test showed that Cronbach's alpha would increase to 0.86 if the item 'vital signs' was deleted.	**Inter-rater reliability** of the MAPS, represented by kappa's statistic, was good to excellent for four categories, ranging from 0.68 to 0.84. Inter-rater reliability for breathing pattern was moderate (0.54). Excellent inter-rater reliability was shown by ICC coefficients for total MAPS (r =0.91), VASobs(r = 0.89), HR (r = 0.99), SBP (r =0.98), DBP (r = 0.98), MAP (r = 0.99) and RR (r =0.98)

Scale	Original Study	Validation Study	Internal Consistency	Reliability
Multidimensional Assessment of Pain Scale (MAPS)	-	Ramelet et al. (Ramelet, Rees et al., 2007)	Internal consistency, represented by Cronbach's alpha coefficient, was 0.62 at baseline, 0.80 at 15 min, 0.37 at 30 min, and 0.26 at 60 min. If item 'vital signs' was deleted, internal consistency of the MAPS at baseline, 15, 30, and 60 min would improve r = 0.64, 0.79, 0.67, and 0.71, respectively	NA
Neonatal Pain, Agitation and Sedation Scale (N-PASS)	Hummel, et al. (Hummel et al., 2008)	-	The mean (s.d.) pain scores were 4.9 (3.4) and 5.5 (3.1) for raters 1 and 2, respectively, and the medians (range) were 5 (0, 12) and 5 (0, 12). Cronbach's alpha for the five-item pain scale was 0.82 and 0.72 for raters 1 and 2, respectively. The mean (s.d.) sedation scores were -2.78 (2.81) and -1.68 (2.23) for raters 1 and 2, respectively and the medians (range) were -2 (8, 0) and -1 (8, 0). Cronbach's alpha for the five-item sedation scale was 0.89 and 0.89 for raters 1 and 2, respectively	Pain scale. - first observation of a neonate (prior to any intervention for pain or sedation), the ICC estimates (95% confidence interval, CI) of the pain scale were 0.95 (0.90, 0.97) and 0.97 (0.95, 0.99) for a single rating and average of two independent ratings, respectively. -second observation of a neonate (after an intervention for pain or sedation), the ICC estimates (95% CI) of the pain scale were 0.92 (0.85, 0.96) and 0.96 (0.92, 0.98) for a single rating and average of two independent ratings, respectively. - F-tests, carried out to test the null hypothesis that the ICC (single rating) was 0.6 or less (versus >0.6) were highly significant (pre-intervention F=9.65, P<0.0001 and post-intervention F= 6.36, P<0.0001). Sedation scale. - ICC estimates (95% CI); first observation of a neonate prior to any intervention for pain or sedation were 0.85 (0.72, 0.92) and 0.92 (0.83, 0.96) for a single rating and average of two independent ratings (F=3.06, =0.001). -the second observation of a neonate after an intervention for pain or sedation were 0.90 (0.80, 0.95) and 0.95 (0.89, 0.97) for a single rating and average of two independent ratings (F= 4.63, P<0.0001).
Neonatal Pain, Agitation and Sedation Scale (N-PASS)	-	Hummel et al. (Hummel et al., 2010)	The mean sham heelstick pain scores were 0.83 (1.25) and 1.09 (1.71) for raters 1 and 2, respectively. The mean heelstick pain scores were 3.74 (2.35) and 3.5 (2.48) for raters 1 and 2, respectively. Cronbach's alpha for the 5-item pain scale was 0.835 and 0.887 for raters 1 and 2, respectively.	***Test- Retest Reliability:*** Scores of 13 videotaped heelstick and sham procedures 1 week apart yielded a Spearman's rho correlation coefficient of 0.874, (P<0.0001). The videotapes were scored again 1 year after the initial evaluation, yielding a Spearman's rho correlation coefficient of 0.846 (P<0.0001). ***Inter-rater Reliability:*** For the heelstick observation of a neonate, the ICC estimates (95% CI) of the pain scale were 0.86 (0.78, 0.92) and 0.93 (0.88, 0.96) for a single rating and average of two independent ratings, respectively. Similarly, for the sham observation, the ICC estimates (95% CI) of the pain scale were 0.79 (0.67, 0.87) and 0.88 (0.80, 0.93) for a single rating and average of

Scale	Original Study	Validation Study	Internal Consistency	Reliability
				two independent ratings, respectively. F-tests, carried out to test the null hypothesis that the ICC (single rating) was 0.6 or less (versus >0.6) were highly significant (heelstick F=13.70, P<0.0001 and sham F=8.56, P<0.0001).
No name	Hughes et al. (Hughes et al., 1994)	-	NA	There was a strong correlation between the assessments the nurse in charge and the research nurse (r=0.94)
Pain Assessment Tool (PAT)	Spence et al. (Spence et al., 2005)	-	NA	***Inter-rater Reliability:*** Using intraclass correlation, the reliability between pain assessors using the PAT was .85. The mean difference was 0.17 on the scale of 1 to 10 (SD 1.73). ***Measurement error:*** As the standard error of measurement (Sw) was 1.22, the repeatability of the PAT score was 3.38 (Bland & Altman, 1996). Comparing Repeatability and Validity Between Groups: The measurement error (Sw) when calculated for each group was similar: surgical = 1.29 and nonsurgical = 1.09; term = 1.16 and preterm = 1.29; on a ventilator = 1.18 and not on a ventilator = 1.24. Bland Plot for reliability: There was a small significant correlation when the mean score was plotted against the mean difference in a Bland and Altman plot (Spearman rho = .17; p < .05).
Pediatric Confusion Assessment Method for Intensive	Smith et al. (Smith et al., 2011)		NA	The pCAM-ICU was completed with very high inter-rater reliability of k coefficient= 0.96 (95% CI, 0.74 -1.0).

Scale	Original Study	Validation Study	Internal Consistency	Reliability
Care Unit (pCAM-ICU)				
Penn State Children's Hospital Sedation Algorithm (PSCHSA)	Popernack et al. (Popernack et al., 2004)	-	NA	NA
PICU Modified Comfort Sedation Scale for the muscle relaxed patient	Razmus et al. (Razmus et al., 2003)	-	NA	NA
Sedation Scale (no name)	Parkinson et al. (Parkinson et al., 1997)	-	NA	There was a very strong correlation between the assessments by the Research Nurse and the nurse in charge of the patient (r2=0.96). Discrepancies only occurred in six assessments in six patients of different ages (1 month-13 years). In each case the assessment varied by 1 level on the sedation scale only
State Behavioral Scale (SBS)	Curley et al. (Curley et al., 2006)	-	NA	**Inter-rater reliability**: not FIRST= first rating per patient Respiratory drive/response to ventilation: Weighted Kappa ALL (95% CI)= .76 (.67-.84) Weighted Kappa FIRST (95% CI)= .77 (.65-.88) Coughing Weighted Kappa ALL (95% CI)= .68(.59-.77) Weighted Kappa FIRST (95% CI)= .76 (.64-.87) Best response to stimulation Weighted Kappa ALL (95% CI)= .71 (.64-.78) Weighted Kappa FIRST (95% CI)= .65 (.54-.76) Attentiveness to care provider Weighted Kappa ALL (95% CI)=.69 (.61-.76) Weighted Kappa FIRST (95% CI)= .67 (.56-.78) Tolerance to care Weighted Kappa ALL (95% CI)= .63 (.55-.71) Weighted Kappa FIRST (95% CI)= .60 (.48-.73) Consolability Weighted Kappa ALL

Scale	Original Study	Validation Study	Internal Consistency	Reliability
				(95% CI)= .44 (.32-.55) Weighted Kappa FIRST (95% CI)= .62 (.46-.77) Movement after consoled Weighted Kappa ALL (95% CI)= .61(.52-.70) Weighted Kappa FIRST (95% CI)= .55(.43-.68)
The Vancouver Sedative Recovery Scale (VSRS)	Macnab et al.(Macnab et al., 1991)	-	To determine whether the 12 items measure the same underlying concept, we calculated Cronbach's. For both rater groups, alpha was 0.87. There were two items with low item-total correlations. Both items pertain to movement: item I (whether movement is central or peripheral, and whether it occurs spontaneously or only in response to stimuli) and item J (presence of tremor or ataxia). alpha Increases slightly when these two items are deleted. This suggests that these two items may not be measuring the same concept as the other items in the scale. However, characteristics such as ataxia, although occurring in only 5% of our assessments, were excellent indicators of residual sedation.	The inter-observer agreement for total VSRS scores as indicated by the intraclass correlation for paired observations (equivalent to intraclass correlation for a one-way random effects model) was very high: 0.90. For graphic assessment of inter-observer agreement the difference plot in Fig 1 illustrates the discrepancies between paired observers (for each of the 90 subjects) against their corresponding averages (assumed to be the closest available approximation to the 'true' level of alertness). Precision seems to be fairly consistent through the range of the scale. At first glance it appears that total VSRS scores in the range of 17 to 19 may be somewhat suspect. However, because there is a high frequency of scores in this range, one would expect a greater range of discrepancies among raters. Also, a difference of 1 or 2 points represents a smaller percentage difference for scoring at the high end. Inter-rater agreement for each of the 12 separate items in the VSRS is indicated by Cohen's kappa . These values range from 0.65 to 0.89.

*assumption made of about which instrument property was measured

Table 2-7

Validity, Responsiveness, and Interpretability of the Instruments for Assessing Pain, Non-Pain Related Distress, Sedation, and Analgesia in Mechanically Ventilated PICU Patient

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
Hartwig Scale	Hartwig, et al. (Hartwig et al., 1991)	-	NA	NA	NA
Comfort Scale	Ambuel et al. (Ambuel et al., 1992)	-	***Structural Validity:*** Correlations among the eight COMFORT Scale variables ranged from: r = 0. 12 to .88, n = 50 <u>Principal Component Factor Loadings for Comfort Scale Factors across the Behavioral Dimension (C1) and physiological dimension (C2)</u> 1.Alertness C1= 0.70 C2= N/A 2.Calmness C1= 1.02 C2= N/A 3.Respiration C1= 0.98 C2= N/A 4.Movement C1=0.95 C2= N/A 5. Facial expression C1= 0.88 C2= N/A	*__Concurrent Validity__: Total COMFORT score correlates r=0.75 distressed (n = 50, p < .01).with the clinical judgment of experienced nurses using the VAS(10cm horizontal line anchored by descriptions “absolutely calm” and “extremely distressed”	* ***Interpretability***: The eight variables have desirable measurement properties. Trained raters used nearly the complete range for each variable as well as the total Comfort score, and means for the eight variables fell near the midpoint (from 1.9 to 3.0).

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
			6. Mean arterial pressure C1= -0.38 C2= 0.46 7. Heart rate C1= N/A C2=0.69 8. Muscle tone C1=-0.55 C2=1.14		
Comfort Scale	-	Marx et al. (Marx et al., 1994)	NA	Study 1- ****Concurrent Validity*****- Intensivists' descriptive sedate adequacy rating scores correlate well with Comfort scores (r2= .662, p<0.001). Agreement between Comfort score and intensivist rating demonstrated a bias of -2.2 % and precision of 16.7 %. -Comfort scores of inadequately sedated, optimally, and excessively sedated were significantly different (all p <0.05). Minimal overlap at 16.2 and 26.8. Study 2: ****Predictive Validity*****- Significant agreement between Comfort scores and intensivist-assigned sedation adequacy scores (p <0.001). Comfort scores adequately predicted intensivist scores in 66. 1% of cases. – In the 5/16 cases where they were not predictive there was also disagreement between two ratings by 2 different intensivists	NA
Comfort Scale and Hartwig Scale	-	Brunow de Carvalho, et al. (Brunow de Carvalho et al., 1999)	***Convergent Validity:*** Observed agreement rate: 63%; p = 0.006; Expected agreement rate: 44%; Kappa coefficient: 0.345238 z= 2	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
Comfort Scale/Comfort-B Scale	-	van Dijk, et al. (M. Van Dijk et al., 2000)	NA	***Current validity:*** Correlations between VAS `before' and Comfort `behavioural' ranged from 0.64 to 0.73. Correlations between VAS `after' and Comfort `behavioural' ranged from 0.79 to 0.83. Congruent validity (concurrent) of the Comfort `behaviour' was implied by high correlations between this latent variable and VAS on all three assessments (0.96, 0.89, and 0.90, respectively) moderately with MAP (0.39, 0.34 and 0.29, respectively) and low with HR (0.13, 0.24 and 0.15, respectively). The correlations between MAP and HR were low (0.05, 0.25 and 0.19, respectively)	NA
Comfort Scale	-	van Dijk, et al. (M. van Dijk et al., 2001)	****Convergent Validity:*** -All correlations were significant (P <0.001, two-tailed).	NA	NA
Comfort Scale	-	Carnevale et al. (Carnevale & Razack, 2002)	***Structural Validity:** Stepwise multiple regression analysis for predictors of total score (i.e., the most highly correlated item was entered first). The principal finding was that 97% of the total score variance was explained by six of the eight items--that is, all items except for heart	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
			rate and blood pressure.		
Comfort-Scale/ Comfort-B Scale	-	Ista et al. (Ista et al., 2005)	NA	***Current Validity:*** Comfort-B scores were significantly different for the three NISS categories (Kruskal-Wallis, chi-square = 237, df = 2, p >.001). The median Comfort scores were 7 (range, 6-14) in the over-sedated NISS category, 11 (range, 6-26) in the adequately sedated category, and 19 (range, 11-29) in the under-sedated category. - In 93 of 843 observations (11%), the impression of nurses was under-sedation (NISS=1), with most Comfort scores between 11 and 22. In 85.5% of all observations, nurses considered sedation as adequate, with COMFORT scores ranging between 6 and 22. In 29 of 843 (3.4%) of all observations, nurses considered infants over-sedated, with most Comfort scores between 6 and 10. The risk of over- or under-sedation with a Comfort score 23 was 0% and 95%, respectively. The risk of over- or under-sedation with a Comfort score less than or equal to 10 was 7.8% and 0%, respectively. With Comfort scores between 11 and 22, patients were under- and over-sedated in 15.4% (75 of 488) and 0.4% (2 of 488) of observations, respectively. ****Predictive validity:*** Patients were considered adequately sedated (NISS =2) in 721 (86%) of all observations. In 63% of these observations, the Comfort score pointed at adequate sedation. Patients were	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
				considered overs-sedated (NISS = 3) in 29 observations in 18 patients. In 91% of these observations, the Comfort score also implied over-sedation. Patients were considered under-sedated (NISS =1) in a total of 93 (11%) observations in 35 patients. In 78.3% of these observations, the Comfort scale also implied under-sedation. These 35 (NISS =1) and 18 (NISS =3) patients did not differ significantly from the total study group with regard to age, diagnosis, gender, or Pediatric Index of Mortality score.	
Comfort Scale and Ramsay Scale	-	Kusahara, et al. (Kusahara et al., 2005)	***Convergent validity:*** The percentage of agreement identified between Comfort and Ramsay scales was of 80.0%, with a Kappa of 0.59 expressing a tendency to concordance.	***Concurrent validity***: The health care professional's assessment showed a similar percentage of agreement with Comfort (47.0%) and with Ramsay scale (50.0%). However, the kappa statistics showed a higher concordance level between the health care professionals, clinical assessment and the scores expressed by the Ramsay (κ=0.66) than with Comfort scale (κ=0.1).	NA
Comfort Scale	-	Grap, et al. (Grap et al., 2006)	***Structural Validity:*** (in brackets is the comparator for correlation) <u>Comfort-Start:</u> r=-0.55 (Comfort end) r=0.07 (behavior state) r=0.13 (arms move) r=-0.07 (legs move) r=-0.03 (head move) r=0.06 (twitch) r=0.32 (facial activity) r= -0.06 (cough) <u>Comfort-End</u>:	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
			r=0.23(behavior state) r=0.17(arms move) r=-0.14 (legs move) r=0.04 (head move) r=-0.27 (twitch) r=-0.02 (facial activity) r= -0.22 (cough) Behavior State r=.30 (arms move) r=0.69* (legs move) r= 0.44(head move) r=-0.69 (twitch)r=0 .16 (facial activity) r=0.53(cough) Arms move : r-0.30 (legs move) r=0.64*(head move) r=0.47* (twitch).r=0.77 (facial activity) r=0.24 (cough) Legs move r=0.71 (head move) r=0 .74 (twitch) r=0 .28 (facial activity) r=0 .54 (cough) Head move: r= 0.58 (twitch) r=0.72 (facial activity) r=0.42 (cough) Twitch r=0 .34 (facial activity) r=.67 (cough) Facial activity: r=0.09 (cough)		
Comfort-B and Modified FLACC	-	Johansson, et al. (Johansson & Kokinsky, 2009)	****Hypothesis Testing:*** Assumption: the investigators don't clearly specify type of construct validity, but appears to be hypotheses testing as predicting that with morphine administration FLACC should decrease. After administration of morphine (n = 20), the FLACC median score	***Concurrent Validity:*** Sedation level and Comfort-gamma correlation coefficient= 0.57 p<0.05 Sedation Level and FLACC- gamma coefficient= 0.50 p<0.05 VAS and FLACC- gamma coefficient= 0.50 p<0.05 VAS and Comfort-B-- gamma coefficient= 0.49 p<0.05 Comfort-B and modified	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
			decreased from 5 to 0-2 ($p < 0.001$)	FLACC gamma coefficient= 0.76, $p < 0.05$	
Comfort-B Scale	-	Valkenburg, et al. (Valkenburg et al., 2011)	***Hypothesis testing***: Median number of Comfort-B scores significantly differed between the 2 groups (P = .023): a median [IQR] of 9 [4-22] in the control group vs. a median [IQR] of 16 [8-22] in the Down syndrome group. Mean Comfort-B score was 12.1 (SD 1.7) in the Down syndrome group vs. 12.3 (SD 1.8) in the control group (P = .31). A total of 7% of the 7439 Comfort-B scores across both groups were 17 or higher. The percentage of Comfort-B scores of 17 or higher was calculated for each patient. Median [IQR] percentage per patient was 8.3 [0-20] in the Down syndrome group vs. 6.7 [0-20] in the control group (P = .48).	***Concurrent validity*** when comparing with NRSobs and NISS as acceptable gold standards NRSOBS pain scores of 4 or higher (an indication for moderate to severe pain) were even more rare (5% of scores). The median percentage of NRSOBS ratings of 4 or higher per patient was 0 in both groups. NRSOBS pain ratings of 4 or higher were seen in 4.8% of all 6954 NRSOBS pain assessments. -The Pearson product moment correlation between mean NRSOBS pain and mean Comfort-B scores per patient was 0.45 for the Down syndrome group (P < .01) and 0.57 for the control group (P < .01) A median [IQR] number of 10 [1-41] NISS assessments in 41.1% of the 542 patients (37.7% in the control group vs. 44.7% in the Down syndrome group) were recorded in the PDMS. The median [IQR] percentage of adequate sedation scores was 90.5% [78-100] in the control group vs. 87.8% [79-100] in the Down syndrome group (P = .33).	* ***Responsiveness:*** For both groups, the clinical cut-off COMFORT-B score of 17 presented with good sensitivity (82% in the Down syndrome group and 83% in the control group) and excellent specificity (92% in the Down syndrome group and 91% in the control group). The positive predictive value was 0.32 in both groups. The negative predictive value was excellent and 0.99 for both groups. The AUC for the Down syndrome group did not statistically significantly differ from the AUC of the control group (P = .85) (Fig. 1).
Comfort-B Scale	-	Boerlage et al. (Boerlage et al., 2011)	NA	NA	****Responsiveness:*** The mean Comfort-B scores for the 2-min observation was 13.5 (SD 3.8); that for the 30-sec observation was 12.7 (SD 3.7). The mean difference therefore was 0.8 (confidence interval 0.6 -1.1,

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
					paired t test, $p < .001$). In 25 observations (10.6%), the 2-min Comfort-B score exceeded 17 whereas the 30-sec score did not. Analysis of variance for repeated measures with observer as a covariate revealed statistically significant different scores for the two observation periods ($F= 3.899$, $p <.05$). The covariate distinguishing between expert and nurse was not statistically significant ($F = 2.463$, $p <.5$). The sensitivity and positive predictive value for the 30-sec Comfort-B scores compared to the 2-min scores were 0.44 and 0.80, respectively. The specificity and negative predictive value were 0.97 and 0.88, respectively. For the NRS-11, the sensitivity and positive predictive value for the 30-sec scores were 0.75 and 0.86, respectively; the specificity and negative predictive value were both 0.99.
Faces, Legs, Activity, Cry, Consolability Scale	_	Manworren et al. (Manworren & Hynan, 2003)	***Hypothesis Testing:*** from discussion- more than 85 % of pts. showed a 4 pt. decrease in FLACC score post-analgesic, 59.9 % had a 6 pt. decrease between pre-analgesic and peak analgesia		

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
(FLACC)			(Time 3)		
Faces, Legs, Activity, Cry, Consolability Scale (FLACC)	-	Voepel-Lewis et al. (Voepel-Lewis et al., 2010)	***Hypothesis testing:*** FLACC pain scores decreased significantly after administration of an analgesic or from painful to non-painful situations (mean, 5.27; SD, 2.3 vs. mean, 0.52; SD, 1.1; P<.001)	***Concurrent validity:*** FLACC scores correlated significantly with CNPI scores, supporting excellent criterion validity in adults ($\rho = 0.963$; $P < .01$). FLACC and Comfort scores were highly correlated ($\rho = 0.849$; $P < .01$), supporting criterion validity in critically ill children. Note: technically CNPI and FLACC are not gold standards. Should actually be convergent validity but they call it criterion	NA
Cardiac Analgesic Assessment Scale (CAAS)	Suominen, P., 2004 (Suominen et al., 2004)	-	****Concurrent Validity:*** The average of the four Spearman rank correlations between CAAS and VAS was low (0.27) indicating that CAAS does not correlate well with the VAS.	NA	****Responsiveness***: Changes in pain status and response to analgesia over time were analysed in 37 patients participating in the placebo-controlled intrathecal morphine trial. There was a statistical difference (2.2 points, 95%CI: 1.9, 2.6) between the mean CAAS scores assessed before rescue intravenous morphine was indicated and the mean score at the time when rescue intravenous morphine was administered. The mean time difference between these two scorings was 102.8 min with a range of 5-315 min The mean CAAS score obtained after the administration of the intravenous morphine bolus and the commencement of the

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
					continuous morphine infusion was significantly lower (1.7 points, 95% CI: 1.1,2.3) than the CAAS before the intervention. The mean time difference between the score indicating the use of rescue intravenous morphine and the score after the intervention was 136.3 min, range 40-330 min.
Multidimensional Assessment of Pain Scale (MAPS)	Ramelet et al. (Ramelet et al., 2007)	-	***Convergent:*** Bland and Altman plot of the differences between the MAPS and the FLACC scales against the mean. The mean of the differences was 0.44 (CI: 0.18-0.71). The limits of agreement were -1.22 to 2.09. The regression line shows that the difference in agreement decreased as pain scores increased, but was not significant (F= 1.55, d.f. =39, P =0.22).	***Concurrent validity***: Comparison of the MAPS with VASobs, using the method described by Bland and Altman, showed a mean of the differences of 0.25 with a CI of 0.02-0.49; the limits of agreement were -1.24 to 1.74. The regression line shows that the difference in agreement decreased as pain scores increased, but was not significant (F =2.83, d.f. =41, P =0.100).	NA
Multidimensional Assessment of Pain Scale (MAPS)	-	Ramelet et al. (Ramelet, Rees et al., 2007)	***Convergent Validity:*** Comparison of the MAPS with FLACC using the method described by Bland and Altman showed a mean of the differences of) 0.12; the limits of agreement were) 3.71 to 3.78. The plot also shows that the difference in agreement is greater at mean pain score 1; the difference tends to	***Concurrent Validity:*** Comparison of the MAPS with the VASobs using the same method showed a mean of the differences of -0.29; the limits of agreement were 1.78 to -2.37.	****Responsiveness:*** Descriptive statistics showed that pain scores > or equal to 4 represented 80% of the scores before and 30% after the administration of a rescue intravenous morphine bolus. The median MAPS pain scores was 5 (IQR 3.21) before administration of a rescue intravenous morphine bolus,

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
			decrease as pain scores increase.		and following administration, 0 (IQR 2.25), 0 (IQR 2.00), and 1 (IQR 2.46) at 15, 30, and 60 min, respectively. Three MAPS scores were under 4 before bolus. Analysis of these three individual boluses showed that MAPS score was 2, 1 and 1, FLACC was 0, 0, and 0 and VASobs was 6, 7, and 5. Results from the Friedman Test showed a 5-point decrease in median pain scores after opioid bolus at time 0 and 15, 30, and 60 min ($P < 0.001$). However, there were no significant differences between pain scores at 15, 30, and 60 min ($P = 0.353$) and at 30 and 60 min ($P = 0.366$) after the administration of a bolus of opioid
Neonatal Pain, Agitation and Sedation Scale (N-PASS)	Hummel et al. (Hummel et al., 2008)	-	***Convergent validity:*** The correlations between the pre-intervention N-PASS pain scale and the PIPP were 0.83 and 0.81 for raters 1 and 2, respectively. Post-intervention, the correlations between the N-PASS pain scale and the PIPP were 0.61 and 0.61 for raters 1 and 2, respectively. Construct validity analysis. The mean (s.d.) pre-intervention pain score was 4.86 (3.38), which fell to 1.81 (1.53) after analgesic	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
			intervention (P<0.0001). There was a decrease in mean sedation scores from -0.85 (1.66) to -2.78 (2.81) after pharmacologic intervention (P<0.0001).		
Neonatal Pain, Agitation and Sedation Scale (N-PASS)	-	Hummel et al. (Hummel et al., 2010)	**Convergent Validity**: Referred to as ***Discriminate validity*** Values: Wilcoxon Signed-ranks test was used to compare heel-stick to sham scores. The mean heel-stick pain score was 3.93 (2.30); the mean sham score was 0.81 (1.21) (Z=6.429, P<0.001). 0 (>35 weeks) -3.391 (0.001) 1 (32-35 weeks) -3.450 (0.001) 2 (28-31 weeks) - 3.089 (0.002) 3 (<28 weeks) -3.193 (0.001) for the Wilcoxon Signed-ranks test calculated for each gestational age group. ***Convergent Validity***: The Spearman rho correlation between the heel-stick scores of the N-PASS pain scale and the PIPP scale was 0.743 (P<0.001).	NA	NA
No name	Hughes et al. (Hughes et al.,	-	NA	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
	1994)				
Pain Assessment Tool (PAT)	Spence et al. (Spence et al., 2005)	-	***Convergent Validity:** There was a strong correlation between the PAT and CRIES scores (0.76 p<0.001) when measured by the clinical research nurses.	***Concurrent:** A moderate correlation was found between the PAT score and the VAS scale used by the infant's mother 0.38 p <0.01).	NA
Pediatric Confusion Assessment Method for Intensive Care Unit (pCAM-ICU)	Smith et al. (Smith et al., 2011)		***Predictive and concurrent validity:*** The reference standard DSM ratings diagnosed delirium in 18 of the 146 (12.3%) patient assessments, or nine of 68 (13.2%) patients, whereas the pCAM-ICU diagnosed delirium in 16 of 146 (11%) patient assessments, or eight of 68 (11.8%) patients. Of the nine patients with delirium during the study period, the mean age was 13.8 (1.9) yrs. and 77.8% were male. -The sensitivity of the pCAM-ICU was 83% (95% CI, 66-93%) and the specificity was 99% (95% CI, 95-100%). This resulted in a positive predictive value of 93% (95% CI, 63-99%) and a negative predictive value of 98% (95% CI, 93-99%) for the pCAM-ICU. The likelihood ratio for the pCAM-ICU in the diagnosis of delirium was 105 (95% CI, 14-748). - Specifically assessed the	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
			validity of feature 2 given that we had modified the original ASE Auditory test for the pCAM-ICU. Of the 146 assessments, inattention was diagnosed by the pCAM-ICU during 16 vs. 18 assessments positive for inattention by reference raters. The sensitivity of the p-CAM-ICU feature 2 for inattention was 83% (95% CI, 44-100%) and the specificity was 99% (95% CI, 98-100%). –subgroups - In patients 12 yrs. of age, the sensitivity and specificity of the pCAM-ICU were both 100% using 55 paired assessments among 34 patients. In patients 12 yrs. of age, the sensitivity and specificity were 80% (95% CI, 59-91%) and 99% (95% CI, 91-100%). For patients' assessments of delirium while on mechanical ventilation, the pCAM-ICU had a sensitivity and specificity of 75% (95% CI, 66-100%) and 92% (95% CI, 67-100%) using 17 paired assessments.		
Penn State Children's Hospital Sedatio	Popernack et al. (Popernack et al., 2004)	-	***Hypothesis testing*** The difference between the rate of unplanned extubations pre-PSCHSA and post-PSCHSA was statistically significant ($p < .001$). There were a total of 36 unplanned	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
n Algorithm (PSCHSA)			extubations in the pre-PSCHSA time period. Of those, 58% patients remained extubated, and 42% required immediate reintubation. In the period of time after the mandatory use of the PSCHSA, there were a total of seven unplanned extubations, and only one patient (14%) required reintubation, whereas 86% remained extubated.		
PICU Modified Comfort Sedation Scale for the muscle relaxed patient	Razmus et al. (Razmus et al., 2003)	-	NA	****concurrent validity*** as comparing two scales meant to measure sedation- likert scale acceptable gold standard** mean for modified Comfort score was 13.13 with a standard deviation of 4.13, mean for adequacy of sedation scale of 3.19 with a standard deviation of 0.69 Newscale Pearson correlation is 0.026	NA
Sedation Scale (no name)	Parkinson et al. (Parkinson et al., 1997)	-	NA	NA	NA
State Behavioral Scale	Curley et al. (Curley et al.,	-	***Structural Validity:*** Using the median scores on the seven dimensions and mean NRS scores for each group, the five	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
(SBS)	2006)		profiles were then aligned to a bipolar numeric scale ranging from -3 to +1. The first cluster group with a mean (95% confidence interval) NRS score of 1.1 (0.7-1.6) was linked to an SBS dimension of -3, the second cluster group with an NRS score of 2.5 (2.1-2.9) was linked to SBS dimension -2, the third cluster group with an NRS score of 4.0 (3.4-4.5) was linked to SBS dimension -1, the fourth cluster group with a mean NRS score of 5.3 (4.9 -5.6) was equated with a SBS dimension of 0, the fifth cluster score with an NRS score of 5.3 (4.9-5.6) was linked to SBS dimension, the 6th cluster score with an NRS score of linked to SBS dimension +1. The authors added a + 2 agitated level because, although rarely observed, and thus not empirically captured in this study, experienced clinicians have cared for patients exhibiting these agitated behaviors in the pediatric ICU.		
The Vancouver Sedative Recovery Scale	Macnab et al. (Macnab et al., 1991)	-	NA	NA	NA

Scale	Original Report	Validation Study	Construct Validity	Criterion Validity	Responsiveness or Interpretability
(VSRS)					

NA= not assessed, VAS=Visual Analog Scale, NISS- Nurse Interpretation of Sedation Score, NRS= Nurse Rating of Sedation

*assumption made by writer with regard to psychometric property assessed as not stated by authors of study.

