Clinimetric properties of the Nociception Coma Scale (-Revised): A systematic review

P. Vink[1,2], C. Lucas[3], J.M. Maaskant[4,5], W.S. van Erp[6,7], R. Lindeboom[3], H. Vermeulen[8,9]

1 Department of Neurology and Neurosurgery, Academic Medical Center (AMC), Amsterdam, The Netherlands

2 Omni Cura Nursing Teaching Research, Amsterdam, The Netherlands

3 Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Medical Faculty (AMC), University of Amsterdam, The Netherlands

4 Emma Children's Hospital, Academic Medical Center Amsterdam, The Netherlands

5 ACHIEVE Centre of Applied Research, Faculty of Health, Amsterdam University of Applied Sciences, The Netherlands

6 Department of Primary and Community Care, Centre for Family Medicine, Geriatric Care and Public Health, Radboud University Medical Centre, Nijmegen, The Netherlands

7 Coma Science Group, Cyclotron Research Center and Neurology Department, University of Liège, Belgium

8 Department of Surgery, Academic Medical Center (AMC), Amsterdam, The Netherlands

9 Radboud University Medical Center, Radboud Institute for Health Sciences, Scientific Center for Quality of Healthcare (IQ Healthcare), Nijmegen, The Netherlands

Correspondence
Peter Vink
E-mail: p.vink@amc.uva.nl

Abstract

The Nociception Coma Scale is a nociception behaviour observation tool, developed specifically for patients with disorders of consciousness (DOC) due to (acquired) brain injury. Over the years, the clinimetric properties of the NCS and its revised version (NCS-R) have been assessed, but no formal summary of these properties has been made. Therefore, we performed a systematic review on the clinimetric properties (i.e. reliability, validity, responsiveness and interpretability) of the NCS(-R). We systematically searched CENTRAL, CINAHL, Embase, PsycInfo and Web of Science until August 2015. Two reviewers independently selected the clinimetric studies and extracted data with a structured form. Included studies were appraised on quality with the COSMIN checklist. Eight studies were found eligible and were appraised with the COSMIN checklist. Although nearly all studies lacked sample size calculation, and were executed by the same group of authors, the methodological quality ranged from fair to excellent. Important aspects of reliability, construct validity and responsiveness have been studied in depth and with sufficient methodological quality. The overview of clinimetric properties in this study shows that the NCS and NCS-R are both valid and useful instruments to assess nociceptive behaviour in DOC patients. The studies provide guidance for the choice in NCS-R cut-off value for possible pain treatment and cautions awareness of interprofessional differences in NCS-R measurements.

Significance: This systematic review provides a structured overview of the clinimetric properties of the Nociception Coma Scale (-Revised) and provides insights for a solid evidence-based nociception behaviour assessment and treatment plan.

1. Introduction

The assessment and treatment of pain is one of the most important areas in both medical and nursing care. Patients who are unable to communicate their pain are at risk of under recognition and under-treatment of their discomforts (Herr et al., 2011).

Among these patients exists a particularly fragile group with disorders of consciousness (DOC), such as unresponsive wakefulness syndrome (UWS) and minimally conscious state (MCS), due to acquired brain injury (ABI). Both patient groups show behavioural sleep–wake cycles, but only MCS patients might interact with objects or people and even give (adequate or inadequate) verbal responses. Both diagnoses are by definition incompatible with a reliable and consistent ability to communicate about pain experiences, while the nature of these conditions is characterized by various factors that can give rise to pain (e.g. spasticity, contractures, etc.; Thibaut et al., 2015).

It has been a subject of discussion whether these patients are capable of experiencing pain in a similar way as conscious patients. These discussions may complicate communication among healthcare professionals, and between medical staff and the patient's family. The general consensus tells us, however, to assess pain behaviour in all of these patients regardless of neurological capacity of nociceptive awareness (Schnakers et al., 2012; Chatelle et al., 2014).

For the assessment of nociceptive behaviour among these patients, the Nociception Coma Scale (NCS) has been developed in 2009 (Schnakers et al., 2009). It is a nociception behaviour observation scale, consisting of four items (motor response, verbal response, facial expression and visual response) with each a range score of 1–3. Over the years the scale has been further validated, revised to the Nociception Coma Scale – Revised (NCS-R), which omits the visual response item, and its practical implications have been discussed. Despite this research, the NCS(-R) has not been implemented in the field of neuroscience nursing, possibly due to the lack of a definitive conclusion on its clinimetric properties and practical implications (Vink et al., 2015).

This systematic review therefore aims to evaluate the clinimetric properties, in terms of reliability and validity, of the NCS and NCS-R as an instrument to assess nociceptive behaviour in patients with DOC. To ensure a systematic approach, we used the Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) checklist (Mokkink et al., 2010a,c). Such an in-depth overview would reveal gaps in validity and reliability and provide guidance for future research. It would also provide conclusions for clinical implementation of the scale(s) and ensure confidence for clinicians and nurses to use the NCS(-R) for the measurement of pain in DOC patients. This is an important prerequisite to determine a solid evidence-based pain assessment and -treatment policy for these patients.

2. Methods

The study was entered in the International prospective register of systematic reviews (PROSPERO) on May 26th 2015.

The recommendations of the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) were used for the reporting of this study (Prisma Statement, 2009).

2.1 Identification of studies

In August 2015, one of the authors (PV) performed a literature search in CENTRAL, CINAHL, Embase, PsycInfo and Web of Science. The search was repeated in August 2016. The search strategy consisted of the terms 'nociception coma scale' without limitations on language or publication date. To ensure a sensitive search, the strategy consisted of free text and did not use any Medical Subject Headings. A clinical librarian was consulted to investigate the sensitivity of the search strategy and no other feasible strategy was found.

