EJP
European Journal of Pain

SYSTEMATIC REVIEW

The clinimetric properties of the COMFORT scale: A systematic review

J. Maaskant[1,2], P. Raymakers-Janssen[3], E. Veldhoen[3], E. Ista[4], C. Lucas[2], H. Vermeulen[5,6]

1 Emma Children's Hospital, Academic Medical Center Amsterdam, The Netherlands
2 Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam, The Netherlands
3 Pediatric Intensive Care Unit, Wilhelmina Children's Hospital, University Medical Center Utrecht, The Netherlands
4 Intensive Care Unit, Department of Pediatrics and Pediatric Surgery, Erasmus MC – Sophia Children's Hospital, Rotterdam, The Netherlands
5 Department of Surgery, Academic Medical Center, Amsterdam, The Netherlands
6 Amsterdam School of Health Professions, Amsterdam, The Netherlands

Correspondence
Jolanda Maaskant
E-mail: j.m.maaskant@amc.nl

Funding sources
None declared.

Conflict of interest
None declared.

Accepted for publication
18 March 2016

doi:10.1002/ejp.880

Abstract

The COMFORT scale is a measurement tool to assess distress, sedation and pain in nonverbal paediatric patients. Several studies have described the COMFORT scale, but no formal assessment of the methodological quality has been undertaken. Therefore, we performed a systematic review to study the clinimetric properties of the (modified) COMFORT scale in children up to 18 years. We searched Central, CINAHL, Embase, Medline, PsycInfo and Web of Science until December 2014. The selection, data extraction and quality assessment were performed independently by two reviewers. Quality of the included studies was appraised using the COSMIN checklist. We found 30 studies that met the inclusion criteria. Most participants were ventilated children up to 4 years without neurological disorders. The results on internal consistency and interrater reliability showed values of >0.70 in most studies, indicating an adequate reliability. Construct validity resulted in correlations between 0.68 and 0.84 for distress, between 0.42 and 0.94 for sedation and between 0.31 and 0.96 for pain. The responsiveness of the (modified) COMFORT scale seems to be adequate. The quality of the included studies ranged from poor to excellent. The COMFORT scale shows overall an adequate reliability in providing information on distress, sedation and pain. Construct validity varies from good to excellent for distress, from moderate to excellent for sedation, and from poor to excellent for pain. The included studies were clinically and methodologically heterogeneous, hampering firm conclusions.

What does this review add?:

- An in-depth assessment of the clinimetric properties of the COMFORT scale.
- The COMFORT scale shows overall an adequate reliability in providing information on distress, sedation and pain. Construct validity varies from good to excellent for distress, from moderate to excellent for sedation, and from poor to excellent for pain.

1. Introduction

Hospitalized children are often confronted with invasive procedures in a hostile environment that cause distress and pain. Distress is defined as 'an organism's response to aversive internal and external stimuli and may include discomfort, anxiety, fear and pain' (Ambuel et al., 1992). Pain is described as 'an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage' (IASP, 2014). Distress and pain are important clinical problems because they result in elevated metabolism, possible deterioration of the immune system, impaired brain development and may even affect morbidity (Stoddard et al., 2006; Brummelte et al., 2012). Also the number of accidental removals of medical devices and complications during invasive procedures increase when distress or pain are present (Taddio et al., 1997; Hermann et al., 2006; Stoddard et al., 2006). Sedatives and analgesics are therefore needed to treat distress and pain in these patients. Sedation seeks to reduce distress, whereas analgesia aims to diminish pain. However, maintaining adequate sedation or pain relief in children is difficult, because of the wide range of ages and stages of development, and the different requirements over the course of the illness. Therefore, the accurate measurement of distress and pain is essential to establish its presence, and to monitor the effectiveness of interventions for relief and prevention. However, despite the many available scales and their potential to support daily practice, the assessment of distress and pain in children is still considered suboptimal (Franck and Bruce, 2009; Johnston et al., 2011; Vet et al., 2013).

Self-report is considered the reference standard for the measurement of distress and pain in adults and verbal children. Preverbal and sedated critically ill children, however, are unable to self-report, which requires observational parameters assessed by proxies or healthcare professionals. In recent decades, a number of observational tools to measure distress and pain have been developed (Von Baeyer and Spagrud, 2007; Dorfman et al., 2014). A well-known multidimensional tool is the COMFORT scale, that was originally developed as a continuous measure of distress in ventilated paediatric patients aged from birth to 18 years (Ambuel et al., 1992). The original COMFORT scale consists of six behavioural and two physiological measures. Later studies showed that the six behavioural items explained most of the variance in scores (Van Dijk et al., 2001; Carnevale and Razack, 2002; Ista et al., 2005; Nolent et al., 2006). For this reason the COMFORT scale was revised, resulting in the COMFORT-Behaviour (COMFORT-B) scale, including only these six items. Literature also describes other modified COMFORT scales like the COMFORT-neo, COMFORT-without muscle tone and the COMFORT-without blood pressure (Wielenga et al., 2004; Lee and Young, 2005; Caljouw et al., 2007).

Since the original development, a number of studies have described the clinimetric properties of the (modified) COMFORT scale measuring distress, sedation or pain in children of different ages, with different health conditions in different clinical contexts (Carnevale and Razack, 2002; Caljouw et al., 2007; De Jong et al., 2010; Valkenburg et al., 2011; Tristão et al., 2013), but formal assessment of the methodological quality of the (modified) COMFORT scale has not been performed. An in-depth appraisal on this topic is needed to support healthcare professionals to decide when the COMFORT scale can be used to obtain reliable and valid information on distress, sedation and pain, ultimately to improve treatment and outcome. Therefore, we performed a systematic review to study the clinimetric properties of the (modified) COMFORT scale as a tool to measure distress, sedation and pain in paediatric patients.

2. Methods

We used the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) recommendations for the reporting of the study (PRISMA, 2014).

2.1 Identification of studies

A literature search was performed in Central, CINAHL, Embase, Medline, PsycInfo and Web of Science to identify relevant studies published until 1 December 2014. We used the search terms 'COMFORT scale' OR 'COMFORT score'. No limitations were imposed on language or publication date.

Two independent reviewers (PRJ, EV or JM) screened the search results on titles and abstracts to assess which studies satisfied the inclusion criteria. This was followed by full-text review of potentially eligible studies. The reference lists of the potentially eligible studies were manually searched to identify additional articles. Studies were included if they met the following criteria:

1) study aim was to evaluate one or more clinimetric properties of the original COMFORT scale or any of the modified versions, as a tool to measure distress, sedation or pain;

2) the study population consisted of children from birth until 18 years, including premature neonates;
3) the studies were published as original articles.

We decided to use the term '(modified) COMFORT' to describe collectively all versions of the scale we found in the literature. Publications were excluded if only abstracts were available, or if it concerned reviews, guidelines, descriptive studies, editorials or poster publications. Disagreements were discussed and solved with help of a third reviewer (JM, HV or EI) when necessary.

2.2 Data extraction

A structured form was used to extract data from the original studies on participants (setting, number of participants and observations, age, diagnosis and intervention), type of outcome (distress, sedation or pain), assessment procedures and results on the clinimetric properties (statistics, outcomes, subgroups analysis). Data on the clinimetric properties included reliability, validity and responsiveness. Reliability refers to the extent to which the instrument produces consistent and reproducible results. Validity is the extent to which an instrument measures what it intends to measure. Responsiveness expresses the ability of an instrument to detect change over time. The data extraction included the following sub-categories: internal consistency, interrater reliability, intrarater reliability, measurement error, content, construct, criterion validity and responsiveness. Descriptions are presented in Table 1. Translations and transcultural validation were not assessed in this systematic review. Two reviewers (PRJ, EV or JM) extracted the data independently. Consensus was reached after discussion and consulting a third reviewer (JM or HV) when necessary.

2.3 Quality assessment

We used the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist to evaluate the methodological quality of the studies (Mokkink et al., 2010a,b,c, 2013). This checklist contains separate boxes each dealing with one clinimetric property. Every box contains several items, assessing the design and statistical methods. The quality score is expressed on a four-point scale: excellent, good, fair or poor. A detailed description of the COSMIN checklist can be found on the website www.cosmin.nl.

Before the quality assessment a pilot was performed to obtain consistency. Therefore, four studies were assessed by four reviewers (JM, PRJ, EV and HV) independently, using the COSMIN checklist. The reviewers compared and discussed the COSMIN scores until consensus was reached on definitions and methods. Subsequently, two reviewers (PRJ, EV or JM) independently scored the remaining publications selected for full text assessment. The results of the quality assessment of all studies were compared and discussed between the reviewers before reaching final conclusions. The reviewers were not blinded for authors, research environment and journals.

2.4 Outcome measurements

Cronbach's alpha between 0.70 and 0.95, and Intra Class Correlation or (weighted) kappa of at least 0.70 were considered adequate (Terwee et al., 2007). We considered correlations lower than 0.40 poor,

Table 1 Description of clinimetric properties (Mokkink et al., 2013).

Reliability	
Internal consistency	The extent to which the different items of a (sub)scale are correlated, thus are measuring the same construct
Reliability	The extent to which the measurement tool produces consistent and reproducible results
Measurement error	Systematic and random error in the scores, that is not attributed to the true changes in the construct
Validity	
Content validity (including face validity)	The extent to which the domain of interest is comprehensively reflected by the items of the measurement tool
Construct validity: structural validity	The extent to which the scores of the measurement tool are an adequate reflection of the dimensionality of the construct to be measured
Construct validity: hypothesis testing	Comparing the scores of the measurement tool to scores of another measurement tool that is considered to measure the same construct (convergent validity) or a different construct (divergent validity)
Criterion validity	The extent to which the scores of the measurement tool relate with a reference standard ('gold standard')
Other	
Responsiveness	The ability of a measurement tool to detect change over time in the construct to be measured

between 0.40 and 0.60 moderate, between 0.61 and 0.80 good and between 0.81 and 1.00 excellent (Sackett et al., 1991). Explained variance outcomes were recalculated by taken square roots, resulting in correlations. We considered a p-value of <0.05 adequate. The results of the Rasch analysis were expressed in mean square values, which are considered adequate when ranging from 0.6 to 1.4 (Wright and Linacre, 1994).