Table 2-8

Quality Assessment of Studies assessing the Reliability and Validity of Instruments which assess Pain, Non-Pain Related Distress, Sedation, and Analgesia using the COSMIN Checklist

Study	CTT/ IRT	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypotheses Testing	Cross-Cultural Validity	Criterion Validity	Responsive-ness	Interpret-ability	Generaliz-ability
Johnasson et al. (Johansson & Kokinsky, 2009)	CTT	Good	*	*	*	*	Fair	*	Good	*	+++	+++
Hummel et al. (Hummel et al., 2008)	CTT	Poor	Good	*	*	*	Good	*	*	*	+	+++
Grap et al. (Grap et al., 2006)	CTT	*	*	*	*	Poor	*	*	*	*	+	+
Curley et al. (Curley et al., 2006)	CTT	*	Excellent	*	*	Excellent	*	*	*	*	++	++++
Spence et al. (Spence et al., 2005)	CTT	*	Excellent	Excellent	*	*	Excellent	*	Excellent	*	+++	+++
Suominen et al. (Suominen et al., 2004)	CTT	*	Fair	*	*	*	*	*	Fair	Fair	++	+++
Popernack et al. (Popernack et al., 2004)	CTT	*	*	*	*	*	Good	*	*	*	++	+++
de Carvalho et al. (Brunow de Carvalho et al., 1999)	CTT	*	*	*	*	*	Fair	*	*	*	+	+++
Marx et al. (Marx et al., 1994)	CTT	*	Good	*	*	*	*	*	Good	*	++	++++
Hartwig et al. (Hartwig et al., 1991)	CTT	*	Poor	*	*	*	*	*	*	*	++	+++
Razmus et al. (Razmus et al.,	CTT	*	*	*	*	*	*	*	Fair	*	++	+

Study	CTT/ IRT	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypotheses Testing	Cross-Cultural Validity	Criterion Validity	Responsive-ness	Interpret-ability	Generaliz-ability
2003)												
van Dijk et al. (M. van Dijk et al., 2001)	CTT	*	*	*	*	*	Good	*	*	*	++	++++
Hummel et al. (Hummel et al., 2010)	CTT	Poor	Good	*	*	*	Good	*	*	*	+	+++
Carnevale et al. (Carnevale & Razack, 2002)	CTT	Fair	*	*	*	Excellent	*	*	*	*	++	+++
Kusahara et al. (Kusahara et al., 2005)	CTT	*	*	*	*	*	Fair	*	Good	*	+	+++
Smith et al , (Smith et al., 2011)	CTT	*	Excellent	*	*	*	*	*	Excellent	*	+++	++++
Ista et al. (Ista et al., 2005)	CTT	Excellent	Fair	*	*	*	*	*	Excellent	*	+++	++++
Macnab et al. (Macnab et al., 1991)	CTT	Excellent	Excellent	*	*	*	*	*	*	*	+++	++
Ambuel et al. (Ambuel et al., 1992)	CTT	Good	Good	*	*	Good	*	*	Good	*	++	++++
Hughes et al. (Hughes et al., 1994)	CTT	*	Fair	*	*	*	*	*	*	*	+	+++
Parkinson et al. (Parkinson et al., 1997)	CTT	*	Good	*	*	*	*	*	*	*	+	+++
(M. Van Dijk et al., 2000)	IRT	Excellent	Excellent	*	*	*	*	*	Excellent	*	++	++++
Valkenburg et al. (Valkenburg et al., 2011)	CTT	Excellent	Excellent	*	*	*	Fair	*	Excellent	Excellent	++	++++
Boerlage et al. (Boerlage et al., 2011)	CTT	*	*	*	*	*	*	*	*	Excellent	++	++
Manworren et al (Manworren	CTT	*	Good	*	*	*	Fair	*	*	Fair	++	++

Study	CTT/ IRT	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypotheses Testing	Cross-Cultural Validity	Criterion Validity	Responsive-ness	Interpret-ability	Generaliz-ability
& Hynan, 2003)												
Ramelet et al, (Ramelet, Rees et al., 2007)	CTT	Good	*	*	*	*	Good	*	Good	Good	++	++++
Ramelet et al, (Ramelet et al., 2007)	CTT	Excellent	Excellent	*	*	*	Good	*	Excellent	*	++	++++
Voepel –Lewis et al. Voepel-Lewis et al., 2010	CTT	Good	Good	*	*	*	Fair	*	Fair	*	+	++

- * Not Assessed; + Limited ; ++ Somewhat; +++ Acceptable; ++++ Superior

Table 2-9

Description and Findings of Studies examining the Efficacy and Effectiveness of Scales which use Physiological and Behavioral cues of pain and non- pain related distress to assess Level of Sedation and/or Analgesia

		Author (Country)		
		Alexander & Carnevale, 2002 (Canada)	Jin et al., 2007 (Korea)	Ista et al., 2009 (Netherlands)
Study Type		Retrospective chart review of Case Studies	CBA *	CBA
Intervention Group		10 ventilated PICU pts., due to a cardiovascular or respiratory problem. Age not provided	21 ventilated PICU pts. Diagnoses: ARDS, pneumonia, bronchiolitis, or other. Mean age 20.1±30.7 mos.	29 PICU pts. (21 ventilated; 8 non-ventilated). Diagnoses: Congenital Heart Disease, Respiratory Failure, Sepsis, or other. Median age 3.1 mos.
Control Group		Same as the intervention group	20 ventilated PICU pts. Diagnoses: ARDS, pneumonia, bronchiolitis, or other, Mean age: 21.7±25.2 mos.	27 PICU pts. (17 ventilated; 10 non-ventilated). Respiratory Failure, Sepsis, or other. Median age 4.1 mos.
Intervention		Protocol-directed Sedation with the Comfort Scale	Protocol-directed sedation with the Comfort Scale	Protocol-Directed Sedation with the Comfort- B Scale **
Assessment done by:		Registered Nurse	Pharmacist in consultation with the attending physician	Physician, registered nurse, or pharmacist
Sedation Protocol		Comfort Scale is used to objectively score patient distress and a standard order sheet that permits adjustment of sedative and analgesic dosages (based on the score) to maintain a specified target level of patient comfort.	Non-pharmacologic treatments were attempted first. Then based on comfort score a medication for analgesia, sedation, delirium, or NMB is given. If patient is comfortable they use Finnegan score to assess how weaning is tolerated.	3 protocols: Basic scheme, hemo-dynamically stable, and hemo-dynamically unstable patient. Comfort-B score is assessed minimum Q8H and prn. The score is then used to determine if sedation is weaned, increased, or kept the same. Within the boundaries of the protocol, nurses were allowed to titrate medications on the guidance of assessments.
Outcome	Analgesic	Significantly higher amounts of fentanyl were received during the time the patient was on the protocol (p=0.024) ; no significant difference in the # of boluses received	Total dose and maximum rate of continuous infusion of fentanyl was significantly lower in intervention group (p=0.02; p<0.01 respectively)	Median morphine doses were significantly higher in intervention period (p=0.004)
	Sedation	Significantly higher amounts of midazolam were received were during the time the patient was on the sedation protocol (p =0.002); no significant difference in the # of boluses received	Duration of sedation was lower (8.0 days vs. 11.5 days, p=0.053), but non-significant in the intervention group; Total dose was lower (non-significant)and maximum rate of continuous infusion of	Median midazolam doses were significantly higher in the intervention period (p=0.001)

			midazolam was significantly lower in intervention group	
	Length of Mechanical Ventilation	NA	Intervention group- 11.0 days, Control group - 12.5 days; (p=0.04 , significant decrease)	NA
	Median length of PICU Stay	NA	Intervention group- 15.0 days Control group - 19.5 days;, (p=0.04 ,significant decrease)	NA
	Adverse Withdrawal Symptom	NA	The overall development of withdrawal symptoms was significantly lower in the intervention group than in the control group (1 vs.7 patients, p=0.02)	NA
	Other	Less incidents of under-sedation when on protocol	N/A	Proportion of Comfort-B scale scores that indicate adequate sedation (11-22) increased from 63% to 72 % from control period to intervention period. Proportion NISS were nearly equal between the two stages

*Retrospective- before group, prospective- after group; ** the control group used Comfort-B assessment without the sedation protocol in this study. CBA= Controlled Before and After Study; NA= Not assessed, N/A= not applicable; NISS= Nurse Interpretation of Sedation Scores

Table 2-10

New Castle Ottawa Quality Assessment (Before and After Quality Assessment Tool) for Studies examining the Efficacy and Effectiveness of Scales which use Physiological and Behavioral cues of pain and non- pain related distress to assess Level of Sedation and/or Analgesia

		Study		
NOS Assessment Category	NOS Assessment :	Carnevale and Razack, 2002	Jin, H.S., et al. , 2007	Ista et al., 2009
Selection	Is the post-intervention group Representative?	YES (1)	YES (1)	YES (1)
	Is the Pre-Intervention Group Representative?	YES (1)	YES (1)	YES (1)
	Are the Pre- and post-intervention groups drawn from the same source?	YES (1)	YES (1)	NO
Comparability	Were the pre-and post- groups comparable on the basis of design and Analysis?	YES-comparable (1)	YES-comparable (1)	YES-comparable (1)
Assessment of Outcome	Was the Assessment of outcome(s) valid?	YES-reference to secure records or validated methods (2)	YES-reference to secure records or validated methods (2)	YES-reference to secure records or validated methods (2)
	Was the assessment of outcome(s) reliable/accurate?	UNCLEAR	UNCLEAR	UNCLEAR
	Was the method of outcome assessment the same for the pre- and post-intervention group?	YES (1)	NO	YES (1)
Intervention	Did the study report the point in time when the intervention occurred?	YES (1)	YES (1)	YES (1)
	Was the intervention clearly described	YES (1)	YES (1)	YES (1)
Pre- and Post-Intervention Periods	Were the data collected during a similar time frame?	YES-same duration, but seasonal bias (1)	YES-same duration, but seasonal bias (1)	YES-same duration, but seasonal bias (1)
	Total Score (out of 13)	10	9	9

Table 2-11

Summary of Recommendations for Assessment of Mechanically Ventilated Pediatric Patients

Assessment of:	**Recommended Instrument Name**	**Instruments that have Potential/ Require Further Research**
Post-operative Pain, Analgesia, Non-Pain Related Distress, and Sedation	Comfort Scale	N/A
Sedation Only (when no pain in present)	Comfort Scale	State Behavioral Scale (SBS)
Post-operative Pain and/or Procedural Pain or brief Painful Events	Faces, Legs, Activity, Cry, Consolability Scale (FLACC) OR Multidimensional Pain Assessment Scale (MAPS)	N/A
Sedation in Muscle Relaxed Patients	Unable to make recommendation	PICU Modified Comfort Scale for Muscle Relaxed Patients
Delirium	Unable to make recommendation	Pediatric Confusion Assessment Method for Intensive Care Unit (pCAM-ICU)
Pain, non-pain related distress, sedation in neonates	Neonatal Pain Agitation Sedation Scale (N-PASS)	N/A

References

Alexander, E., Carnevale, F. A., & Razack, S. (2002). Evaluation of a sedation protocol for intubated critically ill children. *Intensive & Critical Care Nursing, 18*(5), 292-301. doi: 10.1016/S0964339702000502

Ambuel, B., Hamlett, K. W., Marx, C. M., & Blumer, J. L. (1992). Assessing distress in pediatric intensive care environments: The COMFORT scale. *Journal of Pediatric Psychology, 17*(1), 95-109. doi: 10.1093/jpepsy/17.1.95

Boerlage, A. A., Ista, E., de Jong, M., Tibboel, D., & van Dijk, M. (2011). The COMFORT behavior scale: Is a shorter observation period feasible? *Pediatric Critical Care Medicine,* doi: 10.1097/PCC.0b013e3182192d92

Brunow de Carvalho, W., Lucas da Silva, P. S., Paulo, C. S., Fonseca, M. M., & Belli, L. A. (1999). Comparison between the comfort and hartwig sedation scales in pediatric patients undergoing mechanical lung ventilation. *Sao Paulo Medical Journal = Revista Paulista De Medicina, 117*(5), 192-196.

Carnevale, F. A., & Ducharme, C. (1997). Adverse reactions to the withdrawal of opioids and benzodiazepines in paediatric intensive care. *Intensive & Critical Care Nursing, 13*(4), 181-188. doi: 10.1016/S0964-3397(97)80012-2

Carnevale, F. A., & Razack, S. (2002). An item analysis of the COMFORT scale in a pediatric intensive care unit. *Pediatric Critical Care Medicine, 3*(2), 177-180.

Connolly, D., McClowry, S., Hayman, L., Mahony, L., & Artman, M. (2004). Posttraumatic stress disorder in children after cardiac surgery. *Journal of Pediatrics, 144*(4), 480-484. doi: 10.1016/j.jpeds.2003.12.048

Curley, M. A., Harris, S. K., Fraser, K. A., Johnson, R. A., & Arnold, J. H. (2006). State behavioral scale: A sedation assessment instrument for infants and young children supported on mechanical ventilation. *Pediatric Critical Care Medicine, 7*(2), 107-114.

De Jonghe, B., Cook, D., Appere-De-Vecchi, C., Guyatt, G., Meade, M., & Outin, H. (2000). Using and understanding sedation scoring systems: A systematic review. *Intensive Care Medicine, 26*(3), 275-285.

Farias, J., Frutos, F., Esteban, A., Flores, J., Retta, A., Baltodano, A., . . . Johnson, M. (2004). What is the daily practice of mechanical ventilation in pediatric intensive care units? A multicenter study. *Intensive Care Medicine,* (5), 918-925. doi: 10.1007/s00134-004-2225-5

Grap, M. J., Pickler, R. H., & Munro, C. (2006). Observation of behavior in sedated, mechanically ventilated children. *Pediatric Nursing, 32*(3), 216-220.

Harris, P. A., Taylor, R., Thielke, R., Payne, P., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377-381. doi: 10.1016/j.jbi.2008.08.010

Hartwig, S., Roth, B., & Theisohn, M. (1991). Clinical experience with continuous intravenous sedation using midazolam and fentanyl in the paediatric intensive care unit. *European Journal of Pediatrics, 150*(11), 784-788. doi: 10.1007/BF02026712

Hodgkinson, K., Bear, M., Thorn, J., & Van Blaricum, S. (1994). Measuring pain in neonates: Evaluating an instrument and developing a common language. *Australian Journal of Advanced Nursing, 12*(1), 17-22.

Hughes, J., Gill, A., Leach, H. J., Nunn, A. J., Billingham, I., Ratcliffe, J., . . . Choonara, I. (1994). A prospective study of the adverse effects of midazolam on withdrawal in critically ill children. *Acta Paediatrica, International Journal of Paediatrics, 83*(11), 1194-1199. doi: 10.1111/j.1651-2227.1994.tb18280.x

Hummel, P., Lawlor-Klean, P., & Weiss, M. G. (2010). Validity and reliability of the N-PASS assessment tool with acute pain. *Journal of Perinatology, 30*(7), 474-478. doi: 10.1038/jp.2009.185

Hummel, P., Puchalski, M., Creech, S. D., & Weiss, M. G. (2008). Clinical reliability and validity of the N-PASS: Neonatal pain, agitation and sedation scale with prolonged pain. *Journal of Perinatology, 28*(1), 55-60.

Ista, E., De Hoog, M., Tibboel, D., & Van Dijk, M. (2009). Implementation of standard sedation management in paediatric intensive care: Effective and

feasible? *Journal of Clinical Nursing*, *18*(17), 2511-2520. doi: 10.1111/j.1365-2702.2009.02836.x

Ista, E., Van Dijk, M., Tibboel, D., & De Hoog, M. (2005). Assessment of sedation levels in pediatric intensive care patients can be improved by using the COMFORT "behavior" scale. *Pediatric Critical Care Medicine, 6*(1), 58-63. doi: 10.1097/01.PCC.0000149318.40279.1A

Jin, H. S., Yum, M. S., Kim, S. L., Shin, H. Y., Lee, E. H., Ha, E. J., . . . Park, S. J. (2007). The efficacy of the COMFORT scale in assessing optimal sedation in critically ill children requiring mechanical ventilation. *Journal of Korean Medical Science, 22*(4), 693-697. doi: 10.3346/jkms.2007.22.4.693

Johansson, M., & Kokinsky, E. (2009). The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. *Nursing in Critical Care, 14*(3), 122-130. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=19366409

Khemani, R. G., Markovitz, B. P., & Curley, M. A. (2009). Characteristics of children intubated and mechanically ventilated in 16 PICUs. *Chest*, *136*(3), 765-771. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovid

web.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=19542258

Kusahara, D. M., Rego, R. C., Pedreira, M., Peterlini, M., & Carvalho, W. B. (2005). Sedation of children in mechanical pulmonary ventilation: Concordance between clinical assessment, COMFORT and ramsay scales... nursing abstracts from the 2nd conference of the world federation of critical care nurses. *CONNECT: The World of Critical Care Nursing, 4*(2), 1p.

Lamas, A., Lopez-Herce, J., Sancho, L., Mencia, S., Carrillo, A., Santiago, M. J., & Martinez, V. (2008). Assessing sedation in critically ill children by bispectral index, auditory-evoked potentials and clinical scales. *Intensive Care Medicine, 34*(11), 2092-2099. doi: 10.1007/s00134-008-1198-1

Macnab, A. J., Levine, M., Glick, N., Susak, L., & Baker-Brown, G. (1991). A research tool for measurement of recovery from sedation: The vancouver sedative recovery scale. *Journal of Pediatric Surgery, 26*(11), 1263-1267. doi: 10.1016/0022-3468(91)90594-J

Manworren, R. C., & Hynan, L. S. (2003). Clinical validation of FLACC: Preverbal patient pain scale. *Pediatric Nursing, 29*(2), 140-146.

Marx, C. M., Smith, P. G., Lowrie, L. H., Hamlett, K. W., Ambuel, B., Yamashita, T. S., & Blumer, J. L. (1994). Optimal sedation of mechanically ventilated pediatric critical care patients. *Critical Care Medicine, 22*(1), 163-170.

Merkel, S. I., Voepel-Lewis, T., Shayevitz, J. R., & Malviya, S. (1997). The FLACC: A behavioral scale for scoring postoperative pain in young children. *Pediatric Nursing, 23*(3), 293-297.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international delphi study. *Quality of Life Research, 19*(4), 539-549. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20169472; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:20169472&id=doi:&issn=0962-9343&isbn=&volume=19&issue=4&spage=539&pages=539-49&date=2010&title=Quality+of+Life+Research&atitle=The+COSMIN+checklist+for+assessing+the+methodological+quality+of+studies+on+measurement+properties+of+health+status+measurement+instruments%3A+an+international+Delphi+study.&aulast=Mokkink&pid=%3Cauthor%3EMokkink+LB%2CTerwee+CB%2CPatrick+DL%2CAlonso+J%2CStratford+PW%2CKnol+DL%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E20169472%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010b). *COSMIN checklist manual*. The Netherlands: doi: Retrieved from: www.cosmin.nl

Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., . . . de Vet, H. C. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research, 18*(3), 313-333.

Parkinson, L., Hughes, J., Gill, A., Billingham, I., Ratcliffe, J., & Choonara, I. (1997). A randomized controlled trial of sedation in the critically ill. *Paediatric Anaesthesia, 7*(5), 405-410. doi: 10.1046/j.1460-9592.1997.d01-109.x

Popernack, M. L., Thomas, N. J., & Lucking, S. E. (2004). Decreasing unplanned extubations: Utilization of the penn state children's hospital sedation algorithm. *Pediatric Critical Care Medicine, 5*(1), 58-62.

Ramelet, A. S., Rees, N., McDonald, S., Bulsara, M., & Abu-Saad, H. H. (2007). Development and preliminary psychometric testing of the multidimensional assessment of pain scale: MAPS. *Paediatric Anaesthesia, 17*(4), 333-340. doi: 10.1111/j.1460-9592.2006.02115.x

Ramelet, A. S., Rees, N. W., McDonald, S., Bulsara, M. K., & Huijer Abu-Saad, H. (2007). Clinical validation of the multidimensional assessment of pain

scale. *Paediatric Anaesthesia*, *17*(12), 1156-1165. doi: 10.1111/j.1460-9592.2007.02325.x

Ramsay, M. A., Savege, T. M., Simpson, B. R., & Goodwin, R. (1974). Controlled sedation with alphaxalone-alphadolone. *British Medical Journal*, *2*(5920), 656-659. doi: 10.1136/bmj.2.5920.656

Razmus, I. S., Clarke, K. A., & Naufel, K. Z. (2003). Development of a sedation scale for the mechanically ventilated muscle relaxed pediatric critical care patient. *Pediatric Intensive Care Nursing*, *4*(1), 7-11.

Refworks-COS. (2011). *Refworks 2.0 web based bibliographic manager*. Michigian, USA:

Scott-Findlay, S., & Estabrooks, C. A. (2006). Knowledge translation and pain management. In G. A. Finley, P. J. McGrath & C. T. Chambers (Eds.), *Bringing pain relief to children* (1st ed., pp. 199-227). Totowa, New Jersey: Humana Press.

Smith, H. A. B., Boyd, J., Fuchs, D. C., Melvin, K., Berry, P., Shintani, A., . . . Ely, E. W. (2011). Diagnosing delirium in critically ill children: Validity and reliability of the pediatric confusion assessment method for the intensive care unit. *Critical Care Medicine*, *39*(1), 150-157. doi: 10.1097/CCM.0b013e3181feb489

Spence, K., Gillies, D., Harrison, D., Johnston, L., & Nagy, S. (2005). A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. *JOGNN - Journal of Obstetric, Gynecologic, & Neonatal Nursing, 34*(1), 80-86. doi: 10.1177/0884217504272810

Suominen, P., Caffin, C., Linton, S., McKinley, D., Ragg, P., Davie, G., & Eyres, R. (2004). The cardiac analgesic assessment scale (CAAS): A pain assessment tool for intubated and ventilated children after cardiac surgery. *Paediatric Anaesthesia, 14*(4), 336-343.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research, 21*(4), 651-657. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=mesx&AN=21732199; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:21732199&id=doi:&issn=0962-9343&isbn=&volume=21&issue=4&spage=651&pages=651-7&date=2012&title=Quality+of+Life+Research&atitle=Rating+the+methodological+quality+in+systematic+reviews+of+studies+on+measurement+properties%3A+a+scoring+system+for+the+COSMIN+checklist.&aulast=Terwee&pid=%3Cauthor%3ETerwee+CB%2CMokkink+LB%2CKnol+DL%2COst

elo+RW%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E21732199%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Valkenburg, A. J., Boerlage, A. A., Ista, E., Duivenvoorden, H. J., Tibboel, D., & Dijk, M. v. (2011). The COMFORT-behavior scale is useful to assess pain and distress in 0- to 3-year-old children with down syndrome. *Pain,* doi: 10.1016/j.pain.2011.05.001

van Dijk, M., de Boer, J., Koot, H. M., Duivenvoorden, H. J., Passchier, J., Bouwmeester, N., & Tibboel, D. (2001). The association between physiological and behavioral pain measures in 0- to 3-year-old infants after major surgery. *Journal of Pain & Symptom Management, 22*(1), 600-609.

Van Dijk, M., De Boer, J. B., Koot, H. M., Tibboel, D., Passchier, J., & Duivenvoorden, H. J. (2000). The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain, 84*(2-3), 367-377.

Voepel-Lewis, T., Zanotti, J., Dammeyer, J. A., & Merkel, S. (2010). Reliability and validity of the face, legs, activity, cry, consolability behavioral tool in assessing acute pain in critically ill patients. *American Journal of Critical Care, 19*(1), 55-61. doi: 10.4037/ajcc2010624

von Baeyer, C. L., & Spagrud, L. J. (2007). Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. *Pain, 127*(1-2), 140-150. doi: 10.1016/j.pain.2006.08.014

Waltz, C. F., Strickland, O. L., & Lentz, E. R. (2010). *Measurement in nursing and health reasearch* (4th ed.). New York, NY: Springer Publishing Company.

Wells, G., Shea, B., O'Connell, J., Robertson, J., Peterson, V., Welch, V. & et al. (2011). *The newcastle-ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analysis [website]*. Retrieved September 7, 2012, from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Chapter 3: Comparison of Para-Clinical Tests with Validated Objective Systematic Instruments for assessing Sedation in PICU Mechanically Ventilated Patients: A Systematic Review

Tamara Dorfman, RN, BScN, BSc Honors Neuroscience

Dr. Gwen Rempel, PhD, RN

University of Alberta, Edmonton, Alberta, Canada

Potential Venue: Pediatric Critical Care Medicine

The objective of sedation in the pediatric intensive care unit (PICU) is to minimize the distress experienced by critically ill children, thereby optimizing the delivery of care and the child's recovery. Functions of sedation include: 1) facilitation of mechanical ventilation; 2) induction of sleep and control of agitation; 3) induction of amnesia during paralysis and painful procedures, and; 4) decrease of cellular metabolism (Bavdekar, Mahajan, & Chandu, 1999). Over-sedation can lead to prolonged mechanical ventilation, ventilator-associated pneumonias, lung injury, or neuromuscular disorders. Conversely, under-sedation can lead to interference with effective mechanical ventilation, myocardial and cerebral ischemia, and dangerous outcomes such as self-extubation or the removal of other mechanical devices including intravenous lines or chest tubes (Jin et al., 2007). Therefore, well titrated sedation is an essential component of pediatric intensive care.

There is no "gold standard" for the assessment of sedation. Consequently, the clinical judgement of the clinician is the next logical option. This assessment, however, can result in many interpretations and lead to disagreements and variation within the health care provider team. Disagreements can result in significant fluctuations in the administration and discontinuation of sedation (Carnevale & Ducharme, 1997). These fluctuations can predispose the patient to the development of adverse reactions and potentially result in over- or under-sedation. Other factors such as length of ventilation, length of stay, and long-term psychological and neurodevelopmental factors may also be affected by fluctuations in sedation (Carnevale & Ducharme, 1997).

To address these problems, systematic assessment scales, using physiological and behavioral variables of distress, have been developed to standardize or objectively measure the effectiveness of sedation in mechanically ventilated and non-verbal patients (Marx et al., 1994). These tools include the Comfort Scale and its modified form, and the Ramsay Scale and its modified form. The Comfort scale assesses pain, non-pain related distress, and sedation. It assesses the physiological variables of heart rate and mean arterial blood pressure (MAP), muscle tone, and the behavioral variables of alertness, calmness, movement, facial tension, and respiratory response (Ambuel, Hamlett, Marx, & Blumer, 1992). The modified version of the Comfort-Behavioral scale excludes the physiological variables of heart rate and MAP, but assesses the remaining six components (Ista, Van Dijk, Tibboel, & De Hoog, 2005). The Ramsay Scale assesses sedation using the behavioral variables of calmness, agitation, and physical movement. The modified Ramsay Scale assesses the same variables, but considers acceptance of ventilation when examining the variable of agitation and response to loud noise when examining the variable of physical movement (Ramsay, Savege, Simpson, & Goodwin, 1974). Theoretically, these tools provide a more consistent measure of the adequacy of sedation in controlling patient's distress than do a nurse or physician descriptive analog or visual analog scale (Marx et al., 1994). These scales allow for fewer discrepancies between individual assessments of distress and the patient's response sedation.

Another method of assessing sedation is through the use of para-clinical tests (i.e., tests that assess the underlying biochemical and morphological clinical

manifestations of sedation). Para-clinical tests such as the Bispectral index scale (BIS), middle latency auditory-evoked potential index (AEP index), and Skin Conductance have been shown to be indirect quantitative measures of sedation (Aneja, Heard, Fletcher, & Heard, 2003; Courtman, Wardurgh, & Petros, 2003; Gjerstad, Wagner, Henrichsen, & Storm, 2008; Lamas et al., 2008). The BIS, noninvasively, measures the hypnotic effect of anesthetic and sedative drugs on the brain. It is derived from the electroencephalographic data by a computer algorithm that produces a single numeric value, on a scale from 0 to 100, which reflects a patient's state of hypnosis. BIS values of <40 are defined as very deep sedation, 41-60 are defined as deep sedation, 61-80 are defined as moderate sedation, 81-100 are defined as light sedation (Courtman et al., 2003; Crain, Slonim, & Pollack, 2002).The BIS monitor was designed to provide a quantitative measurement of the level of hypnosis (Twite, Zuk, Gralla, & Friesen, 2005).