To identify eligible studies, the search results were screened on titles and abstracts by two authors (PV and HV) independently. Disagreements were resolved by discussion. Articles were included if they met the inclusion criteria. Full-text of the article was reviewed when title/abstract did not provide sufficient information. Reference lists of the potentially eligible studies were manually searched to identify additional articles. We also contacted experts in the field to detect possible studies. Again, no limitations were imposed on language, publication date or publication status.

Studies were eligible for inclusion if:
(1) The aim of the study was to evaluate one or more clinimetric properties of the NCS or NCS-R as a tool to measure nociceptive behaviour.
(2) The study population consisted of adult patients (>18 years) with DOC due to acquired brain injury.
(3) The study was published as original article.

In the absence of a golden standard reference for the measurement of pain in non-communicative patients with DOC, studies comparing the NCS or NCS-R to instruments measuring the same construct were considered eligible. For the same reasons, the authors carefully considered the inclusion of studies aimed to evaluate the correlation between the NCS or NCS-R to a physiologic phenomenon known to be present during nociception.

Reviews, guidelines, descriptive studies, editorials or poster publications were excluded. Publications of which full-texts were unavailable to university libraries were also excluded. Disagreements were solved by discussion.

2.2 Data extraction

A structured form was used to extract data from original studies on in- and exclusion criteria, number of patients, number of observations, patient characteristics (age, diagnoses), methods of painful stimuli, the researched scale (NCS/NCS-R) and context (interventions, setting). Context data on clinical setting, observation technique and observers were extracted by one of the authors (PV). Age was extracted as provided by the article or, when all data was available, calculated into a median with range. Data on clinimetric properties included internal consistency, interrater reliability, intrarater reliability, measurement error, content, construct, criterion validity and responsiveness. The definitions used are presented in Table 1. Data on clinimetric properties were extracted by two authors (PV and CL) independently, whereas disagreements were resolved by discussion. If no consensus could be reached, a third author (JM) was consulted. As one researcher (PV) was the author of one of the publications (Vink et al., 2014), a third, independent researcher (JM) carried out quality assessment in this case.

2.3 Quality assessment

The COSMIN checklist was used to assess methodological quality of the studies (Mokkink et al., 2010a,b,c). Two authors (PV and CL)

independently assessed the methodological quality of the eligible studies. Disagreements were resolved by discussion and consultation of the COSMIN checklist manual (Mokkink et al., 2013) or a third author (JM). The reviewers provided each clinimetric property with an overall quality score, based on the 4-point scale (excellent, good, fair or poor) of the corresponding quality criteria of the COSMIN checklist. An overall rating was obtained by consensus of all involved reviewers. The reviewers were not blinded for authors, research environments and journals.

2.4 Outcome measurements

For reliability outcomes, the authors maintained magnitude criteria as described by Terwee et al. Reported Cronbach's alpha were considered adequate if above 0.70 and further classified as *unacceptable* (>0.5), *poor* (0.5–0.59), *questionable* (0.6–0.69), *acceptable* (0.7–0.79), *good* (0.8–0.89) or *excellent* (≥ 0.9; Terwee et al., 2007). We classified strength of agreement by means of Intraclass Correlation (ICC) or (weighted) kappa as *slight* (0.00–0.20), *fair* (0.21–0.40), *moderate* (0.41–0.60), *substantial* (0.61–0.80) or *excellent* (0.81–1.00; Landis and Koch, 1977).

3. Results

3.1 Identification of studies

The results of the literature searches are summarized in the flow diagram in Fig. 1. Among the five searches, a total of 95 references were assessed for eligibility on title and abstract. On the basis of title

Table 1 Definitions of clinimetric properties.

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 s 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

Definitions of clinimetric properties (Mokkink et al., 2013).

or abstract, 26 found references were excluded on type of publication: conference publication ($n = 15$), meeting/editorial ($n = 8$), book ($n = 3$). Among the remaining 69 references, the reviewers identified 29 unique studies, among which eight were found eligible for review by both reviewers independently (Schnakers et al., 2009; Chatelle et al., 2012, 2014, 2015; Sattin et al., 2013; Vink et al., 2014; Riganello et al., 2015; de Tomasso et al., 2015). The eligibility of one other study (Thibaut et al., 2015) was questioned by both reviewers and eventually excluded based on the full-text. One study (Suraseranivongse et al., 2015) was excluded because it could not be obtained by online databases, university libraries or by contacting the authors. A complete list of the in- and excluded studies is provided in Appendix S1 (Supporting Information).

3.2 Description of included studies

One study (Vink et al., 2014) reported on internal consistency, three on reliability (Schnakers et al., 2009; Vink et al., 2014; Riganello et al., 2015), one on content validity (Schnakers et al., 2009), two on cross-cultural validity (Sattin et al., 2013; Vink et al., 2014), three on construct validity (Schnakers et al., 2009; Vink et al., 2014; de Tomasso et al., 2015) and three on responsiveness (Schnakers et al., 2009; Chatelle et al., 2012, 2015; Vink et al., 2014).