3. Results

3.1 Identification of studies

The results of the literature search and selection procedure are summarized on the flow diagram in Fig. 1. The literature search identified 747 studies. Review of title and abstract resulted in the exclusion of 39 duplicates and 659 ineligible studies. We read the full texts of the remaining 49 studies and excluded 19, because the aim of the study was not the COMFORT scale (n = 8), the articles were editorials (n = 4), poster presentations (n = 4) or implementation studies (n = 2). We could not obtain the full text of one additional study, and the publisher did not respond after multiple requests. After exclusion, 30 studies were eligible for further review (Ambuel et al., 1992; Marx et al., 1994; Blauer and Gerstmann, 1998; Brunow de Carvalho et al., 1999; Van Dijk et al., 2000; Carnevale and Razack, 2002; Crain et al., 2002; Courtman et al., 2003; Wielenga et al., 2004; Ista et al., 2005; Lee and Young, 2005; Triltsch et al., 2005; Twite et al., 2005; Bear and Ward-Smith, 2006; Nolent et al., 2006; Caljouw et al., 2007; Gjerstad et al., 2008; Lamas et al., 2008; Johansson and Kokinsky, 2009; Van Dijk et al., 2009; De Jong et al., 2010; Franck et al., 2011; Valkenburg et al., 2011; Bai et al., 2012; De Jong et al., 2012; Cury et al., 2013; Da Costa Silva et al., 2013; Tristão et al., 2013; Boerlage et al., 2014; Tschiedel et al., 2015).

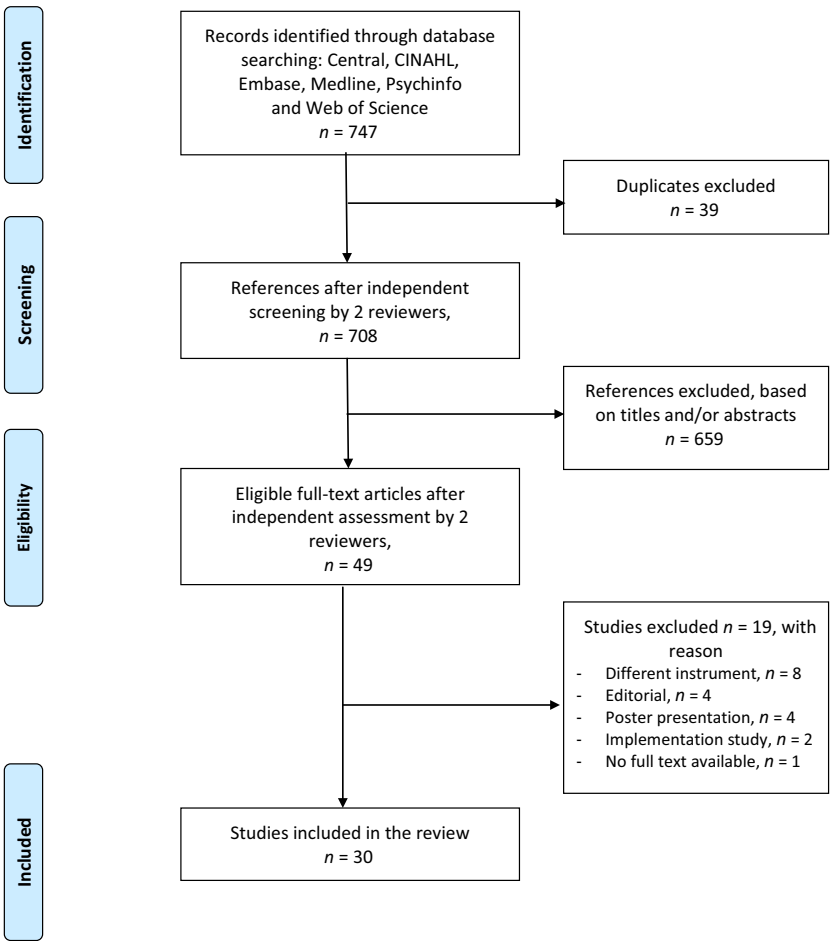

Figure 1 Flow diagram.

3.2 Description of included studies

Of the 30 included studies 20 took place in highly specialized paediatric intensive care units (PICUs) (Ambuel et al., 1992; Marx et al., 1994; Brunow de Carvalho et al., 1999; Van Dijk et al., 2000; Carnevale and Razack, 2002; Crain et al., 2002; Courtman et al., 2003; Ista et al., 2005; Triltsch et al., 2005; Twite et al., 2005; Bear and Ward-Smith, 2006; Nolent et al., 2006; Gjerstad et al., 2008; Lamas et al., 2008; Johansson and Kokinsky, 2009; Valkenburg et al., 2011; Bai et al., 2012; Da Costa Silva et al., 2013; Cury et al., 2013; Boerlage et al., 2014) and six in Neonatal Intensive Care Units (NICUs) (Blauer and Gerstmann, 1998; Wielenga et al., 2004; Lee and Young, 2005; Caljouw et al., 2007; Van Dijk et al., 2009; Franck et al., 2011). The remaining studies were performed in an operation room (Tschiedel et al., 2015), a specialized Burn Hospital (De Jong et al., 2010, 2012) and on a maternity ward (Tristão et al., 2013).

Sixteen studies investigated the original COMFORT scale (Ambuel et al., 1992; Marx et al., 1994; Blauer and Gerstmann, 1998; Brunow de Carvalho et al., 1999; Carnevale and Razack, 2002; Crain et al., 2002; Courtman et al., 2003; Triltsch et al., 2005; Twite et al., 2005; Bear and Ward-Smith, 2006; Gjerstad et al., 2008; Lamas et al., 2008; Franck et al., 2011; Cury et al., 2013; Tristão et al., 2013; Tschiedel et al., 2015) and the COMFORT-B was investigated in eight studies (Ista et al., 2005; Johansson and Kokinsky, 2009; De Jong et al., 2010; Valkenburg et al., 2011; Bai et al., 2012; De Jong et al., 2012; Da Costa Silva et al., 2013; Boerlage et al., 2014). Two studies investigated both the original COMFORT scale and the COMFORT-B (Van Dijk et al., 2000; Nolent et al., 2006). Other modified versions were the COMFORT-without muscle tone (Lee and Young, 2005), COMFORT-neo (Wielenga et al., 2004; Van Dijk et al., 2009) and the COMFORT-without blood pressure (Caljouw et al., 2007).

Distress was studied in four studies (Ambuel et al., 1992; Carnevale and Razack, 2002; Wielenga et al., 2004; Gjerstad et al., 2008) and 11 studies report on sedation (Marx et al., 1994; Brunow de Carvalho et al., 1999; Crain et al., 2002; Courtman et al., 2003; Ista et al., 2005; Lee and Young, 2005; Triltsch et al., 2005; Twite et al., 2005; Lamas et al., 2008; Da Costa Silva et al., 2013; Tschiedel et al., 2015).

All studies on distress and sedation included ventilated patients, with exception of one study that assessed the scale capacity to measure distress in patients receiving short-term sedation during diagnostic open muscle biopsy (Tschiedel et al., 2015). Pain was studied in 11 studies (Blauer and Gerstmann, 1998; Van Dijk et al., 2000; Bear and Ward-Smith, 2006; Caljouw et al., 2007; De Jong et al., 2010; Franck et al., 2011; Valkenburg et al., 2011; Bai et al., 2012; De Jong et al., 2012; Cury et al., 2013; Tristão et al., 2013). Eight studies on pain included ventilated patients, who were therefore also sedated (Van Dijk et al., 2000; Bear and Ward-Smith, 2006; Johansson and Kokinsky, 2009; Franck et al., 2011; Valkenburg et al., 2011; Bai et al., 2012; Cury et al., 2013; Boerlage et al., 2014). Studies on pain included study populations with postoperative pain (Van Dijk et al., 2000; Bear and Ward-Smith, 2006; Johansson and Kokinsky, 2009; Bai et al., 2012) or procedural pain (Blauer and Gerstmann, 1998; Caljouw et al., 2007; De Jong et al., 2010; Franck et al., 2011; Valkenburg et al., 2011;

	1	2	3	4	5
Alertness	Deeply asleep	Lightly asleep	Drownsy	Full awake and alert	Hyperalert
Calmness/agitation	Calm	Slightly anxious	Anxious	Very anxious	Panicky
Respiratory response (ventilated children)	No coughing and no spontaneous respiration	Spontanious respiration with little or no response to ventilation	Occasional cough or resistance to ventilator	Breathes against ventilator or coughs regulary	Fights ventilator, cough or choking
Physical movement	No movement	Occasional, slight movements	Frequent, slight movements	Vigorous movement limited to extremities	Vigorous movements including torso and head
Muscle tone	Muscles totally relaxed, no muscle tone	Reduced muscle tone	Normal muscle tone	Increased muscle tone and flexion of fingers and toes	Extreme muscle rigidity and flexion of fingers and toes
Facial tension	Facial muscle totally relaxed	Facial muscle tone normal; no facial muscle tension evident	Tension evident in some facial muscles	Tension evident throughout facial muscles	Facial muscles contorted and grimacing
Blood pressure	Blood pressure below baseline	Blood pressure consistently at baseline	Infrequent elevations ≥15% above baseline	Frequent elevations ≥15% above baseline	Sustained elevations ≥15% above baseline
Heart rate	Heart rate below baseline	Heart rate consistently at baseline	Infrequent elevations ≥15% above baseline	Frequent elevations ≥15% above baseline	Sustained elevations ≥15% above baseline

Figure 2 The COMFORT scale.