The AEP index provides a means for quantitatively estimating the level of sedation in patients. However, it only permits monitoring to be performed intermittently. The AEP index is elicited with headphones producing a bilateral click stimulus at a set intensity and duration. Acoustic stimuli induce distinct changes in the electroencephalogram (EEG) that can be used to assess level of sedation (Lamas et al., 2008). Skin conductance (SC) fluctuations have been used to evaluate pain or increased stress through changes in the sympathetic nervous system. Efferent skin nerve bursts occur in a random manner, but both the amplitude and the number of bursts rise when stimulated. An increase in the number of skin conductance fluctuations (NSCF) and the amplitude of SC

fluctuations (ASCF) can be interpreted as increased activity in this part of the sympathetic nervous system (Gjerstad et al., 2008).

Objective assessment of sedation is essential to the evaluation of distress in mechanically ventilated, non-verbal pediatric patients. The objectives of this systematic review is to compare validated systematic observational scales utilizing physiological and behavioural cues of distress to assess adequacy of sedation with para-clinical tests for assessing sedation.

Methods

Nine studies were identified from a larger systematic review evaluating the characteristics and psychometric properties of observational scales utilizing physiological and behavioural cues of pain and non-pain related distress to assess adequacy sedation and sedation (see companion article). These studies were separated from the studies included in this larger review because they served the dual purpose of validating the use of para-clinical tests in assessing sedation in PICU mechanically ventilated patients and comparing these tools to previously validated systematic assessment scales for assessing sedation. A flow diagram of the review is presented in Figure 3-1.

In identifying these studies, an information specialist was consulted to develop a comprehensive search strategy for published articles, theses and dissertations. Based on her recommendations, MEDLINE, CINAHL, EMBASE, Web of Science (WOS), BIOSIS Previews, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Database of Abstracts of Review and Effects (DARE), and Scopus, and Proquest Dissertations

were searched from January 1970 to June 2011. The search was limited to literature published during this timeframe because previous systematic reviews of pain and sedation observational assessment measures did not reveal any published scales for children or adults before 1970 (De Jonghe et al., 2000; von Baeyer & Spagrud, 2007). The search strategy used is described in the companion article. All identified sources of information were entered into the citation database Refworks and duplicates were removed (Refworks-COS, 2011).

The studies appropriate for analysis, in this systematic review, met the following inclusion criteria: 1) study population of inpatient PICU, invasively mechanically intubated patients, age 0 days to 18 years; 2) the study compared a systematic assessment instrument to measure physiological and behavioral cues of distress and sedation with a para-clinical test; and; 3) the systematic assessment instrument is previously validated and does not require self-report from patient or the rater's global impression of the patient's level of sedation.

Studies published in any language were eligible for the review. Two reviewers (TD and SI) independently screened all abstracts to ensure they met the criteria for inclusion. When no abstract was available, subject headings and an extract from the text were used. The full texts of all studies potentially meeting our inclusion criteria, as determined by either reviewer, were retrieved. The retrieved articles were independently screened by two reviewers (TD and either SI or GR) using the same criteria to confirm inclusion. Disagreements between the reviewers were resolved through discussion between the two initial reviewers,

and, when needed, the involvement of the third reviewer (SI or GR depending on who did the initial review).

Methodological quality was assessed using the COSMIN checklist with four point scale (Mokkink et al., 2010; Terwee et al., 2012). The nine studies identified assessed the convergent validity of the para-clinical tests and the systematic assessment scales. Therefore, the quality of the studies was assessed using the hypothesis testing box (contains convergent validity questions), interpretability box, and the generalizability box of the COSMIN checklist with four point scale by two independent reviewers (TD and GR or ES) (Appendix A, B) (Mokkink et al., 2010; Terwee et al., 2012). Disagreements between the reviewers were resolved through discussion and, when necessary, the involvement of the third reviewer. The four point scale of the COSMIN checklist allowed the reviewer to assign the measurement property a score of poor, fair, good, or excellent. There is no scoring system for interpretability and generalizability developed for the COSMIN four point scale (Terwee et al., 2012). The two independent reviewers, therefore, used the COSMIN checklist questions for interpretability and generalizability as a guide and assigned scores of limited generalizability/ interpretability, somewhat generalizable/ interpretable, acceptable generalizability/ interpretability, and superior generalizability/ interpretability.

Study data were collected and managed using REDCap (Research Electronic Data Capture) hosted at [University of Alberta's Women and Children's Health Research Institute] (Harris et al., 2009). REDCap is a secure, web-based application designed to support data capture for research studies, providing: 1) an

intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources. Data abstraction forms were initially piloted on three studies, independently, by reviewer TD and ES in Microsoft Word. Revisions and modifications were made based on the pilot and the forms were entered into REDCap (Appendix D). Data from all eligible studies was abstracted by the first reviewer (TD) and then independently verified by a second reviewer (ES). Disagreements on abstracted data were resolved by the two reviewers through discussion and consensus. Data abstraction included coder information, study characteristics, baseline characteristics of the study population, information on interventions, and author interpretations.

Results

Of the nine studies identified for this analysis, seven of those studies used BIS as the comparator, one study used BIS and AEP, and one study used skin conductance. Validated scales compared with para-clinical tests included the Comfort Scale, the Comfort-B scale, the Ramsay Scale, and the modified Ramsay Scale (Table 3-1) (Aneja et al., 2003; Bustos Bu, Fuentes, & C, 2007; Courtman et al., 2003; Crain et al., 2002; Froom et al., 2008; Gjerstad et al., 2008; Lamas et al., 2008; Triltsch et al., 2005; Twite et al., 2005). The quality of these studies, assessed using the COSMIN Checklist, is presented in Table 3-2. Quality ranged from poor to excellent. Sample size and presence of a deducible hypothesis had the greatest influence on quality rating.

Eight of the studies showed moderate ($r > 0.5$) to good ($r > 0.8$) correlations between BIS and the Comfort Scale in non-paralyzed patients, indicating convergent validity between the two measures of sedation (Bustos Bu et al., 2007; Courtman et al., 2003; Crain et al., 2002; Froom et al., 2008; Lamas et al., 2008; Triltsch et al., 2005; Twite et al., 2005). However, this correlation decreased during periods of stimulation (i.e., physiotherapy) (Froom et al., 2008). The modified Ramsay score and the Ramsay score also correlated significantly with BIS scores (Aneja et al., 2003; Lamas et al., 2008). Though, the level of agreement between BIS and the Modified Ramsay score and BIS and the Comfort score was 78 % (κ=0.559) and 64% (κ=0.280), respectively in non-paralyzed patients (Lamas et al., 2008). This finding indicates higher concordance between the Modified Ramsay Score and BIS than the Comfort Score and BIS.

In contrast, the level of agreement decreased in paralyzed patients, and the correlation was found to be non-significant (Lamas et al., 2008). BIS scores, however, determined the degree of sedation in paralyzed patients more effectively than both the modified and original Ramsay score, and the Comfort Score, suggesting its usefulness in assessing sedation patients who are chemically paralyzed (Aneja et al., 2003; Lamas et al., 2008).

In general, BIS values of <40 are defined as very deep sedation, 41-60 are defined as deep sedation, 61-80 are defined as moderate sedation, 81-100 are defined as light sedation (Courtman et al., 2003; Crain et al., 2002). Crain et al. (2002) determined corresponding COMFORT scale mean measurements in each category to be 15.8 $\pm$0.6, 16.2 $\pm$0.6, 18.1 $\pm$1.3, and 22.3 $\pm$1.4 (r^2 =0.89). These

findings suggest that BIS may be more accurate at identifying patients in the very deep sedation range who appear clinically similar to patients in the deep sedation range and may be at risk for the effects of over-sedation. Similar findings were also found in three other studies (Table 3-1) (Froom et al., 2008; Triltsch et al., 2005; Twite et al., 2005). Aneja et al. (2003) determined that the same was true when BIS was compared with the Ramsay Scale (Table 3-1). The Comfort Scale and the Ramsay Scale, therefore, lack clinical usefulness in differentiating between patients in very deep sedation and deep sedation.

There is minimal published literature comparing other para-clinical tests to systematic assessment scales for sedation. The number of skin conductance fluctuations (NSCF), mean skin conductance and the amplitude of skin conductance (ASCF) was compared with the Comfort-B Scale in one of the studies identified for inclusion in the review. The NSCF before, during, and after tracheal suctioning correlated with changes in the Comfort-B scale, but ASCF and mean skin conductance did not (Gjerstad et al., 2008). Similarly, the AEP index was compared with the Modified Ramsay Scale and the Comfort Scale in one of the studies identified. It correlated well with both scales in non-paralyzed patients, but not in paralyzed patients (Lamas et al., 2008). Moderate sedation was determined to be $\geq$ 30 but $\leq$60, and deep sedation was determined to be AEP $\geq$ 60 (Lamas et al., 2008).

Discussion

The results of this review indicate that the Comfort Scale and the Ramsay Scale (original and modified versions) have been shown to correlate with BIS in

PICU medically ventilated patients (Aneja et al., 2003; Bustos Bu et al., 2007; Courtman et al., 2003; Crain et al., 2002; Froom et al., 2008; Lamas et al., 2008; Triltsch et al., 2005; Twite et al., 2005). This correlation, however, is weaker with respect to deep sedation (Aneja et al., 2003; Crain et al., 2002; Froom et al., 2008; Triltsch et al., 2005; Twite et al., 2005). Even still, convergent validity exists been the measures. The Comfort-B scale has been shown to correlate with NSCF, and the Comfort Scale has been shown to correlate with AEP (Gjerstad et al., 2008; Lamas et al., 2008). However, there is limited published research on the convergent validity of AEP and skin conductance and these scales in assessing sedation.

In retrospect, however, a perfect correlation cannot be expected between para-clinical tests and these scales as they measure different variables. For instance, BIS measures the level of hypnosis (sedation), whereas the Comfort scale is designed to measure overall distress including sedation, pain, and agitation (Bustos Bu et al., 2007; Twite et al., 2005). A correlation is found because these variables are related to each other. For example, a patient may be awake and alert with a high BIS value and comfortable with a moderately low Comfort score. In contrast, a stimulus may awake a patient who is sedated thereby changing the BIS value but not affecting the Comfort score to a great extent (Twite et al., 2005). Therefore, these variables can never perfectly correlate with each other.

Pros and Cons of the Measures

The decision of whether to use these para-clinical verus systematic assessment scales of sedation may be better determined in terms of their pros and cons. It can be argued that BIS, AEP, and skin conductance are more objective measures of sedation. This is because systematic assessment scales contain some subjective measures and require interpretation of physiological measures (which can vary significantly during the rating period for the observation). For instance, four out of eight of the variables (assessment of alertness, calmness, muscle tone, and facial expression) of the Comfort scale are subjective (Twite et al., 2005). Additionally, BIS has shown more usefulness in assessing degree of sedation in mechanically ventilated PICU paralyzed patients, whereas these other three scales are not applicable in this subgroup of PICU patients (Aneja et al., 2003; Lamas et al., 2008). Lastly, BIS is able to differentiate between very deep sedation and deep sedation, whereas, the Comfort Scale and Ramsay Scales are not able to make this differentation (Aneja et al., 2003; Crain et al., 2002; Froom et al., 2008).

Para-clinical tests, however, also have their disadvantages. As mentioned in the introduction, they require the use of electrodes and additional monitors and machines which add to the clutter of an already busy PICU bedside. Similarly, impedance can decrease their ability to effectively measure sedation as well as their correlation with systematic assessment scales (Triltsch et al., 2005). These tools are also more costly than using systematic assessment scales and require more training for use by bedside nurse or attending physician than do systematic

assessment scales. More research is needed on both types of measures to determine which is the superior method of measuring sedation.

Conclusion

The comparison of systematic assessment scales with para-clinical tests provides more insight into the clinical validity of systematic assessment scales as well the clinical validity of para-clinical tests in the assessment of sedation. This comparison, however, has only been done with well established sedation scales. Other sedation scales identified in the larger portion of this systematic review (see companion paper) should also be compared with para-clinical tests in an effort to assess both groups' clinical validity. Lastly, the question of whether para-clinical can be used in routine everyday practice needs to be answered. Specifically, can they become a part of the practice of bedside nurses and physicians?

Figure 3-1: Flow Diagram of Review #2

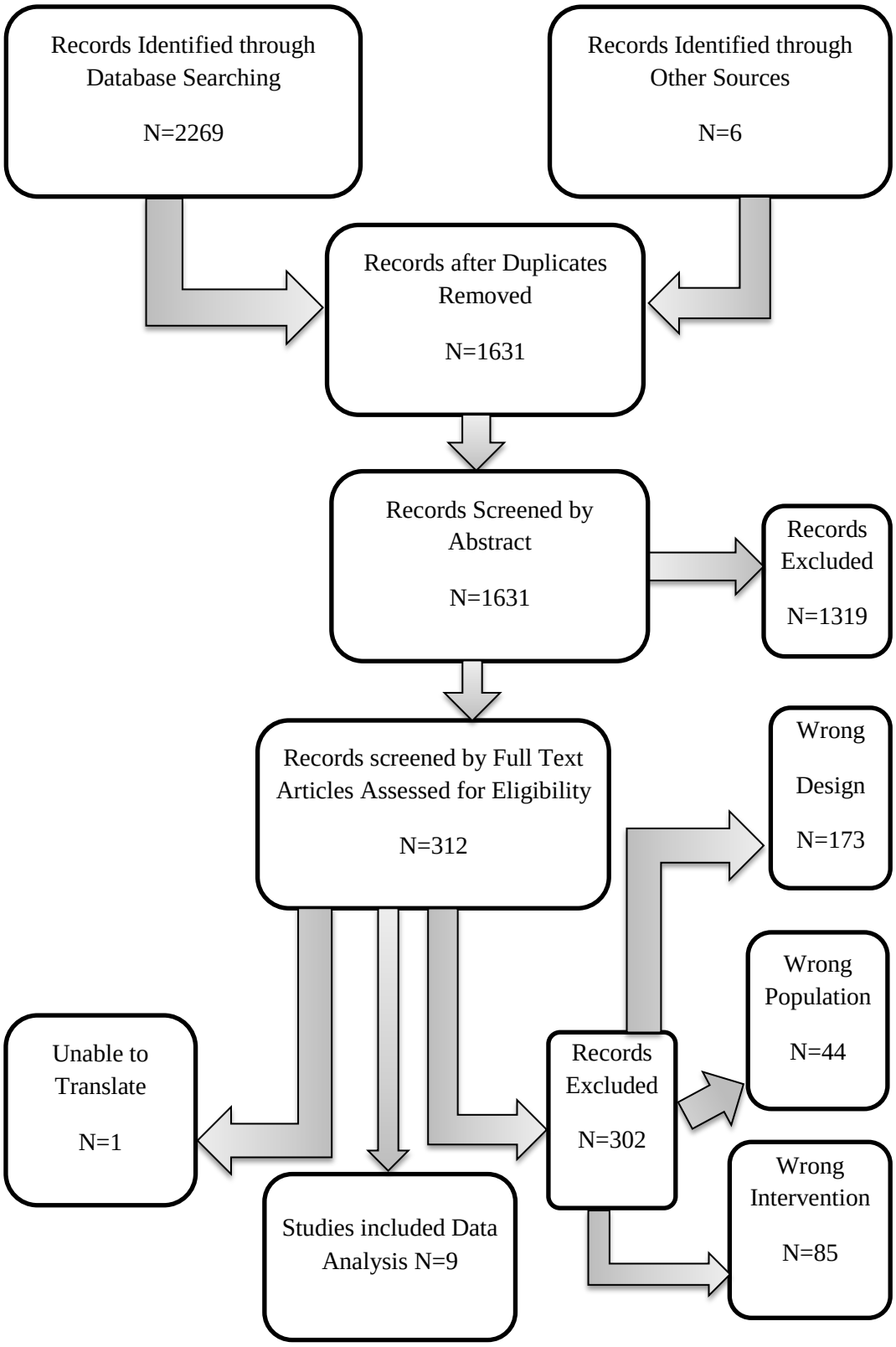


Table 3-1

Comparison for instruments for measuring sedation with para-clinical tests

Study (Country)	Population (ventilated PICU pts.)	Study Design	Para-Clinical Test /Scale	Para-clinical Test scored by /Scale Scored by	Length of assessment (# of assessments)	Correlation/ Agreement
Bustos et al. (Bustos Bu et al., 2007) (Chile)	N= 9 median age - 50 mos. (range 9-168), mean age- 69.8 mos.	Prospective – blinded observational study	BIS/ CS	Investigator A/ Investigator B	BIS measurements and CS scores obtained for 24 hours. (90 BIS and 90 Paired CS)	BIS vs. CS in the individual measurements r = 0.74 Categorized BIS measurements vs. CS r= 0.92.
Gjerstad et al. (Gjerstad et al., 2008) (Norway)	N=20 Age range -1 day to 11 yrs.	Cohort study	SC/ CB-S	Investigator A/ Investigator B (20 paired CS-B and BIS done before, during , and after suctioning (60 total))	NSCF, ASCF, Mean SC level, CS-B were measured for 2 minutes before, during, and 10mins after endotracheal suctioning	The regression analysis showed: NSCF & CB-S (before to during sxn) r^2 = 0.61 (P <.0005) r^2= 0.46 (P =.001) NSCF & CB-S (during to after sxn) ASCF & CB-S (before to during sxn) r^2=0.04 (p=0.41 ASCF & CB-S (during to after sxn) r^2=0.00 (p=0.99) Mean SC level & CS (before to during sxn) r^2=0.13 (p=0.12) Mean SC level & CS (during to after sxn) r^2=0.05 (p=0.33)
Lamas et al. (Lamas et al., 2008)(Spain)	N=77 Age range 15 days and 228 mos. (median age-8 mos.)	Prospective observational study	BIS & AEP/ M-RS & CS	Investigator/ Beside Nurse	M-RS, CS, BIS, AEP measured once a day for a maximum of 5 days (234 observations, 140 in non-paralyzed 94 observations in paralyzed patients)	*Non-paralyzed pts.:* BIS vs. AEP Overall (n=140) r=0.61, p<0.001 <6 mos.(n=52) r=0.661, p<0.001 >6 mos.-2yrs. . (n= 56) 0.618, p<0.001 >2 yrs. (n=32) 0.482** , p <0.001 BIS-RS Overall (n=140) r =-0.701, p<0.001 <6 mos. (n=52) r=-0.823, p<0.001

Study (Country)	Population (ventilated PICU pts.)	Study Design	Para-Clinical Test /Scale	Para-clinical Test scored by /Scale Scored by	Length of assessment (# of assessments)	Correlation/ Agreement
						>6 mos.-2yrs. (n=56) r=-0.626, p <0.001 > 2 yrs. (n=32)r= -0.624, p<0.001 <u>BIS-CS</u> Overall (n=140) r= 0.482, p<0.001 <6 mos. (n=52)r= 60.662, p <0.001 >6 mos.-2yrs. (n=56) r= 0.410, p<0.001 >2 yrs. (n=32) r=0.299, p>0.001, p<0.01 <u>AEP index-RS</u> Overall (n=140) r =-0.588, p<0.001 <6 mos. (n=52) r=-0.611, p<0.001 >6 mos.-2yrs.(n=56) r=-0.639, p<0.001 >2yrs. (n=32) r= -0.443, p>0.001 <u>AEP index-CS</u> Overall (n=140) r= 0.532, p<0.001 < 6mos. (n=52) r=0.601, p<0.001 >6 mos.-2yrs. .-2yrs.(n= 56) r=0.457, p< 0.001 >2 yrs. (n=32) r=0.624, p<0.001 <u>RS-CS</u> Overall (n=140) r =-0.729, p<0.001 < 6 mos. (n=52) r=-0.746, p<0.001 >6 mos.-2yrs. (n=56) r=-0.699, p<0.001 >2yrs. (n=32) r=-0.734, p<0.001 <u>*Paralyzed pts.:*</u> BIS-AEP index (n=94) r = 0.200, p>0.01 BIS-RS (n=94) r = -0.165, p>0.01 BIS-CS(n=94) r = 0.161, p >0.01 AEP index-RS (n=94) r = -0.234, p>0.01 AEP index-CS (n=94) r = 0.190 ,

Study (Country)	Population (ventilated PICU pts.)	Study Design	Para-Clinical Test /Scale	Para-clinical Test scored by /Scale Scored by	Length of assessment (# of assessments)	Correlation/ Agreement
						p>0.01 RS-CS: r = -0.411, P< 0.01 <u>Level of Agreement (κ)/ Agreement Percentage:</u> BIS-AEP index κ= 0.514, 76% BIS-RS κ =0.559, 78 % BIS-CS κ = 0.280, 64 % AEP index-RS κ = 0.473, 74 % AEP index-CS κ =0.393, 69% RS-COMFORT κ = 0.517, 77 % with BIS and AEP index
Froom et al. (Froom et al., 2008) (England)	N=19 Median age-1.5 yrs. (age range-0.55-9.7 yrs.)	Prospective single-blinded observational study	BIS (collected from right and left hemisphere)/	Investigator/ Beside PICU Nurse/ CS	BIS and CS measured Pre-physiotherapy (10 mins. before), during physiotherapy and after physiotherapy.* (28 sets of data (included a pre-physiotherapy, during physiotherapy, and post-physiotherapy periods). 83 Comfort scores, 1 missing))	<u>Un-stimulated periods</u> Mean BIS to CS r=0.603 , p <0.001 Left BIS to CS r= 0.569 , p<0.001 Right BIS to CS r=0.584, p< 0.001 <u>Stimulated period</u> Peak mean BIS to CS r= 0.383, p= 0.044 Peak left BIS to CS r=0.459, p= 0.014 Peak right BIS to CS r=0.223, p=0.253. -CS were separated into four BIS categories which determined different depths of sedation: 81-100 were defined as light sedation, 61-80 as moderate sedation, 41-60 as deep sedation, and 40 as very deep sedation. Mean CS were calculated for each category of sedation. Kruskal-Wallis multi-group analysis between the groups showed a significant (P <0.001) overall difference.
Twite et al. (Twite et al., 2005) (USA)	N=75 median age 10 mos., range 1 mo.- 12 yrs.	Prospective observational study	BIS/ CS	Investigator A/ Investigator B	The study ended after 12 scores of BS and CS were obtained or upon extubation. CS	The overall correlation coefficient for all time points was r= 0.54 (p < .0001).

Study (Country)	Population (ventilated PICU pts.)	Study Design	Para-Clinical Test /Scale	Para-clinical Test scored by /Scale Scored by	Length of assessment (# of assessments)	Correlation/ Agreement
					obtained every 30 mins. (869 valid paired observations of BIS values and COMFORT scores)	Time 1-(n=75)r= 0.62 Time 2- (n=75) r= 0.53 Time 3 (n=75) r= 0.44 Time 4 (n=74) r=0.51 Time 5 (n=73) r=0.44 Time 6 (n=74) r= 0.57 Time 7 (n=74) r= 0.50 Time 8 (n=73) r= 0.67 Time 9 (n=71) r= 0.57 Time 10 (n=72) r= 0.52 Time 11 (n=68) r= 0.54 Time 12 (n=65) r= 0.52 <u>*By Age group:*</u> Time 1 < 6 mos. (n=19) r= 0.75 > 6 mos. (n=56) r= 0.60 Time 2 < 6 mos. (n=19) r= 0.54 > 6 mos. (n=56)r= 0.53 Time 3 < 6 mos. (n=19) r= 0.57 >6 mos. (n=56) r= 0.39 Time4 < 6 mos. (n=18) r= 0.44 >6 mos. (n=56) r=0.54 Time 5 < 6 mos. (n=18) r=0.50 >6 mos. (n=55) r= 0.43 Time 6 < 6 mos. (n=19) r= 0.78 >6 mos. (n=55) r= 0.51 Time 7 < 6 mos. (n=19) r=0.40 >6 mos. (n=55) r= 0.52 Time 8 <6 mos. (n=18)r=0.42 >6 mos. (n= 55) r= 0.70 Time 9 < 6 mos. (n=19) r= 0.58 >6 mos. (n=52) r= 0.52 Time 10

Study (Country)	Population (ventilated PICU pts.)	Study Design	Para-Clinical Test /Scale	Para-clinical Test scored by /Scale Scored by	Length of assessment (# of assessments)	Correlation/ Agreement
						< 6 mos.(n=19) r=0.53 >6 mos. (n=53) r= 0.52 Time 11 < 6 mos. (n=19) r= 0.59 >6 mos. (n=49) r= 0.53 Time 12 < 6 mos. (n=19) r= 0.27 >6 mos. (n=46) r= 0.57 The results of the repeated measures analysis showed that the correlation coefficient between CS scores and BIS values, averaged over subjects, was 0.56 (p< 0.0001). The correlation coefficient between CS scores and BIS values, averaged over time, was r=0.61 (p < 0.0001).
Triltsch et al. (Triltsch et al., 2005) (Germany)	N=40 Age- 5.6 mos. (21 days-16 years)	Prospective, blinded study	BIS/ CS	Investigator B/ Investigator B	Every patient assessed 3 times (BIS and CS measured) with a minimum of 1 hour between assessments. (120 paired assessments)	<u>BIS & CS</u> n = 40, P = 0.001; Spearman's rho: r = 0.651, r^2 = 0.42 For patients without ketamine n = 38, P = 0.001; Spearman's rho: r = 0.668; r^2 = 0.45 This total percentage is derived from correct predictions in 90% of deeply sedated patients (CS score <17, n = 29) and in 55% of lightly sedated patients (CS score 17-26, n = 11).
Courtman et al. (10) (England)	N=40 Age- range- 1 mo. to 16 yrs. (mean age-3.9 yrs.)	Prospective convenience sample	BIS/ CS	Investigator/ Bedside Nurses	The study period was continued for a minimum of 2 h up to 16 h. CS and BIS completed throughout study period. (373 paired assessments)	The correlation coefficient between the BIS scores and the CS observed in this study of R=0.5 (r2=0.25, p<0.0001)
Aneja et al. (Aneja et al.,	N= 48 Phase 1- sedation	Prospective observational	BIS/ RS	Investigator/Bedside Nurse	20 hrs. , CS and BIS scores done hourly	<u>Phase 1- sedation only</u>:

Study (Country)	Population (ventilated PICU pts.)	Study Design	Para-Clinical Test /Scale	Para-clinical Test scored by /Scale Scored by	Length of assessment (# of assessments)	Correlation/ Agreement
2003) (USA)	only- mean age- 6.3 yrs. (range-1-16 yrs.). Phase 3- sedation +paralyzed- mean age 8.4 yrs. (range- 0.5-19 yrs.).	study			(phase 1- sedation group only 458 paired phase 2- sedation+ paralysis- 475 paired)	BIS score & RS: r=0.77 ($p < .0001$). Analysis of variance, with subsequent t-tests demonstrated significant BIS score differences between each RS group ($F = 230$, $p< 0.0001$). ROC analysis of patients with an RS of 1 had an area under the curve of 0.91. Receiver operator characteristic analysis on those patients who were considered to be over-sedated by RS assessment (RS 6) resulted in an area under the curve of 0.93. <u>Phase 2- sedated and paralyzed:</u> The nurses tended to assess the paralyzed patients as being either over-sedated or comfortable. For each of the three nurse assessment groups, there were a wide range of BIS scores.
Crain et al. (Crain et al., 2002) (USA)	N= 31 Mean age- 53 mos. (median age-25 mos.).	Prospective convenience sample	BIS/ CS	Investigator A/ Investigator B	Daily paired BIS and CS assessments up to 5 days. (144 paired assessments)	The mean BIS and CS measurements were 60+/- 1.6 and 17+/- 0.3, respectively The correlation between the BIS and COMFORT scale measurements was $R^2 =.26$

BIS=Bispectral Index; AEP= Auditory Evoked Potentials Index; SC= Skin Conductance; NSCF= (NSCF), ASCF=amplitude of skin conductance fluctuations, NMB= Neuromuscular Blocker, SXN=suction, M-RS= Modified Ramsay Scale; RS= Ramsay Scale; CS= Comfort Scale; CB-S Comfort-Behavioral scale, ROC= Receiver operator characteristic

*Physiotherapy: tracheal suctioning, physiotherapy fibrillation, percussion, and patient turning

Table 3-2

Quality assessment of studies comparing systematic assessment scales with para-clinical tests using the COSMIN checklist

Study	CTT/IRT	Hypotheses Testing	Interpretability	Generalizability
Gjerstad et al. (Gjerstad et al., 2008)	CTT	Poor	+	+++
Lamas et al. (Lamas et al., 2008)	CTT	Excellent	++++	++++
Froom et al. (Froom et al., 2008)	CTT	Good	+	++++
Twite et al. (Twite et al., 2005)	CTT	Excellent	+++	++++
Triltsch et al. (Triltsch et al., 2005)	CTT	Good	++	++++
Courtman et al. (Courtman et al., 2003)	CTT	Good	+++	++++
Aneja et al. (Aneja et al., 2003)	CTT	Good	++++	++
Crain et al. (Crain et al., 2002)	CTT	Fair	+++	+++
Bustos Bu et al. (Bustos Bu et al., 2007)	CTT	Fair	+++	++

+ Limited; ++ Somewhat; +++ Acceptable; ++++ Superior

References

Ambuel, B., Hamlett, K. W., Marx, C. M., & Blumer, J. L. (1992). Assessing distress in pediatric intensive care environments: The COMFORT scale. *Journal of Pediatric Psychology, 17*(1), 95-109. doi: 10.1093/jpepsy/17.1.95

Aneja, R., Heard, A. M., Fletcher, J. E., & Heard, C. M. (2003). Sedation monitoring of children by the bispectral index in the pediatric intensive care unit. *Pediatric Critical Care Medicine, 4*(1), 60-64.