The included studies were published between 2009 and 2015 and originated from Belgium ($n = 5$), Italy ($n = 2$) and The Netherlands ($n = 1$). From the eight included studies, three investigated clinimetric properties of the NCS (Schnakers et al., 2009; Sattin et al., 2013; Riganello et al., 2015), three of the NCS-R (Chatelle et al., 2014, 2015; de Tomasso
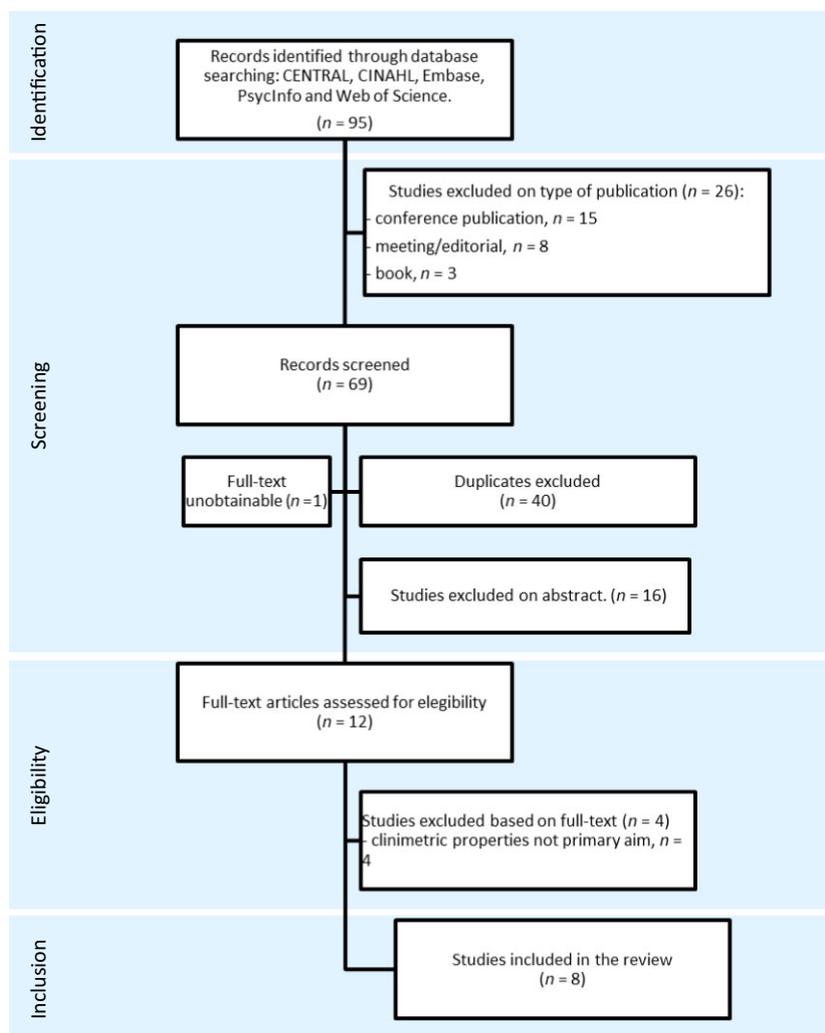

Figure 1 PRISMA flow diagram of search.

et al., 2015) and two of both NCS and NCS-R (Chatelle et al., 2012; Vink et al., 2015). The developer of the original NCS was involved in six of the included studies. All studies used the NCS(-R) as an instrument to assess nociceptive behaviour. The study details are summarized in Table 2.

The patient samples from seven of the included studies are a good representation of ABI patients with DOC, which are prevalent in both (semi)acute and long-term settings. Two studies included patients from acute or semiacute settings such as Intensive Care Units (ICU), neurology and neurosurgery hospital wards (Vink et al., 2014; Chatelle et al., 2015). Three studies included patients from both (semi-) acute settings and long-term care facilities (Schnakers et al., 2009; Chatelle et al., 2012; Riganello et al., 2015). All studies with the exception of Vink et al. (2014) included patients with confirmed diagnosis of UWS, also known as vegetative state (VS), or MCS by means of the Coma Recovery Scale – Revised (Kalmar and Giacino, 2005). None of the studies included intubated patients. The patients of the included studies show a variation in age, reasons for admission, diagnosis and time since onset. A summary of these characteristics are presented in Table 2. Five studies used observations after standardized administration of a painful stimulus to the nailbed during at least 5-s until a behavioural response was observed (Schnakers et al., 2009; Chatelle et al., 2012, 2014; Vink et al., 2014; Riganello et al., 2015). The observers consisted of neuropsychologists, physiotherapists and (neuroscience) nurses. The assessment of methodological quality by means of the COSMIN checklist is rated as *poor, fair, good* or *excellent*. The ratings are summarized in Table 3 and further explained in the following paragraphs.

3.3 Reliability

A total of three studies reported on reliability of the NCS and/or NCS-R (Schnakers et al., 2009; Vink et al., 2014; Riganello et al., 2015). A summary of the psychometric values on internal consistency (Cronbach's α) and interrater agreement (*kappa* or ICC) is presented in Tables 4 and 5.

3.3.1 Internal consistency

Only the study of Vink et al. (2014) investigated the internal consistency of the NCS and NCS-R. The methodological quality for this item of reliability was rated as *good*, because more thorough testing of the

item response theory might have been desirable. Although a limited number of patients ($n = 10$) were included, the number of observations ($n = 270$) was increased using video observations and a large pool of observers ($n = 27$). Three video observations (during rest, tactile and noxious stimuli) per patients were shown to nurses from different hospitals with different levels of education (minimum of Associate in Nursing Degree) and experience with ABI patients (median 7 years, range 1–28). The found Cronbach's α was *questionable* for both NCS (α = 0.68) total scores and NCS-R (α = 0.61) total scores. To test internal consistency Cronbach's α was recalculated when removing each sub score item from the scale. The result remained *questionable*, with the exception of facial expression. Cronbach's α when removing this item was *poor* with α 0.54.

3.3.2 Reliability

In the study of Schnakers et al. (2009) neuropsychologists, with experience in patients with DOC, assessed the NCS during noxious stimulus of 15 patients. The methodological quality for this item was rated as *fair*, due to a moderate sample ($n = 48$) and an unclear indication of observer independency (e.g. blinding and adequate time interval). The found interrater agreement was *substantial* ($k = 0.61$) for total scores, *substantial* for facial expression and visual response ($k = 0.73$) and *excellent* for motor response and verbal response ($k = 0.93$).

