

Assessing pain intensity in children with chronic pain: Convergent and discriminant validity of the 0 to 10 numerical rating scale in clinical practice

Danielle Ruskin PhD CPsych^{1,2}, Chitra Laloo BHSc PhD Candidate³, Khushnuma Amaria PhD CPsych², Jennifer N Stinson RN PhD CPNP^{1,4,5}, Erika Kewley Hon BA¹, Fiona Campbell MD^{1,6}, Stephen C Brown MD^{1,6}, Michael Jeavons MD¹, Patricia A McGrath PhD CPsych⁶

D Ruskin, C Laloo, K Amaria, et al. Assessing pain intensity in children with chronic pain: Convergent and discriminant validity of the 0 to 10 numerical rating scale in clinical practice. *Pain Res Manag* 2014;19(3):141-148.

BACKGROUND: In clinical practice, children are often asked to rate their pain intensity on a simple 0 to 10 numerical rating scale (NRS). Although the NRS is a well-established measure for adults, no study has yet evaluated its validity for children with chronic pain.

OBJECTIVES: To examine the convergent and discriminant validity of the NRS as it is used within regular clinical practice to document pain intensity for children with chronic pain. Interchangeability between the NRS and an analogue pain measure was also assessed.

METHODS: A cohort of 143 children (mean [\pm SD] age 14.1 \pm 2.4 years; 72% female) rated their pain intensity (current, usual, lowest and strongest levels) on a verbally administered 0 to 10 NRS during their first appointment at a specialized pain clinic. In a separate session that occurred either immediately before or after their appointment, children also rated their pain using the validated 0 to 10 coloured analogue scale (CAS).

RESULTS: NRS ratings met a priori criteria for convergent validity ($r > 0.3$ to 0.5), correlating with CAS ratings at all four pain levels ($r = 0.58$ to 0.68 ; all $P < 0.001$). NRS for usual pain intensity differed significantly from an affective pain rating, as hypothesized ($Z = 2.84$; $P = 0.005$), demonstrating discriminant validity. The absolute differences between NRS and CAS pain scores were small (range 0.98 ± 1.4 to 1.75 ± 1.9); however, the two scales were not interchangeable.

CONCLUSIONS: The present study provides preliminary evidence that the NRS is a valid measure for assessing pain intensity in children with chronic pain.

Key Words: Children; Chronic pain; Clinical pain assessment; Numerical rating scale; Pain intensity scales

Self-report measures are a key source of information about pain for children who understand the measure and are capable of communicating with caregivers (1,2). It is important to balance children's self-reports of pain with knowledge of the clinical context, the child's developmental level and current state, as well as behavioural observation (1,3). Many analogue scales, facial scales, word checklists, pain diaries, clinical interviews and pain questionnaires are used clinically to assess children's pain (2,4-7). However, the numerical rating scale (NRS) is one of the most commonly used self-report scales for measuring children's pain, likely due to its ease of use (it requires no specialized equipment) and because its 0 to 10 metric is preferred by health care professionals who assess children's pain (8,9). While clinicians use a variety of verbal anchors for this scale (10), patients are typically

L'évaluation de l'intensité de la douleur chez les enfants atteints de douleur chronique : la validité convergente et discriminante de l'échelle d'évaluation numérique de 0 à 10 en pratique clinique

HISTORIQUE : En pratique clinique, on demande souvent aux enfants d'évaluer l'intensité de leur douleur sur une simple échelle d'évaluation numérique (ÉÉN) de 0 à 10. Même si cette échelle est une mesure bien établie chez les adultes, aucune étude n'en a encore évalué la validité chez les enfants atteints de douleur chronique.

OBJECTIFS : Examiner la validité convergente et discriminante de l'ÉÉN telle qu'elle est utilisée en pratique clinique normale pour corroborer l'intensité de la douleur chez les enfants atteints de douleur chronique. Évaluer également l'interchangeabilité entre l'ÉÉN et une mesure analogue de la douleur.

MÉTHODOLOGIE : Une cohorte de 143 enfants (âge moyen [\pm ÉT] de 14,1 \pm 2,4 ans, 72 % de femmes) ont évalué verbalement le niveau d'intensité de la douleur (niveau actuel, habituel, minimal et maximal) selon une ÉÉN de 0 à 10 lors de leur premier rendez-vous à une clinique de douleur spécialisée. Lors d'une autre séance, juste avant ou juste après leur rendez-vous, les enfants ont également évalué leur douleur à l'aide de l'échelle analogique colorée (ÉAC) validée de 0 à 10.