Bavdekar, S. B., Mahajan, M. D., & Chandu, K. V. (1999). Analgesia and sedation in paediatric intensive care unit. *Journal of Postgraduate Medicine, 45*(3), 95-102. doi: Retrieved from: http://www.jpgmonline.com/text.asp?1999/45/3/95/338

Bustos Bu, R., Fuentes, S., & C. (2007). Bispectral index and COMFORT scale melation in the evaluation of sedation on paediatrics intensive care units. [Correlacion entre analisis biespectral y escala COMFORT en la evaluacion de sedacion en la Unidad de Cuidados Intensivos Pediatricos.] *Revista Chilena De Pediatria, 78*(6), 592-598.

Carnevale, F. A., & Ducharme, C. (1997). Adverse reactions to the withdrawal of opioids and benzodiazepines in paediatric intensive care. *Intensive & Critical Care Nursing, 13*(4), 181-188. doi: 10.1016/S0964-3397(97)80012-2

Courtman, S. P., Wardurgh, A., & Petros, A. J. (2003). Comparison of the bispectral index monitor with the comfort score in assessing level of sedation of critically ill children. *Intensive Care Medicine, 29*(12), 2239-2246. doi: 10.1007/s00134-003-1997-3

Crain, N., Slonim, A., & Pollack, M. M. (2002). Assessing sedation in the pediatric intensive care unit by using BIS and the COMFORT scale... bispectral index. *Pediatric Critical Care Medicine, 3*(1), 11-14.

De Jonghe, B., Cook, D., Appere-De-Vecchi, C., Guyatt, G., Meade, M., & Outin, H. (2000). Using and understanding sedation scoring systems: A systematic review. *Intensive Care Medicine, 26*(3), 275-285.

Froom, S. R., Malan, C. A., Mecklenburgh, J. S., Price, M., Chawathe, M. S., Hall, J. E., & Goodwin, N. (2008). Bispectral index asymmetry and COMFORT score in paediatric intensive care patients. *British Journal of Anaesthesia, 100*(5), 690-696. doi: 10.1093/bja/aen035

Gjerstad, A. C., Wagner, K., Henrichsen, T., & Storm, H. (2008). Skin conductance versus the modified COMFORT sedation score as a measure of discomfort in artificially ventilated children. *Pediatrics, 122*(4), e848-53. doi: 10.1542/peds.2007-2545

Harris, P. A., Taylor, R., Thielke, R., Payne, P., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research

informatics support. *Journal of Biomedical Informatics, 42*(2), 377-381. doi: 10.1016/j.jbi.2008.08.010

Ista, E., Van Dijk, M., Tibboel, D., & De Hoog, M. (2005). Assessment of sedation levels in pediatric intensive care patients can be improved by using the COMFORT "behavior" scale. *Pediatric Critical Care Medicine, 6*(1), 58-63. doi: 10.1097/01.PCC.0000149318.40279.1A

Jin, H. S., Yum, M. S., Kim, S. L., Shin, H. Y., Lee, E. H., Ha, E. J., . . . Park, S. J. (2007). The efficacy of the COMFORT scale in assessing optimal sedation in critically ill children requiring mechanical ventilation. *Journal of Korean Medical Science, 22*(4), 693-697. doi: 10.3346/jkms.2007.22.4.693

Lamas, A., Lopez-Herce, J., Sancho, L., Mencia, S., Carrillo, A., Santiago, M. J., & Martinez, V. (2008). Assessing sedation in critically ill children by bispectral index, auditory-evoked potentials and clinical scales. *Intensive Care Medicine, 34*(11), 2092-2099. doi: 10.1007/s00134-008-1198-1

Marx, C. M., Smith, P. G., Lowrie, L. H., Hamlett, K. W., Ambuel, B., Yamashita, T. S., & Blumer, J. L. (1994). Optimal sedation of mechanically ventilated pediatric critical care patients. *Critical Care Medicine, 22*(1), 163-170.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010). *COSMIN checklist manual*. The Netherlands: doi: Retrieved from: www.cosmin.nl

Ramsay, M. A., Savege, T. M., Simpson, B. R., & Goodwin, R. (1974). Controlled sedation with alphaxalone-alphadolone. *British Medical Journal, 2*(5920), 656-659. doi: 10.1136/bmj.2.5920.656

Refworks-COS. (2011). *Refworks 2.0 web based bibliographic manager*. Michigian, USA:

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research, 21*(4), 651-657. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=mesx&AN=21732199; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:21732199&id=doi:&issn=0962-9343&isbn=&volume=21&issue=4&spage=651&pages=651-7&date=2012&title=Quality+of+Life+Research&atitle=Rating+the+methodological+quality+in+systematic+reviews+of+studies+on+measurement+properties%3A+a+scoring+system+for+the+COSMIN+checklist.&aulast=Terwee&pid=%3Cauthor%3ETerwee+CB%2CMokkink+LB%2CKnol+DL%2COstelo+RW%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E21732199%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Triltsch, A. E., Nestmann, G., Orawa, H., Moshirzadeh, M., Sander, M., Grosse, J., . . . Spies, C. D. (2005). Bispectral index versus COMFORT score to determine the level of sedation in paediatric intensive care unit patients: A prospective study. *Critical Care (London, England), 9*(1), R9-17. doi: 10.1186/cc2977

Twite, M. D., Zuk, J., Gralla, J., & Friesen, R. H. (2005). Correlation of the bispectral index monitor with the COMFORT scale in the pediatric intensive care unit. *Pediatric Critical Care Medicine, 6*(6), 648-653. doi: 10.1097/01.PCC.0000185482.76715.D2

von Baeyer, C. L., & Spagrud, L. J. (2007). Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. *Pain, 127*(1-2), 140-150. doi: 10.1016/j.pain.2006.08.01

Chapter 4: General Discussion and Conclusions

In the previous chapters, I presented results from two systematic reviews conducted for my paper-based thesis. The aim of this knowledge synthesis project was to: 1) identify available scales appropriate for measuring physiological and behavioural cues of pain, non-pain related distress, and the adequacy of analgesia and sedation in PICU and NICU patients who are mechanically ventilated; 2) describe the instruments in terms of how they were developed, the elements of behavioural and physiological cues of pain and non-pain related distress, analgesia and sedation they assess, and the results of any validity and reliability testing completed on these scales; and, 3) determine the efficacy/effectiveness of these instruments on patient care outcomes, including total use of analgesics and sedatives, fluctuations in analgesia and sedation between nursing shifts, length of ventilation and PICU/NICU stay, and any adverse withdrawal effects. In the following sections, I will summarize the key findings of papers one and two, discuss the implications of the reviews for future research and clinical practice, discuss limitations of the reviews, and reflect on the use of the COSMIN checklist to assess the quality of the studies meeting the inclusion criteria for objectives 1 and 2, and use of REDCap for abstracting and organizing data for this review.

Key Findings from Paper 1

Twenty-eight articles were included in this review. From those 28 articles, 17 instruments were determined to be appropriate measures of physiological and behavioural cues of pain, non-pain related distress, sedation, and/or analgesia in mechanically ventilated PICU or NICU patients (Appendix E). Of these scales,

three assessed the pain, non-pain related distress, and sedation: the Comfort Scale and its modified form, the Comfort-Behavioral Scale (Comfort-B), and the Neonatal Pain, Agitation, and Sedation Scale (N-PASS) (Ambuel, Hamlett, Marx, & Blumer, 1992; Hummel, Puchalski, Creech, & Weiss, 2008; Ista, Van Dijk, Tibboel, & De Hoog, 2005; Van Dijk et al., 2000). The N-PASS is age-dependent as its validity has only been established in neonates (Hummel et al., 2008). However, it does have validity in measuring procedural pain (Hummel, Lawlor-Klean, & Weiss, 2010).

Five scales were identified that assess pain only: the Multi-dimensional Assessment Pain Scale (MAPS), the Faces, Legs, Activity, Cry, and Consolability Scale (FLACC; in its original and modified form), Pain Assessment Tool (PAT), and the Cardiac Analgesic Assessment Scale (CAAS) (Johansson & Kokinsky, 2009; Ramelet, Rees, McDonald, Bulsara, & Abu-Saad, 2007; Spence, Gillies, Harrison, Johnston, & Nagy, 2005; Suominen et al., 2004; Voepel-Lewis, Zanotti, Dammeyer, & Merkel, 2010). Of these scales, the MAPS and FLACC Scale have shown validity for measuring procedural pain (Ramelet et al., 2007; Ramelet, Rees, McDonald, Bulsara, & Huijer Abu-Saad, 2007). Like the N-PASS, PAT has only been validated in neonates (Spence et al., 2005). The CAAS did not exhibit strong validity lacks generalizability as it was developed specifically for post-cardiac surgery patients who are mechanically ventilated (Suominen et al., 2004).

Seven scales were identified for assessing sedation only, the Vancouver Sedative Recovery Scale (VSRS), the Hughes et al. unnamed sedation scale, the Parkinson et al. unamed sedation scale, the State Behavioral Sale (SBS), the Penn

State Children's Hospital Sedation Algorithm (PSCHSA), the Hartwig Scale, and the Ramsay Scale (in its original and slightly modified form) (Curley, Harris, Fraser, Johnson, & Arnold, 2006; Hartwig, Roth, & Theisohn, 1991; Hughes et al., 1994; Macnab, Levine, Glick, Susak, & Baker-Brown, 1991; Parkinson et al., 1997; Popernack, Thomas, & Lucking, 2004; Ramsay, Savege, Simpson, & Goodwin, 1974). Of these scales the Hartwig and SBS instruments are the most researched (Curley et al., 2006; Hartwig et al, 1991). The validity of the Ramsay scale and its modified form requires more investigation in the pediatric population.

Only one scale was identified for assessing sedation in PICU patients who are both mechanically ventilated and muscle relaxed (chemically paralyzed): the PICU Modified Comfort Sedation Scale for the Muscle Relaxed Patient (Razmus, Clarke, & Naufel, 2003). This scale, however, requires further investigation in terms of its reliability and validity in this population.

Lastly, one published instrument was found for assessing delirium (a component of non-pain related distress) in PICU patients, the Pediatric Confusion Method for the Intensive Care Unit (p-CAM-ICU) (Smith et al., 2011). A diagnostic algorithm for pediatric delirium was presented by another group of authors, but it has not been validated so it did not meet the inclusion criteria for this review. Like the p-CAM-ICU, this algorithm uses Richmond Agitation-Sedation Scale (RASS) as apart of its algorithm, but also employs the Pediatric Anesthesia Emergence Delirium Scale (PAED) (Schieveld et al., 2009).

As described in paper 1, three studies were found which examined the effect of systematic scales for the assessment of pain, non-pain distress, sedation, and analgesia on patient outcomes. In all three studies, these scales were used to guide the rater in following a sedation protocol (Alexander, Carnevale, & Razack, 2002; Ista, De Hoog, Tibboel, & Van Dijk, 2009; Jin et al., 2007). Two studies showed an increase in the amount of pain and sedation medication adminstered following the implementation of the sedation directed protocol guided by these scales and one showed a decrease (Alexander et al., 2002; Ista et al., 2009; Jin et al., 2007). This diffference may be related to the sedation protocol given that each of the three protocols clearly direct the user as to when sedation and anaglesia should be administered and which medication should be used. The use of the sedation directed protocols did however, have a positive impact on patient outcomes as the intervention was shown to decrease length of ventilation, decrease PICU stay, decrease the development of adverse withdrawal effects in one study (Jin et al., 2007). It also appeared to improve the assessment of sedation as there were fewer incidents of undersedation in one of the studies and more patients' level of sedation was rated as adequate in another study (Alexander et al., 2002; Ista et al., 2009).

Key Findings from Paper 2

The Comfort Scale, the Ramsay Scale, and their modified forms, the Comfort-B Scale, and the Modified Ramsay Scale, respectively, have been compared with para-clinical tests (Bispectral index scale (BIS), middle latency auditory-evoked potential index (AEP index), and Skin Conductance (SC), which

are alternative measures of sedation and are thought to be more objective than observational systematic scales of sedation (Aneja, Heard, Fletcher, & Heard, 2003; Bustos Bu, Fuentes, & C, 2007; Courtman, Wardurgh, & Petros, 2003; Crain, Slonim, & Pollack, 2002; Froom et al., 2008; Gjerstad, Wagner, Henrichsen, & Storm, 2008; Lamas et al., 2008; Triltsch et al., 2005; Twite, Zuk, Gralla, & Friesen, 2005). Eight of the studies showed moderate to good correlations between BIS and the Comfort Scale in non-paralyzed patients, indicating convergent validity between the two measures of sedation (Bustos Bu et al., 2007; Courtman et al., 2003; Crain et al., 2002; Froom et al., 2008; Lamas et al., 2008; Triltsch et al., 2005; Twite et al., 2005). This correlation appears to decrease during periods of stimulation (i.e., physiotherapy), possibly due to impedance of the BIS measurement (Froom et al., 2008). The modified Ramsay Score and the Ramsay Score also correlated significantly with BIS scores (Aneja et al., 2003; Lamas et al., 2008). The degree of correlation between BIS and the Ramsay and Comfort scales, however, decreased when observed in paralyzed patients in two of the studies (Aneja et al., 2003; Lamas et al., 2008).

The number of skin conductance fluctuations (NSCF), mean skin conductance level, and the amplitude of skin conductance (ASCF) was compared with the Comfort-B Scale in one of the studies identified for inclusion in the review. The NSCF before, during, and after tracheal suctioning correlated with changes in the Comfort-B scale, but ASCF and mean SC did not (Gjerstad et al., 2008). AEP index was compared with the Modified Ramsay Scale and the

Comfort Scale in one of the studies identified. It correlated well with both scales in non-paralyzed patients, but not in paralyzed patients (Lamas et al., 2008).

A perfect correlation cannot be expected between the the para-clinical tests and these scales as they measure different variables. For instance, BIS measures the level of hypnosis (sedation), whereas the Comfort scale is designed to measure overall distress including sedation, pain, and agitation (Bustos Bu et al., 2007; Twite et al., 2005). A correlation is found because these variables are related to each other. For example, a patient may be awake and alert with a high BIS value and comfortable with a moderately low Comfort score. In contrast, a stimulus may awake a patient who is sedated thereby changing the BIS value but not affecting the Comfort score to a great extent (Twite et al., 2005). The comparison of para-clinical tests and observational systematic scales of sedation is useful, however, as it provides insight into the validity of both tools in measuring sedation. Therefore, further comparison of para-clinical tests with the other sedation scales identified in this review would be useful.

Implications for Future Research

The Comfort Scale and Comfort- B Scale have been extensively researched. Further research, however, is required for the remaining 15 scales to determine their reliability and validity in clinical practice. The comparison of para-clinical tests and observational systematic scales of sedation is useful as it provides insight into the validity of both of tools in measuring sedation. Further comparison of para-clinical tests with the other sedation scales identified in this review is needed. Additional research is needed to asses the effect of instruments

on patient care outcomes. Where possible these scales should be assessed without the co-intervention of a sedation protocol. Lastly, none of the studies investigated assessed fluctuations of analgesia and sedation between nursing shifts, which may be a more conclusive method of assessing how these scales affect the amount of analgesia and sedation patients receive when they are assesed using these scales. Fluctuations in analgesia and sedation between nursing shifts indicates if the patient is in a steady state of analgesia or sedation or if these levels are flutuating. It may also indicate if the patient is becoming resisant to their current medications for analgesia and sedation.

Implications for Clinical Practice

This knowledge synthesis project provides pracitioners in PICUs and NICUs with a comprehensive list of instruments for measuring pain, non-pain related distress, and adequacy of sedation and analgesia. It further identifies which of those components the scale measures, the structure, and scoring of the instruments. Additionally, the evaluation of the psychometric properties is included allowing the clinical team to examine the validity and reliability of the instruments in clinical practice. Lastly, the efficacy of using some of these scales is identified based on the effect on patient care outcomes. The information provided in this systematic review, therefore, allows the clinical team to determine which instrument(s) may be both applicable and useful in their clinical setting (Ambuel et al., 1992).

Limitations of the Review

1. Due to time restraints and workloads, three different second reviewers were involved at the the various stages of the review (see Table 4-1).

Table 4-1

Reviewers Identified by Phase

Reviewer	Phase 1 Screening by abstract and title	Phase 2 Full text	Quality assessment	Data abstraction
TD (first reviewer all phases)	X	X	X	X
SI	X	X		
ES			X	X
GR - supervisor		X	X	

This may influence the results as the different reviewers varied in level of education, training, and familarilty with the review. It, however, can be beneficial for the same reason. The first reviewer (TD) remained consistent throughout each phase, and ensured that each reviewer was versed on the review topic and objectives, defintions of terms and the use of REDCap.

2. Only studies using validated instruments were included for Objective 3. This resulted in three studies being excluded from objective 3. One of these scales was the PSCHSA as the study was determined to fit better with objective 1 and 2 as its validity and reliability was not assessed prior to use and the instrument was only assessed in terms of what was assumed to be hypothesis testing.

3. Due to lack of an available translator, one Bulgarian study could not be assessed for inclusion at the full text level.We did receive assistance with the screening of French, Spanish, Russian, and Portugese from available translators. None of these articles met the inclusion criteria for paper 1. One of Spanish articles met the inclusion criteria for paper 2 and was translated using google translator .

4. Authors of several of the articles did not specify or define the pyschometric properities they were assessing. As a result the reviewers made assumptions as to which properties were being assessed when completing the quality assessment of the studies for objectives 1 and 2 using the COSMIN checklist. All assumptions were made with the consensus of at least two of the reviewers (one of which being myself). The defintions of the different psychometric properties as outlined by the COSMIN checklist manual and from Waltz et al. (2010) were used in making this assumptions (Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010b; Waltz, Strickland, & Lentz, 2010).

5. Only studies using validated instruments were included for part two of paper one. This resulted in approximately 3 studies being excluded from objective 3. One of these scales was the PSCHSA as the study was determined to fit better with objective 1 and 2 as its validity and reliability was not assessed prior to use and the instrument was only assessed in terms of what was assumed to be hypothesis testing.

6. A non-validated tool was used to assess the quality of the studies included in objective 3 :Before and After Quality assessment tool adapated from the New Castle Ottawa Scale by the by Alberta Research Centre for Health Evidence (ARCHE) (Wells et al., 2011). This scale was use because it was simple to understand and efficient.

Reflection on the use of the COSMIN checklist

The COSMIN checklist was originally developed to evaluate health- related patient- reported outcomes. However, it can also be used to evaluate the quality of studies on the measurement properties of other measurement instruments (Mokkink et al., 2009; Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010a). For the most part, it proved adaptable to evaluating the quality of studies assessing the psychometric properties of instruments for assessing pain, non-pain related distress, sedation, and analgesia. There were, however, some disadvantages. The COSMIN checklist does not separate out convergent validity from hypothesis testing when assessing the measurment property of construct validity (Terwee et al., 2012). Instead they are assessed an one aspect of construct validity. As a result, many of the studies in the reviews were rated poor or fair as no hypothesis was provided by the author of the article. Similarly, in all of the categories assessed by the COSMIN checklist the sample size influenced its rating even in some studies where a power calculation was completed for sample size, indicating that the sample size was appropriate for that study. COSMIN instead assigns ratings of poor for sample sizes less than 30, moderate for sample sizes of 30-49, good for sample sizes of 50-99, and excellent

for sample sizes greater than 100 (Appendix B) (Terwee et al., 2012). Lastly, a 4-point rating scale is not provided for the dimensions of interpretability and generalizability, despite the suggestion that these properties should be evaluated for every study (Terwee et al., 2012). As a result we developed a method of rating these properties.

Advantages of using the COSMIN checklist are: 1) it is a well researched checklist and has been validated and proven useful in studies assesing measurement properties (Mokkink et al., 2006; Mokkink et al., 2009; Mokkink et al., 2010; Mokkink, Terwee, Knol, Stratford, Alonso, Patrick, Bouter, & de Vet et al., 2010; Mokkink, Terwee, Knol, Stratford, Alonso, Patrick, Bouter, & de Vet, 2010; Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010a; Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010c; Terwee et al., 2012), and 2) a manual and description of the scoring system is available for use (Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, Bouter, & de Vet, 2010b; Terwee et al., 2012). Disadvantages of the checklist are: 1) it not simple to learn as the user needs to be familiar with the psychometric properties it assesses and be able to identify them when the author of the study does not make this identification, and, 2) it is time consuming to learn and use.

Reflections on use of REDCap

Study data for this review were collected and managed using REDCap (Research Electronic Data Capture) hosted at University of Alberta's Women and Children's Health Research Institute. REDCap is a secure, web-based application designed to support data capture for research studies providing: 1) an intuitive

interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and, 4) procedures for importing data from external sources (Harris et al., 2009). The advantages of using REDCap are articles could be organized and tracked easily. Unfamiliar users (such as second reviewer, ES) were able to learn the program quickly and found data verification relatively user friendly. Specifically, the articles could be assigned a matching record number to the one given to the article by Refworks and the second reviewer could mark the data abstraction as complete when it was verified allowing for the first reviewer to track the progress of this phase of the review. Additionally, the rating form could be prepared in a way that certain fields would not appear if they were not applicable to the article. The creation of the rating forms was also easy to accomplish and user friendly. The online designer allows one to choose from fields such as long and short text boxes, multiple choice fields (in the form of radio buttons or lists), and yes/no fields that meet the needs of data to be abstracted.

Although, it has several advantages such as usefulness in terms of organizing and tracking articles, REDCap was not developed originally to be used for systematic reviews. As a result, there are some disadvantages to using it for this purpose. If records (articles with an assigned reference ID) are entered into the database at the screening level and the article is excluded following screening this record cannot be deleted when using subsequent rating forms. Additionally, when data is exported to an Excel template for tabling, records that are incomplete

cannot be separated from records that are complete and specific records cannot be selected. As a result, the user must do a significant amount of cleaning of the data before using it. Another disadvantage is there is no field that allows you to make a table within REDCap. Consequently, this must be done after data abstraction. Lastly, REDCap simply allows for the organization and entry of data not data analysis or synthesis. Consequently the user must export the data from REDCap and do this separately.

Conclusion

According to the systematic review conducted for this Master of Nursing thesis project, there are 17 instruments available for assessing pain, non-pain related distress, adequacy of analgesia and sedation in PICU and NICU patients who are mechanically ventilated. Of the 17 scales evaluated the Comfort Scale and its modified form the Comfort-B scale are the most widely researched scales. Despite the rigorous quality assessment that was undertaken, it is difficult to identify which is the most superior tool to use in this population. When choosing a instrument to use in PICU or NICU, clinical teams have the option of choosing measure(s) that are easy to use and assess the condition(s) that they are evaluating whether it be acute pain, procedural pain, agitation, sedation, delirium, analgeisa, or another form of non-pain related distress. They also need to choose a scale that is generalizable to the various sub-populations of mechanically ventilated patients is pediatric critical care (e.g., neonates, children with Down Syndrome, children with congenital or acquired heart disease, children who are post-operative) and has strong reliability, validity, responsiveness, and interpretability. Future

research on all these scales may further clinicans' ability to effectively choose between these instruments and may demonstrate the value of their use related to the improvement of patient outcomes.

References

Alexander, E., Carnevale, F. A., & Razack, S. (2002). Evaluation of a sedation protocol for intubated critically ill children. *Intensive & Critical Care Nursing, 18*(5), 292-301. doi: 10.1016/S0964339702000502

Ambuel, B., Hamlett, K. W., Marx, C. M., & Blumer, J. L. (1992). Assessing distress in pediatric intensive care environments: The COMFORT scale. *Journal of Pediatric Psychology, 17*(1), 95-109. doi: 10.1093/jpepsy/17.1.95

Aneja, R., Heard, A. M., Fletcher, J. E., & Heard, C. M. (2003). Sedation monitoring of children by the bispectral index in the pediatric intensive care unit. *Pediatric Critical Care Medicine, 4*(1), 60-64.

Bustos Bu, R., Fuentes, S., & C. (2007). Bispectral index and COMFORT scale melation in the evaluation of sedation on paediatrics intensive care units. [Correlacion entre analisis biespectral y escala COMFORT en la evaluacion de sedacion en la Unidad de Cuidados Intensivos Pediatricos.] *Revista Chilena De Pediatria, 78*(6), 592-598.

Courtman, S. P., Wardurgh, A., & Petros, A. J. (2003). Comparison of the bispectral index monitor with the comfort score in assessing level of sedation of critically ill children. *Intensive Care Medicine, 29*(12), 2239-2246. doi: 10.1007/s00134-003-1997-3

Crain, N., Slonim, A., & Pollack, M. M. (2002). Assessing sedation in the pediatric intensive care unit by using BIS and the COMFORT scale... bispectral index. *Pediatric Critical Care Medicine, 3*(1), 11-14.

Curley, M. A., Harris, S. K., Fraser, K. A., Johnson, R. A., & Arnold, J. H. (2006). State behavioral scale: A sedation assessment instrument for infants and young children supported on mechanical ventilation. *Pediatric Critical Care Medicine, 7*(2), 107-114.

Froom, S. R., Malan, C. A., Mecklenburgh, J. S., Price, M., Chawathe, M. S., Hall, J. E., & Goodwin, N. (2008). Bispectral index asymmetry and COMFORT score in paediatric intensive care patients. *British Journal of Anaesthesia*, *100*(5), 690-696. doi: 10.1093/bja/aen035

Gjerstad, A. C., Wagner, K., Henrichsen, T., & Storm, H. (2008). Skin conductance versus the modified COMFORT sedation score as a measure of discomfort in artificially ventilated children. *Pediatrics, 122*(4), e848-53. doi: 10.1542/peds.2007-2545

Harris, P. A., Taylor, R., Thielke, R., Payne, P., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377-381. doi: 10.1016/j.jbi.2008.08.010

Hartwig, S., Roth, B., & Theisohn, M. (1991). Clinical experience with continuous intravenous sedation using midazolam and fentanyl in the paediatric intensive care unit. *European Journal of Pediatrics, 150*(11), 784-788. doi: 10.1007/BF02026712

Hughes, J., Gill, A., Leach, H. J., Nunn, A. J., Billingham, I., Ratcliffe, J., . . . Choonara, I. (1994). A prospective study of the adverse effects of midazolam on withdrawal in critically ill children. *Acta Paediatrica, International Journal of Paediatrics, 83*(11), 1194-1199. doi: 10.1111/j.1651-2227.1994.tb18280.x

Hummel, P., Lawlor-Klean, P., & Weiss, M. G. (2010). Validity and reliability of the N-PASS assessment tool with acute pain. *Journal of Perinatology, 30*(7), 474-478. doi: 10.1038/jp.2009.185

Hummel, P., Puchalski, M., Creech, S. D., & Weiss, M. G. (2008). Clinical reliability and validity of the N-PASS: Neonatal pain, agitation and sedation scale with prolonged pain. *Journal of Perinatology, 28*(1), 55-60.

Ista, E., De Hoog, M., Tibboel, D., & Van Dijk, M. (2009). Implementation of standard sedation management in paediatric intensive care: Effective and feasible? *Journal of Clinical Nursing*, *18*(17), 2511-2520. doi: 10.1111/j.1365-2702.2009.02836.x

Ista, E., Van Dijk, M., Tibboel, D., & De Hoog, M. (2005). Assessment of sedation levels in pediatric intensive care patients can be improved by using

the COMFORT "behavior" scale. *Pediatric Critical Care Medicine, 6*(1), 58-63. doi: 10.1097/01.PCC.0000149318.40279.1A

Jin, H. S., Yum, M. S., Kim, S. L., Shin, H. Y., Lee, E. H., Ha, E. J., . . . Park, S. J. (2007). The efficacy of the COMFORT scale in assessing optimal sedation in critically ill children requiring mechanical ventilation. *Journal of Korean Medical Science, 22*(4), 693-697. doi: 10.3346/jkms.2007.22.4.693

Johansson, M., & Kokinsky, E. (2009). The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. *Nursing in Critical Care, 14*(3), 122-130. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=19366409

Lamas, A., Lopez-Herce, J., Sancho, L., Mencia, S., Carrillo, A., Santiago, M. J., & Martinez, V. (2008). Assessing sedation in critically ill children by bispectral index, auditory-evoked potentials and clinical scales. *Intensive Care Medicine, 34*(11), 2092-2099. doi: 10.1007/s00134-008-1198-1

Macnab, A. J., Levine, M., Glick, N., Susak, L., & Baker-Brown, G. (1991). A research tool for measurement of recovery from sedation: The vancouver sedative recovery scale. *Journal of Pediatric Surgery, 26*(11), 1263-1267. doi: 10.1016/0022-3468(91)90594-J

Merkel, S. I., Voepel-Lewis, T., Shayevitz, J. R., & Malviya, S. (1997). The FLACC: A behavioral scale for scoring postoperative pain in young children. *Pediatric Nursing, 23*(3), 293-297.