Vink et al. (2014) investigated the interrater agreement between hospital (neuroscience) nurses. The methodological quality for this item was rated as *good*. The sample size was moderate, but the number of observations was increased as described in the above section 'internal consistency'. The found single-measure intraclass coefficient (ICC) was *substantial* for both NCS (ICC = 0.67) and NCS-R (ICC = 0.69). An average measure ICC of 0.95 (*excellent*) was found for both scales. No statistically significant differences were found in ICC estimates when comparing groups of nurses based on educational level or years of experience.

Riganello et al. (2015) tested interrater agreement between two observers. One observer had a background in neuropsychology and had used the NCS for 6 months. The other observer had a background in physiotherapy and had used the NCS for 2 months. Both observers had experience working with severely brain-injured patients. The methodological quality for this item was rated as *good*, because a moderate sample size was included

Table 2 Summary of included clinical trials.

Author, year Country	Scale	Inclusion/ exclusion criteria	No. of patients/ controls	Reason for admission (patients)	Diagnosis of DOC (patients)	Age in years[a]	No. of observations	Observation technique	Context: time since onset	Context: setting
Schnakers et al., 2009 Belgium	NCS	A, B, C, D, / F, G, H	Patients: 48	TBI: 17 AN: 10 IS: 7 HS: 7 Other: 7	UWS: 28 MCS: 20	[20–82]	2 per patient, total 96	Observation during: baseline, non-noxious stimulus, standardized noxious stimulus to nailbed	<1 month (n = 31) 1 month to 6 years (n = 17)	ICU, Neurology Wards, Neurorehabilitation and nursing homes
Chatelle et al., 2012 Belgium	NCS NCS-R	A, B, C, D, / F, G, H,	Patients: 64	TBI: 22 AN: 20 HS: 8 IS: 3 Other: 11	UWS: 27 MCS: 37	[20–82]	3 per patient, total 192	Standardized noxious stimulus to nailbed	<1 month (n = 21) 1 month to 6 years (n = 43)	ICU, Neurology Wards, Neurorehabilitation and nursing homes
Chatelle et al., 2014 Belgium	NCS-R	None described	Patients: 49	TBI: 21 AN: 17 SAH: 3 HS/IS: 4 Other: 4	UWS: 18 MCS: 31	40 [21–83]	1 per patient, total 49	Observation during: standardized noxious stimulus to nail bed + PET-scan same day	<1 year (n = 26) >1 year (n = 23)	Not specified
Vink et al., 2014 The Netherlands	NCS NCS-R	A, B, C / I	Patients: 10	TBI: 1 HS/IS: 7 Other: 2	GCS*: 10 [6–12]	56 [26–75]	Total 270	Video recordings during: baseline, non-noxious stimulus, noxious stimulus (varies)	Not specified	Neurology/Neurosurgery Ward
Chatelle et al., 2015 Belgium	NCS-R	A, B, D, E, / G, H, I	Patients: 39	TBI: 15 HS: 13 AN: 6 IS: 3 Other: 2	UWS: 12 MCS: 27	61 [21–93]	2 per patient, total 78	Observation: During/before potentially painful nursing cares, + Before and after analgesic treatment	<1 month (n = 36) 1 month to 6 years (n = 3)	Acute Care: ICU and Neurology Ward
de Tomasso et al., 2015 Italy	NCS-R	D	Patients: 9 Controls: 11	TBI: 4 HS: 2 AN: 2 IS: 1	UWS: 5 MCS: 4	Patients: 60.6 [44–76] Controls: 60.2 [50–68]	2 series of 25 laser, electrical, auditory and visual stimuli per patient/control	EEG and EOG	<1 month (n = 0) 1 month to 6 years (n = 9) [3 month to 4 year]	Long-term Care

Table 2 (*Continued*)

Author, year Country	Scale	Inclusion/ exclusion criteria	No. of patients/ controls	Reason for admission (patients)	Diagnosis of DOC (patients)	Age in years[a]	No. of observations	Observation technique	Context: time since onset	Context: setting
Riganello et al., 2015 Italy	NCS	A, B, C, D, / F, G, H	Patients: 44	TBI: 19 HS: 9 IS: 16	UWS: 26 MCS: 18	53.6 (19.3)	Total 176	Observation during: standardized noxious stimulus to nailbed	<1 month (n = 0) 1 month to 6 years (n = 44)	Semi-ICU hospital (n = 16) and Long-term Care (n = 28)

Inclusion Criteria

A. Age over 18 years;
B. No administration of neuromuscular blockers or sedative agents within 24 h prior to assessment;
C. Documented periods of eye-opening;
D. Diagnosis of UWS or MCS based on the Coma Recovery Scale – Revised;
E. Presence of potential pain during care interventions

Exclusion Criteria

F. Fractures or flaccid paralysis of the upper limbs;
G. History of premorbid brain injuries;
H. Developmental, psychiatric or neurological disorders prior to admission;
I. Intubation

NCS, Nociception Coma Scale; NCS-R, Nociception Coma Scale Revised; TBI, traumatic brain injury; HS, Haemorrhagic Stroke, including subarachnoid haemorrhage and intracranial haemorrhage; AN, anoxia; IS, ischaemic stroke; GCS, Glasgow Coma Scale; UWS, unresponsive wakefulness syndrome; MCS, minimally conscious state; EEG, electroencephalographic; EOG, electro-oculogram; ICU, Intensive Care Unit.

[a]Numbers are presented as [range], median [range] or mean (SD).