De Jong et al., 2012; Cury et al., 2013; Tristão et al., 2013). Three studies report results on background or prolonged pain (Van Dijk et al., 2009; De Jong et al., 2010, 2012). Four studies report on both distress, sedation and/or pain (Nolent et al., 2006; Johansson and Kokinsky, 2009; Van Dijk et al., 2009; Boerlage et al., 2014).

In all studies the mean/median age of the participants was under 4 years. The study population in seven studies consisted of (premature) neonates (Blauer and Gerstmann, 1998; Wielenga et al., 2004; Lee and Young, 2005; Caljouw et al., 2007; Van Dijk et al., 2009; Franck et al., 2011; Tristão et al., 2013). Whereas patients with neurological impairments were excluded from most studies, one study investigated the clinimetric properties of the COMFORT scale to measure pain in children with Down syndrome (Valkenburg et al., 2011). The characteristics of the included studies are summarized in Table 2.

3.3 Reliability

In total 18 studies report on the reliability of the (modified) COMFORT scale (Ambuel et al., 1992; Blauer and Gerstmann, 1998; Van Dijk et al., 2000; Carnevale and Razack, 2002; Crain et al., 2002; Wielenga et al., 2004; Ista et al., 2005; Lee and Young, 2005; Bear and Ward-Smith, 2006; Nolent et al., 2006; Caljouw et al., 2007; Johansson and Kokinsky, 2009; Van Dijk et al., 2009; De Jong et al., 2010; Franck et al., 2011; Valkenburg et al., 2011; Da Costa Silva et al., 2013; Boerlage et al., 2014). Results on reliability are summarized in Tables 3 (distress), 4 (sedation) and 5 (pain). The studies that report on more than one of these concepts are presented in both tables.

3.3.1 Internal consistency

The internal consistency of the (modified) COMFORT scale measuring distress is presented in three studies (Ambuel et al., 1992; Carnevale and Razack, 2002; Van Dijk et al., 2009). These studies report a Cronbach's alpha/correlation between 0.84 and 0.99, representing an adequate internal consistency. The internal consistency was studied in different study populations: ventilated patients on a PICU and (non-)ventilated neonates on a NICU. The internal consistency of the (modified) COMFORT scale measuring sedation was studied in three studies (Ista et al., 2005; Lee and Young, 2005; Nolent et al., 2006). These studies report a Cronbach's alpha between 0.76 and 0.80. The internal consistency is adequate for different study populations: (non-)ventilated patients on a PICU and ventilated neonates on a NICU. The internal consistency of the (modified) COMFORT scale measuring pain was studied in nine studies (Blauer and Gerstmann, 1998; Van Dijk et al., 2000; Bear and Ward-Smith, 2006; Nolent et al., 2006; Caljouw et al., 2007; Van Dijk et al., 2009; De Jong et al., 2010; Franck et al., 2011; Valkenburg et al., 2011). Eight studies report a Cronbach's alpha between 0.76 and 0.92. Again, these results represent an adequate internal consistency for different populations: (non-)ventilated patients on a PICU, (non-)ventilated neonates on a NICU and patients admitted to a specialized Burn Hospital. One study performed on a NICU found a less adequate result: correlation 0.52 (Blauer and Gerstmann, 1998).

3.3.2 Reliability

Interrater reliability of the (modified) COMFORT scale measuring distress was reported in four studies (Ambuel et al., 1992; Wielenga et al., 2004; Van Dijk et al., 2009; Boerlage et al., 2014). All studies report adequate interrater reliability: ICC/weighted kappa/correlation between 0.79 and 0.96 for ventilated patients on a PICU and (non-)ventilated neonates on a NICU. Interrater reliability of the (modified) COMFORT scale measuring sedation was reported in five studies (Crain et al., 2002; Ista et al., 2005; Lee and Young, 2005; Johansson and Kokinsky, 2009; Da Costa Silva et al., 2013). Three studies report adequate reliability in ventilated patients on a PICU: ICC/weighted kappa/correlation 0.71–1.00. Two studies report less adequate reliability: 0.56 in ventilated patient on a PICU (Da Costa Silva et al., 2013) and 0.63 in ventilated patient on a NICU (Lee and Young, 2005). Interrater reliability of the (modified) COMFORT scale measuring pain was reported in nine studies (Van Dijk et al., 2000; Bear and Ward-Smith, 2006; Caljouw et al., 2007; Johansson and Kokinsky, 2009; Van Dijk et al., 2009; De Jong et al., 2010; Franck et al., 2011; Valkenburg et al., 2011; Boerlage et al., 2014). The results of eight studies present adequate reliability: ICC/weighted kappa/correlation between 0.70 and 0.96. These results represent different populations: (non-)ventilated patients on a PICU, (non-)ventilated neonates on a NICU and patients admitted to a specialized Burn Hospital. Only Caljouw et al. (2007) report lower reliability in non-ventilated neonates on a NICU: weighted kappa 0.62. None of the studies examined intrarater reliability.

Table 2 Summary of the included studies.

Author, year	No. of patients	No. of observations	Age	Reason for admission	Type of COMFORT scale; construct measured	Context; interventions	Exclusion criteria	Raters
Distress								
Ambuel et al. (1992)	37	50	Mean 37.1 months (SD ± 52.7)	Cardiac and respiratory diseases	COMFORT; distress	PICU; intermittent mandatory ventilation or continuous positive airway pressure	Compromised neurologic status, profound mental retardation, multiple trauma <72 h, altered muscle tone or contractures; temporarily excluded when experiencing severe acute pain	Investigator, research assistant, nurses
Carnevale and Razack (2002)	18	514	Range <1 month–7 years	Respiratory dysfunction and cardiac surgery	COMFORT; distress	PICU; mechanically ventilation, intubation	Not specified	Nurses
Gjerstad et al. (2008)	20	20	1 day–11 years	Thoracic or abdominal surgery, pulmonary insufficiency	COMFORT; distress	PICU, mechanically ventilation	Neuromuscular blockers, anticholinergics, sympatholytics	Investigators
Wielenga et al. (2004)	19	Not specified	Median gestational age 30 weeks (range 26–36)	Idiopathic respiratory distress syndrome, infection, respiratory insufficiency	COMFORT-neo; distress	NICU; ventilation	Congenital and neurological abnormalities	Clinical expert and nurses
Sedation								
Brunow de Carvalho et al. (1999)	18	30	Mean 16.5 months (SD ± 7.3)	Respiratory and infectious diseases	COMFORT; sedation	PICU; intermittent mandatory ventilation or continuous airway pressure	Mental dysfunction	Specialists
Courtman et al. (2003)	43	373	Mean 3.9 years (SD ± 4.5)	Respiratory, head injury, neurology, sepsis, oncology, metabolic and cardiac	COMFORT; sedation	PICU, mechanically ventilation and a sedative agent	Receipt of ketamine	Nurses

Table 2 (*Continued*)

Author, year	No. of patients	No. of observations	Age	Reason for admission	Type of COMFORT scale; construct mea sured	Context; interventions	Exclusion criteria	Raters
Crain et al. (2002)	31	144	Median 25 months	Pneumonia, bronchioloitis, congenital heart disease, airway compromise	COMFORT; sedation	PICU; endotracheal intubation	History of seizures, traumatic brain injury, encephalopathy, history of neurological impairment	Investigators
Da Costa Silva et al. (2013)	11	35	1 month–16 years	Various medical problems	COMFORT-B; sedation	PICU, mechanically ventilation, sedation	Hypoxic-ischemic injury, receipt of ketamine, conditions affecting the forehead	Doctors and nurses
Ista et al. (2005)	78	843	Median 17 months (range 0–223)	Respiratory failure, cardiac congenital, cardiac other, sepsis/septic shock	COMFORT-B; sedation	PICU; 85% ventilated patients and 83% received midazolam	Severe mental retardation, hypotonia. neuromuscular blockers	Nurses
Lamas et al. (2008)	77	Not specified	Median 8 months (15 days to 228 months)	Cardiac and other surgeries Subgroups: 40% paralyzed patients 60% non-paralyzed patients	COMFORT; sedation	PICU; mechanical ventilation, continuous sedation	Neurologic diseases, no need for mechanical ventilation or interruption within 24 h after admission	Nurses
Lee and Young (2005)	15	27	Range 23–54 weeks gestational age	Respiratory distress syndrome, pneumonia, @meconium aspiration syndrome	COMFORT-without muscle tone; sedation	NICU; mechanical ventilation, intubation	Neurological impairment, unstable cardiovascular status, neuromuscular blockers, neuromuscular disorders	Nurse, senior nurse and doctor

Table 2 (*Continued*)

Author, year	No. of patients	No. of observations	Age	Reason for admission	Type of COMFORT scale; construct mea sured	Context; interventions	Exclusion criteria	Raters
Marx et al. (1994)	Study 1: 34	100	Mean 15.9 months (SD ± 22.7; range 0–82)	Cardiac surgery, respiratory diseases	COMFORT; sedation	PICU; mechanical ventilation, intubated endotracheal or tracheostomy	Head injury, ischaemic encephalopathy, stroke, profound mental retardation, multiple trauma <72 h, surgery <24 h, abnormality muscle function, neuromuscular blockade, chronic cough, ability of the patient to self report pain	Research assistants
	Study 2: 30	96	Mean 20.1 months (SD ± 30.9; range 0–102)	Cardiac surgery, respiratory diseases	COMFORT; sedation	PICU; sedated with opiates and/or benzodiazepines and/or barbiturates	Conform Marx study 1	Research assistants
	Study 3: 21	120	Mean 27.5 months (SD ± 30.9; range 0–84)	Cardiac surgery, respiratory diseases	COMFORT; sedation	PICU; sedated with opiates and/or benzodiazepines and/or barbiturates	Conform Marx study 1	Research assistants
Triltsch et al. (2005)	40	40	5.6 months (21 days–16 years)	Cardia, gastrointestinal, other	COMFORT; sedation	PICU, mechanically ventilation	Brain trauma, use of muscle relaxation or persistent postoperative relaxation, intractable agitation	Investigators