Mokkink, L. B., Terwee, C. B., Gibbons, E., Stratford, P. W., Alonso, J., Patrick, D. L., . . . de Vet, H. C. (2010). Inter-rater agreement and reliability of the COSMIN (COnsensus-based standards for the selection of health status measurement instruments) checklist. *BMC Medical Research Methodology, 10*, 82. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20860789; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:20860789&id=doi:&issn=1471-2288&isbn=&volume=10&issue=&spage=82&pages=82&date=2010&title=BMC+Medical+Research+Methodology&atitle=Inter-rater+agreement+and+reliability+of+the+COSMIN+%28COnsensus-based+Standards+for+the+selection+of+health+status+Measurement+Instruments%29+checklist.&aulast=Mokkink&pid=%3Cauthor%3EMokkink+LB%2CTerwee+CB%2CGibbons+E%2CStratford+PW%2CAlonso+J%2CPatrick+DL%2CKnol+DL%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E20860789%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., . . . de Vet, H. C. (2006). Protocol of the COSMIN study: COnsensus-based standards for the selection of health measurement INstruments. *BMC Medical Research Methodology, 6*, 2. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16433905; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:16433905&id=doi:&issn=1471-2288&isbn=&volume=6&issue=&spage=2&pages=2&date=2006&title=BMC+Medical+Research+Methodology&atitle=Protocol+of+the+COSMIN+study%3A+COnsensus-based+Standards+for+the+selection+of+health+Measurement+INstruments.&aulast=Mokkink&pid=%3Cauthor%3EMokkink+LB%2CTerwee+CB%2CKnol+DL%2CStratford+PW%2CAlonso+J%2CPatrick+DL%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E16433905%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., . . . de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10*, 22. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20298

572;
http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:20298572&id=doi:&issn=1471-2288&isbn=&volume=10&issue=&spage=22&pages=22&date=2010&title=BMC+Medical+Research+Methodology&atitle=The+COSMIN+checklist+for+evaluating+the+methodological+quality+of+studies+on+measurement+properties%3A+a+clarification+of+its+content.&aulast=Mokkink&pid=%3Cauthor%3EMokkink+LB%2CTerwee+CB%2CKnol+DL%2CStratford+PW%2CAlonso+J%2CPatrick+DL%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E20298572%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., . . . de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10*, 22. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medc&AN=20298572;
http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:20298572&id=doi:&issn=1471-2288&isbn=&volume=10&issue=&spage=22&pages=22&date=2010&title=BMC+Medical+Research+Methodology&atitle=The+COSMIN+checklist+for+evaluating+the+methodological+quality+of+studies+on+measurement+pro

perties%3A+a+clarification+of+its+content.&aulast=Mokkink&pid=%3Cauthor%3EMokkink+LB%2CTerwee+CB%2CKnol+DL%2CStratford+PW%2CAlonso+J%2CPatrick+DL%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E20298572%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international delphi study. *Quality of Life Research, 19*(4), 539-549. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20169472; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:20169472&id=doi:&issn=0962-9343&isbn=&volume=19&issue=4&spage=539&pages=539-49&date=2010&title=Quality+of+Life+Research&atitle=The+COSMIN+checklist+for+assessing+the+methodological+quality+of+studies+on+measurement+properties+of+health+status+measurement+instruments%3A+an+international+Delphi+study.&aulast=Mokkink&pid=%3Cauthor%3EMokkink+LB%2CTerwee+CB%2CPatrick+DL%2CAlonso+J%2CStratford+PW%2CKnol+DL%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E20169472%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010b). *COSMIN checklist manual*. The Netherlands: doi: Retrieved from: www.cosmin.nl

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010c). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737-745. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20494804; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:20494804&id=doi:&issn=0895-4356&isbn=&volume=63&issue=7&spage=737&pages=737-45&date=2010&title=Journal+of+Clinical+Epidemiology&atitle=The+COSMIN+study+reached+international+consensus+on+taxonomy%2C+terminology%2C+and+definitions+of+measurement+properties+for+health-related+patient-reported+outcomes.&aulast=Mokkink&pid=%3Cauthor%3EMokkink+LB%2CTerwee+CB%2CPatrick+DL%2CAlonso+J%2CStratford+PW%2CKnol+DL%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E20494804%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., . . . de Vet, H. C. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research, 18*(3), 313-333.

Parkinson, L., Hughes, J., Gill, A., Billingham, I., Ratcliffe, J., & Choonara, I. (1997). A randomized controlled trial of sedation in the critically ill. *Paediatric Anaesthesia, 7*(5), 405-410. doi: 10.1046/j.1460-9592.1997.d01-109.x

Popernack, M. L., Thomas, N. J., & Lucking, S. E. (2004). Decreasing unplanned extubations: Utilization of the penn state children's hospital sedation algorithm. *Pediatric Critical Care Medicine, 5*(1), 58-62.

Ramelet, A. S., Rees, N., McDonald, S., Bulsara, M., & Abu-Saad, H. H. (2007). Development and preliminary psychometric testing of the multidimensional assessment of pain scale: MAPS. *Paediatric Anaesthesia, 17*(4), 333-340. doi: 10.1111/j.1460-9592.2006.02115.x

Ramelet, A. S., Rees, N. W., McDonald, S., Bulsara, M. K., & Huijer Abu-Saad, H. (2007). Clinical validation of the multidimensional assessment of pain scale. *Paediatric Anaesthesia*, *17*(12), 1156-1165. doi: 10.1111/j.1460-9592.2007.02325.x

Ramsay, M. A., Savege, T. M., Simpson, B. R., & Goodwin, R. (1974). Controlled sedation with alphaxalone-alphadolone. *British Medical Journal, 2*(5920), 656-659. doi: 10.1136/bmj.2.5920.656

Razmus, I. S., Clarke, K. A., & Naufel, K. Z. (2003). Development of a sedation scale for the mechanically ventilated muscle relaxed pediatric critical care patient. *Pediatric Intensive Care Nursing, 4*(1), 7-11.

Schieveld, J. N. M., van der Valk, J. A., Smeets, I., Berghmans, E., Wassenberg, R., Leroy, P. L. M. N., . . . van Os, J. (2009). Diagnostic considerations regarding pediatric delirium: A review and a proposal for an algorithm for pediatric intensive care units. *Intensive Care Medicine,* , 1-7. doi: 10.1007/s00134-009-1652-8

Smith, H. A. B., Boyd, J., Fuchs, D. C., Melvin, K., Berry, P., Shintani, A., . . . Ely, E. W. (2011). Diagnosing delirium in critically ill children: Validity and reliability of the pediatric confusion assessment method for the intensive care unit. *Critical Care Medicine, 39*(1), 150-157. doi:10.1097/CCM.0b013e3181feb489

Spence, K., Gillies, D., Harrison, D., Johnston, L., & Nagy, S. (2005). A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. *JOGNN - Journal of Obstetric, Gynecologic, & Neonatal Nursing, 34*(1), 80-86. doi: 10.1177/0884217504272810

Suominen, P., Caffin, C., Linton, S., McKinley, D., Ragg, P., Davie, G., & Eyres, R. (2004). The cardiac analgesic assessment scale (CAAS): A pain assessment tool for intubated and ventilated children after cardiac surgery. *Paediatric Anaesthesia, 14*(4), 336-343.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research, 21*(4), 651-657. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=mesx&AN=21732199; http://resolver.library.ualberta.ca/resolver?sid=OVID:medline&id=pmid:21732199&id=doi:&issn=0962-9343&isbn=&volume=21&issue=4&spage=651&pages=651-7&date=2012&title=Quality+of+Life+Research&atitle=Rating+the+methodological+quality+in+systematic+reviews+of+studies+on+measurement+properties%3A+a+scoring+system+for+the+COSMIN+checklist.&aulast=Terwee&pid=%3Cauthor%3ETerwee+CB%2CMokkink+LB%2CKnol+DL%2COstelo+RW%2CBouter+LM%2Cde+Vet+HC%3C%2Fauthor%3E%3CAN%3E21732199%3C%2FAN%3E%3CDT%3EJournal+Article%3C%2FDT%3E

Triltsch, A. E., Nestmann, G., Orawa, H., Moshirzadeh, M., Sander, M., Grosse, J., . . . Spies, C. D. (2005). Bispectral index versus COMFORT score to

determine the level of sedation in paediatric intensive care unit patients: A prospective study. *Critical Care (London, England), 9*(1), R9-17. doi: 10.1186/cc2977

Twite, M. D., Zuk, J., Gralla, J., & Friesen, R. H. (2005). Correlation of the bispectral index monitor with the COMFORT scale in the pediatric intensive care unit. *Pediatric Critical Care Medicine, 6*(6), 648-653. doi: 10.1097/01.PCC.0000185482.76715.D2

Van Dijk, M., De Boer, J. B., Koot, H. M., Tibboel, D., Passchier, J., & Duivenvoorden, H. J. (2000). The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain, 84*(2-3), 367-377.

Voepel-Lewis, T., Zanotti, J., Dammeyer, J. A., & Merkel, S. (2010). Reliability and validity of the face, legs, activity, cry, consolability behavioral tool in assessing acute pain in critically ill patients. *American Journal of Critical Care, 19*(1), 55-61. doi: 10.4037/ajcc2010624

Waltz, C. F., Strickland, O. L., & Lentz, E. R. (2010). *Measurement in nursing and health reasearch* (4th ed.). New York, NY: Springer Publishing Company.

Wells, G., Shea, B., O'Connell, J., Robertson, J., Peterson, V., Welch, V. & et al. (2011). *The newcastle-ottawa scale (NOS) for assessing the quality of*

nonrandomised studies in meta-analysis [website]. Retrieved September 7, 2012, from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Appendix A

COSMIN Checklist

The COSMIN checklist

Contact
CB Terwee, PhD
VU University Medical Center
Department of Epidemiology and Biostatistics
EMGO Institute for Health and Care Research
1081 BT Amsterdam
The Netherlands
Website: www.cosmin.nl, www.emgo.nl
E-mail: cb.terwee@vumc.nl

Step 1. Evaluated measurement properties in the article

	Internal consistency	Box A
	Reliability	Box B
	Measurement error	Box C
	Content validity	Box D
	Structural validity	Box E
	Hypotheses testing	Box F
	Cross-cultural validity	Box G
	Criterion validity	Box H
	Responsiveness	Box I
	Interpretability	Box J

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models		**yes**	**no**	**?**
1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	☐	☐	
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	☐	☐	
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	☐	☐	
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	☐	☐	☐

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency

		yes	no	?
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	☐	☐	☐
Design requirements		**yes**	**no**	**?**
2	Was the percentage of missing items given?	☐	☐	
3	Was there a description of how missing items were handled?	☐	☐	
4	Was the sample size included in the internal consistency analysis adequate?	☐	☐	☐
5	Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	☐	☐	
6	Was the sample size included in the unidimensionality analysis adequate?	☐	☐	☐
7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	☐	☐	☐
8	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**NA**
9	for Classical Test Theory (CTT): Was Cronbach's alpha calculated?	☐	☐	☐
10	for dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	☐	☐	☐
11	for IRT: Was a goodness of fit statistic at a global level calculated? e.g. χ^2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	☐	☐	☐

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)

Design requirements		**yes**	**no**	**?**
1	Was the percentage of missing items given?	☐	☐	
2	Was there a description of how missing items were handled?	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐
4	Were at least two measurements available?	☐	☐	
5	Were the administrations independent?	☐	☐	☐
6	Was the time interval stated?	☐	☐	
7	Were patients stable in the interim period on the construct to be measured?	☐	☐	☐

8	Was the time interval appropriate?		☐	☐	☐
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions		☐	☐	☐
10	Were there any important flaws in the design or methods of the study?		☐	☐	
Statistical methods		**yes**	**no**	**NA**	**?**
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	☐	☐	☐	
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	☐	☐	☐	
13	for ordinal scores: Was a weighted kappa calculated?	☐	☐	☐	☐
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	☐	☐	☐	

Box C. Measurement error: absolute measures

Design requirements		**yes**	**no**	**?**
1	Was the percentage of missing items given?	☐	☐	
2	Was there a description of how missing items were handled?	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐
4	Were at least two measurements available?	☐	☐	
5	Were the administrations independent?	☐	☐	☐
6	Was the time interval stated?	☐	☐	
7	Were patients stable in the interim period on the construct to be measured?	☐	☐	☐
8	Was the time interval appropriate?	☐	☐	☐
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	☐	☐	☐
10	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**?**
11	for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	☐	☐	

Box D. Content validity (including face validity)

General requirements		**yes**	**no**	**?**
1	Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	☐	☐	☐
2	Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	☐	☐	☐
3	Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	☐	☐	☐
4	Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	☐	☐	☐
5	Were there any important flaws in the design or methods of the study?	☐	☐	

Box E. Structural validity

		yes	**no**	**?**
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	☐	☐	☐
Design requirements		**yes**	**no**	**?**
2	Was the percentage of missing items given?	☐	☐	
3	Was there a description of how missing items were handled?	☐	☐	
4	Was the sample size included in the analysis adequate?	☐	☐	☐
5	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**NA**
6	for CTT: Was exploratory or confirmatory factor analysis performed?	☐	☐	☐
7	for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	☐	☐	☐

Box F. Hypotheses testing

Design requirements		**yes**	**no**	**?**
1	Was the percentage of missing items given?	☐	☐	
2	Was there a description of how missing items were handled?	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐

4	Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	☐	☐	☐
		yes	**no**	**NA**
5	Was the expected *direction* of correlations or mean differences included in the hypotheses?	☐	☐	☐
6	Was the expected absolute or relative *magnitude* of correlations or mean differences included in the hypotheses?	☐	☐	☐
7	for convergent validity: Was an adequate description provided of the comparator instrument(s)?	☐	☐	
8	for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	☐	☐	
9	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**NA**
10	Were design and statistical methods adequate for the hypotheses to be tested?	☐	☐	☐

Box G. Cross-cultural validity				
Design requirements		**yes**	**no**	**?**
1	Was the percentage of missing items given?	☐	☐	
2	Was there a description of how missing items were handled?	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	☐	☐	
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	☐	☐	
6	Did the translators work independently from each other?	☐	☐	☐
7	Were items translated forward and backward?	☐	☐	☐
8	Was there an adequate description of how differences between the original and translated versions were resolved?	☐	☐	
9	Was the translation reviewed by a committee (e.g. original developers)?	☐	☐	
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	☐	☐	

11	Was the sample used in the pre-test adequately described?	☐	☐	
12	Were the samples similar for all characteristics except language and/or cultural background?	☐	☐	☐
13	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**NA**
14	for CTT: Was confirmatory factor analysis performed?	☐	☐	☐
15	for IRT: Was differential item function (DIF) between language groups assessed?	☐	☐	☐

Box H. Criterion validity

Design requirements		**yes**	**no**	**?**
1	Was the percentage of missing items given?	☐	☐	
2	Was there a description of how missing items were handled?	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐
4	Can the criterion used or employed be considered as a reasonable 'gold standard'?	☐	☐	☐
5	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**NA**
6	for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	☐	☐	☐
7	for dichotomous scores: Were sensitivity and specificity determined?	☐	☐	☐

Box I. Responsiveness

Design requirements		**yes**	**no**	**?**
1	Was the percentage of missing items given?	☐	☐	
2	Was there a description of how missing items were handled?	☐	☐	
3	Was the sample size included in the analysis adequate?	☐	☐	☐
4	Was a longitudinal design with at least two measurement used?	☐	☐	
5	Was the time interval stated?	☐	☐	
6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	☐	☐	

7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	☐	☐	
Design requirements for hypotheses testing		**yes**	**no**	**?**
For constructs for which a gold standard was not available:				
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	☐	☐	☐
		yes	**no**	**NA**
9	Was the expected *direction* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	☐	☐	☐
10	Were the expected absolute or relative *magnitude* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	☐	☐	☐
11	Was an adequate description provided of the comparator instrument(s)?	☐	☐	
12	Were the measurement properties of the comparator instrument(s) adequately described?	☐	☐	
13	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**NA**
14	Were design and statistical methods adequate for the hypotheses to be tested?	☐	☐	☐
Design requirement for comparison to a gold standard		**yes**	**no**	**?**
For constructs for which a gold standard was available:				
15	Can the criterion for change be considered as a reasonable gold standard?	☐	☐	☐
16	Were there any important flaws in the design or methods of the study?	☐	☐	
Statistical methods		**yes**	**no**	**NA**
17	for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	☐	☐	☐
18	for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	☐	☐	☐

Box J. Interpretability				
		yes	**no**	**?**
1	Was the percentage of missing items given?	☐	☐	
2	Was there a description of how missing items were handled?	☐	☐	

3	Was the sample size included in the analysis adequate?	☐	☐	☐
4	Was the distribution of the (total) scores in the study sample described?	☐	☐	
5	Was the percentage of the respondents who had the lowest possible (total) score described?	☐	☐	
6	Was the percentage of the respondents who had the highest possible (total) score described?	☐	☐	
7	Were scores and change scores (i.e. means and SD) presented for relevant (sub) groups? e.g. for normative groups, subgroups of patients, or the general population	☐	☐	
8	Was the minimal important change (MIC) or the minimal important difference (MID) determined?	☐	☐	
9	Were there any important flaws in the design or methods of the study?	☐	☐	

Step 4: Determining the Generalisability of the results

Box Generalisability				
		yes	**no**	**NA**
	Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of:			
1	median or mean age (with standard deviation or range)?	☐	☐	
2	distribution of sex?	☐	☐	
3	important disease characteristics (e.g. severity, status, duration) and description of treatment?	☐	☐	☐
4	setting(s) in which the study was conducted? e.g. general population, primary care or hospital/rehabilitation care	☐	☐	
5	countries in which the study was conducted?	☐	☐	
6	language in which the HR-PRO instrument was evaluated?	☐	☐	
7	Was the method used to select patients adequately described? e.g. convenience, consecutive, or random	☐	☐	
		yes	**no**	**?**
8	Was the percentage of missing responses (response rate) acceptable?	☐	☐	☐

Appendix B

COSMIN Checklist with 4-Point Scale

COSMIN checklist with 4-point scale

Contact
CB Terwee, PhD
VU University Medical Center
Department of Epidemiology and Biostatistics
EMGO Institute for Health and Care Research
1081 BT Amsterdam
The Netherlands
Website: www.cosmin.nl, www.emgo.nl
E-mail: cb.terwee@vumc.nl

Instructions

This version of the COSMIN checklist is recommended for use in systematic reviews of measurement properties. With this version it is possible to calculate overall methodological quality scores per study on a measurement property. A methodological quality score per box is obtained by taking the lowest rating of any item in a box ('worse score counts'). For example, if for a reliability study one item in the box 'Reliability' is scored poor, the methodological quality of that reliability study is rated as poor. The Interpretability box and the Generalizability box are mainly used as data extraction forms. We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box (e.g. norm scores, floor-ceiling effects, minimal important change) of the instruments under study from the included articles. Similar, we recommend to use the Generalizability box to extract data on the characteristics of the study population and sampling procedure. Therefore no scoring system was developed for these boxes.

This scoring system is described in this paper:

Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research 2011, July 6 [epub ahead of print].

Step 1. Evaluated measurement properties in the article

	Internal consistency	Box A
	Reliability	Box B
	Measurement error	Box C
	Content validity	Box D
	Structural validity	Box E
	Hypotheses testing	Box F
	Cross-cultural validity	Box G
	Criterion validity	Box H
	Responsiveness	Box I

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models					
		excellent	**good**	**fair**	**poor**
1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	IRT model adequately described	IRT model not adequately described		
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	Software package adequately described	Software package not adequately described		
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	Method of estimation adequately described	Method of estimation not adequately described		
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	assumptions of the IRT model checked	assumptions of the IRT model partly checked	assumptions of the IRT model not checked or unknown	

To obtain a total score for the methodological quality of studies that use IRT methods, the 'worse score counts' algorithm should be applied to the IRT box in combination with the box of the measurement property that was evaluated in the IRT study. For example, if IRT methods are used to study internal consistency and item 4 in the IRT box is scored fair, while the items in the internal consistency box (box A) are all scored as good or excellent, the methodological quality score for internal consistency will be fair. However, if any of the items in box A is scored poor, the methodological quality score for internal consistency will be poor.

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency				
	excellent	**good**	**fair**	**poor**
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
Design requirements				
2 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4 Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
5 Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was performed in a similar study population	Authors refer to another study in which factor analysis was performed, but not in a similar study population	Factor analysis NOT performed and no reference to another study
6 Was the sample size included in the unidimensionality analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 6-7* #items but <100	5* #items but <100	<5* #items

7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	Internal consistency statistic calculated for each subscale separately			Internal consistency statistic NOT calculated for each subscale separately
8	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
9	for Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	Cronbach's alpha calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated
10	for CTT, dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated
11	for IRT: Was a goodness of fit statistic at a global level calculated? E.g. χ^2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistic at a global level calculated			Goodness of fit statistic at a global level NOT calculated

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)				
	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
13	for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

Box C. Measurement error: absolute measures					
		excellent	**good**	**fair**	**poor**
Design requirements					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4	Were at least two measurements available?	At least two measurements			Only one measurement
5	Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6	Was the time interval stated?	Time interval stated		Time interval NOT stated	
7	Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
11	for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population

Box D. Content validity (including face validity)					
		excellent	**good**	**fair**	**poor**
General requirements					
1	Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of the construct to be measured poorly described AND this was not taken into consideration	NOT assessed if all items refer to relevant aspects of the construct to be measured

2	Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	Assessed if all items are relevant for the study population in adequate sample size (≥10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
3	Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
4	Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehen-sively reflect the construct to be measured
5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

Box E. Structural validity				
	excellent	**good**	**fair**	**poor**
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
Design requirements				
2 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4 Was the sample size included in the analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 5-7* #items but <100	5* #items but <100	<5* #items
5 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. rotation method not described)	Other important methodological flaws in the design or execution of the study (e.g. inappropriate rotation method)

Statistical methods					
6	for CTT: Was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information	Exploratory factor analysis performed while confirmatory would have been more appropriate		No exploratory or confirmatory factor analysis performed
7	for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	IRT test for determining (uni)dimension-ality performed			IRT test for determining (uni)dimension-ality NOT performed

Box F. Hypotheses testing					
		excellent	**good**	**fair**	**Poor**
Design requirements					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)

4	Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
5	Was the expected *direction* of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
6	Was the expected absolute or relative *magnitude* of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
7	for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
8	for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)

9 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods				
10 Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Box G. Cross-cultural validity				
	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	

3	Was the sample size included in the analysis adequate?	CTT: 7* #items and ≥100 IRT: ≥200 per group	CTT: 5* #items and ≥100 OR 5-7* #items but <100 IRT: ≥200 in 1 group and 100-199 in 1 group	CTT: 5* #items but <100 IRT: 100-199 per group	CTT: <5* #items IRT: (<100 in 1 or both groups
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	Both source language and target language described			Source language NOT known
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	Expertise of the translators described with respect to disease, construct, and language	Expertise of the translators with respect to disease or construct poor or not described	Expertise of the translators with respect to language not described	
6	Did the translators work independently from each other?	Translators worked independent	Assumable that the translators worked independent	Unclear whether translators worked independent	Translators worked NOT independent
7	Were items translated forward and backward?	Multiple forward and multiple backward translations	Multiple forward translations but one backward translation	One forward and one backward translation	Only a forward translation
8	Was there an adequate description of how differences between the original and translated versions were resolved?	Adequate description of how differences between translators were resolved	Poorly or NOT described how differences between translators were resolved		

9	Was the translation reviewed by a committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)	Translation NOT reviewed by (such) a committee		
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	Translated instrument pre-tested in the target population	Translated instrument pre-tested, but unclear if this was done in the target population	Translated instrument pre-tested, but NOT in the target population	Translated instrument NOT pre-tested
11	Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described		Sample used in the pre-test NOT (adequately) described	
12	Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language /culture	Stated (but not shown) that samples were similar for all characteristics except language /culture	Unclear whether samples were similar for all characteristics except language /culture	Samples were NOT similar for all characteristics except language /culture
13	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

Statistical methods		
14 for CTT: Was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed	Multiple-group confirmatory factor analysis NOT performed
15 for IRT: Was differential item function (DIF) between language groups assessed?	DIF between language groups assessed	DIF between language groups NOT assessed

Box H. Criterion validity

	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'

5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
6	for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated		Correlations or AUC NOT calculated
7	for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated		Sensitivity and specificity NOT calculated

Box I. Responsiveness				
	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Was a longitudinal design with at least two measurement used?	Longitudinal design used			No longitudinal design used
5 Was the time interval stated?	Time interval adequately described			Time interval NOT described

6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the interim period	
7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided)	NO evidence provided, but assumable that part of the patients were changed	Unclear if part of the patients were changed	Patients were NOT changed
Design requirements for hypotheses testing					
	For constructs for which a gold standard was not available:				
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
9	Was the expected *direction* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
10	Were the expected absolute or relative *magnitude* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		

11 Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
12 Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	NO information on the measurement properties of the comparator instrument(s)
13 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods				
14 Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Design requirement for comparison to a gold standard				
For constructs for which a gold standard was available:				
15 Can the criterion for change be considered as a reasonable gold standard?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
16 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
17 for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated
18 for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Interpretability

We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box of the instruments under study from the included articles.

Box Interpretability	
Percentage of missing items	
Description of how missing items were handled	
Distribution of the (total) scores	
Percentage of the respondents who had the lowest possible (total) score	
Percentage of the respondents who had the highest possible (total) score	
Scores and change scores (i.e. means and SD) for relevant (sub) groups, e.g. for normative groups, subgroups of patients, or the general population	
Minimal Important Change (MIC) or Minimal Important Difference (MID)	

Generalizability

We recommend to use the Generalizability box to extract data on the characteristics of the study populations and sampling procedures of the included studies.

Box Generalisability	
Median or mean age (with standard deviation or range)	
Distribution of sex	
Important disease characteristics (e.g. severity, status, duration) and description of treatment	
Setting(s) in which the study was conducted (e.g. general population, primary care or hospital/rehabilitation care)	
Countries in which the study was conducted	
Language in which the HR-PRO instrument was evaluated	
Method used to select patients (e.g. convenience, consecutive, or random)	
Percentage of missing responses (response rate)	

Appendix C

Before-after quality assessment tool (BAQA)

Before-after quality assessment tool (BAQA)

Selection

1. Is the post-intervention group representative?

a) Consecutive or obviously representative series of participants from the target population (YES) * (1)
b) Potential for selection biases or not stated (Selected group of users) (NO)
c) No description of the derivation of the sample (UNCLEAR)

2. Is the pre-intervention sample representative?

a) Consecutive or obviously representative series of participants from target population (YES) * (1)
b) Potential for selection biases or not stated (Selected group of users) (NO)
c) No description of the derivation of the sample (UNCLEAR)

3) Are the pre- and post- intervention groups drawn from the same source?

a) Pre and post groups are drawn from the same source (YES) * (1)
b) Pre and post groups are drawn from different source (NO)
c) No description of the source of the groups (UNCLEAR)

Comparability:

4) Were the pre- and post- groups comparable on the basis of the design or analysis?

a) Pre- and post- intervention groups are comparable regarding main characteristics (YES) *, (1) or
b) Analysis was adjusted for differences in pre- and post- groups (YES) * (1)
c) No attempt to control for differences between the groups in the design or the analysis (NO)
d) No description of the comparability of the pre- and post- intervention groups (UNCLEAR)

Assessment of outcome:

5) Was the assessment of outcome(s) valid?

a) Confirmation of the outcome by reference to secure records, or validated methods (YES) ** (2)
b) Confirmation of the outcome in a sample of cases (YES) * (1)

c) No confirmation or clearly non-validated outcome (NO)
d) No description (UNCLEAR)

6) Was the assessment of outcome(s) reliable/accurate?

a) Independent assessment of the outcome by a second reviewer in all cases (YES) ** (2)
b) Independent assessment of the outcome by second reviewer in a sample of cases (YES) * (1)
c) No confirmation of outcomes (NO)
d) No description (UNCLEAR)

7) Was the method of outcome assessment the same for the pre- and post-intervention groups?

a) YES * (1)
b) NO
c) UNCLEAR

Intervention:

8) Did the study report the point in time when the intervention occurred?

a) The study reported that intervention occurred at a clearly defined point in time (YES) * (1)
b) The study reported that intervention did not occur at a clearly defined point in time, or not reported in the paper (NO).

9) Was the intervention clearly described?

a) YES * (1)
b) NO

Pre and post intervention periods:

10) Were the data collected during a similar timeframe?

a) Pre and post intervention periods for study are the same (e.g. June 1-30, 2004 and June 1-30, 2005) (YES) ** (2)
b) Pre and post intervention duration, however, seasonal bias may have occurred (e.g. June 1-30, 2004 and July 1-30, 2004) (YES) * (1)

c) Data collection during pre and post intervention periods for study was not conducted during similar timeframes (e.g. June 1-December 30, 2004 and January 1 –May 30, 2005) (NO)
d) It is not clear in the paper, e.g. dates of collection are not mentioned in the text (UNCLEAR).

Appendix D

REDCap Forms for Screening and Data Abstraction

Rating Form

Metadata

Record ID

_____
(Article number followed by your initials in lower case, with no spaces, e.g. "123ab")

Form completed on

_____
(YYYY-MM-DD)

Inclusion/Exclusion Form for Objective 1 and 2

1) Report of primary research published after 1970.

☐ 1 Yes
☐ 2 No
☐ 3 Unclear

2a) The study population includes inpatient, critical care, invasively mechanically intubated patients, age 0 days to 18 years.

☐ 1 Yes
☐ 2 No
☐ 3 Unclear

2b) Setting

☐ 1 PICU
☐ 2 NICU

2c) Age range

_____

3a) The study describes the development and evaluation of a systematic assessment instrument to measure physiological and behavioural cues of pain and/or non-pain related distress and/or level and effectiveness of analgesia and/or sedation or evaluation of a previously published instrument (i.e., construct validity of FLACC).