(*n* = 44). The assessments were performed by two observers independently and repeated after 1 week. The interrater agreement during the first measurement occasion was *fair*, *k* = 0.40 (0.21–0.47 subscores) and *moderate*, *k* = 0.57 (0.33–0.62 subscores) at the second measurement occasion. This study also reported intrarater reliability, which ranged from *substantial* (*k* = 0.66) to *moderate* (*k* = 0.57) among their two raters. Observations were obtained with 1-week interval.

3.4 Validity

3.4.1 Content validity

The content validity was only assessed by the developers of the NCS in the original article of Schnakers et al. (2009). The methodological quality for this item was rated as *excellent*. The subscore items were selected from earlier scientific publications on pain assessment in non-communicative patients. Vital signs, such as breathing, respiration and heart rate, were not incorporated in the NCS on the basis of a pilot study and previous scientific research. Social behaviours, such as interpersonal interaction, were excluded from the NCS, due to the behavioural limitations of UWS/MCS patients.

3.4.2 Cross-cultural validity

Two studies investigated part of the cross-cultural validity. Sattin et al. (2013) described the translation of the NCS to Italian. The methodological quality for this item was rated as *poor*. The scale was translated according to international standards with multiple forward and backward translations, but after translation no further group analysis was performed. Vink et al. (2014) translated the NCS to Dutch, also according to international standards but again without further group analysis. The methodological quality for this item was therefore also rated as *poor*. The authors state that the behavioural descriptions of the NCS are such common terminology in neurological care, that further testing of cross-cultural validity was not necessary.

3.4.3 Construct validity

One study investigated construct validity of the NCS and two of the NCS-R. In the 2009 study of Schakers et al., the NCS was correlated with other pain behaviour measurement instruments. The methodological quality for this item was rated as *good*, mainly due to a moderate sample size (*n* = 48). Convergent validity

Table 3 Methodological quality of primary studies.

Author, year	Internal consistency	Reliability	Content validity	Cross-cultural	Construct validity	Responsiveness
Schnakers et al., 2009	–	Fair	Excellent	–	Good	Good
Chatelle et al., 2012	–	–	–	–	–	Excellent
Sattin et al., 2013	–	–	–	Poor	–	–
Chatelle et al., 2014;	–	–	–	–	Good	–
Vink et al., 2014	Good	Good	–	Poor	–	Excellent
Chatelle et al., 2015	–	–	–	–	–	Fair
de Tomasso et al., 2015	–	–	–	–	Fair	–
Riganello et al., 2015	–	Good	–	–	–	–

Table 4 Summary of internal consistency coefficients.

	Item rest correlation	Cronbach's α (if item deleted)[a]
Motor response	0.39	0.69
Verbal response	0.49	0.63
Visual response	0.48	0.61
Facial expression	0.58	0.54
Total NCS score	–	0.68
Total NCS-R score	–	0.61

Internal consistency from Vink et al. (2014).

[a]Cronbach's α if item deleted values from NCS with four items. Values for total NCS and NCS-R scores denote total scale internal consistency.

Table 5 Summary of interrater agreement coefficients.

	Unweighted Cohen's Kappa Schnakers et al., 2009	Single-measure ICC Vink et al., 2014	Unweighted Cohen's Kappa (week 1 & 2) Riganello et al., 2015
Motor response	0.93	0.68	0.21 & 0.33
Verbal response	0.93	0.62	0.47 & 0.62
Visual response	0.73	0.42	0.37 & 0.41
Facial expression	0.73	0.61	0.34 & 0.38
Total NCS	0.61	0.67	0.40 & 0.57
Total NCS-R	–	0.69	–

Interrater agreement coefficients from Schnakers et al. (2009), Vink et al. (2014) and Riganello et al. (2015).

was established between the NCS and the Neonatal Infant Pain Scale (NIPS; Lawrence et al., 1993), the Faces Legs Activity Cry Consolability pain assessment tool (FLACC; Merkel et al., 1997), the Pain Assessment In Advanced Dementia Scale (PAINAD; Warden et al., 2003) and the Checklist of Non-verbal Pain Indicators (CNPI; Feldt, 2000). The Spearman rank correlations were all above 0.71 on total scores and ranged from 0.26 to 0.79 on individual items. All total score correlations were statistically significant ($p < 0.05$). From the subscore items (motor, verbal, face and visual) only the motor response had no statistically significant correlation with the NIPS, FLACC and CNPI.

The study of Chatelle et al. (2014) correlated the NCS-R to metabolism in the anterior cingulate cortex (ACC), which is known to be involved in pain processing. The methodological quality for this item was rated as *good*. Although the study was of excellent design, the sample size was moderate ($n = 49$) and there was insufficient information to determine the appropriateness of the statistical analysis. The NCS-R was recorded during a standardized painful stimulus on the nailbed, after which a PET-scan was performed. A relation was found between the NCS-R total scores and the posterior part of the ACC ($Z = 2.76$; corrected p-value=0.18). A statistically

significant relation was found between the ACC metabolism and the level of consciousness, as measured by the Coma Recovery Scale – Revised, suggesting that the NCS-R only reflects the construct nociception (convergent validity).

One study by de Tomasso et al. (2015) compared electroencephalography (EEG) and electro-oculography (EOG) results during visual, auditory, non-noxious and noxious laser stimulation between UWS/MCS patients and healthy controls. The methodological quality for this item was rated as *fair*, because the study population was small ($n = 20$; 9 patients, 11 controls). Also, although statistical analysis for comparing EEG/EOG results of UWS patients with the control group was sufficiently described, no correlation coefficients, distributions or p-values were provided. It is therefore not possible to assess the strength of the correlations with the NCS-R. The cortical response to noxious laser stimuli was reported to be uncorrelated to NCS-R scores.