Table 2 (*Continued*)

Author, year	No. of patients	No. of observations	Age	Reason for admission	Type of COMFORT scale; construct mea sured	Context; interventions	Exclusion criteria	Raters
Tschiedel et al. (2015)	30	204	9–18 months	Muscle biopsy	COMFORT; sedation	Diagnostic procedure; intervention and sedation time <30 min	<6 months of age, structural brain damage, intolerance to remifentanil or propofol, mitochondrial disease or upper airway abnormalities	Investigator
Twite et al. (2005)	75	869	Median 10 months (range 1 month–12 years)	Respiratory diagnoses (50%), cardiac problems (24%), elective surgical procedures, various medical problems	COMFORT; sedation	PICU, intubated and mechanically ventilated children	History of seizures, encephalopathy, neuromuscular blocking drugs, receipt of ketamine, conditions affecting the forehead	Investigators
Pain								
Bai et al. (2012)	170	2815	Median 8 months (range 0.5–72)	Congenital heart diseases	COMFORT-B; postoperative pain	PICU; mechanical ventilation, cardiac surgery	Intellectual disability, ECMO support	Investigator
Bear and Ward-Smith (2006)	55	Not specified	Median 27 months (range 1 month–18 years)	(Post)operative patients	COMFORT; pain (82% postoperative)	PICU; mechanical ventilation	Extremely uncontrolled pain, frequent intubation, ability to self-report	Investigator, nurses
Blauer and Gerstmann (1998)	33	1428	Gestational age range 24–40 weeks	Not specified	COMFORT; procedural pain	NICU; endo-tracheal intubation, intravenous catheter insertion, endotracheal suctioning, diaper change	Recent surgery	Investigator
Caljouw et al. (2007)	57	Not specified	Mean 3.3 days (SD ± 1.8)	Prematurity	COMFORT-without blood pressure; procedural pain	NICU; non ventilated children, capillary heel puncture	Not specified	Research nurses

Table 2 (*Continued*)

Author, year	No. of patients	No. of observations	Age	Reason for admission	Type of COMFORT scale; construct measured	Context; interventions	Exclusion criteria	Raters
Cury et al. (2013)	16	Not specified	Median 60 days (range 4–410)	Cardiac surgery	COMFORT; procedural pain	PICU; intubation, sedation with midazolam and fentanyl Painful procedure: suctioning of endotracheal tube	Neuromuscular blockers, immediate postoperative extubation, diagnosis of genetic syndrome, early postoperative death, severe hemodynamic instability, no postoperative sedation, unavailable researcher	Investigator
De Jong et al. (2010)	154	Not specified	Mean 20 months (SD ± 11; range 1–56)	Wound care (burns, scalds)	COMFORT-B; procedural pain, background pain	Specialized Burn Hospitals; wound care	Developmental delay	Nurses
De Jong et al. (2012)	154	3884	Mean 20 months (SD ± 11; range 1–56)	Wound care (burns, scalds)	COMFORT-B; procedural pain, background pain	Specialized Burn Hospitals; wound care	Not specified	Nurses

Table 2 (*Continued*)

Author, year	No. of patients	No. of observations	Age	Reason for admission	Type of COMFORT scale; construct measured	Context; interventions	Exclusion criteria	Raters
Franck et al. (2011)	207 (81)	Not specified	Mean 12.1 days (SD ± 8.6)	Congenital heart diseases	COMFORT; procedural pain	NICU; cardiac surgery, sedation with morphine	APGAR ≤6, neurological impairment, major genetic anomaly, administration of anticholinergic or beta adrenergic antagonist drugs within 24 h of surgery, previous surgery, resuscitation in past, pre- or postnatal steroid use, postoperative cardiac pacing	Research nurses
Tristão et al. (2013)	36	36	Mean gestational age 38.9 (SD ± 1.4) weeks	Normal clinical labour procedure	COMFORT; procedural pain	Maternity unit	Postnatal age<24 h, Apgar <7, intracranial haemorrhage, metabolic, respiratory, circulatory, congenital disorders, anagesics, opioid use during pregnancy	Examinators

Table 2 (*Continued*)

Author, year	No. of patients	No. of observations	Age	Reason for admission	Type of COMFORT scale; construct mea sured	Context; interventions	Exclusion criteria	Raters
Valkenburg et al. (2011)	76 Down syndrome versus 466 control	1163–6276	Down syndrome: median 81 days (range 42–273) Controls: median 119 days (range 22–355)	Cardiac, gastro-intestinal, ear-nose-throat, craniofacial and other surgery	COMFORT-B; procedural pain	PICU; Down syndrome: mechanical ventilation 74% sedation with morphine 62% sedation with midazolam 68% Control: mechanical ventilation 47% sedation with morphine 45% sedation with midazolam 51%	>36 months, <2 assessments available	Nurses
Van Dijk et al. (2000)	158	Not specified	0–4 weeks: 35%, 1–6 months: 30%, 7–12 months: 15%, 1–3 years: 20%	Congenital anomalies and acquired diseases	COMFORT and COMFORT-B; postoperative pain	PICU; abdominal or thoracic surgery, mechanical ventilation 39%	Severe neurological problems, usage of medication that influences behavioural assessment	Nurses
Distress, sedation and/or pain								
Boerlage et al. (2014)	180	747	0.4 years (IQR 0.1–2.0)	Both surgical and non-surgical treatment	COMFORT-B; distress, pain	PICU; 95% mechanically ventilation or ECMO treatment	None	Nurses
Johansson and Kokinsky (2009)	40	119	Median 4 months (range 0–108)	Cardiac surgery, gastrointestinal malformations, congenital surgery	COMFORT-B; postoperative pain, sedation	PICU; intubation, mechanical ventilation	Severe physical and mental handicap, neuromuscular blockers	Research nurses and nurses
Nolent et al. (2006)	20	55	Median 4.7 weeks (range 1–15)	Medical, surgery	COMFORT and COMFORT-B; sedation, pain	PICU; intubation and ventilation 55%, non-invasieventilation 5%, 40 non-ventilation	Neuromuscular disease, neuromuscular blockers	Nurses
Van Dijk et al. (2009)	174	Not specified	Gestational age range 24–43 weeks	Respiratory insufficiency, prematurity, congenital cardiac defects, perinatal asphyxia	COMFORT-neo; distress, prolonged pain	NICU; continuously ventilated 24%, partly ventilated 44%, non-ventilated 32%	Neuromuscular blockers	Nurses

PICU, paediatric intensive care unit; NICU, neonatal intensive care unit; SD, standard deviation; IQR, interquartile range.

Table 3 The reliability of the included studies on distress.

Author, year	Type of COMFORT scale	Internal consistency	Interrater reliability
Ambuel et al. (1992)	COMFORT	Cronbach's alpha 0.90	Correlation coefficient 0.84 Per item • Alertness 0.73 • Calmness 0.69 • Respiratory response 0.70 • Physical movement 0.75 • Blood pressure 0.51 • Heart rate 0.66 • Muscle tone 0.52 • Facial tension 0.51
Boerlage et al. (2014)	COMFORT-B		Intra Class Correlation 0.96 Weighted kappa 0.72–0.86
Carnevale and Razack (2002)	COMFORT	Explained variance • Total 99% • Deleted HR & MAP 97%	
Van Dijk et al. (2009)	COMFORT-neo	Not ventilated patients: • Cronbach's alpha 0.88 • Corrected item total correlation muscle 0.52 and calmness 0.80 Ventilated patients: • Cronbach's alpha 0.84 • Corrected item total correlation respiratory 0.49 and calmness 0.72	Weighted kappa median 0.79 Per item 0.65–0.97
Wielenga et al. (2004)	COMFORT-neo		Weighted kappa median 0.84 Intra Class Correlation 0.94 Per item • Alertness 0.96 • Calmness 0.86 • Respiratory response 0.79 • Physical movement 0.71 • Arterial pressure 0.64 • Heart rate 1.00 • Muscle tone 0.56 • Facial tension 0.85

MAP, Mean arterial pressure; HR, heart rate.

3.4 Validity

In total 24 studies report on the validity of the (modified) COMFORT scale (Ambuel et al., 1992; Marx et al., 1994; Brunow de Carvalho et al., 1999; Van Dijk et al., 2000; Crain et al., 2002; Courtman et al., 2003; Wielenga et al., 2004; Ista et al., 2005; Triltsch et al., 2005; Twite et al., 2005; Nolent et al., 2006; Caljouw et al., 2007; Gjerstad et al., 2008; Lamas et al., 2008; Johansson and Kokinsky, 2009; Van Dijk et al., 2009; De Jong et al., 2010; Valkenburg et al., 2011; Bai et al., 2012; De Jong et al., 2012; Cury et al., 2013; Da Costa Silva et al., 2013; Tristão et al., 2013; Tschiedel et al., 2015). The results on validity are summarized in Tables 6 (distress), 7 (sedation) and 8 (pain). The studies that report on more than one of these concepts are presented in both tables.

3.4.1 Content validity

The only study that established content validity was the study by Ambuel et al. (1992), the developers of the COMFORT scale. In this study, eight dimensions were selected from the behavioural science and medical literature, and nurses' expert opinion: mean arterial blood pressure, heart rate, muscle tone, facial tension, alertness, calmness/agitation, respiratory behaviour and physical movement. The preliminary tool was tested and revised to clarify the verbal descriptions of each dimension.

Table 4 The reliability of the included studies on sedation.