☐ 1 Yes
☐ 2 No
☐ 3 Unclear

3b) Measures

☐ Pain
☐ Non-pain related distress
☐ Level and effectiveness of sedation
☐ Level and effectiveness of analgesia

3c) Tool name

_____

3d) Comments

_____

4) The validity of the scale is assessed in terms of: --its psychometric properties: internal consistency, validity, reliability, and/or responsiveness OR --its comparison with para-clinical tests (e.g., Electroencephalography (EEG) bispectral analysis (BIS), auditory evoked potentials, or heart rate variability).

☐ 1 Yes
☐ 2 No
☐ 3 Unclear

5) The instrument is made up of one or more items based on physiological or behavioral cues of pain, and/ or non-pain related distress, and/or analgesia, and/or sedation.

☐ 1 Yes
☐ 2 No
☐ 3 Unclear

6) Each item on the instrument has a number of response options, which may be measured as categorical variables, either numerical (e.g., with a 5 or 7 point scale) or non-numerical.

☐ 1 Yes
☐ 2 No
☐ 3 Unclear

7) Instrument does NOT require self-report from patient.
- ☐ 1 Yes (does NOT require)
- ☐ 2 No (DOES require)

8a) Instrument is NOT a global rating scale, meaning it does not require the observer's global impression of the patient's pain, non-pain related distress, analgesia, or sedation.
- ☐ 1 Yes (is NOT a global rating scale)
- ☐ 2 No (IS a global rating scale)

8b) Type of scale
- ☐ 1 NRS (numerical rating scales)
- ☐ 2 VAS (visual analog scales)
- ☐ 3 Faces Scale

9) No para-clinical tests (e.g., EEG, BIS, auditory evoked potentials, or heart rate variability) to measure the level of analgesia or sedation is required for use of the tool.
- ☐ 1 Yes (para-clinical tests NOT required)
- ☐ 2 No (para-clinical tests ARE required)

10) Comments _____

11a) Reviewer's Decision
- ☐ 1 Include
- ☐ 2 Exclude
- ☐ 3 Unsure

11b) Reason for exclusion _____

12) Final Decision
- ☐ 1 Include
- ☐ 2 Exclude
- ☐ 3 Unsure

Inclusion/Exclusion Form for Objective 3

1a) The type of study is Randomized control trial (RCT), Controlled Clinical Trial (CCT), Controlled before and after study (CBA), Cohort Study, Case control study, retrospective chart review, crossover study, or Quasi-experimental design. Exclude: Quality initiatives, editorials, qualitative studies, correlation studies, cross-sectional studies, longitudinal studies, literature reviews.
- ☐ 1 Yes
- ☐ 2 No
- ☐ 3 Unclear

1b) Specify other type of study _____

2a) The study population includes inpatient, critical care, invasively mechanically intubated patients, age 0 days to 18 years.
- ☐ 1 Yes
- ☐ 2 No
- ☐ 3 Unclear

2b) Setting
- ☐ 1 PICU
- ☐ 2 NICU

2c) Age range _____

3aa) Use of an objective, systematic assessment tool is used to measure physiological and behavioural cues of pain and/or non-pain related distress and/or level and effectiveness of analgesia and/or sedation is the intervention in at least one group of mechanically ventilated patients in the study.
- ☐ 1 Yes
- ☐ 2 No
- ☐ 3 Unclear

3a) A validated objective systematic assessment tool is used to measure physiological and behavioural cues of pain and/or non-pain related distress and/or level and effectiveness of analgesia and/or sedation of a least one group of mechanically ventilated patients in the study.
- ☐ 1 Yes
- ☐ 2 No
- ☐ 3 Unclear

(Validated means: It has been tested previously or as a part of this study in terms of validity and reliability (i.e. psychometric properties). Refer to reference for assessment tool if unsure.)

3b) Measures

- ☐ Pain
- ☐ Non-pain related distress
- ☐ Level and effectiveness of sedation
- ☐ Level and effectiveness of analgesia

3c) Tool name _____

3d) Comments _____

4) Each item on the instrument has a number of response options, which may be measured as categorical variables, either numerical (e.g. with a 5 or 7 point scale) or non-numerical.

- ☐ 1 Yes
- ☐ 2 No
- ☐ 3 Unclear

5) Instrument does NOT require self-report from patient.

- ☐ 1 Yes (does NOT require)
- ☐ 2 No (DOES require)

6a) Instrument is NOT a global rating scale, meaning it does not require the observer's global impression of the patient's pain, non-pain related distress, analgesia, or sedation.

- ☐ 1 Yes (is NOT a global rating scale)
- ☐ 2 No (IS a global rating scale)

6b) Type of scale

- ☐ 1 NRS (numerical rating scales)
- ☐ 2 VAS (visual analog scales)
- ☐ 3 Faces Scale

7a) Numeric data is reported on at least one of the five outcomes.

- ☐ 1 Yes
- ☐ 2 No
- ☐ 3 Unclear

(Note: These outcomes reflect efficacy in practice, meaning they influence patient care outcomes.)

7b) Outcome(s) reported

- ☐ Total use of analgesic and sedatives
- ☐ Fluctuations in analgesia and sedation between nursing shifts
- ☐ Length of ventilation
- ☐ Length of PICU stay
- ☐ Adverse withdrawal effects
- ☐ Other (specify below)

7c) Specify other _____

8) Comments _____

9a) Reviewer's Decision

- ☐ 1 Include
- ☐ 2 Exclude
- ☐ 3 Unsure

9b) Reason for exclusion _____

10) Final Decision

- ☐ 1 Include
- ☐ 2 Exclude
- ☐ 3 Unsure

Quality Assessment Form for Objective I & II (cosmin)

Ref ID _____

Date of Quality Assessment _____

Step 1. Evaluated measurement properties in the article

- ☐ Internal consistency
- ☐ Reliability
- ☐ Measurement error
- ☐ Content validity
- ☐ Structural validity
- ☐ Hypotheses testing
- ☐ Cross-cultural validity
- ☐ Criterion validity
- ☐ Responsiveness
- ☐ Interpretability

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

1. Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)
   - ☐ Yes
   - ☐ No

2. Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED
   - ☐ Yes
   - ☐ No

3. Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)
   - ☐ Yes
   - ☐ No

4. Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))
   - ☐ Yes
   - ☐ No

Comments _____

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency

1.Does the scale consist of effect indicators, i.e. is it based on a reflective model?
- ☐ Yes
- ☐ No
- ☐ Unclear

Design Requirements: 2. Was the percentage of missing items given?
- ☐ Yes
- ☐ No

3. Was there a description of how missing items were handled?
- ☐ Yes
- ☐ No

4. Was the sample size included in the internal consistency analysis adequate?
- ☐ Yes
- ☐ No
- ☐ Unclear

5. Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?
- ☐ Yes
- ☐ No

6. Was the sample size included in the unidimensionality analysis adequate?
☐ Yes
☐ No
☐ Unclear

7. Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?
☐ Yes
☐ No
☐ Unclear

8. Were there any important flaws in the design or methods of the study?
☐ Yes
☐ No

Statistical methods: 9. For Classical Test Theory (CTT): Was Cronbach's alpha calculated?
☐ Yes
☐ No
☐ N/A

10. for dichotomous scores: Was Cronbach's alpha or KR-20 calculated?
☐ Yes
☐ No
☐ N/A

11. for IRT: Was a goodness of fit statistic at a global level calculated? e.g. $\chi 2$, reliability coefficient of estimated latent trait value (index of (subject or item) separation)
☐ Yes
☐ No
☐ Unclear

comments _____

Assessment of Quality:
☐ Poor
☐ Fair
☐ Good
☐ Excellent

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)

Design requirements: 1. Was the percentage of missing items given?
☐ Yes
☐ No

2. Was there a description of how missing items were handled?
☐ Yes
☐ No

3. Was the sample size included in the analysis adequate?
☐ Yes
☐ No
☐ Unclear

4. Were at least two measurements available?
☐ Yes
☐ No

5. Were the administrations independent?
☐ Yes
☐ No
☐ Unclear

6. Was the time interval stated?
☐ Yes
☐ No

7. Were patients stable in the interim period on the construct to be measured?
☐ Yes
☐ No
☐ Unclear

8. Was the time interval appropriate?
☐ Yes
☐ No
☐ Unclear

9. Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions
☐ Yes
☐ No
☐ Unclear

10. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

Statistical methods: 11. for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?
- ☐ Yes
- ☐ No
- ☐ N/A

12. for dichotomous/nominal/ordinal scores: Was kappa calculated?
- ☐ Yes
- ☐ No
- ☐ N/A

13. for ordinal scores: Was a weighted kappa calculated?
- ☐ Yes
- ☐ No
- ☐ N/A
- ☐ Unclear

14. for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic
- ☐ Yes
- ☐ No
- ☐ N/A

comments _____

Assessment of Quality
- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Box C. Measurement error: absolute measures

Design requirements: 1. Was the percentage of missing items given?
- ☐ Yes
- ☐ No

2. Was there a description of how missing items were handled?
- ☐ Yes
- ☐ No

3. Was the sample size included in the analysis adequate?
- ☐ Yes
- ☐ No
- ☐ Unknow

4. Were at least two measurements available?
- ☐ Yes
- ☐ No

5. Were the administrations independent?
- ☐ Yes
- ☐ No
- ☐ unclear

6. Was the time interval stated?
- ☐ Yes
- ☐ No

7. Were patients stable in the interim period on the construct to be measured?
- ☐ Yes
- ☐ No
- ☐ unclear

8. Was the time interval appropriate?
- ☐ Yes
- ☐ No
- ☐ Unclear

9. Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions
- ☐ Yes
- ☐ No
- ☐ Unclear

10. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

Statistical methods: 11. for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?
- ☐ Yes
- ☐ No

comments _____

Assessment of Quality
- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Box D. Content validity (including face validity)

General requirements: 1.Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?
- ☐ Yes
- ☐ No
- ☐ Unclear

2. Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)
- ☐ Yes
- ☐ No
- ☐ Unclear

3. Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)
- ☐ Yes
- ☐ No
- ☐ Unclear

4. Was there an assessment of whether all items together comprehensively reflect the construct to be measured?
- ☐ Yes
- ☐ No
- ☐ Unclear

5. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

comments _____

Assessment of Quality
- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Box E. Structural validity

1. Does the scale consist of effect indicators, i.e. is it based on a reflective model?
- ☐ Yes
- ☐ No
- ☐ Unclear

Design requirements: 2. Was the percentage of missing items given?
- ☐ Yes
- ☐ No

3. Was there a description of how missing items were handled?
- ☐ Yes
- ☐ No

4. Was the sample size included in the analysis adequate?
- ☐ Yes
- ☐ No
- ☐ Unclear

5. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

Statistical methods: 6. for CTT: Was exploratory or confirmatory factor analysis performed?
- ☐ Yes
- ☐ No
- ☐ N/A

7.for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?
- ☐ Yes
- ☐ No
- ☐ N/A

comments _____

Assessment of Quality
- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Box F. Hypotheses testing

Design requirements: 1.Was the percentage of missing items given?
- ☐ Yes
- ☐ No

2. Was there a description of how missing items were handled?
- ☐ Yes
- ☐ No

3.Was the sample size included in the analysis adequate?
- ☐ Yes
- ☐ No
- ☐ Unclear

4. Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?
- ☐ yes
- ☐ no
- ☐ unclear

5. Was the expected direction of correlations or mean differences included in the hypotheses?
- ☐ Yes
- ☐ No
- ☐ N/A

6. Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?
- ☐ Yes
- ☐ No
- ☐ N/A

7.for convergent validity: Was an adequate description provided of the comparator instrument(s)?
- ☐ Yes
- ☐ No

8. for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?
- ☐ Yes
- ☐ No

9. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

Statistical methods: 10. Were design and statistical methods adequate for the hypotheses to be tested?
- ☐ Yes
- ☐ No
- ☐ N/A

comments _____

Assessment of Quality
- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Box G. Cross-cultural validity

Design requirements: 1. Was the percentage of missing items given?	☐ Yes ☐ No
2. Was there a description of how missing items were handled?	☐ Yes ☐ No
3. Was the sample size included in the analysis adequate?	☐ Yes ☐ No ☐ Unclear
4. Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	☐ Yes ☐ No
5. Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	☐ Yes ☐ No
6.Did the translators work independently from each other?	☐ Yes ☐ No ☐ Unclear
7.Were items translated forward and backward?	☐ Yes ☐ No ☐ Unclear
8. Was there an adequate description of how differences between the original and translated versions were resolved?	☐ Yes ☐ No
9. Was the translation reviewed by a committee (e.g. original developers)?	☐ Yes ☐ No
10. Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	☐ Yes ☐ No
11. Was the sample used in the pre-test adequately described?	☐ Yes ☐ No
12. Were the samples similar for all characteristics except language and/or cultural background?	☐ Yes ☐ No ☐ Unclear
13. Were there any important flaws in the design or methods of the study?	☐ Yes ☐ No
Statistical methods: 14.for CTT: Was confirmatory factor analysis performed?	☐ Yes ☐ No ☐ N/A
15. for IRT: Was differential item function (DIF) between language groups assessed?	☐ Yes ☐ No ☐ N/A
comments	_____

Box H. Criterion validity

Design requirements: 1. Was the percentage of missing items given?
- ☐ Yes
- ☐ No

2. Was there a description of how missing items were handled?
- ☐ Yes
- ☐ No

3. Was the sample size included in the analysis adequate?
- ☐ Yes
- ☐ No
- ☐ Unclear

4. Can the criterion used or employed be considered as a reasonable 'gold standard'?
- ☐ Yes
- ☐ No
- ☐ Unclear

5. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

Statistical methods: 6.for continuous scores: Were correlations, or the area under the receiver operating curve calculated?
- ☐ Yes
- ☐ No
- ☐ N/A

7. for dichotomous scores: Were sensitivity and specificity determined?
- ☐ Yes
- ☐ No
- ☐ N/A

comments _____

Assessment of Quality
- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Box I. Responsiveness

Design requirements: 1. Was the percentage of missing items given?
- ☐ Yes
- ☐ No

2. Was there a description of how missing items were handled?
- ☐ Yes
- ☐ No

3. Was the sample size included in the analysis adequate?
- ☐ Yes
- ☐ No
- ☐ Unclear

4. Was a longitudinal design with at least two measurement used?
- ☐ Yes
- ☐ No

5. Was the time interval stated?
- ☐ Yes
- ☐ No

6. If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?
- ☐ Yes
- ☐ No

7. Was a proportion of the patients changed (i.e. improvement or deterioration)?
- ☐ Yes
- ☐ No

For constructs for which a gold standard was not available:

8. Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?
- ☐ Yes
- ☐ No
- ☐ Unclear

9. Was the expected direction of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?
- ☐ Yes
- ☐ No
- ☐ N/A

10. Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?
- ☐ Yes
- ☐ No
- ☐ N/A

11. Was an adequate description provided of the comparator instrument(s)?
- ☐ Yes
- ☐ No

12. Were the measurement properties of the comparator instrument(s) adequately described?
- ☐ Yes
- ☐ No

13. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

Statistical methods: 14. Were design and statistical methods adequate for the hypotheses to be tested?
- ☐ Yes
- ☐ No
- ☐ N/A

Hypotheses testing:

Design requirement for comparison to a gold standard

For constructs for which a gold standard was available: 15.Can the criterion for change be considered as a reasonable gold standard?
- ☐ Yes
- ☐ No
- ☐ Unclear

16. Were there any important flaws in the design or methods of the study?
- ☐ Yes
- ☐ No

Statistical methods: 17. for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?
- ☐ Yes
- ☐ No
- ☐ N/A

18. for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?
- ☐ Yes
- ☐ No
- ☐ N/A

comments _____

Assessment of Quality
- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Box J. Interpretability

1. Was the percentage of missing items given?
☐ Yes
☐ No

2. Was there a description of how missing items were handled?
☐ Yes
☐ No

3. Was the sample size included in the analysis adequate?
☐ Yes
☐ No
☐ Unclear

4. Was the distribution of the (total) scores in the study sample described?
☐ Yes
☐ No

5. Was the percentage of the respondents who had the lowest possible (total) score described?
☐ Yes
☐ No

6. Was the percentage of the respondents who had the highest possible (total) score described?
☐ Yes
☐ No

7. Were scores and change scores (i.e. means and SD) presented for relevant (sub) groups? e.g. for normative groups, subgroups of patients, or the general population
☐ Yes
☐ No

8. Was the minimal important change (MIC) or the minimal important difference (MID) determined?
☐ Yes
☐ No

9. Were there any important flaws in the design or methods of the study?
☐ Yes
☐ No

comments _____

Assessment of Quality
☐ Poor
☐ Fair
☐ Good
☐ Excellent

Step 4: Determining the Generalisability of the results

Box Generalisability

Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of: 1.median or mean age (with standard deviation or range)?
☐ Yes
☐ No

Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of: 2. distribution of sex?
☐ Yes
☐ No

Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of: 3. important disease characteristics (e.g. severity, status, duration) and description of treatment?
☐ Yes
☐ No
☐ N/A

Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of: 4. setting(s) in which the study was conducted? e.g. general population, primary care or hospital/rehabilitation care
☐ Yes
☐ No

Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of: 5. countries in which the study was conducted?

- ☐ Yes
- ☐ No

Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of: 6. language in which the HR-PRO instrument was evaluated?

- ☐ Yes
- ☐ No

7. Was the method used to select patients adequately described? e.g. convenience, consecutive, or random

- ☐ Yes
- ☐ No

8. Was the percentage of missing responses (response rate) acceptable?

- ☐ Yes
- ☐ No
- ☐ Unclear

comments _____

Assessment of Quality

- ☐ Poor
- ☐ Fair
- ☐ Good
- ☐ Excellent

Objective I & II Data Abstraction Form

Ref ID _____

Date of Data Abstraction _____

Date of Data Abstraction Verification _____

Basic Study Characteristics

First Author _____

Year of Publication _____

Journal of Publication _____

Country _____

Study Design _____

Publication Type _____

Publication Language _____

Title of Article _____

Funding _____

Is this the original report on the instrument or a validation/reliability study or both:

- ☐ Original Report
- ☐ Validation/Reliability Study
- ☐ Both
- ☐ Unclear

Comments _____

Does the Second Reviewer Agree with Basic Study Characteristics Abstracted?

- ☐ Yes
- ☐ No

Second Reviewer comments on Basic Study Characteristics: _____

Instrument/Scale Characteristics

How many scale are assessed in the study

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4

Name of Instrument 1 #: _____

Instrument #1 Assesses:

- ☐ Pain
- ☐ Non-pain related distress
- ☐ Level of Sedation
- ☐ Level of Analgesia

Specify type of non-pain related distress: _____
(e.g.anxiety, agitation, delirium, confusion, fear, ect.)

Name of Instrument 2 #: _____

Scale # 2 Assesses

- ☐ Pain
- ☐ Non-pain related distress
- ☐ Level of Sedation
- ☐ Level of Analgesia

Specify type of non-pain related distress

_____
(e.g.anxiety, agitation, delirium, confusion, fear, ect.)

Name of Instrument 3 #:

_____

Instrument #3 assesses:

- ☐ Pain
- ☐ Non-pain related distress
- ☐ Level of Sedation
- ☐ Level of Anxiety

Specify Type of non-pain related distress:

_____
(e.g.anxiety, agitation, delirium, confusion, fear, ect.)

Name of Instrument 4 #:

_____

Instrument # 4 assesses:

- ☐ Pain
- ☐ Non-pain related distress
- ☐ Level of Sedation
- ☐ Level of Analgesia

Specifiy Type of non-pain related distress:

_____
(e.g.anxiety, agitation, delirium, confusion, fear, ect.)

How was the instrument tested/ Validated?

- ☐ Comparison with non-systematic Instrument
- ☐ Comparison with para-clinical testing
- ☐ Comparison with another systematic instrument (Examples of non-systematic rating scales are the NRS=numerical rating scale, NIS=Nurse Interpretation Scale, VAS= Visual analog scale; examples of para-clinical tests are EEG, bispectral analysis (BIS), auditory evoked potentials, heart rate variability, or skin conductance; if compared to another systematic scale than it must be one not being validated in the study)

Describe non-systematic scale (Global rating instrument)

_____

Describe para-clinical testing being used:

_____

Describe systematic scale used for comparison

_____

How was the instrument # 1developed?

_____

How was the instrument # 2 developed ?

_____

How was the instrument # 3 developed?

_____

How was the instrument # 4 developed?

_____

What indicators are assessed by the instrument #1?

(e.g. facial grimace, muscle, tone, etc.)

What indicators are assessed by the instrument #2?

(e.g. facial grimace, muscle, tone, etc.)

What indicators are assessed by the instrument #3?

(e.g. facial grimace, muscle, tone, etc.)

What indicators are assessed by the instrument #4? (e.g. facial grimace, muscle, tone, etc.)

Does Instrument #1 have subscales?
☐ Yes
☐ No

Describe Subscales _____

Does Instrument #2 have subscales?
☐ Yes
☐ No

Describe Subscales _____

Does Instrument #3 have subscales?
☐ Yes
☐ No

Describe subscales _____

Does Instrument #4 have subscales?
☐ Yes
☐ No

Describe Subscales _____

Number of Items for Instrument #1? _____

Desription of items _____

Number of Items for Instrument #2? _____

Description of items _____

Number of Items for Instrument #3? _____

Description of items _____

Number of Items for Instrument #4? _____

Description of items _____

What is the lowest posssible score and Highest possible score for Instrument #1? _____

What is the lowest posssible score and Highest possible score for Instrument #2? _____

What is the lowest posssible score and Highest possible score for Instrument #3? _____

What is the lowest posssible score and Highest possible score for Instrument #4? _____

How is Instrument #1 Scored? _____
(ie. summed, averaged, weighted, overall or subscale?)

How is Instrument #2 Scored? _____
(ie. summed, averaged, weighted, overall or subscale?)

How is Instrument #3 Scored? _____
(ie. summed, averaged, weighted, overall or subscale?)

How is Instrument # 4 Scored? _____
(ie. summed, averaged, weighted, overall or subscale?)

What does the score mean for instrument #1? (i.e. low score less pain; high score more pain)

What does the score mean for instrument #2? (i.e. low score less pain; high score more pain)

What does the score mean for instrument #3? (i.e. low score less pain; high score more pain)

What does the score mean for instrument #4? (i.e. low score less pain; high score more pain)

Other comments on the instrument(s): _____

How was the actual assessment done for scale #1? (i.e. length of assessment, who completed it, etc.)

How was the actual assessment done for scale #2? (i.e. length of assessment, who completed it, etc.)

How was the actual assessment done for scale #3? (i.e. length of assessment, who completed it, etc.)

How was the actual assessment done for scale #4? (i.e. length of assessment, who completed it, etc.)

Number of Assessments completed during the validation process scale #1? _____

Number of Assessments completed during the validation process scale #2? _____

Number of Assessments completed during the validation process scale #3? _____

Number of Assessments completed during the validation process scale #4? _____

Second Reviewer agreement with instrument characteristics section?
- ☐ Yes
- ☐ No

Second Reviewer Comments for Instrument Characteristics Section _____

Statistical Measures and Results used to Validate the Instrument(s)

Is reliability of the instrument(s) measured?
- ☐ Yes
- ☐ No

(See chart for subtypes)

Which Specific Subgroups of reliability are measured?
- ☐ Internal Consistency
- ☐ Test-retest Reliability
- ☐ Inter-rater Reliability
- ☐ Intra-rater Reliability
- ☐ Measurement Error
- ☐ Not defined by subgroup

What value(s) of internal consistency is given? _____

Authors interpretation on internal consistency of instrument(s) _____

What value(s) for test-retest reliability are given? _____

Authors Interpretation of test-retest reliability of instrument(s) _____

What value(s) of Inter-rater reliability are given? _____

Authors interpretation of Inter-rater reliability of instrument(s) _____

What value(s) of intra-rater reliablity are given? _____

Authors Interpretation of intra-rater reliability of the instrument (s) _____

What value(s) of measurement error are given? _____

Authors Interpretation of measurement error of the instrument(s) _____

Values for reliability (not defined by subgroup) _____

Authors Interpretation of reliability of the instrument(s) _____

Is the validity of the instrument(s) evaluated?
- ☐ Yes
- ☐ No

(See chart for subtypes)

Which subtypes of validity of evaluated?
- ☐ Content Validity
- ☐ Criterion Validity
- ☐ Construct Validity
- ☐ Not specified

Which aspect of content validity evaluated?
- ☐ Face Validity
- ☐ Not specified

Values of Content Validity Provided _____

Authors Interpretation of Content Validity of instrument(s) _____

What aspect of criterion validity is measured?
- ☐ Predictive
- ☐ Concurrent
- ☐ Not specified

Values of criterion validity provided _____

Authors interpretation of criterion validity of instrument(s) _____

What aspect of construct validity is evaluated?
- ☐ Convergent Validity
- ☐ Structural Validity
- ☐ Hypotheses Validity
- ☐ Cross-cultural Validity
- ☐ Unspecified

If unspecified what does it appear to be? ☐(see definitions table)

Values of construct validity provided _____

Authors interpretation fo construct validity of instrument(s) _____

Is the responsiveness of the instrument measured?
- ☐ Yes
- ☐ No

Values of responsiveness provided _____

Authors interpretation of responsiveness of instrument(s) _____

Is the interpretability of the instrument evaluated? ☐ Yes ☐ No

Values of interpretability provided _____

Authors interpretation of interpretability of instrument(s) _____

What is the author's assessment of the instrument #1? (i.e. is the instrument measuring what it is supposed to be

What is the author's assessment of the instrument #2 (i.e. is the instrument measuring what it is supposed to be

What is the author's assessment of the instrument #3? (i.e. is the instrument measuring what it is supposed to be

What is the author's assessment of the instrument #4? (i.e. is the instrument measuring what it is supposed to be

What are the limitations of the study? _____

Other comments about the assessment/evaluation of the scale _____

Does the Second Reviewer agree with the data abstraction for the statistical measures/evaluation section? ☐ Yes ☐ No

Second Reviewer comments for the statistical measures/evaluation section _____

Description of the Study Population

How many males? _____

How many females? _____

Age (mean, median, and range if provided) _____

Diagnoses and # in each diagnostic group _____

Number of Ventilated patients _____

Number of non-ventilated patients _____

Other patient characteristics _____

Inclusion/Exclusion Criteria for patient population in the study _____

Comments on patient characteristics _____

Time to extract _____

Does the second reviewer agree with the data abstraction for the patient characteristic section? ☐ Yes ☐ No

Second Reviewer Comments for the Patient Characteristics Section _____

Time for second reviewer to review _____

Quality Assessment Form for Objective III (NOS)

Ref ID _____

Date of Quality Assessment _____

What type of study design is employed?