3.5 Responsiveness

A total of four studies reported on responsiveness of the NCS and/or NCS-R for an increase in nociception. An overview of the resulting cut-off values is presented in Table 6.

Table 6 Summary of discriminative values.

Study	Patient groups	Group Mean ± SD/Cut-Off Value	Sensitivity	Specificity	Definition
Schnakers et al., 2009	All	NCS 2.5 ± 1.5	NA	NA	No nociception
	All	NCS 5.1 ± 1.7	NA	NA	Light nociception
	All	NCS 8.0 ± 1.0	NA	NA	Moderate nociception
Chatelle et al., 2012	All	NCS ≥4	46%	97%	Noxious stimulation present
	MCS	NCS-R ≥ 4	83%	95%	Noxious stimulation present
	UWS	NCS-R ≥ 3	96%	89%	Noxious stimulation present
Vink et al., 2014	All	NCS ≥2	74%	68%	Possible presence of pain
	All	NCS ≥3	72%	67%	Probable presence of pain
	All	NCS-R ≥ 1	77%	75%	Possible presence of pain
	All	NCS-R ≥ 2	74%	73%	Probable presence of pain

MCS, minimally conscious state; UWS, unresponsive wakefulness syndrome.
Group mean values from Schnakers et al. (2009) and cut-off values from Chatelle et al. (2012) and Vink et al. (2014).

Schnakers et al. (2009) compared the NCS total scores with nociception thresholds as provided by the Checklist of Non-verbal Pain Indicators (CNPI). The methodological quality for this item was rated as *good*, mainly due to a moderate sample size ($n = 48$) and lack of a priori hypotheses descriptions. A significant difference ($p < 0.05$) was found between mean NCS total scores, grouped according to the CNPI threshold for no nociception (NCS 2.5 ± 1.5), light nociception (NCS 5.1 ± 1.7) and moderate nociception (NCS 8.0 ± 1.0). No individual t-values or p-values were provided.

The study of Chatelle et al. (2012) investigated the changes of the NCS total and subscores between resting observation (baseline), tactile/non-noxious stimulus and noxious stimulus. Although the sample size ($n = 64$) could have been bigger, the methodological quality for this item was rated as *excellent* because observation techniques and analyses were well executed. The total score and motor, verbal and facial subscores all increased significantly between these conditions. The visual response item was not significantly different between non-noxious and noxious stimulus. An ROC analysis of the NCS revealed an optimal sensitivity of 46% and specificity of 97%, at a cut-off value of 4. Because of the lack of discriminative abilities of the visual response item, the authors decided to remove it from the scale and thereby created the NCS-R. ROC analysis for the NCS-R resulted in a sensitivity of 96% and specificity of 89%, also with a cut-off value of 4. Separate analysis for MCS and UWS patients revealed a NCS-R cut-off value of 4 for MCS patients (sensitivity 83% and, specificity 95%) and 3 for UWS patients (sensitivity 96% and specificity 89%).

Vink et al. (2014) investigated the cut-off values for the NCS and NCS-R to differentiate between the absence of pain (no stimulus), the possible presence of pain (none versus tactile stimulus) and the probable presence of pain (tactile versus noxious stimulus). The methodological quality for this item was rated as *excellent*, because methodological flaws were limited by the use of video recordings. An ROC analysis revealed a cut-off value of NCS ≥3 for the probable presence of pain (sensitivity 72% and specificity 67%). For the NCS-R, the ROC analysis revealed a cut-off value of ≥2 for the probable presence of pain (sensitivity 74% and specificity 74%).

Chatelle et al. (2015) investigated responsiveness of the NCS-R in an acute care setting. Patients with an NCS-R score of 4 or higher, measured by trained nurses, during potentially painful nursing care interventions were and found eligible for the study. After analgesic treatment, the NCS-R was reassessed during similar conditions. The methodological quality for this item was rated as *fair*, because of a moderate sample ($n = 39$), the lack of blinding, comparison to a control and a vague description of the time interval ('within 24 h'). The NCS-R total scores decreased significantly after analgesic treatment (5.2 ± 1.3 vs. 3.7 ± 1.9; $z = 4.37$; $p < 0.0001$) and so did the subscores of motor response (2 ± 0.7 vs. 1.5 ± 0.9; $z = 3.09$; $p = .002$), verbal response (1.2 ± 1.1 vs. 1 ± 1; $z = 2.22$; $p = 0.027$) and facial expression (2 ± 0.5 vs. 1.2 ± 0.9; $z = 3.92$; $p < .0001$). This did not differ according to aetiology or the level of consciousness, suggesting the NCS-R only measures nociception.

4. Discussion

The results of our systematic review show that all clinimetric properties of the NCS and most of the NCS-R have been studied and tested. This systematic review shows that the aspects of reliability, construct validity and responsiveness have been studied

sufficiently for the NCS/NCS-R. The methodological quality of the studies investigating these aspects ranged from *fair* to *excellent*. In comparison to similar systematic reviews on pain observation scales, such as the COMFORT pain scale, the quality of the studies is rather high (Maaskant et al., 2016). For clinical implications and future research, some comments must be made.

The reliability of the scale(s) is one of the best studied aspects of validity. Three studies, all on the NCS, used observers from different professions and different methods of observation and analyses (Schnakers et al., 2009; Vink et al., 2014; Riganello et al., 2015). The results of these studies tend towards *substantial* to *excellent* interrater agreement, although Riganello et al. (2015) reports a range from *fair* to *moderate*. The different results in this last study might be due to the limited number of observers (*n* = 2) and a difference in background (physiotherapy) of one of the observers. Although we believe that the NCS produces consistent and reproducible results, further research on inter- and in particular intrarater agreement between different disciplines could strengthen this statement. In clinical practice, the possible difference in measurement between professions should be kept in mind and if possible eliminated by interdisciplinary training/education.