Author, year	Type of COMFORT scale	Internal consistency	Interrater reliability
Crain et al. (2002)	COMFORT		Cronbach's alpha >0.90
Da Costa Silva et al. (2013)	COMFORT-B		Kappa 0.56–0.75
Ista et al. (2005)	COMFORT-B	Cronbach's alpha 0.78 Cronbach's alpha • Deleted HR 0.79 • Deleted MAP 0.80 • Deleted HR & MAP 0.84	Intra Class Correlation 0.99 Weighted kappa 0.77–1.00
Johansson and Kokinsky (2009)	COMFORT-B		Weighted kappa 0.71 Per item • Alertness 0.69 • Calmness 0.54 • Respiratory response 0.78 • Physical movement 0.69 • Muscle tone 0.43 • Facial tension 0.54
Lee and Young (2005)	COMFORT-without muscle tone	Cronbach's alpha 0.80 Per item • Alertness 0.71 • Calmness 0.46 • Respiratory 0.63 • Movement 0.70 • Blood pressure 0.38 • Heart rate 0.51 • Facial expression 0.52	Physician – nurse: Correlation coefficient total 0.63 Per item • Alertness 0.62 • Calmness 0.36 • Respiratory response 0.70 • Physical movement 0.52 • Blood pressure 0.71 • Heart rate 0.51 • Facial tension 0.37 Doctor – senior Nurse: Correlation coefficient total 0.76 Per item • Alertness 0.72 • Calmness 0.32 • Respiratory response 0.72 • Physical movement 0.79 • Blood pressure 0.77 • Heart rate 0.82 • Facial tension 0.55 Senior nurse – nurse: Correlation coefficient total 0.78 Per item • Alertness 0.70 • Calmness 0.29 • Respiratory response 0.76 • Physical movement 0.56 • Blood pressure 0.58 • Heart rate 0.68 • Facial tension 0.43
Nolent et al. (2006)	COMFORT and COMFORT-B	Cronbach's alpha 0.76 Cronbach's alpha • Deleted heart rate 0.77 • Deleted MAP 0.82 • Deleted HR & MAP 0.84	

MAP, Mean arterial pressure; HR, heart rate.

Table 5 The reliability of the included studies on pain.

Author, year	Type of COMFORT scale	Internal consistency	Interrater reliability
Bear and Ward-Smith (2006)	COMFORT	Cronbach's alpha 0.85	Correlation coefficient 0.79 Per item • Alertness 0.68 • Calmness 0.55 • Respiratory response 0.67 • Physical movement 0.79 • Blood pressure 0.67 • Heart rate 0.78 • Muscle tone 0.60 • facial tension 0.40
Blauer and Gerstmann (1998)	COMFORT	Correlation coefficient 0.52 Explained variance 27%	
Boerlage et al. (2014)	COMFORT-B		Intra Class Correlation 0.96 Weighted kappa 0.72–0.86
Caljouw et al. (2007)	COMFORT-without blood pressure	Cronbach's alpha • Pre-intervention 0.76 • Postintervention 0.86 Item total correlation • Pre-intervention 0.24–0.65 • Postintervention 0.44–0.81	Weighted kappa 0.62–0.84 Intra Class Correlation • Pre-intervention 0.85 • Postintervention 0.93
De Jong et al. (2010)	COMFORT-B	Cronbach's alpha •Background pain 0.77 •Procedural pain 0.80	Intra Class Correlation • Background pain 0.83 • Procedural pain 0.82
Franck et al. (2011)	COMFORT	Correlation coefficient 0.77 Explained variance 60% • Behavioural items 45% • Physiological items 15%	Intra Class Correlation 0.84–0.96
Johansson and Kokinsky (2009)	COMFORT-B		Weighted kappa 0.71 Per item • Alertness 0.69 • Calmness 0.54 • Respiratory response 0.78 • Physical movement 0.69 • Muscle tone 0.43 • Facial tension 0.54
Nolent et al. (2006)	COMFORT and COMFORT-B	Cronbach's alpha 0.76 Cronbach's alpha • Deleted heart rate 0.77 • Deleted MAP 0.82 • Deleted HR & MAP 0.84	
Valkenburg et al. (2011)	COMFORT-B	Down syndrome: • Cronbach's alpha per item 0.54–0.72 • Unstandardized 0.84 • Standardized 0.86 Control group: • Cronbach's alpha per item 0.57–0.76 • Unstandardized: 0.87 • Standardized 0.88	Median kappa 0.81

Table 5 (*Continued*)

Author, year	Type of COMFORT scale	Internal consistency	Interrater reliability
Van Dijk et al. (2000)	COMFORT and COMFORT-B	Cronbach's alpha 0.90–0.92	Weighted kappa 0.70 Per item • Alertness 0.74 • Calmness 0.69 • Respiratory response 0.54 • Crying 0.70 • Physical movement 0.70 • Blood pressure 0.93 • Heart rate 0.93 • Muscle tone 0.66 • Facial tension 0.63
Van Dijk et al. (2009)	COMFORT-neo	Not ventilated patients: • Cronbach's alpha 0.88 • Corrected item total correlation muscle 0.52 and calmness 0.80 Ventilated patients: • Cronbach's alpha 0.84 • Corrected item total correlation respiratory 0.49 and calmness 0.72	Weighted kappa median 0.79 Per item 0.65–0.97

MAP, mean arterial pressure; HR, heart rate.

Table 6 The construct validity of the included studies on distress.

Author, year	Comparison	Results
Ambuel et al. (1992)	COMFORT with VAS-obs	Correlation coefficient 0.75
Gjerstad et al. (2008)	COMFORT-B with skin conductance	Regression analysis before and during suctioning: $R^2 = 0.61$, correlation coefficient 0.78 during and after suctioning: $R^2 = 0.46$, correlation coefficient 0.68
Van Dijk et al. (2009)	$COMFORT_{neo}$ with NRS-obs	Correlation coefficient: • NRS-obs 0.83 • Mean NRS-obs 0.75 Cut-off points $COMFORT_{neo}$ 14: • Sensitivity 0.81 • Specificity 0.90
Wielenga et al. (2004)	COMFORT with expert opinion	Correlation coefficient 0.84 Cut-off point 20: • Sensitivity 1.00 • Specificity 0.77 • AUC 0.95

AUC, area under the curve; NRS-obs, numeric rating scale observers; VAS-obs, visual analogue scale observer.

3.4.2 Construct validity

The included 24 studies addressing validity all compare the (modified) COMFORT scale with another measurement tool which is considered to measure the same construct. As no reference standard exists to measure distress, sedation and pain in nonverbal and sedated children, we classified this as construct validity (hypothesis testing). A variety of tools were used as comparators: the Auditory Evoked Potentials (AEP), Bispectral Index Score (BIS), Cardiac Analgesic Assessment Scale (CAAS), expert opinion, Faces Legs Activity Cry Consolability (FLACC), Hartwig scale, Numeric Rating Scale-observer (NRS-obs),

Table 7 The construct validity of the included studies on sedation.

Author, year	Comparison	Results
Brunow de Carvalho et al. (1999)	COMFORT with Hartwig	Kappa 0.35
Courtman et al. (2003)	COMFORT with BIS	Overall: Correlation coefficient 0.50, $R^2 = 0.25$ Neurological diagnosis: Correlation coefficient 0.26, $R^2 = 0.06$
Crain et al. (2002)	COMFORT with BIS	Individual measurements: Correlation coefficient 0.51, $R^2 = 0.26$, Categorized measurements: Correlation coefficient 0.94, $R^2 = 0.89$,
Da Costa Silva et al. (2013)	COMFORT-B with BIS	Correlation coefficient (four raters) 0.42–0.52
Ista et al. (2005)	COMFORT with NISS	Differences COMFORT scores with the NISS categories Kruskal–Wallis $p < 0.01$
Johansson and Kokinsky (2009)	COMFORT with NISS	Correlation coefficient 0.57
Lamas et al. (2008)	COMFORT with BIS COMFORT with AEP COMFORT with Ramsay scale	Correlation coefficient • COMFORT with BIS 0.48 • COMFORT with AEP 0.53 • COMFORT with Ramsay scale 0.73
Marx et al. (1994)	COMFORT with expert	Explained variance $R^2 = 0.63$, 0.66 and 0.82 Correlation coefficient 0.79, 0.81 and 0.91
Nolent et al. (2006)	COMFORT with VAS-obs	Correlation coefficient 0.53
Triltsch et al. (2005)	COMFORT with BIS	Spearman correlation, explained variance All patients: 0.65, $R^2 = 0.42$ Patients <6 months: 0.78, $R^2 = 0.61$ Patients >6 months: 0.47, $R^2 = 0.22$
Tschiedel et al. (2015)	COMFORT with BIS	Spearman correlation: 0.59
Twite et al. (2005)	COMFORT with BIS	Spearman correlation Averaged over patients: 0.56 Averaged over time: 0.61

AEP, auditory evoked potentials; BIS, bispectral index score; NISS, nurse interpretation of sedation score; VAS-obs, visual analogue scale observer.

Nurse Interpretation of Sedation Score (NISS), Objective Pain Scale (OPS), Pain Observation Scale for Young Children (POCIS), Ramsey Score (RS), skin conductance and Visual Analogue Scale-observer (VAS-obs).

Construct validity of the (modified) COMFORT scale measuring distress was studied in four studies (Ambuel et al., 1992; Wielenga et al., 2004; Gjerstad et al., 2008; Van Dijk et al., 2009). Two studies show that the construct validity is good for ventilated patients on a PICU with correlations between 0.68 and 0.78 (Ambuel et al., 1992; Gjerstad et al., 2008). Studies performed with (non-)ventilated neonates on a NICU show excellent validity: correlations 0.83 and 0.84 (Wielenga et al., 2004; Van Dijk et al., 2009).