RÉSULTATS : Les évaluations de l'ÉÉN respectaient des critères *a priori* de validité convergente ($r > 0,3$ à $0,5$), qui présentaient une corrélation avec les évaluations d'ÉAC aux quatre niveaux de douleur ($r = 0,58$ à $0,68$; tous $P < 0,001$). Tel qu'on l'avait postulé, l'ÉÉN pour l'intensité de la douleur habituelle différerait de manière significative de l'évaluation de la douleur affective ($Z = 2,84$; $P = 0,005$), en démontrant la validité discriminante. Les différences absolues entre les résultats de douleur de l'ÉÉN et de l'ÉAC étaient peu importantes (plage de $0,98 \pm 1,4$ à $1,75 \pm 1,9$). Cependant, les deux échelles n'étaient pas interchangeables.

CONCLUSIONS : La présente étude fournit des preuves préliminaires selon lesquelles l'ÉÉN est une mesure valide d'évaluation de l'intensité de la douleur chez les enfants atteints d'une douleur chronique.

asked, "How strong is your pain, where 0 is no pain and 10 is the strongest or worst pain you can imagine?". Although the psychometric properties of the NRS have been well established for adults with acute and chronic pain conditions (11-14), studies have only recently explicitly examined the psychometric properties of the NRS as a pain measure for children with acute pain (15-19).

Miró et al (15) evaluated the correlation between the NRS and the Faces Pain Scale – Revised (FPS-R) for two cohorts: 175 schoolchildren (eight to 12 years of age) and 63 children postsurgery (six to 16 years of age). Children rated their strongest pain on both scales (during the previous three months for schoolchildren, and during first postoperative week for the patient group). As hypothesized, the NRS and FPS-R ratings correlated strongly for both cohorts ($r = 0.78$ and

¹Department of Anesthesia and Pain Medicine; ²Department of Psychology, The Hospital for Sick Children, Toronto; ³Medical Sciences Graduate Program, Faculty of Health Sciences, McMaster University, Hamilton; ⁴Department of Child Health Evaluative Sciences, The Hospital for Sick Children; ⁵Lawrence S Bloomberg Faculty of Nursing; ⁶Department of Anesthesia, Faculty of Medicine, University of Toronto, Toronto, Ontario

Correspondence: Dr Danielle Ruskin, Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8. Telephone 416-813-8960, fax 416-813-7646, e-mail danielle.ruskin@sickkids.ca

$r=0.93$, respectively). The NRS also showed adequate discriminant validity in relation to the Facial Affective Scale (FAS) for both cohorts ($r=0.58$ and $r=0.66$, respectively). Bailey et al (18) demonstrated the concurrent, construct and content validity of a verbal NRS for a cohort of 202 children (eight to 17 years of age) with acute pain who presented to an emergency department. Children's NRS ratings correlated strongly ($r=0.93$) with their ratings on a 10 mm visual analogue scale (VAS). Connelly and Neville (20) demonstrated high correlations (all $r>0.80$) among the NRS, FPS-R and VAS among 29 children postsurgery. In addition, von Baeyer et al (17) demonstrated agreement and functional equivalence for acute pain ratings on the NRS, the VAS and the FPS-R in three samples of children and adolescents seven to 17 years of age. More recently, Sánchez-Rodríguez et al (16) compared the NRS and the coloured analogue scale (CAS) in a sample of 126 schoolchildren six to eight years of age. Children were asked to recall their most frequent pain in the past three months and rate the maximal intensity of this pain. The reported correlation between the NRS and CAS was $r=0.79$; however, the scales were not interchangeable (16). These studies provide evidence of the validity of the NRS for assessing the intensity of acute pain in children.

The NRS should also be an appropriate scale for assessing chronic pain in children. Chronic pain describes pain that has persisted beyond the normal tissue healing time (typically considered to be three months) (21). Unlike acute pain, which has a generally predictable time course, the intensity of chronic pain is dynamic and varies widely over a prolonged period. To capture meaningful information regarding the intensity of chronic pain, it is important to ask children about the pattern of pain and document intensity at a variety of levels – for example, at its usual, lowest and strongest levels over a specified time period, rather than asking only about one discrete level. The NRS is ideally suited as a convenient and flexible scale for obtaining multiple ratings in clinical practice. However, although the NRS is often used as an outcome measure in studies of children's chronic pain (22-24), no study has yet examined aspects of its validity for use in children with chronic pain conditions.

Furthermore, most studies examining the validity of the NRS have been conducted in highly controlled settings where the assessment protocol (eg, question phrasing, choice of anchors, time of administration) is standardized. However, these validation data may not be easily generalizable to routine clinical practice, where the assessment protocol is subject to 'real world' factors such as time constraints, varying interview styles of different health care professionals (eg, modification of top scale anchor) and the time period between administration of comparator scales (25).

To address these gaps, the present study had the following aims: to examine NRS ratings obtained during a routine interdisciplinary team appointment to assess children's chronic pain intensity; to determine the convergent and discriminant validity of the NRS as used to assess children's chronic pain; and to determine the interchangeability or equivalency of the NRS with an analogue pain rating scale. Additionally, we explored differences in the validity of the NRS according to children's age, sex, pain duration and diagnosis.