- ☐ Before-After
- ☐ Cohort
- ☐ Both (retrospective chart reviews are cohort studies when studied alone)

NOS Before and After

1. Is the post-intervention group representative?

- ☐ YES (1)
- ☐ NO
- ☐ UNCLEAR

comments _____

2. Is the per-intervention sample representative?

- ☐ YES (1)
- ☐ NO
- ☐ UNCLEAR

comments _____

3. Are the pre- and post-intervention groups drawn from the same source?

- ☐ YES (1)
- ☐ NO
- ☐ UNCLEAR

comments: _____

Comparability

4. Were the pre- and post-groups comparable on the basis of the design and analysis?

- ☐ YES-comparable (1)
- ☐ YES-Adjusted (1)
- ☐ NO
- ☐ UNCLEAR

comments _____

Assessment of Outcome

5. Was the assessment of outcome(s) valid?

- ☐ YES-reference to secure records or validated methods (2)
- ☐ YES-confirmation in a sample of cases (1)
- ☐ NO
- ☐ UNCLEAR

Comments _____

6. Was the assessment of outcome(s) reliable/accurate?

- ☐ YES-independent assessment by second reviewer-all cases (2)
- ☐ Yes- independent assessment by second reviewer-sample of cases (1)
- ☐ NO
- ☐ UNCLEAR

Comments _____

7. Was the method of outcome assessment the same for pre- and post-intervention groups?

☐ YES (1)
☐ NO
☐ UNCLEAR

Comments _____

8. Did the study report the point in time when the intervention occurred?

☐ YES (1)
☐ NO
☐ UNCLEAR

Comments _____

9. Was the intervention clearly described?

☐ YES (1)
☐ NO

Comments _____

10. Were the data collected during a similar time frame?

☐ YES-same (2)
☐ YES-same duration, but seasonal bias (1)
☐ NO
☐ UNCLEAR

Comments _____

Total Score _____
(Highest possible score=13; lowest 0)

NOS Cohort Studies- Selection

1) Representativeness of the exposed cohort

☐ truly representative of the average described in the community(1)
☐ somewhat representative of the average described in the community(1)
☐ selected group of users, e.g. nurses, volunteers
☐ no description of the derivation of the cohort

comments: _____

2) Selection of the non-exposed cohort

☐ drawn from the same community as the exposed cohort (1)
☐ drawn from a different source
☐ no description of the derivation of the non-exposed cohort

comments: _____

3) Ascertainment of exposure

☐ secure record (e.g.) surgical records (1)
☐ structured interview (1)
☐ written self-report
☐ no description

comments: _____

4) Demonstration that outcome of interest was not present at start of study

☐ Yes (1)
☐ No

comments: _____

Comparability

1) Comparability of cohorts on the basis of the design or analysis

- ☐ study control for _____ (select most important factor) (1)
- ☐ study controls for any additional factor (1)

comments: _____

Outcome

1) Assessment of outcome

- ☐ independent assessment (1)
- ☐ record linkage (1)
- ☐ self report
- ☐ no description

comments: _____

2) Was the follow-up long enough for outcomes to occur

- ☐ yes (1) (select an adequate follow up period for outcome of interest to occur)
- ☐ no

comments: _____

3) Adequacy of follow up of cohorts

- ☐ complete follow up- all subjects accounted for (1)
- ☐ subjects lost to follow up unlikely to introduce bias- small number lost>__% (select an adequate %) follow up, or description provided for the loss) (1)
- ☐ follow up rate < ___% and no description of lost
- ☐ no statement
  (80 % adequate)

comments: _____

Total Score _____
(highest score 9;lowest 0)

Objective III Data Abstraction Form

Refworks ID # _____
(Refworks ID #)

Study Characteristics

First Author _____

Year of Publication _____

Title _____

Country Trial Conducted in: _____

Publication Type _____

Publication Language _____

Journal of Publication _____

Funding _____

Type of ICU
- ☐ PICU (combined med/surg/cardiac)
- ☐ General systems PICU (med/surg)
- ☐ Cardiac PICU
- ☐ NICU
- ☐ Unclear
- ☐ Other

If other, specify type of ICU _____

Number of Beds in PICU _____

Comments (Additional Information about the PICU)

Type of Study
- ☐ Randomized Control Trial (RCT)
- ☐ Controlled Clinical Trial (CCT)
- ☐ Controlled before and after Study (CBA)
- ☐ Cohort Study
- ☐ Case Control Study
- ☐ Retrospective Chart Review
- ☐ Crossover Study
- ☐ Quasi-experimental Design
- ☐ Other

Comments on Study Design: _____

Blinding
- ☐ Yes
- ☐ No

(Was anyone blinded, assume no unless otherwise stated)

Comments on blinding: _____

Who was blinded in the study?
- ☐ Parents
- ☐ Physician
- ☐ Outcome Assessor

(Assume were not blinded unless stated)

Sampling Style _____

Name of Scale used: _____
(Full name and abbreviated form)

Scale Measures:
- ☐ Pain
- ☐ Non-pain related distress
- ☐ Level of Sedation
- ☐ Level of Analgesia

Type of Non-pain related distress
- ☐ anxiety
- ☐ fear
- ☐ agitation
- ☐ delirium
- ☐ confusion
- ☐ other

Who developed the original and Scale and/or adapted it? _____
(Name and date of publication)

What a Sedation Protocol used?
- ☐ Yes
- ☐ No

Describe the sedation protocol
(Include pg. # or figure/table #)

Additional Comments about Scale or Sedation Protocol: _____

Inclusion Criteria _____

Exclusion Criteria _____

Type of Statistical Analysis Employed _____

Second Reviewer Agreement for Study Characteristics Section
- ☐ Yes
- ☐ No

Second Reviewer Comments Study Characteristics Section: _____

Baseline Characteristics of Patients Involved in Study

Number of Patient Groups
- ☐ 1
- ☐ 2
- ☐ 3

Name of Group A _____
(Enter N for Randomized and Assessed)

Name of Group B _____
(Enter N for Randomized and Assessed)

Name of Group C _____
(Enter N for Randomized and Assessed)

Number of Males Group A _____
(Enter N for Randomized and Assessed)

Number of Males Group B _____
(Enter N for Randomized and Assessed)

Number of Males Group C _____
(Enter N for Randomized and Assessed)

Number of females Group A _____
(Enter N for Randomized and Assessed)

Number of females Group B	_____ (Enter N for Randomized and Assessed)
Number of females Group C	_____ (Enter N for Randomized and Assessed)
Age Group A	(Enter age group categories and N for Randomized and As
Age Group B	(Enter age group categories and N for Randomized and As
Age Group C	(Enter age group categories and N for Randomized and As
Mean age and Standard Deviation or Standard Error for Group A	_____
Mean age and Standard Deviation or Standard Error for Group B	_____
Mean age and Standard Deviation or Standard Error for Group C	_____
Ethnicity Group A	(Enter N for Randomized and Assessed and type)
Ethnicity Group B	(Enter N for Randomized and Assessed and type)
Ethnicity Group C	(Enter N for Randomized and Assessed and type)
Number of Patients who were ventilated Group A	_____ (Enter N for Randomized and Assessed)
Number of Patients who were ventilated Group B	_____ (Enter N for Randomized and Assessed)
Number of Patients who were ventilated Group C	(Enter N for Randomized and Assessed)
Diagnosis of Patients in Group A and N for each Diagnosis	_____ (Enter N for Randomized and Assessed)
Diagnosis of Patients in Group B and N for each Diagnosis	_____ (Enter N for Randomized and Assessed)
Diagnosis of Patients in Group C and N for each Diagnosis	_____ (Enter N for Randomized and Assessed)
Average Weight (kg) for Group A	_____ (Enter N for Randomized and Assessed)
Average Weight (kg) for Group B	_____ (Enter N for Randomized and Assessed)
Average Weight (kg) for Group C	_____ (Enter N for Randomized and Assessed)
Pediatric Risk of Mortality Score (PRISM) Mean/ Median and Min-Max for Group A	_____ (Enter N for Randomized and Assessed)
Pediatric Risk of Mortality Score (PRISM) Mean/ Median and Min-Max for Group B	_____ (Enter N for Randomized and Assessed)
Pediatric Risk of Mortality Score (PRISM) Mean/ Median and Min-Max for Group C	_____ (Enter N for Randomized and Assessed)

Pediatric Index of Mortality (PIM) Score Group A (Mean/median and/or Min-Max) _____ (Enter N for Randomized and Assessed)

Pediatric Index of Mortality (PIM) Score Group B (Mean/median and/or Min-Max) _____ (Enter N for Randomized and Assessed)

Pediatric Index of Mortality (PIM) Score Group C (Mean/median and/or Min-Max) _____ (Enter N for Randomized and Assessed)

Other Patient Characteristics Group A (Enter N for Randomized and Assessed)

Other Patient Characteristics Group B (Enter N for Randomized and Assessed)

Other Patient Characteristics Group C (Enter N for Randomized and Assessed)

Comments on patient characteristics (i.e. was there a signficant difference in demographic cha

Second Reviewer Agreement with Patient Characteristics Section
☐ Yes
☐ No

Second Reviewer Comments on Patient Characteristics Section _____

Interventions Section

Scale used group A
☐ Yes
☐ No

Scale used group B
☐ Yes
☐ No

Scale used group C
☐ Yes
☐ No

Specify Scales used Group A (Enter all scales used)

Specify Scales used Group B (Enter all scales used)

Specify Scales used Group C (Enter all scales used)

Sedation Protocol Used Group A
☐ Yes
☐ No

Sedation Protocol Used Group B
☐ Yes
☐ No

Sedation Protocol Used Group C
☐ Yes
☐ No

Describe Sedation Protocol Group A: _____

Describe Sedation Protocol Group B: _____

Describe Sedation Protocol Group C: _____

Medications used Group A (include names, routes, if they were continuous, breakthr

Medications used Group B — (include names, routes, if they were continuous, breakthro

Medications used Group C — (include names, routes, if they were continuous, breakthro

Personnel involved/performing intervention (i.e. assessing pain and sedation and making decisions about medication adminstration) Group A
- ☐ Physician
- ☐ Nurse Practitioner
- ☐ Registered Nurse
- ☐ Pharmacist
- ☐ Other

Other specify: _____

Personnel involved/performing intervention (i.e. assessing pain and sedation and making decisions about medication adminstration) Group B
- ☐ Physician
- ☐ Nurse Practitioner
- ☐ Registered Nurse
- ☐ Pharmacist
- ☐ Other

Other specifiy: _____

Personnel involved/performing intervention (i.e. assessing pain and sedation and making decisions about medication adminstration) Group C
- ☐ Physician
- ☐ Nurse Practitioner
- ☐ Registered Nurse
- ☐ Pharmacist
- ☐ Other

Other Specify _____

Co-interventions Group A _____

Co-interventions Group B _____

Co-interventions Group C _____

Comments intervention section: _____

Second Review Agreement with Interventions Section
- ☐ Yes
- ☐ No

Second Reviewer Comments on Interventions Section _____

Outcomes Measures (What was Measured?)

Total usage of Analgesics
- ☐ Yes
- ☐ No

Page in Publication — (identify pg. #, paragraph, and/or tables or figures)

Was the ouctome pre-specified by the authors?
- ☐ Yes
- ☐ No
  (must be stated in the methods)

Definition of Outcome — (includes units of measurement)

How was the outcome measured? — (e.g. as a total, median, range/sum, ect.)

Was the ouctome defined as a primary or secondary outcome in the article?
- ☐ Primary
- ☐ Seconday
  (Considered a primary outcome unless stated otherwise)

Was the sample size based on the study outcomes?
☐ Yes
☐ No

Was the Sample Size Calculated or Study Power calculated?
☐ Yes
☐ No
(Must be stated in the article)

Data Sources and Data Collection
_____

Total Usuage of Sedatives
☐ Yes
☐ No

Page in publication
(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?
☐ Yes
☐ No
(must be stated in the methods)

Definition of Outcome
(includes units of measurement)

How was the outcome measured?
(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary outcome in the article?
☐ Primary
☐ Secondary
(Considered a primary outcome unless stated otherwise)

Was the sample size calculated or the study power calculated?
☐ Yes
☐ No

Was sample size based on the study outcomes?
☐ Yes
☐ No

Data Sources and Data Collection
_____

Fluctuations in analgesia between nursing shifts
☐ Yes
☐ No

Page in Publication
(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?
☐ Yes
☐ No
(must be stated in the methods)

Defintion of Outcome
(includes units of measurement)

How was the outcome measured?
(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary outcome in the article?
☐ Primary
☐ Secondary
(Considered a primary outcome unless stated otherwise)

Was the sample size calculated or study power calculated?
☐ Yes
☐ No
(Must be stated in the article)

Was sample size based on the study outcomes?
☐ Yes
☐ No

Data Sources and Data Collection
_____

Fluctuations in sedation between nursing shifts

☐ Yes
☐ No

Page in Publication

(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?

☐ Yes
☐ No
(must be stated in the methods)

Definition of Outcome

(includes units of measurement)

How was the Outcome Measured?

(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary outcome in the article?

☐ Primary
☐ Seconday
(Considered a primary outcome unless stated otherwise)

Was the size calculated or study power calculated?

☐ Yes
☐ No
(Must be stated in the article)

Was sample size based on the study outcomes?

☐ Yes
☐ No

Data Sources and Data Collection

_____

Length of Venilation

☐ Yes
☐ No

Page in Publication

(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?

☐ Yes
☐ No
(must be stated in the methods section)

Definition of Outcome

(includes units of measurement)

How was the outcome measured?

(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary outcome in the article?

☐ Primary
☐ Secondary
(Considered a primary outcome unless stated otherwise)

Was the sample size calculated or study power calculated?

☐ Yes
☐ No
(Must be stated in the article)

Was the sample size based on the study outcomes?

☐ Yes
☐ No

Data Sources and Data Collection

_____

Length of PICU stay

☐ Yes
☐ No

Page in Publication

(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?
- ☐ Yes
- ☐ No

(must be stated in the methods)

Definition of Outcome

(includes units of measurement)

How outcome was measured?

(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary outcome in the article?
- ☐ Primary
- ☐ Seconday

(Considered a primary outcome unless stated otherwise)

Was the sample size or study power calculated?
- ☐ Yes
- ☐ No

Was sample size based on the study outcomes?
- ☐ Yes
- ☐ No

Data Sources and Data Collection

_____

Adverse Withdrawal effects
- ☐ Yes
- ☐ No

Page in Publication

(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?
- ☐ Yes
- ☐ No

(must be stated in the methods section)

Definition of Outcome

(includes units of measurement)

How was the outcome measured?

(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as primary or secondary outcome in the article?
- ☐ Primary
- ☐ Secondary

(Considered a primary outcome unless stated otherwise)

Was the sample size or study power calculated?
- ☐ Yes
- ☐ No

(Must be stated in the article)

Was sample size based on the study outcomes?
- ☐ Yes
- ☐ No

Data sources and Data Collection

_____

How many other outcomes were assessed?
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 0

Specify Outcome 1

_____

Page in Publication

(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?
- ☐ Yes
- ☐ No

(must be stated in the methods section)

Definition of Outcome

(includes units of measurement)

How outcome was measured?

(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary ouctome in the article?

- ☐ Primary
- ☐ Secondary

(Considered a primary outcome unless stated otherwise)

Was the sample size or study power calculated?

- ☐ Yes
- ☐ No

(Must be stated in the article)

Was the sample size based on study outcomes?

- ☐ Yes
- ☐ No

Data Sources and Data Collection

_____

Specify Outcome 2

_____

Page in Publication

(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?

- ☐ Yes
- ☐ No

Definition of Outcome

(includes units of measurement)

How outcome was measured?

(e.g. as a total, median, range/sum, ect.)

Was the sample size or study power calculated?

- ☐ Yes
- ☐ No

(Must be stated in the article)

Was the sample size based on the study outcomes?

- ☐ Yes
- ☐ No

Data sources and Data Collection

_____

Specify Outcome 3

_____

Page in Publication

(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specified by the authors?

- ☐ Yes
- ☐ No

(must be stated in the methods section)

Definition of outcome

(includes units of measurement)

How was the outcome measured?

(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondray outcome in the article?

- ☐ Primary
- ☐ Secondary

(Considered a primary outcome unless stated otherwise)

Was the sample size or study power calculated?

- ☐ Yes
- ☐ No

(Must be stated in the article)

Was the sample sized based on the study outcomes?
☐ Yes
☐ No

Data Sources and Data Collection
_____

Specify Outcome 4
_____

Page in Publication
(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specifief by the authors?
☐ Yes
☐ No
(must be stated in the methods section)

Definition of Outcome
(includes units of measurement)

How was the outcome measured?
(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary outcome in the article?
☐ Primary
☐ Secondary

Was the sample size or study power calculated?
☐ Yes
☐ No
(Must be stated in the article)

Was the sample size based on the study outcomes?
☐ Yes
☐ No

Data Sources and Data Collection
_____

Specify Outcome 5
_____

Page in Publication
(identify pg. #, paragraph, and/or tables or figures)

Was the outcome pre-specificied by the authors?
☐ Yes
☐ No
(must be stated in the methods section)

Definition of the Outcome
(includes units of measurement)

How was the outcome measured?
(e.g. as a total, median, range/sum, ect.)

Was the outcome defined as a primary or secondary outcome in the article?
☐ Primary
☐ Secondary
(Considered a primary outcome unless stated otherwise)

Was the sample size or study power calculated?
☐ Yes
☐ No
(Must be stated in the article)

Was sample size based on the study outcomes?
☐ Yes
☐ No

Data Sources and Data Collection
_____

Are there any results of Subgroups defined by the author of the article?
☐ Yes
☐ No

Subgroup results
(Specifify subgroups and pg./table/figure of results)

Comments on outcomes or subgroups:
_____

Time to extract by first reviewer _____

Second Review Agreement for Outcomes section
- ☐ Yes
- ☐ No

Second Reviewer comments for Outcomes Section _____

Time to Review by Second Reviewer _____

Appendix E

Systematic Assessment Scales

Figure 1: Cardiac Analgesic Assessment Scale (CAAS) (Suominen et al., 2004)

	Scoring		
Variables	*0*	*1*	*2*
Pupillary size[a]	≤2 mm (pinpoint)	3–4 mm (midsize)	>4 mm (dilated)
Heart rate[b]	Within baseline[c]	5–15% Increase	>15% Increase
Blood pressure (mean)	Within baseline[c]	5–15% Increase	>15% Increase
Respiratory and motor response[d]	No response	Cough and minimal movement settles after removal of stimulus	Cough and/or excessive movement >1 min after removal of stimuli

Response to suctioning or turning: [a]Pupils: score to the nearest size. [b]Patients with pacemaker on fixed rate or with junctional ectopic tachycardia (JET) are scored 1. [c]Within baseline of <5% increase of BP or heart rate. [d]Paralysed patients are scored 1.

Figure 2: Hartwig Sedation Scale (Hartwig, Roth, & Theisohn, 1991)

	1	2	3	4	5
A. Motor response	No spontaneous movements	Spontaneous movements with pain	Spontaneous movements of extremities	Spontaneous global movements	Continous spontaneous movements, restless
B. Mimic	No reaction	Grimacing only with pain	Cries only when with pain, rapid return to rest	Cries even when without pain, but soon returns to rest	Cries, difficult to soothe
C. Eyes	Permanently closed	Opening only with pain	Opening when manipulated, quickly falls asleep again	Spontaneous opening, soon returns to sleep	Spontaneous opening, awake for long periods, sweating
D. Respiration			Easy, spontaneous breathing, fully synchronized	Mechanical respiration not disturbed by spontaneous breathing	Spontaneous breathing not synchronous with machine, tachypnoea
E. Aspiration		No reaction when aspirated	Grimacing only, no movements of extremities	Little coughing or retching	Strong opposition, intense coughing, straining

Figure 3: Ramsay Scale (Ramsay, Savege, Simpson, & Goodwin, 1974)

Awake	Patient anxious or agitated or both	1
Levels:	Patient cooperative, oriented and tranquil	2
	Patient responds to commands only	3
Asleep	A brisk response to a light glabellar tap	4
Levels:	A sluggish response to a light glabellar tap	5
	No response	6

Figure 3a: Modified Ramsay Score (Ramsay et al., 1974)

Modified Ramsey Sedation Scale.

1. Anxious, Agitated, Restless

2. Cooperative, Oriented, Tranquil
Accepts mechanical ventilation.

3. Responds to commands only

4. Brisk response to light glabellar tap or loud noise.

5. Sluggish response to light glabellar tap or loud noise.

6. No Response.

Figure 4: Comfort Scale (Ambuel, Hamlett, Marx, & Blumer, 1992)

ALERTNESS		
Deeply asleep	1	
Lightly asleep	2	
Drowsy	3	
Fully awake and alert	4	
Hyper-alert	5	
CALMNESS		
Calm	1	
Slightly anxious	2	
Anxious	3	
Very anxious	4	
Panicky	5	
RESPIRATORY RESPONSE		
No coughing and no spontaneous respiration	1	
Spontaneous respiration with little or no response to ventilation	2	
Occasional cough or resistance to ventilator	3	
Actively breathes against ventilator or coughs regularly	4	
Fights ventilator; coughing or choking	5	
CRYING		
Quiet breathing, no crying	1	
Sobbing or gasping	2	
Moaning	3	
Crying	4	
Screaming	5	
PHYSICAL MOVEMENT		
No movement	1	
Occasional, slight movement	2	
Frequent, slight movements	3	
Vigorous movement limited to extremities	4	
Vigorous movements including torso and head	5	
MUSCLE TONE		
Muscles totally relaxed; no muscle tone	1	
Reduced muscle tone	2	
Normal muscle tone	3	
Increased muscle tone and flexion of fingers and toes	4	
Extreme muscle rigidity and flexion of fingers and toes	5	
FACIAL TENSION		
Facial muscles totally relaxed	1	
Facial muscle tone normal; no facial muscle tension evident	2	
Tension evident in some facial muscles	3	
Tension evident throughout facial muscles	4	
Facial muscles contorted and grimacing	5	
BLOOD PRESSURE (MAP) BASELINE		
Blood pressure below baseline	1	
Blood pressure consistently at baseline	2	
Infrequent elevations of 15% or more above baseline (1-3 during 2 minutes observation)	3	
Frequent elevations of 15% or more above baseline (> 3 during 2 minutes observation)	4	
Sustained elevations of 15% or more	5	
HEART RATE BASELINE		
Heart rate below baseline	1	
Heart rate consistently at baseline	2	
Infrequent elevations of 15% or more above baseline (1-3 during 2 minutes observation)	3	
Frequent elevations of 15% or more above baseline (> 3 during 2 minutes observation)	4	
Sustained elevations of 15% or more	5	

Figure 5: Comfort-Behavioral (Comfort-B) Scale (Johansson & Kokinsky, 2009)

Alertness
1. Deeply asleep
2. Lightly asleep
3. Drowsy
4. Fully awake and alert
5. Hyperalert

Calmness/agitation
1. Calm
2. Slightly anxious
3. Anxious
4. Very anxious
5. Panicky

Respiratory response (ventilated children)
1. No coughing and no spontaneous respiration
2. Spontaneous respiration with little or no response to ventilation
3. Occasional cough or resistance to ventilator
4. Actively breathes against ventilator or coughs regularly
5. Fights ventilator, cough or choking

Cry (non-ventilated children)
1. Quiet breathing, no crying
2. Sobbing or gasping
3. Moaning
4. Crying
5. Screaming

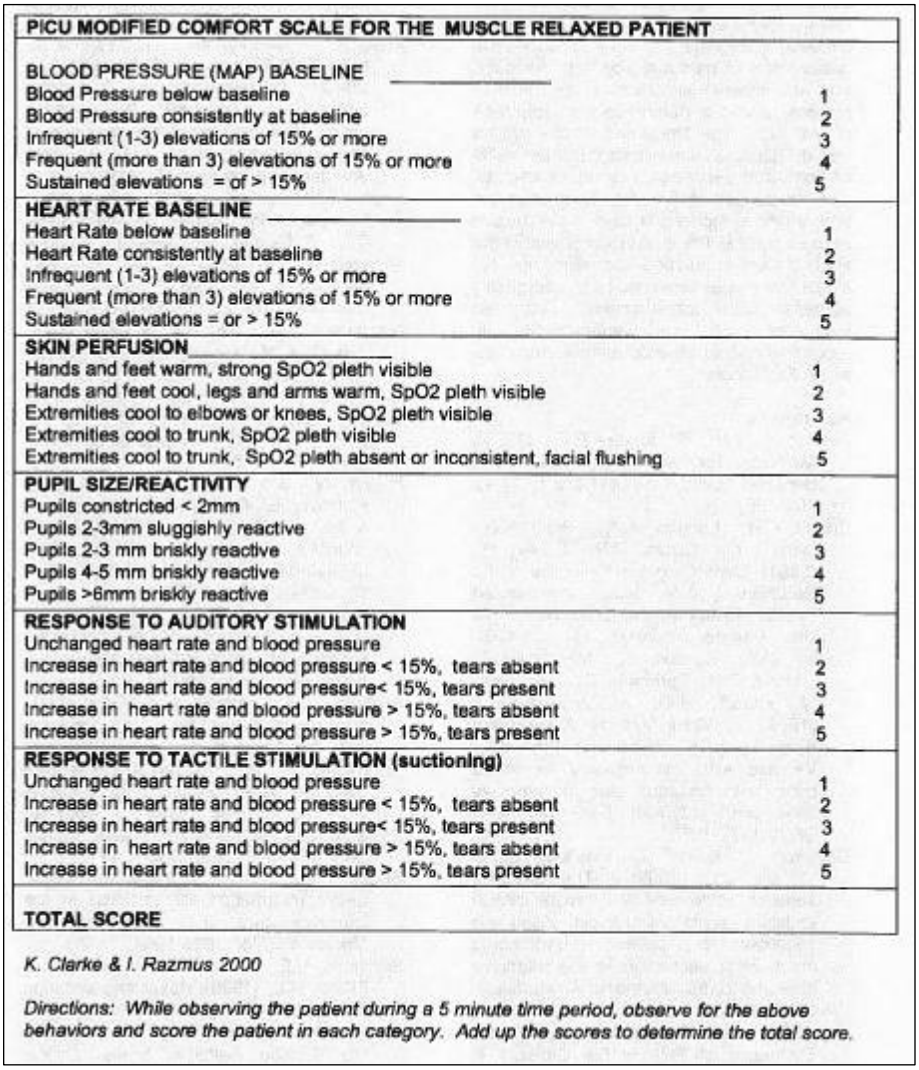
Physical movement
1. No movement
2. Occasional, slight movements
3. Frequent, slight movements
4. Vigorous movement limited to extremities
5. Vigorous movements including torso and head

Muscle tone
1. Muscles totally relaxed, no muscle tone
2. Reduced muscle tone
3. Normal muscle tone
4. Increased muscle tone and flexion of fingers and toes
5. Extreme muscle rigidity and flexion of fingers and toes

Facial tension
1. Facial muscle totally relaxed
2. Facial muscle tone normal; no facial muscle tension evident
3. Tension evident in some facial muscles
4. Tension evident throughout facial muscles
5. Facial muscles contorted and grimacing

Total

Figure 6: PICU Modified Comfort Scale for Muscle Relaxed Patients (Razmus, Clarke, & Naufel, 2003)



PICU MODIFIED COMFORT SCALE FOR THE MUSCLE RELAXED PATIENT

BLOOD PRESSURE (MAP) BASELINE _____	
Blood Pressure below baseline	1
Blood Pressure consistently at baseline	2
Infrequent (1-3) elevations of 15% or more	3
Frequent (more than 3) elevations of 15% or more	4
Sustained elevations = of > 15%	5
HEART RATE BASELINE _____	
Heart Rate below baseline	1
Heart Rate consistently at baseline	2
Infrequent (1-3) elevations of 15% or more	3
Frequent (more than 3) elevations of 15% or more	4
Sustained elevations = or > 15%	5
SKIN PERFUSION _____	
Hands and feet warm, strong SpO2 pleth visible	1
Hands and feet cool, legs and arms warm, SpO2 pleth visible	2
Extremities cool to elbows or knees, SpO2 pleth visible	3
Extremities cool to trunk, SpO2 pleth visible	4
Extremities cool to trunk, SpO2 pleth absent or inconsistent, facial flushing	5
PUPIL SIZE/REACTIVITY	
Pupils constricted < 2mm	1
Pupils 2-3mm sluggishly reactive	2
Pupils 2-3 mm briskly reactive	3
Pupils 4-5 mm briskly reactive	4
Pupils >6mm briskly reactive	5
RESPONSE TO AUDITORY STIMULATION	
Unchanged heart rate and blood pressure	1
Increase in heart rate and blood pressure < 15%, tears absent	2
Increase in heart rate and blood pressure< 15%, tears present	3
Increase in heart rate and blood pressure > 15%, tears absent	4
Increase in heart rate and blood pressure > 15%, tears present	5
RESPONSE TO TACTILE STIMULATION (suctioning)	
Unchanged heart rate and blood pressure	1
Increase in heart rate and blood pressure < 15%, tears absent	2
Increase in heart rate and blood pressure< 15%, tears present	3
Increase in heart rate and blood pressure > 15%, tears absent	4
Increase in heart rate and blood pressure > 15%, tears present	5
TOTAL SCORE	

K. Clarke & I. Razmus 2000

Directions: While observing the patient during a 5 minute time period, observe for the above behaviors and score the patient in each category. Add up the scores to determine the total score.

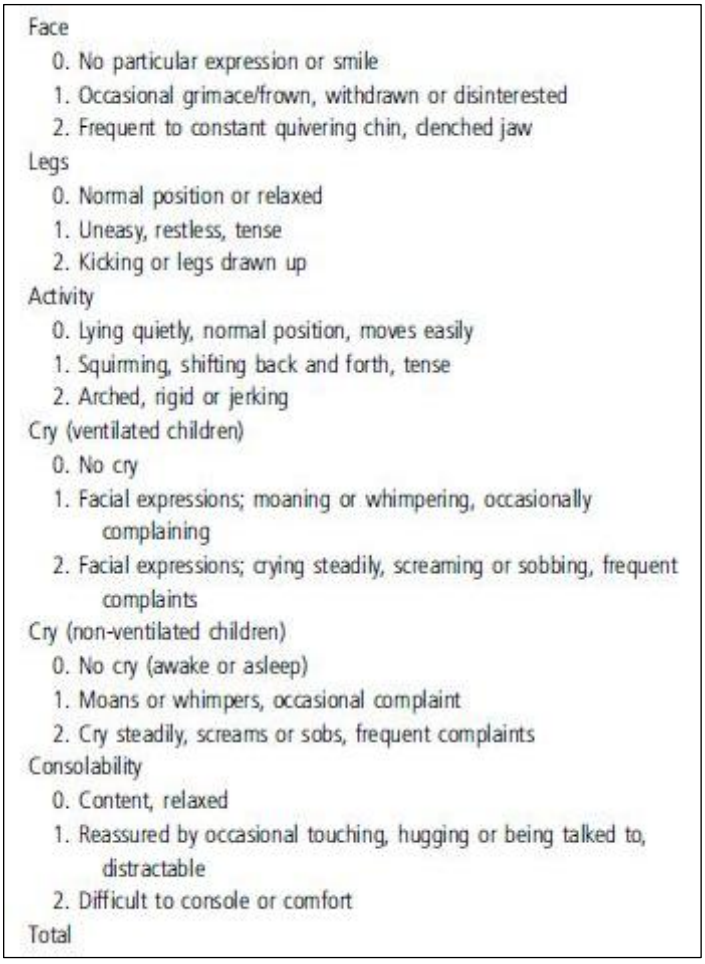
Figure 7: Parkinson et al. Sedation Scale (Parkinson et al., 1997)

1. **No response to tracheal suction.**
   There is no movement when tracheal suctioning is performed. The patient is asleep.
2. **Cough and small limb movement and/or grimace in response to tracheal suction.**
   There is brief movement of limbs or face when tracheal suctioning is performed, but *not* eye opening. The patient rapidly settles back to presuction state (asleep).
3. **Agitation with major limb movement or crying in response to tracheal suction.**
   There is marked movement of limbs which may be purposeful (hand(s) up towards tube or face) with crying (lacrimation) and/or eye opening when tracheal suctioning is performed. The patient may take longer to settle after suction (1–2 min). The patient is asleep when he/she is not being stimulated.
4. **Patient is awake and moving but is not agitated when not disturbed.**
   The patient is awake for part or all of the time. He/she may move, open eyes or be trying to communicate, but does not seem distressed when not stimulated (i.e. not being suctioned or having blood taken, etc.).
5. **Patient is awake and restless or distressed when not disturbed.**
   The patient is restless, moving around with or without eyes open; he or she may be crying, sweating and unable to settle, lines and tubes may be pulled. He or she is awake for a major part of the time.