Construct validity of the (modified) COMFORT scale measuring sedation was studied in 12 studies (Marx et al., 1994; Brunow de Carvalho et al., 1999; Crain et al., 2002; Courtman et al., 2003; Ista et al., 2005; Triltsch et al., 2005; Twite et al., 2005; Nolent et al., 2006; Lamas et al., 2008; Johansson and Kokinsky, 2009; Da Costa Silva et al., 2013; Tschiedel et al., 2015). Most studies show that the construct validity is moderate: correlations between 0.42 and 0.59 (Crain et al., 2002; Courtman et al., 2003; Twite et al., 2005; Nolent et al., 2006; Lamas et al., 2008; Johansson and Kokinsky, 2009; Da Costa Silva et al., 2013; Tschiedel et al., 2015). Three studies report good validity with correlations between 0.65 and 0.79 (Marx et al., 1994; Triltsch et al., 2005; Lamas et al., 2008). Excellent validity is reported in one study: correlations of 0.81 and 0.91 (Marx et al., 1994). One study report a less adequate result: kappa 0.35 (Brunow de Carvalho et al., 1999). The results all apply to (non-)ventilated patients on a PICU.

Construct validity of the (modified) COMFORT scale measuring pain was studied in 11 studies (Van Dijk et al., 2000; Nolent et al., 2006; Caljouw et al., 2007; Johansson and Kokinsky, 2009; Van Dijk et al., 2009; De Jong et al., 2010; Valkenburg et al., 2011; Bai et al., 2012; De Jong et al., 2012; Cury et al., 2013; Tristão et al., 2013). Six studies report a construct validity for (non-)ventilated patients on a PICU (Van Dijk et al., 2000; Nolent et al., 2006; Johansson and Kokinsky, 2009; Valkenburg et al., 2011; Bai et al., 2012; Cury et al., 2013). Most stud-

Table 8 The construct validity of the included studies on pain.

Author, year	Comparison	Results
Bai et al. (2012)	COMFORT-B with VAS-obs COMFORT-B with FLACC	Correlation coefficient: • COMFORT with VAS-obs 0.31 • COMFORT with FLACC 0.51 COMFORT-B with VAS-obs Cut-off point 13: • Sensitivity 0.86 • Specificity 0.83 • AUC 0.93
Caljouw et al. (2007)	$COMFORT_{without\ blood\ pressure}$ with VAS-obs	Correlation coefficient • Pre-test 0.09–0.49 • Post-test 0.44–0.74 Cut-off point 17: • Sensitivity 0.93 • Specificity 0.80 • AUC 0.97
Cury et al. (2013)	COMFORT with CAAS	CAAS cut-off 4: kappa 0.37 CAAS cut-off 3: kappa 0.73
De Jong et al. (2010)	COMFORT-B with POCIS	Correlation coefficient • Background pain 0.45 • Procedural pain 0.88
De Jong et al. (2012)	COMFORT-B with POCIS	Item difficulty logit (infit mean square): • Alertness −4.53 (1.63) • Calmness 2.72 (0.74) • Crying 2.77 (0.98) • Physical movement −0.16 (0.96) • Muscle tone −1.62 (0.79) • Facial tension 0.82 (0.46)
Johansson and Kokinsky (2009)	COMFORT with VAS-obs COMFORT with FLACC	Correlation coefficient: • COMFORT with VAS-obs 0.49 • COMFORT with FLACC 0.50
Nolent et al. (2006)	COMFORT with OPS	Correlation coefficient 0.54
Tristão et al. (2013)	COMFORT-B with skin conductance	Correlation coefficient at 15, 30 and 180 s after procedure: • Overall: 0.42, 0.38, 0.50 • Alertness: 0.58, 0.47, 0.44 • Calmness: 0.45, 0.41, 0.33 • Crying: 0.33, 0.35. 0.44 • Movement: 0.56, 0.53, 0.61 • Muscle tone: 0.38, 0.40, 0.41 • Facial tension: 0.42, 0.39, 0.42
Valkenburg et al. (2011)	COMFORT-B with NRS-obs	Down syndrome Correlation coefficient 0.45 Cut-off point 17: • Sensitivity 0.82 • Specificity 0.92 Controls Correlation coefficient 0.57 Cut-off point 17: • Sensitivity 0.83 • Specificity 0.91
Van Dijk et al. (2000)	COMFORT-B with VAS-obs	Correlation coefficient 0.96, 0.89 and 0.90

Table 8 (*Continued*)

Author, year	Comparison	Results
Van Dijk et al. (2009)	$COMFORT_{neo}$ with NRS-obs	Correlation coefficient: • NRS-obs 0.54 • Mean NRS-obs 0.51 Cut-off points $COMFORT_{neo}$ 14: • Sensitivity 0.72 • Specificity 0.80

AUC, area under the curve; CAAS, cardiac analgesic assessment scale; FLACC, faces legs activity cry consolability; NRS-obs, numeric rating scale observer; OPS, objective pain scale; POCIS, pain observation scale for young children; VAS-obs, visual analogue scale observer.

ies report moderate validity: correlation/kappa between 0.45 and 0.54 (Nolent et al., 2006; Johansson and Kokinsky, 2009; Valkenburg et al., 2011; Bai et al., 2012). One study report a poor validity with a correlation of 0.31 (Bai et al., 2012), while in one additional study the results show an excellent validity with correlations between 0.89 and 0.90 (Van Dijk et al., 2000). Two studies report a moderate and good construct validity for (non-)ventilated neonates NICU: correlations between 0.44 and 0.74 (Caljouw et al., 2007; Van Dijk et al., 2009). One study was executed on a Maternity Ward and presents poor to moderate correlations: 0.38 and 0.50 (Tristão et al., 2013). One study investigated the construct validity in patients admitted to a Burn Hospital and report a moderate validity for background pain, correlation 0.45, and excellent validity for procedural pain, correlation 0.88 (De Jong et al., 2010). In one study, the construct validity was studied by using Rasch analysis (De Jong et al., 2012). The Rasch analysis includes characteristics of the respondents and the items on the scale, based on the idea that the probability of getting a certain response is determined by the respondents' ability and by the difficulty of the items. This study reports item difficulty between −4.53 and 2.77 logits, and infit mean square between 0.46 and 1.63, suggesting the COMFORT-B scale measures pain adequately.

3.5 Responsiveness

We identified eight studies investigating this clinimetric property (Blauer and Gerstmann, 1998; Caljouw et al., 2007; Van Dijk et al., 2009; De Jong et al., 2010; Franck et al., 2011; Cury et al., 2013; Tristão et al., 2013; Boerlage et al., 2014). In these studies changes of the (modified) COMFORT scores were investigated before and after painful interventions, or pharmacological interventions to relieve distress (sedatives) or pain (opioids). All studies report significant increases or decreases of the COMFORT scores according to the expectations, indicating that the COMFORT scale is able to measure change. Results on responsiveness are summarized in Table 9.

3.6 Quality assessment

The results of the methodological quality assessments are shown in Table 10. The quality of the studies that report on internal consistency and interrater reliability varies between poor ($n = 1$), fair ($n = 15$), good ($n = 11$) and excellent ($n = 1$). The quality of the studies that report on construct validity varies between poor ($n = 5$), fair ($n = 15$) and good ($n = 4$). The studies that report on responsiveness are of fair ($n = 2$), good ($n = 4$) and excellent ($n = 2$) quality. We report concerns related to blinding, missing items and sample size.

3.6.1 Blinding

Blinding is important wherever items are to be compared without influences from the preferences or expectations from researchers or participants. The absence of blinding usually leads to an overestimation of the results. Blinding was a major problem in 16 studies, because the researcher, nurse or physician who rated distress, sedation or pain with the instruments under investigation were unblinded or this item was unclear described (Marx et al., 1994; Blauer and Gerstmann, 1998; Brunow de Carvalho et al., 1999; Van Dijk et al., 2000; Carnevale and Razack, 2002; Crain et al., 2002; Courtman et al., 2003; Ista et al., 2005; Nolent et al., 2006; Van Dijk et al., 2009; Franck et al., 2011; Valkenburg et al., 2011; Bai et al., 2012; De Jong et al., 2012; Cury et al., 2013; Tschiedel et al., 2015).

3.6.2 Missing items

A high number of missing items can introduce bias, resulting in an overestimation or underestimation in

Table 9 The responsiveness.

Author, year	Type of COMFORT scale	Description	Results
Pain			
Blauer and Gerstmann (1998)	COMFORT	Scores 2 min before (T0), 2 min during (T1) and 3 min after (T2): (1) Intubation (2) Intravenous insertion procedures (3) Endotracheal suctioning (4) Diaper change	Scores increased between T0 and T1 and decreased between T1 and T2: (1) Intubation: 19/20–28/29–19/20 (2) Intravenous insertion: 15–23/24–16/18 (3) Endotracheal suctioning: 15/16–20/21–16 (4) Diaper change: 15/16–22–19
Caljouw et al. (2007)	COMFORT-without blood pressure	Scores 10 min before and 1 min after a heel puncture	The mean scores increased from 13.2 (SD 2.74) to 23.3 (SD 5.47)
Cury et al. (2013)	COMFORT	Scores before and after endotracheal suctioning on the first, second and third day after cardiac surgery	The median scores increased during the endotracheal suctioning procedure, but only significantly on the first day after surgery: 16 versus 20, $p = 0.02$ (first day), 16 versus 17, $p = 0.90$ (second day), 18 versus 22, $p = 0.36$ (third day)
De Jong et al. (2010)	COMFORT-B	Scores 1 h before and directly after wound care.	Scores increased after wound care: 12.61 (SD 2.95) versus 18.54 (SD 4.12), $p < 0.001$
Franck et al. (2011)	COMFORT	Scores immediately prior to and 3 min after painful procedures	Scores increased with 27% (95% CI 7–51%, $p < 0.001$), explained variance $R^2 = 22\%$.
Tristão et al. (2013)	COMFORT	Scores 3 min before (T0), 3 min after (T1) and 6 min after (T2) a heel puncture	Scores increased between T0 and T1, mean score difference 23 (range 13–30, $p < 0.01$). Scores decreased between T1 and T2, mean difference 9 (range 6–25, $p < 0.01$)
Distress and pain			
Boerlage et al. (2014)	COMFORT-B	Scores before and after (within 120 min) pharmacological interventions	Mean scores decreased from 20.0 (SD 3.7) to 14.1 (SD 4.7), $p < 0.001$
Van Dijk et al. (2009)	COMFORT-neo	Scores: (1) Before and after pain or distress reducing interventions (2) In situations no pain and suspected for pain (3) In situations adequate sedation and suspected for oversedation	(1) Mean scores decreased from 19.8 (SD 3.8) to 12.0 (SD 3.4), $p \leq 0.001$ (2) Mean scores increased from 11.4 (SD 2.6) to 18.4 (SD 3.8), $p \leq 0.001$ (3) Mean scores decreased from 10.9 (SD 1.7) to 9.7 (SD 1.8), $p \leq 0.001$