Construct validity proves to be one of the most difficult aspects of validity to investigate for these scale(s). Because there is no golden standard for pain measurement to compare the NCS(-R) with, the researchers had to correlate the NCS(-R) scores to other pain behaviour measurement tools for the same construct or a physiologic phenomenon known or suspected to be present during nociception. This resulted in the support of convergent validity by Schnakers et al. (2009) and divergent validity by Chatelle et al. (2014), both showing good methodological quality. We believe that there is sufficient evidence to use either NCS or NCS-R to assess nociceptive behaviour in UWS/MCS patients. New knowledge and technologies to measure (neuro)-physiological structures involved in nociception and nociceptive awareness could strengthen this statement in future research.

The responsiveness of the NCS and NCS-R are of great importance to daily practice for both nurses and physicians to evaluate the adequacy of pain management. The results of four studies show that the NCS and NCS-R increase during nociception, thus detecting change over time in nociceptive behaviour (Schnakers et al., 2009; Chatelle et al., 2012, 2015; Vink et al., 2014). For clinical practice, we suggest a pain protocol that combines a cut-off value with the clinical judgement of the healthcare professional to initiate diagnostic and/or pain reducing interventions, pharmaceutical or non-pharmaceutical.

Three studies, varying from *fair* to *excellent* methodological quality, have tried to find an overall cut-off value for the presence of nociception (Schnakers et al., 2009; Chatelle et al., 2012; Vink et al., 2014). However, the choice for a cut-off value might prove difficult for clinical practice as cut-off values differ among the studies and may even differ between MCS and UWS patients (Schnakers et al., 2012). First, Chatelle et al. found no statistically significant difference in visual subscores between non-noxious and noxious stimulation conditions. They did find a statistically significant difference between baseline versus non-noxious and baseline versus noxious stimulation. In the NCS-revised (NCS-R), the visual item was omitted, although the item could prove of importance when the patient shows nociception behaviour without non-noxious stimulation. Secondly, in the treatment of severely affected and hypocommunicative patients, it is generally accepted to 'err on the safe side' when it comes to pain treatment, i.e. to regard the possibility of treating non-existing pain acceptable in order to prevent undertreatment of pain.

Following this, we suggest to assess nociception behaviour with the NCS, but maintain the lowest found cut-off value of ≥2 for the NCS-R (Vink et al., 2014). This study did not differentiate between MCS and UWS patients, therefore the cut-off values of NCS-R ≥ 4 (MCS) and ≥3 (UWS) as found by Chatelle et al. (2012) might prove more valuable for settings with fully diagnosed patients. Whichever cut-off value is chosen, it can only be used as a general guideline since each individual patient can have different neurological or motor limitations to show nociceptive behaviour. For example, the use of neuromuscular function blockers or the presence of a tracheostomy will limit the patient's ability to reach a score on NCS motor or verbal item, respectively. A score below any given cut-off value is thereby no guarantee for the absence of nociception and in patients with low baseline scores, any increase in NCS should give rise to assessment of possible discomfort, rather than waiting for a general threshold score to be reached. When a pain reducing intervention is administered, either pharmaceutical or non-pharmaceutical, intrapatient changes in NCS scores should be used to assess the

effectiveness of the chosen treatment and determine a future treatment plan.

The quality of cross-cultural validation was *poor* in both studies investigating this aspect (Sattin et al., 2013; Vink et al., 2014). We believe that this item of validity might require further research, considering all included studies have been conducted in West-European countries. Although nociceptive behaviour does not differ among cultures, the observation and assessment of such pain-related behaviour might be subject to the observers (cultural) perception on pain.

Overall, the studies are of sufficient quality for an evaluation of the clinimetric properties of the NCS and NCS-R. A recurring limitation of these studies is the lack of sample size calculation, except for one (Vink et al., 2014). This may be due to the low incidence and prevalence of DOC (van Erp et al., 2014) or insufficient knowledge of the COSMIN guidelines. The low prevalence of prolonged DOC might also be the reason why the group of authors on the subject is relatively small. Many of the articles are written by the same group of authors and samples in the studies might partially even consist of the same patients. Further research in different countries and settings by different authors is recommended. Another possible limitation is the lack of blinding, though it can be discussed if blinding the observer of the painful stimuli is possible or even necessary. We believe that the NCS or NCS-R will be used in situations where the healthcare professional will either notice or suspect a painful stimulus. By not blinding the observers in the studies they have therefore closely joined and simulated daily practice.

In conclusion, we believe that the overview of clinimetric properties in this study shows that the NCS and NCS-R are both valid and useful instruments to assess nociceptive behaviour in DOC patients. Future research on cross-cultural validity and intrarater agreement will further strengthen this statement. Until a gold standard is available to determine the actual conscious perception of pain in DOC patients, healthcare professionals can use the NCS/NCS-R scores to determine a treatment strategy.

Acknowledgements

Author contributions

P.V. conceptualized and designed the study; analysed and interpreted the data; drafted and revised the article, and approved the final version of the manuscript to be published.

C.L. conceptualized and designed the study; analysed and interpreted the data; revised the article, and approved the final version of the manuscript to be published.

J.M.M. conceptualized and designed the study; analysed and interpreted the data; revised the article, and approved the final version of the manuscript to be published.

W.v.E. conceptualized and designed the study; revised the article, and approved the final version of the manuscript to be published.

R.L. interpreted the data; revised the article, and approved the final version of the manuscript to be published.

H.V. conceptualized and designed the study; analysed and interpreted the data; revised the article, and approved the final version of the manuscript to be published.