Figure 8: Faces, Legs, Activity, Cry, Consolability (FLACC) Scale (Merkel, Voepel-Lewis, Shayevitz, & Malviya, 1997)

Face 0 - No particular expression or smile 1 - Occasional grimace or frown, withdrawn, disinterested 2 - Frequent to constant quivering chin, clenched jaw	
Legs 0 – Normal position or relaxed 1 – Uneasy, restless, tense 2 – Kicking, or legs drawn up	
Activity **0** – Lying quietly, normal position, moves easily **1** – Squirming, shifting back and forth, tense **2** – Arched, rigid or jerking	
Cry **0** – No cry (awake or asleep) **1** – Moans or whimpers; occasional complaint **2** - Crying steadily, screams or sobs, frequent complaints	
Consolability **0** – Content, relaxed **1** – Reassured by occasional touching, hugging or being talked to, distractible **2** – Difficult to console or comfort	
TOTAL SCORE	

Figure 9: Modified Faces, Legs, Activity, Cry, Consolability (FLACC) scale (Johansson & Kokinsky, 2009)



Face
- 0. No particular expression or smile
- 1. Occasional grimace/frown, withdrawn or disinterested
- 2. Frequent to constant quivering chin, clenched jaw

Legs
- 0. Normal position or relaxed
- 1. Uneasy, restless, tense
- 2. Kicking or legs drawn up

Activity
- 0. Lying quietly, normal position, moves easily
- 1. Squirming, shifting back and forth, tense
- 2. Arched, rigid or jerking

Cry (ventilated children)
- 0. No cry
- 1. Facial expressions; moaning or whimpering, occasionally complaining
- 2. Facial expressions; crying steadily, screaming or sobbing, frequent complaints

Cry (non-ventilated children)
- 0. No cry (awake or asleep)
- 1. Moans or whimpers, occasional complaint
- 2. Cry steadily, screams or sobs, frequent complaints

Consolability
- 0. Content, relaxed
- 1. Reassured by occasional touching, hugging or being talked to, distractable
- 2. Difficult to console or comfort

Total

Figure 10: Multidimensional Assessment of Pain (MAP) Scale (Ramelet, Rees, McDonald, Bulsara, & Abu-Saad, 2007)

Categories	*0*	*1*	*2*	*Score*
Vital signs HR and/or BP	Within baseline	More than 10 b·min^{-1} increase and/or more than 10 mmHg increase	More than 10 b·min^{-1} decrease and/or more than 10 mmHg decrease	
Breathing pattern	No change	Development or increase of respiratory distress	Severe respiratory distress	
Facial expressions	Relaxed	Grimace	Grimace associated with silent or weak cry	
Body movements	No movements or purposeful movements	Restless	Rigid and/or limited body movements	
State of arousal	Calm or asleep	Hyperreactive	Shut down	
Total score				

HR, heart rate; BP, blood pressure.

Definitions of the each category of the Multidimensional Assessment of Pain Scale

Vital signs

Record baseline measures when the child has been stabilized and is comfortable

Score 0 if heart rate and blood pressure measures are consistently within baseline measures

Score 1 if heart rate increased more than 10 b·min^{-1} and/or blood pressure (SBP or DBP) increased more than 10 mmHg compared with baseline measures

Score 2 if heart rate decreased more than 10 b·min^{-1} and/or blood pressure (SBP or DBP) decreased more than 10 mmHg compared with baseline measures

Breathing pattern

Record baseline measure at the beginning of each shift when the child is obviously asleep or comfortable with no external stimuli

Score 0 if there is no change in breathing pattern compared with baseline measure

Score 1 if there is an increase in rate and work of breathing and/or paradoxical breathing. For the intubated patient, it may also include fighting the ventilator (↑ respiratory distress)

Score 2 if respiratory distress is severe and compromising oxygenation

Facial expressions

Score 0 if the infant's face is relaxed, with eyes closed relaxed, mouth closed relaxed, and smooth forehead

Score 1 if the infant grimaces (frown, eyes tightly closed, and distinct vertical pull at the corners of the mouth with lips partially or completely parted)

Score 2 if the infant shows frequent grimace is associated with silent cry (tears associated with angular stretch mouth in the intubated patient) or weak cry (nonintelligible and low-pitched sounds; moaning)

Please note that silent cry that is associated with the departure of a parent cannot be scored

Body movements

Score 0 if the infant demonstrates purposeful or no movements of head, torso, arms, hands, legs, and feet

Score 1 if the infant is restless demonstrating jerky leg movements, twitching hands and feet, twisting or writhing

Score 2 if the infant shows rigid body movements that are stiff, firm, and unbending (including pulling knees up, spreading feet, fist clenched, twisting/writhing, and/or limited torso movements) or limited body movements that are naturally limited in its purpose and/or slow in motion

State of arousal

Score 0 if the infant is calm or asleep (no external stimulus is required to elicit purposeful or calm movements, awakens easily)

Score 1 if the infant is hyperreactive: highly or excessively responsive or reactive to a nonpainful stimulus (noise, touch, light, etc.) (increased state of arousal)

Score 2 if the infant's behavior is shutdown: decreased physical alertness and activity; does not communicate; or no eyes contact

SBP, systolic blood pressure; DBP, diastolic blood pressure.

Figure 11: Multidimensional Assessment of Pain (MAP) scale-Revised (Ramelet, Rees, McDonald, Bulsara, & Huijer Abu-Saad, 2007)

Categories	*0*	*1*	*2*	*Score*
Vital signs HR and/or BP	Within baseline	More than 10% increase	More than 20% increase	
Breathing pattern	No change	Development or increase of respiratory distress	Increased respiratory distress with silent or weak cry	
Facial expressions	Relaxed	Grimace	Grimace associated with silent or weak cry	
Body movements	No movements or purposeful movements	Restless	Rigid and/or limited body movements	
State of arousal	Calm or asleep	Hyperreactive	Shut down	
Total score				

Definitions:

Vital signs

Note. Record baseline measures at the beginning of each shift or 3 h postoperatively when the child is obviously asleep or comfortable with no external stimuli.

☐ Score 0 if heart rate and blood pressure measures are consistently within baseline measures.

☐ Score 1 if heart rate and/or blood pressure increased more than 10% compared with baseline measures.

☐ Score 2 if heart rate and/or blood pressure increased more than 20% compared with baseline measures.

Breathing pattern

Note. Record baseline measure at the beginning of each shift when the child is obviously asleep or comfortable with no external stimuli.

☐ Score 0 if there is no change in breathing pattern compared with baseline measure.

☐ Score 1 if there is an increase in rate and work of breathing and/or paradoxical breathing. For the intubated patient, it may also include fighting the ventilator (increased respiratory distress).

☐ Score 2 if increased respiratory distress is associated with silent cry (tears associated with angular stretch mouth in the intubated patient) or weak cry (nonintelligible, low-pitched sounds; moaning).

Facial expressions

☐ Score 0 if the infants' face is relaxed, with eyes closed relaxed, mouth closed relaxed, and smooth forehead.

☐ Score 1 if the infant shows occasional grimace (frown, eyes tightly closed, and distinct vertical pull at the corners of the mouth with lips partially or completely parted).

☐ Score 2 if the infant shows frequent grimace associated or not with silent or weak cry.

Please note that silent cry that is associated with the departure of a parent cannot be scored.

Body movements

☐ Score 0 if the infant demonstrates purposeful or no movements of head, torso, arms, hands, legs and feet.

☐ Score 1 if the infant is restless demonstrating jerky leg movements, twitching hands and feet, twisting or writhing

☐ Score 2 if the infant shows rigid body movements that are stiff, firm and unbending (including pulling knees up, spreading feet, fist clenched, twisting/writhing and/or limited torso movements) or limited body movements that are naturally limited in its purpose and or slow in motion.

State of arousal

☐ Score 0 if the infant is calm or asleep (no external stimulus is required to elicit purposeful or calm movements, awakens easily).

☐ Score 1 if the infant is hyperreactive: highly or excessively responsive or reactive to a nonpainful stimulus (noise, touch, light, etc.) (Increased state of arousal).

☐ Score 2 if the infant's behavior is shut down: decreased physical alertness and activity; does not communicate; or no eye contact.

Figure 12: Hughes et al. Sedation Scale (Hughes et al., 1994)

	Score	Comments
Eye opening	1–4	(1) None; (2) to pain; (3) to speech; (4) spontaneous
Verbal response	1–5	(1) None; (2) incomprehensible; (3) inappropriate; (4) confused; (5) orientated
	1–3 (infants)	(1) None; (2) cry; (3) alert and appropriate
Motor response to stimulus	1–6	(1) None; (2) abnormal extension; (3) abnormal flexion; (4) withdrawal from pain; (5) localization of pain; (6) obeys commands
	1–4 (infants)	(1) None; (2) abnormal extension; (3) abnormal flexion; (4) withdrawal from pain
Level of consciousness	1–4	(1) Unconscious; (2) semi-conscious; (3) sleepy but rousable; (4) fully awake
Pattern of breathing	Regular/ Irregular	
Respiration	Shallow/ Normal	
Respiratory rate		Note ventilator rate
Cough in response to suction	Yes/No	

Figure 13: Vancouver Sedation Recovery Scale (Macnab, Levine, Glick, Susak, & Baker-Brown, 1991)

	Scoring Options
Response	
A. i) Awake/alert	(4)
ii) Awake/drowsy	(3)
iii) Asleep/easily aroused	(2)
iv) Asleep/difficult to arouse	(1)
v) Asleep/unable to arouse	(0)
NOTE. IF CHILD SCORES (0) ON ABOVE, DO NOT PROCEED	
B. i) Responds fully to stimuli in an age-appropriate manner	(2)
ii) Delayed response to stimuli	(1)
iii) Absent response to stimuli	(0)
C. i) "Alert" facial expression	(1)
ii) "Flat" facial expression	(0)
Eyes	
D. i) Bright eyes	(1)
ii) Dull eyes; glazed	(0)
E. i) Looks "at you"	(1)
ii) Looks "through you"	(0)
F. i) Accommodates	(2)
ii) No attempt to accommodate	(1)
iii) Unable to accommodate	(0)
G. i) Recognition of stimulus	(1)
ii) Limited or no recognition of stimulus	(0)
H. i) Purposeful and spontaneous eye movement	(1)
ii) Little or no spontaneous or purposeful eye movement	(0)
Movement	
I. i) Spontaneous and varied central activity	(4)
ii) Spontaneous and varied peripheral activity	(3)
iii) Central activity in response to stimuli	(2)
iv) Peripheral activity in response to stimuli	(1)
v) No movement	(0)
J. i) Absence of tremor or ataxia	(1)
ii) Ataxia or tremor on being moved	(0)
K. i) Coordinated spontaneous movement	(2)
ii) Weak/coarse spontaneous movement	(1)
iii) No purposeful spontaneous movement	(0)
L. i) Shows age-appropriate manual dexterity	(2)
ii) Awkward or clumsy hand movement	(1)
iii) No fine hand movement	(0)
NOTE. Maximum score: 22 (fully awake); minimum score: 0 (unable to arouse).	

Figure 14: State Behavioral Scale broken down by Dimensions (Curley, Harris, Fraser, Johnson, & Arnold, 2006)

Dimensions	Levels
Respiratory Drive	1. No spontaneous respiratory effort 2. Spontaneous but ineffective exhaled tidal volume (Patient specific: <4cc/kg) 3. Spontaneous and effective exhaled tidal volume (Patient specific: > 4cc/kg)
Response to Ventilation	1. No spontaneous breathing 2. Easy spontaneous breathing (fully synchronized with mechanical ventilation) 3. Having difficulty synchronizing with ventilator 4. Unsynchronized with mechanical ventilation, compromising oxygenation/ventilation
Coughing	1. No cough with suction 2. Coughs only when suctioned 3. Coughs when repositioned 4. Occasional spontaneous cough 5. Frequent spontaneous coughing that does not resolve with suctioning 6. Bronchospastic or choking
Best Response to Stimulation	1. No response to noxious stimuli 2. Responds to noxious stimulus 3. Responds to touch 4. Responds to voice 5. No external stimulus is required to elicit response
Attentiveness to Care Provider	1. Unable to pay attention to care provider 2. Able to pay attention to care provider but drifts off after stimulation 3. Spontaneously pays attention to care provider (Infant – fixes and follows) 4. Vigilant of care provider/Eyes follow care provider – watchful 5. Hyper-vigilant to care provider/Panicky when care providers approach patient
Tolerance to Care	1. Does not distress with any procedure including noxious 2. Will distress with noxious procedures 3. Distresses with procedures (i.e., repositioning) 4. Distressed (e.g., picking at tubes, pulling at restraints, etc.) 5. Patient intermittently unsafe (e.g., biting ETT) 6. Patient unsafe (e.g., attempting to pull at ETT/catheters, cannot be left alone)
Consolability	1. Self-regulates/modulates own behavior 2. Able to calm with comforting touch or voice when stimulus removed; distractible 3. Does not consistently calm despite a 5-minute attempt to console 4. Unable to console
Movement after Consoled	1. Does not move 2. Occasional movement of extremities or shifting of position in bed 3. Increased movement (restless, squirming) 4. Excessive movement (thrashing side to side, kicking legs, arched, rigid) 5. Combative

Figure 14a: State Behavioral Scale score as patient's response to voice, then gentle touch, then noxious stimuli (planned endotracheal suctioning or < 5 secs of nail-bed pressure) (Curley et al., 2006)

Score	Description	Definition
−3	Unresponsive	No spontaneous respiratory effort No cough or coughs only with suctioning No response to noxious stimuli Unable to pay attention to care provider Does not distress with any procedure (including noxious) Does not move
−2	Responsive to noxious stimuli	Spontaneous yet supported breathing Coughs with suctioning/repositioning Responds to noxious stimuli Unable to pay attention to care provider Will distress with a noxious procedure Does not move/occasional movement of limbs or shifting of position
−1	Responsive to gentle touch or voice	Spontaneous but ineffective nonsupported breaths Coughs with suctioning/repositioning Responds to touch/voice Able to pay attention but drifts off after stimulation Distresses with procedures Able to calm with comforting touch or voice when stimulus removed Occasional movement of limbs or shifting of position
0	Awake and able to calm	Spontaneous and effective breathing Coughs when repositioned/occasional spontaneous cough Responds to voice/no external stimulus is required to elicit response Spontaneously pays attention to care provider Distresses with procedures Able to calm with comforting touch or voice when stimulus removed Occasional movement of limbs or shifting of position/increased movement (restless, squirming)
+1	Restless and difficult to calm	Spontaneous effective breathing/having difficulty breathing with ventilator Occasional spontaneous cough Responds to voice/no external stimulus is required to elicit response Drifts off/spontaneously pays attention to care provider Intermittently unsafe Does not consistently calm, despite 5-min attempt/unable to console Increased movement (restless, squirming)
+2	Agitated	May have difficulty breathing with ventilator Coughing spontaneously No external stimulus required to elicit response Spontaneously pays attention to care provider Unsafe (biting endotracheal tube, pulling at catheters, cannot be left alone) Unable to console Increased movement (restless, squirming, or thrashing side-to-side, kicking legs)

Figure 15: Penn State Children's Hospital Sedation Algorithm

Table 1. Penn State Children's Hospital Sedation Algorithm for ventilated children

Level 1	Goal: Awake and interactive with environment; i.e., watches television, communicates (generally for more mature children with neuromuscular cause for assisted ventilation) Action: PRN anxiolytics/analgesics
Level 2	Goal: Sleepy, arouses to light stimulation, becomes excited with nursing care/suctioning, moves spontaneously, turns head, consistently breathes above ventilator Action: PRN anxiolytics/analgesics, with or without continuous anxiolytics/ analgesics; paralytics only if PRN sedatives fail
Level 3	Goal: Asleep most of the time, arouses to pain, coughs with suctioning, breathes above ventilator, little spontaneous movement or head turning Action: PRN anxiolytics/analgesics with or without continuous anxiolytics/analgesics; paralytics only if PRN sedatives fail
Level 4	Goal: Asleep, arouses to pain, coughs with suctioning, returns to sleep immediately, does not consistently breathe above ventilator, little spontaneous movement, no head turning Action: continuous anxiolytics/analgesics, PRN anxiolytics/analgesics; paralytics only if PRN sedatives fail
Level 5	Goal: Asleep, minimal response to pain or suctioning, no respiratory effort, no sustained spontaneous movements Action: continuous anxiolytics/analgesics, PRN anxiolytics/analgesics; liberal use of paralytics if PRN sedatives fail
Level 6	Goal: Asleep, continuous paralysis, level of paralysis assessed by nerve stimulator or by observing minor motor movements between supplemental doses Action: continuous anxiolytics/analgesics, continuous paralytics; PRN anxiolytics/analgesics titrated to vital signs. Utilize train of four-nerve stimulator for serial assessments, or observe minor motor movements between supplemental doses.

PRN, as needed.

Table 2. Protocol for using the levels of sedation algorithm

1. After intubation, the level is established by the team to create an individualized patient behavior goal. The level is written as a physician order.
2. With appropriate medications prescribed, the nurse uses clinical assessment skills to administer pharmacological and age-appropriate psychological support to achieve the established goal.
3. The staff explains the desired level and the plans for implementation to the patient's family. Nonverbal communication is used for levels 1 through 3 based on the patient's cognition and age.
4. Respiratory therapists are responsible to help identify changes in patient status that deviate from the set goal.
5. An evaluation and adjustments are made daily and when dynamic changes occur.
6. Documentation:
   - Physician order on the plan of care.
   - Medications administered.
   - Patient data indicating therapeutic efficacy or required adjustments.
   - Assessment in the progress notes.

Figure 16: Pain Assessment Tool (PAT) (Spence, Gillies, Harrison, Johnston, & Nagy, 2005)

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

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

Figure 17: Neonatal Pain, Agitation and Sedation Scale (N-PASS) (Hummel, Puchalski, Creech, & Weiss, 2008; www.N-PASS.com)

Assessment Criteria	Sedation		Sedation/Pain	Pain / Agitation	
	-2	-1	0/0	1	2
Crying Irritability	No cry with painful stimuli	Moans or cries minimally with painful stimuli	No sedation/ No pain signs	Irritable or crying at intervals Consolable	High-pitched or silent-continuous cry Inconsolable
Behavior State	No arousal to any stimuli No spontaneous movement	Arouses minimally to stimuli Little spontaneous movement	No sedation/ No pain signs	Restless, squirming Awakens frequently	Arching, kicking Constantly awake or Arouses minimally / no movement (not sedated)
Facial Expression	Mouth is lax No expression	Minimal expression with stimuli	No sedation/ No pain signs	Any pain expression intermittent	Any pain expression continual
Extremities Tone	No grasp reflex Flaccid tone	Weak grasp reflex ¯ muscle tone	No sedation/ No pain signs	Intermittent clenched toes, fists or finger splay Body is not tense	Continual clenched toes, fists, or finger splay Body is tense
Vital Signs HR, RR, BP, SaO_2	No variability with stimuli Hypoventilation or apnea	< 10% variability from baseline with stimuli	No sedation/ No pain signs	↑ 10-20% from baseline SaO_2 76-85% with stimulation – quick recovery	↑ 20% from baseline SaO_2 £ 75% with stimulation – slow recovery Out of sync with vent

N-PASS Scoring Criteria

Crying / Irritability

-2 No response to painful stimuli

No cry with needle sticks

No reaction to ETT or nares suctioning

No response to care giving

-1 Moans, sighs, or cries (audible or silent) minimally to painful stimuli, e.g. needle sticks, ETT or nares suctioning, care giving

0 No sedation signs or No pain/agitation signs

+1 Infant is irritable/crying at intervals – but can be consoled

If intubated – intermittent silent cry

+2 Any of the following:

Cry is high-pitched

Infant cries inconsolably

If intubated – silent continuous cry

Behavior / State

-2 Does not arouse or react to any stimuli:

Eyes continually shut or open

No spontaneous movement

-1 Little spontaneous movement, arouses briefly and/or minimally to any stimuli:

Opens eyes briefly

Reacts to suctioning

Withdraws to pain

0 No sedation signs or No pain/agitation signs

+1 Any of the following:

Restless, squirming

Awakens frequently/easily with minimal or no stimuli

+2 Any of the following:

Kicking

Arching

Constantly awake

No movement or minimal arousal with stimulation (not sedated, inappropriate for gestational age or clinical situation)

Facial Expression

Facial expression of physical distress and pain in the infant

-2 Any of the following:

Mouth is lax

Drooling

No facial expression at rest or with stimuli

-1 Minimal facial expression with stimuli

0 No sedation signs or No pain/agitation signs

+1 Any pain face expression observed intermittently

+2 Any pain face expression is continual

Extremities / Tone

-2 Any of the following:

No palmar or planter grasp can be elicited

Flaccid tone

-1 Any of the following:

Weak palmar or planter grasp can be elicited

Decreased tone

0 No sedation signs or No pain/agitation signs

+1 Intermittent (<30 seconds duration) observation of toes and/or hands as clenched or fingers splayed

Body is not tense

+2 Any of the following:

Frequent (≥30 seconds duration) observation of toes and/or hands as clenched, or fingers splayed

Body is tense/stiff

Vital Signs: HR, BP, RR, & O2 Saturations

-2 Any of the following:

No variability in vital signs with stimuli

Hypoventilation

Apnea

Ventilated infant – no spontaneous respiratory effort

-1 Vital signs show little variability with stimuli – less than 10% from baseline

0 No sedation signs or No pain/agitation signs

+1 Any of the following:

HR, RR, and/or BP are 10-20% above baseline

With care/stimuli infant desaturates minimally to moderately (SaO2 76-85%) and recovers quickly (within 2 minutes)

+2 Any of the following:

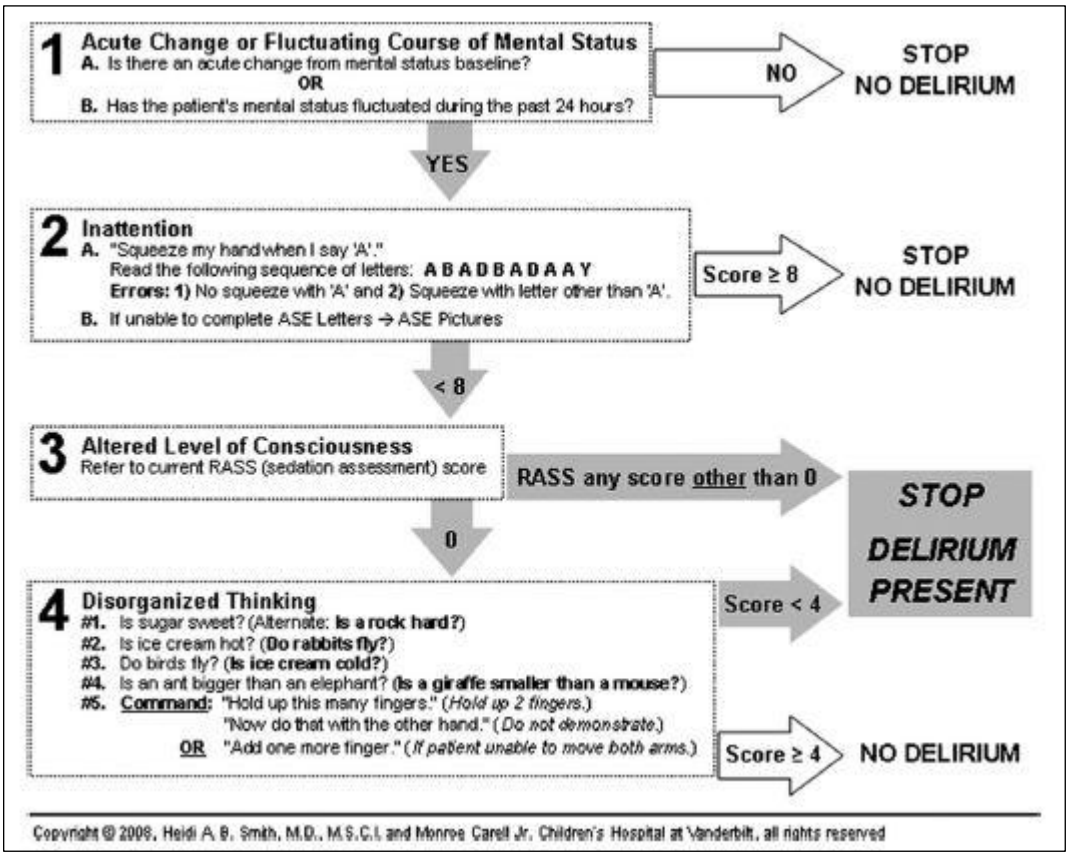
HR, RR, and/or BP are > 20% above baseline

With care/stimuli infant desaturates severely (SaO2 < 75%) and recovers slowly (> 2 minutes)

Out of sync/fighting ventilator

Figure 18: Pediatric Confusion Assessment Method for Intensive Care Unit (pCAM–ICU)

Note: pCAM-ICU uses Richmond Agitation Sedation Scale (RASS) in step 3 (Smith et al., 2011)



References

Ambuel, B., Hamlett, K. W., Marx, C. M., & Blumer, J. L. (1992). Assessing distress in pediatric intensive care environments: The COMFORT scale. *Journal of Pediatric Psychology, 17*(1), 95-109. doi: 10.1093/jpepsy/17.1.95

Curley, M. A., Harris, S. K., Fraser, K. A., Johnson, R. A., & Arnold, J. H. (2006). State behavioral scale: A sedation assessment instrument for infants and young children supported on mechanical ventilation. *Pediatric Critical Care Medicine, 7*(2), 107-114.

Hartwig, S., Roth, B., & Theisohn, M. (1991). Clinical experience with continuous intravenous sedation using midazolam and fentanyl in the paediatric intensive care unit. *European Journal of Pediatrics, 150*(11), 784-788. doi: 10.1007/BF02026712

Hughes, J., Gill, A., Leach, H. J., Nunn, A. J., Billingham, I., Ratcliffe, J., . . . Choonara, I. (1994). A prospective study of the adverse effects of midazolam on withdrawal in critically ill children. *Acta Paediatrica, International Journal of Paediatrics, 83*(11), 1194-1199. doi: 10.1111/j.1651-2227.1994.tb18280.x

Hummel, P., Puchalski, M., Creech, S. D., & Weiss, M. G. (2008). Clinical reliability and validity of the N-PASS: Neonatal pain, agitation and sedation scale with prolonged pain. *Journal of Perinatology, 28*(1), 55-60.

Johansson, M., & Kokinsky, E. (2009). The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. *Nursing in Critical Care, 14*(3), 122-130. Retrieved from http://login.ezproxy.library.ualberta.ca/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=19366409

Macnab, A. J., Levine, M., Glick, N., Susak, L., & Baker-Brown, G. (1991). A research tool for measurement of recovery from sedation: The vancouver sedative recovery scale. *Journal of Pediatric Surgery, 26*(11), 1263-1267. doi: 10.1016/0022-3468(91)90594-J

Merkel, S. I., Voepel-Lewis, T., Shayevitz, J. R., & Malviya, S. (1997). The FLACC: A behavioral scale for scoring postoperative pain in young children. *Pediatric Nursing, 23*(3), 293-297.

Parkinson, L., Hughes, J., Gill, A., Billingham, I., Ratcliffe, J., & Choonara, I. (1997). A randomized controlled trial of sedation in the critically ill. *Paediatric Anaesthesia, 7*(5), 405-410. doi: 10.1046/j.1460-9592.1997.d01-109.x

Ramelet, A. S., Rees, N., McDonald, S., Bulsara, M., & Abu-Saad, H. H. (2007). Development and preliminary psychometric testing of the multidimensional assessment of pain scale: MAPS. *Paediatric Anaesthesia, 17*(4), 333-340. doi: 10.1111/j.1460-9592.2006.02115.x

Ramelet, A. S., Rees, N. W., McDonald, S., Bulsara, M. K., & Huijer Abu-Saad, H. (2007). Clinical validation of the multidimensional assessment of pain scale. *Paediatric Anaesthesia, 17*(12), 1156-1165. doi: 10.1111/j.1460-9592.2007.02325.x

Ramsay, M. A., Savege, T. M., Simpson, B. R., & Goodwin, R. (1974). Controlled sedation with alphaxalone-alphadolone. *British Medical Journal, 2*(5920), 656-659. doi: 10.1136/bmj.2.5920.656

Razmus, I. S., Clarke, K. A., & Naufel, K. Z. (2003). Development of a sedation scale for the mechanically ventilated muscle relaxed pediatric critical care patient. *Pediatric Intensive Care Nursing, 4*(1), 7-11.

Smith, H. A. B., Boyd, J., Fuchs, D. C., Melvin, K., Berry, P., Shintani, A., . . . Ely, E. W. (2011). Diagnosing delirium in critically ill children: Validity and reliability of the pediatric confusion assessment method for the intensive care unit. *Critical Care Medicine, 39*(1), 150-157. doi: 10.1097/CCM.0b013e3181feb489

Spence, K., Gillies, D., Harrison, D., Johnston, L., & Nagy, S. (2005). A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. *JOGNN - Journal of Obstetric, Gynecologic, & Neonatal Nursing, 34*(1), 80-86. doi: 10.1177/0884217504272810

Suominen, P., Caffin, C., Linton, S., McKinley, D., Ragg, P., Davie, G., & Eyres, R. (2004). The cardiac analgesic assessment scale (CAAS): A pain

assessment tool for intubated and ventilated children after cardiac surgery. *Paediatric Anaesthesia, 14*(4), 336-343.