SD, standard deviation.

the results of the study. Therefore, the number of missing items, whether the missing items were at random and how they were handled should be described. In 21 studies the number of missing items was not mentioned, or not explained further (Ambuel et al., 1992; Marx et al., 1994; Blauer and Gerstmann, 1998; Brunow de Carvalho et al., 1999; Carnevale and Razack, 2002; Crain et al., 2002; Wielenga et al., 2004; Lee and Young, 2005; Twite et al., 2005; Bear and Ward-Smith, 2006; Caljouw et al., 2007; Gjerstad et al., 2008; Johansson and Kokinsky, 2009; De Jong et al., 2010; Valkenburg et al., 2011; Bai et al., 2012; De Jong et al., 2012; Cury et al., 2013; Da Costa Silva et al., 2013; Tristão et al., 2013; Tschiedel et al., 2015).

3.6.3 Sample size

An adequate sample size is important to make inferences about a population from a sample. Small sample sizes generally lead to imprecise results. Following the criteria of Mokkink et al. (2013) we considered a sample size of less than 30 poor and a sample size over 100 excellent. In total five studies resulted in low scores on this item (Brunow de Carvalho et al., 1999; Wielenga et al., 2004; Lee and Young, 2005; Gjerstad et al., 2008; Cury et al., 2013).

4. Discussion

In this systematic review, we studied the clinimetric properties of the (modified) COMFORT scale as a

Table 10 The methodological quality.

Author, year	Internal consistency	Interrater reliability	Content validity	Construct validity	Responsiveness
Distress					
Ambuel et al. (1992)	Fair	Fair	Poor	Poor	
Carnevale and Razack (2002)	Fair				
Gjerstad et al. (2008)				Fair	
Wielenga et al. (2004)		Fair		Fair	
Sedation					
Brunow de Carvalho et al. (1999)				Poor	
Courtman et al. (2003)				Good	
Crain et al. (2002)		Fair		Fair	
Da Costa Silva et al. (2013)		Fair		Fair	
Ista et al. (2005)	Good	Fair		Fair	
Lamas et al. (2008)				Fair	
Lee and Young (2005)	Fair	Fair			
Marx et al. (1994)				Poor	
Triltsch et al. (2005)				Fair	
Tschiedel et al. (2015)				Good	
Twite et al. (2005)				Good	
Pain					
Bai et al. (2012)				Fair	
Bear and Ward-Smith (2006)	Fair	Good			
Blauer and Gerstmann (1998)	Poor				Fair
Caljouw et al. (2007)	Good	Fair		Poor	Good
Cury et al. (2013)				Poor	Good
De Jong et al. (2010)	Fair	Good		Fair	Fair
De Jong et al. (2012)				Good	
Franck et al. (2011)	Good	Fair			Excellent
Tristão et al. (2013)				Fair	Good
Valkenburg et al. (2011)	Good	Fair		Fair	
Van Dijk et al. (2000)	Good	Excellent		Fair	
Distress, sedation and/or pain					
Boerlage et al. (2014)		Good			Good
Johansson and Kokinsky (2009)		Good		Fair	
Nolent et al. (2006)	Good			Fair	
Van Dijk et al. (2009)	Fair	Good		Fair	Excellent

tool to measure distress, sedation and pain in children from birth to 18 years old. We included 30 studies that report on the reliability, validity and/or responsiveness. These studies report overall an adequate reliability for distress, sedation and pain. Construct validity varies from good to excellent for distress, from moderate to excellent for sedation and from poor to excellent for pain. Finally, the COMFORT scale shows adequate responsiveness, regardless of the construct it is being claimed to measure. None of the included studies resulted in high quality scores on the clinimetric properties under consideration, according to the COSMIN criteria.

Distress and pain are difficult to discriminate; these experiences may occur simultaneously, influence each other and present with comparable responses. Pain frequently results in distress, however, distress may have causes other than pain. Despite their close association, distinguishing between the two is clinically important as they are treated differently. Unfortunately, up to date there is no tool (either physiological or behavioural) available that is able to differentiate between distress and pain.

The COMFORT scale was originally developed as a continuous measure of distress in children aged from birth to 18 years receiving ventilation in an intensive care environment. Studies from later date confirmed the validity of the COMFORT scale for this originally intended population and clinical context. This systematic review also presents studies that investigated the validity of the (modified) COMFORT scale in the assessment of sedation. Although distress and sedation are not the same, it can be argued these concepts are in the same continuum. Most studies on sedation included ventilated children in an intensive care environment, as in the studies on

distress. The studies on sedation report a construct validity between moderate and excellent, suggesting that the COMFORT scale might be helpful in the assessment of sedation. Other studies investigating the clinimetric properties of the (modified) COMFORT scale focus on pain assessment. The COMFORT scale was not developed to measure pain, but it is obvious that distress and pain are related, which is reflected in the similarity of the content of pain instruments and the (modified) COMFORT scale. The studies that compare the (modified) COMFORT scale with a pain measurement tool are executed in patients with different health conditions and in different clinical contexts. For example, some studies on pain included ventilated patients, who were therefore also sedated. The variety in the results suggest the (modified) COMFORT is less capable in the assessment of pain, but the heterogeneity in patient population and circumstances hampers clear conclusions.

There has been debate about the relevance of physiological parameters in assessing distress, sedation and pain. The original COMFORT scale combines behavioural and physiological items. Behavioural parameters (e.g. body movement, facial expression) rely on subjective observations and interpretations and are therefore questioned. Physiological parameters (e.g. heart rate, blood pressure) might be more objective, but are often influenced by the disease and medical interventions. Heart rate and blood pressure were removed from the original COMFORT, as they have been shown to have low item total correlations, indicating the internal consistency would improve if these items were excluded (Van Dijk et al., 2000; Carnevale and Razack, 2002; Ista et al., 2005; Nolent et al., 2006). On contrary, other studies investigating the COMFORT scale show a combination of behavioural and physiological items is preferable, because both account for a significant proportion of the variance in scores (Van Dijk et al., 2001; Franck et al., 2011). Furthermore, the correlation between physiological and behavioural items improved with a higher intensity of pain, suggesting a combination of both dimensions is more useful to diagnose severe, but not moderate pain (Van Dijk et al., 2001). Up to date, the correlation between behavioural and physiological variables remain an imperfectly solved problem. Therefore, in daily practice a careful interpretation of all information on the patient's situation is considered essential. For example, information on disease, treatment or previous experiences of the patient might be valuable for an accurate assessment of distress and pain (Von Baeyer and Spagrud, 2007; Schiavenato and Craigh, 2010).

In this review, the COSMIN checklist was used for the assessment of methodological quality (Mokkink et al., 2013). This checklist became available for researchers in 2010. Most publications in this review are from earlier date, and were not guided by the suggestions and explanations the COSMIN offers. This might explain the limited studies with good or excellent quality scores in this review. A formal assessment to study the relationship between the quality rates and results does not exist and visual inspection did not reveal any relationship between the quality of the studies and the magnitude of the results. Therefore, we decided to present all selected studies in this review, also the ones with low quality scores. However, it must be kept in mind a low quality score represent a high risk of bias, meaning there might be an important flaw in the study that might have resulted in an underestimation or overestimation of the results. Unfortunately, it is impossible to estimate to what extent the risk of bias affect the results of a particular study.

The included studies were clinically and methodologically heterogeneous. We found differences in the age of the included patients, reasons for admission and treatments. In addition, different primary endpoints were reported (distress, sedation and pain) and the clinimetric properties of the COMFORT scale were studied using different comparators, methodology and statistics. As a result of this heterogeneity, data synthesis was impossible.

This review identified a considerable number of studies on the clinimetric properties of the (modified) COMFORT scale. Still, more research is needed. Firstly, additional studies are warranted for specific patients, like children with neurological impairments and physically or mentally disabled children as they are excluded from studies so far. Also studies that include children older than 4 years are limited up to date. Secondly, studies that investigated the validity of the (modified) COMFORT scale show mixed results, especially in identifying pain. Therefore, more studies of high quality are needed to establish the validity of the COMFORT scale in measuring pain. Finally, more research on the responsiveness is needed to investigate the ability of the COMFORT to measure changes at the extremes of distress/sedation or pain.

5. Conclusions

The COMFORT scale shows overall an adequate reliability in providing information on distress, sedation and pain. Construct validity varies from good to excellent for distress, from moderate to excellent for

sedation, and from poor to excellent for pain. The included studies were clinically and methodologically heterogeneous, hampering firm conclusions.

Author contributions

The authors have made substantial contributions to conception and design of the study (J.M., C.L., H.V.), acquisition of data (J.M., P.R.J., E.V., H.V.), analysis and interpretation of data (J.M., P.R.J., E.V., E.I., H.V.) or drafting the article (J.M., P.R.J., E.V., E.I., H.V.). All authors discussed the results, commented on the manuscript and approved the final version to be submitted (J.M., P.R.J., E.V., E.I., C.L., H.V.).