References

Chatelle, C., Majerus, S., Whyte, J., Laureys, S., Schnakers, C. (2012). A sensitive scale to assess nociceptive pain in patients with disorders of consciousness. *J Neurol Neurosurg Psychiatry* 83, 1233–1237.

Chatelle, C., Thibaut, A., Bruno, M., Boly, M., Beranrd, C., Hustinx, R., Schnakers, C., Laureys, S. (2014). Nocicepton coma scale-revised scores correlate with metabolism in the anterior cingulate cortex. *Neurorehabil Neural Repair* 28, 149–152.

Chatelle, C., De Val, M., Catano, A., Chaskis, C., Seeldrayers, P., Laureys, S., Biston, P., Schnakers, C. (2015). Is the nociception coma scale-revised a useful clinical tool for managing pain in patients with disorders of consciousness? *Clin J Pain* 32, 321–326.

van Erp, W., Lavrijsen, J., van de Laar, F., Vos, P., Laureys, S., Koopmans, R. (2014). The vegetative state/unresponsive wakefulness syndrome: A systematic review of prevalence studies. *Eur J Neurol* 21, 1–8. https://doi.org/10.1111/ene.12483.

Feldt, K.S. (2000). The checklist of nonverbal pain indicators (CNPI). *Pain Manag Nurs* 1, 13–21.

Herr, K., Coyne, P., McCaffery, M., Manworren, R., Merkel, S. (2011). Pain assessment in the patient unable to self-report: Position statement with clinical practice recommendations. *Pain Manag Nurs* 12, 230–250.

Kalmar, K., Giacino, J.T. (2005). 'The JFK coma recovery scale—revised. *Neuropsychol Rehabil* 15, 454–460.

Landis, J., Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Lawrence, J., Alcock, D., McGrath, P., Kay, J., MacMurray, S.B., Dulberg, C. (1993). The development of a tool to assess neonatal pain. *Neonatal Netw* 12, 59–66.

Maaskant, J., Raymakers-Janssen, P., Veldhoen, E., Ista, E., Lucas, C., Vermeulen, H. (2016). The clinimetric properties of the COMFORT scale: A systematic review. *Eur J Pain* 20, 1587–1611. https://doi.org/10.1002/ejp.880.

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

Mokkink, L., Terwee, C., Gibbons, E., Stratford, P., Alonso, J., Patrick, D., De Vet, H. (2010a). Interrater agreement and reliability of the COSMIN (consensus-based standards for the measurement and selection of health status measurement instruments) checklist. *BMC Med Res Methodol* 10, 1–11.

Mokkink, L., Terwee, C., Knol, D., Alonso, J., Patrick, D., De Vet, H. (2010b). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol* 10, 1–8.

Mokkink, L., Terwee, C., Patrick, D., Stratford, P., Knol, D., De Vet, H. (2010c). 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., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., De Vet, H. (2013). *The COSMIN checklist manual* [Online]. Available at: www.cosmin.nl [November 2015].

Prisma Statement (2009). *Preferred Reporting Items for Systematic reviews and Meta-Analysis*, [Online]. Available at: http://www.prisma-statement.org/ [January 2015].

Riganello, F., Cortese, M., Arcuri, F., Candelieri, A., Guglielmino, F., Dolce, G., Sannita, W., Schnakers, C. (2015). A study of the reliability of the Nociception Coma Scale. *Clin Rehabil* 29, 388–393.

Sattin, D., Pagani, M., Covelli, V., Giovannetti, A., Leonardi, M. (2013). The Italian version of the Nociception Coma Scale. *Int J Rehabil Res* 36, 182–186.

Schnakers, C., Chatelle, C., Vanhaudenhuyse, A., Majerus, S., Ledouc, D. et al. (2009). The Nociception Coma Scale: A new tool to assess nociception in disorders of consciousness. *Pain* 148, 215–219.

Schnakers, C., Chatelle, C., Demertzi, A., Majerus, S., Laureys, S. (2012). What about pain in disorders of consciousness? *AAPS J* 14, 437–444.

Suraseranivongse, S., Yuvopoositanont, P., Srisakkrapikoop, P., Pommul, R., Phaka, W., Itthimathin, P. (2015). A comparison of pain scales in patients with disorders of consciousness following craniotomy. *J Med Assoc Thai* 98, 684–692.

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

Thibaut, A., Chatelle, C., Wannez, S., Deltombe, T., Stender, J., Schnakers, C., Laureys, S., Gosseries, O. (2015). Spasticity in disorders of consciousness: A behavioral study. *Eur J Phys Rehabil Med* 51, 389–397.

de Tomasso, M., Navarro, J., Lanzillotti, C., Ricci, K., Buonocunto, F., Livrea, P., Lancioni, G. (2015). Cortical responses to salient nociceptive and not nociceptive stimuli in vegetative and minimal conscious state. *Front Hum Neurosci* 9, 1–14.

Vink, P., Eskes, A., Lindeboom, R., van den Munckhof, P., Vermeulen, H. (2014). Nurses Assessing pain with the nociception coma scale: Interrater reliability and validity. *Pain Manag Nurs* 15, 881–887.

Vink, P., Verweij, L., van Erp, W.S., Lucas, C., Vermeulen, H. (2015). A survey on pain assessment in patients with disorders of consciousness in Dutch hospitals and nursing homes. *Br J Neurosci Nurs* 11, 881–887. https://doi.org/10.12968/bjnn.2015.11.1.34.

Warden, V., Hurley, A.C., Volicer, L. (2003). Development and psychometric evaluation of the pain assessment in advanced dementia (PAINAD) scale. *J Am Med Dir Assoc* 4, 9–15.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. List of included and excluded primary studies, databases in which they were found and if abstract and/or full-text were read before in/exclusion.