References

Ambuel, B., Hamlett, K.W., Marx, C.M., Blumer, J.L. (1992). Assessing distress in pediatric intensive care environments: The COMFORT scale. *J Pediatr Psychol* 17, 95–109.

Bai, J., Hsu, L., Tang, Y., van Dijk, M. (2012). Validation of the COMFORT behavior scale and the FLACC scale for pain assessment in chinese children after cardiac surgery. *Pain Manag Nurs* 13, 18–26.

Bear, L.A., Ward-Smith, P. (2006). Interrater reliability of the COMFORT scale. *Pediatr Nurs* 32, 427–434.

Blauer, T., Gerstmann, D. (1998). A simultaneous comparison of three neonatal pain scales during common NICU procedures. *Clin J Pain* 14, 39–47.

Boerlage, A.A., Ista, E., Duivenvoorden, H.J., de Wildt, S.N., Tibboel, D., van Dijk, M. (2014).The COMFORT behaviour scale detects clinical meaningful effects of analgesics and sedative treatment. *Eur J Pain* 19, 473–479.

Brummelte, S., Grunau, R.E., Chau, V., Poskitt, K.J., Brant, R., Vinall, J., Gover, A., Synnes, A., Miller, S.P. (2012). Procedural pain and brain development in premature newborns. *Ann Neurol* 71, 385–396.

Brunow de Carvalho, W., Da Silva, L.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. *São Paulo Med J* 117, 192–196.

Caljouw, M.A., Kloos, M.A., Olivier, M.Y., Heemskerk, I.W., Pison, W.C., Stigter, G.D., Verhoef, A.M.J. (2007). Measurement of pain in premature infants with a gestational age between 28 to 37 weeks: Validation of the adapted COMFORT scale. *J Neonatal Nurs* 13, 13–18.

Carnevale, F.A., Razack, S. (2002). An item analysis of the COMFORT scale in a pediatric intensive care unit. *Pediatr Crit Care Med* 3, 177–180.

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 Med* 29, 2239–2246.

Crain, N., Slonim, A., Murray, M., Pollack, M.M. (2002). Assessing sedation in the pediatric intensive care unit by using BIS and the COMFORT scale. *Pediatr Crit Care Med* 3, 11–14.

Cury, M.R., Martinez, F.E., Carlotti, A.P. (2013). Pain assessment in neonates and infants in the post-operative period following cardiac surgery. *Postgrad Med J* 89, 63–67.

Da Costa Silva, C., Osorio Alves, M.M., Dos Santos El Halal, M.G., Dos Santos Pinheiro, S., Carvalho, P.R.A. (2013). A comparison of gradual sedation levels using the Comfort-B scale and bispectral index in children on mechanical ventilation in the pediatric intensive care unit. *Rev Bras Ter Intensiva* 25, 306–311.

De Jong, A., Baartmans, M., Bremer, M., Van Komen, R., Middelkoop, E., Tuinebreijer, W., Van Loey, N. (2010). Reliability, validity and clinical utility of three types of pain behavioural observation scales for young children with burns aged 0–5 years. *Pain* 150, 561–567.

De Jong, A.E., Tuinebreijer, W.E., Bremer, M., Van Komen, R., Middelkoop, E., Van Loey, N. (2012). Construct validity of two pain behaviour observation measurement instruments for young children with burns by rasch analysis. *Pain* 153, 2260–2266.

Dorfman, T.L., Schellenberg, E.S., Remple, G.R., Scott, S.D., Hartling, L. (2014). An evaluation of instruments for scoring physiological and behavioral cues of pain, non-pain related distress, and adequacy of analgesia and sedation in pediatric mechanically ventilated patients: A systematic review. *Int J Nur Stud* 51, 654–676.

Franck, L.S., Bruce, S. (2009). Putting pain assessment into practice: Why is it so painful. *Pain Res Manag* 14, 13–20.

Franck, L.S., Ridout, D., Howard, R., Peters, J., Honour, J.W. (2011). A comparison of pain measures in newborn infants after cardiac surgery. *Pain* 152, 1758–1765.

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, e848–e853.

Hermann, C., Hohmeister, J., Demirakca, S., Zohsel, K., Flor, H. (2006). Long-term alteration of pain sensitivity in school-aged children with early pain experiences. *Pain* 125, 278–285.

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. *Pediatr Crit Care Med* 6, 58–63.

Johansson, M., Kokinsky, E. (2009). The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. *Nurs Crit Care* 14, 122–130.

Johnston, C., Barrington, K.J., Taddio, A., Carbajal, R., Filion, F. (2011). Pain in Canadian NICUs: Have we improved over the past 12 years? *Clin J Pain* 27, 225–232.

Lamas, A., López-Herce, J., Sancho, L., Mencía, S., Carrillo, A., Santiago, M.J., Martínez, V. (2008). Assessing sedation in critically ill children by bispectral index, auditory-evoked potentials and clinical scales. *Intensive Care Med* 34, 2092–2099.

Lee, W.K., Young, B.W.Y. (2005). Measuring the sedation level of mechanically ventilated infants by a modified COMFORT scale. *HK J Paediatr (new Series)* 10, 189–195.

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. *Crit Care Med* 22, 163–170.

Mokkink, L.B., Terwee, C.B., Knol, D.L., Stratford, P.W., Alonso, J., Patrick, D.L., De Vet, H.C. (2010a). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodo* 10, 23–30.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., De Vet, H.C. (2010b). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res* 19, 539–549.

Mokkink, L.B., Terwee, C.B., Gibbons, E., Stratford, P.W., Alonso, J., Patrick, D.L., De Vet, H.C. (2010c). Interrater agreement and reliability of the COSMIN (consensus-based standards for the selection of health status measurement instruments) checklist. *BMC Med Res Methodol* 10, 82–93.

Nolent, P., Nanquette, M.C., Carbajal, R., Renolleau, S. (2006). Which sedation scale should be used in the paediatric intensive care unit? A comparative prospective study. *Arch Pédiatr* 13, 32–37.

Sackett, D., Hayes, R., Guyatt, G., Tugwell, P. (1991). *Clinical Epidemiology: A Basic Science for Clinical Medicine*. (Boston: Brown & Co).

Schiavenato, M., Craigh, K.D. (2010). Pain assessment as a social transaction – beyond the "gold standard". *Clin J Pain* 26, 667–676.

Stoddard, F.J., Ronfeldt, H., Kagan, J., Drake, J.E., Snidman, N., Murphy, M.J., Saxe, G., Burns, J., Sheridan, R.L. (2006). Young burned children: The course of acute stress and physiological and behavioural responses. *Am J Psychiatry* 163, 1084–1090.

Taddio, A., Katz, J., Ilersich, A.L., Koren, G. (1997). Effect of neonatal circumcision on pain response during subsequent routine vaccination. *Lancet* 349, 599–603.

Terwee, C.B., Bot, S.D.M., De Boer, M.R., Van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M., De Vet, H.C.W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60, 34–42.

Triltsch, A.E., Nestmann, G., Orawa, H., Moshirzadeh, M., Sander, M., Groβe, J., Genähr, A., Konertz, W., Spies, C.D. (2005). Bispectral index versus COMFORT score to determine the level of sedation in paediatric intensive care unit patients: A prospective study. *Crit Care* 9, R9–R17.

Tristão, R.M., Garcia, N.V.M., Lacerda de Jesus, J.A., Tomaz, C. (2013). Comfort behaviour scale and skin conductance activity: What are they really measuring? *Acta Paediatr* 102, e402–e406.

Tschiedel, E., Müller, O., Schara, U., Felderhoff-Müser, U., Dohna-Schwake, C. (2015). Sedation monitoring during open muscle biopsy in children by COMFORT Score and Bispectral Index–a prospective analysis. *Paediatr Anesth* 25, 265–271.

Twite, M.D., Zuk, J., Gralla, J., Friesen, R.H. (2005). Correlation of the Bispectral Index Monitor with the COMFORT scale in pediatric intensive care unit. *Pediatr Crit Care Med* 6, 648–653.

Valkenburg, A.J., Boerlage, A.A., Ista, E., Duivenvoorden, H.J., Tibboel, D., Van Dijk, M. (2011). The COMFORT-behavior scale is useful to assess pain and distress in 0-to 3-year-old children with down syndrome. *Pain* 152, 2059–2064.

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, 367–377.

Van Dijk, M., de Boer, J.B., 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. *J Pain Symptom Manage* 22, 600–609.

Van Dijk, M., Roofthooft, D.W., Anand, K.J., Guldemond, F., de Graaf, J., Simons, S., de Jager, Y., van Goudoever, J.B., Tibboel, D. (2009). Taking up the challenge of measuring prolonged pain in (premature) neonates: The COMFORTneo scale seems promising. *Clin J of Pain* 25, 607–616.

Vet, N.J., Ista, E., De Wildt, S.N., Van Dijk, S.N., Tibboel, D., De Hoog, M. (2013). Optimal sedation in paediatric intensive care patients: A systematic review. *Intensive Care Med* 39, 1524–1534.

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, 140–150.

Wielenga, J.M., De Vos, R., de Leeuw, R., De Haan, R.J. (2004). COMFORT scale: A reliable and valid method to measure the amount of stress of ventilated preterm infants. *Neonatal Netw* 23, 39–44.

Wright, B.D., Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Meas Trans* 8, 370.

Web references

International Association for the Study of Pain (IASP). www.iasp-pain.org (accessed December 2014).

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., De Vet, H.C. The COSMIN checklist manual. www.cosmin.nl (accessed January 2013).

Preferred Reporting Items for Systematic reviews and Meta-Analysis (Prisma) recommendations for the reporting of the study. www.prisma-statement.org (accessed July 2014).