METHOD

Participants and procedures

The present cohort study involved a convenience sample of children and adolescents eight to 17 years of age who were newly referred to the Chronic Pain Clinic at The Hospital for Sick Children (Toronto, Ontario) from March 19, 2008 to September 30, 2009. The institutional review board approved the present research as a retrospective study, which enabled the collection and study of children's pain ratings as they were typically obtained. Thus, demand characteristics and risk of recall bias that may have influenced children's pain ratings when they learned through the consent process that their pain ratings would be compared was minimized.

An interdisciplinary team (anesthesiologist, nurse, physiotherapist, psychologist and psychiatrist) was present for each child's pain

assessment appointment, which lasted for 45 min to 90 min. The assessment was led by the team anesthesiologist – a specialist in pediatric chronic pain – who used a semistructured assessment interview format to enable children to describe different aspects of their pain (ie, quality, temporal pattern and impact). Typically, NRS ratings were elicited by the anesthesiologist by asking the stem question, "How strong is your pain, where 0 is no pain and 10 is the strongest or worst pain you can imagine?". Wording was occasionally modified if a child had difficulty understanding the question. Additionally, while the goal was to obtain children's ratings at four pain levels, this did not always occur either because of time constraints during the interview or because the child did not provide a response for a specific pain level.

In an individual session that occurred either immediately before or after the team assessment (timing based on clinic schedule), children used the CAS (26) to rate their pain intensity and the FAS (26) to rate their pain affect. The CAS is used in the clinic because it is validated for children with chronic pain and is more easily administered and scored than the VAS (26). In addition, the CAS has ratio properties, which permits valid comparisons of pain scores within a child (eg, post-treatment) and across groups of children (versus category scales such as word, object or facial scales). See Ruskin et al (6) for further discussion of the benefits of ratio scales. CAS and FAS data were collected as part of the psychosocial clinical database by the clinic project coordinator using a script. The team and the CAS administrator were both blinded to the pain ratings children provided on the other pain measure. A 45 min to 90 min gap separated NRS ratings elicited during the team assessment and CAS/FAS ratings.

Measures

The 0 to 10 NRS: Children were asked to rate the intensity of their pain using any number between 0 and 10, where 0 is 'no pain' and 10 is 'the strongest or worst pain you can imagine'. Children in the Pain Clinic typically rate their pain at its current, usual, lowest and strongest levels.

CAS: The CAS (26) has been validated for children with acute and chronic pain (26-28). This analogue scale consists of a triangle 100 mm in length, increasing in width (from 10 mm to 30 mm) and colour (from white to dark red), mounted in plastic with a sliding marker. The bottom of the scale is labelled 'no pain' and the top is labelled 'most pain' and the back of the scale is a ruler ranging from 0 to 10 in 0.25 increments. Children slide the marker along the scale to the level that depicts their pain strength; they do not know the number corresponding to the marker. The marker is returned to the bottom of the scale after each rating. Children used the CAS to rate the intensity of their pain at time of assessment (current pain) followed by their usual, lowest and strongest pain levels. The CAS has demonstrated high test-retest reliability ($r=0.97$) as well as construct and convergent validity among children with pain (26-28).

FAS: The FAS (26) is a self-report measure used to assess the unpleasantness of a child's pain experience and has been validated in a sample of healthy children and children with chronic pain (26). It consists of nine faces depicting varying levels of distress, ranging from happy to sad. Numerical values for faces vary from 0.04 to 0.97, based on children's assigned affective ratings (29). Children are asked to choose the face that appears similar to how they feel deep down inside when they have pain, not necessarily the face that they show to others. For the present study, FAS ratings were provided for children's 'usual' pain.

Data analysis plan

The completion rates, distribution, means and SDs were first examined for children's NRS and CAS ratings to determine whether ratings differed according to order of administration (ie, whether the CAS was administered before or after the clinic appointment) or sex of participant, using independent-sample *t* tests (two-tailed) for comparison.

The primary objective of the present study was to examine the construct validity (convergent and discriminant) of the NRS for children with chronic pain (see Table 1 for a summary of this psychometric

TABLE 1
Summary of psychometric evaluation

Property	Operational definition	Method used
Construct validity	Determines the validity of abstract variables that cannot be directly observed, such as pain. These constructs are assessed by their relationships with other variables (30)	Convergent and discriminant validity were assessed as two components of construct validity, as described below
Convergent validity	Evaluates how well items on a pain scale correlate with other measures of the same construct or related variables (30)	Calculated by correlating children's NRS and CAS pain ratings for each pain level using two-tailed Pearson correlation coefficients. Spearman rank-order correlations were used to adjust for violations to normality
Discriminant validity	Evaluates how items on a pain scale correlate with other measures that are theoretically unrelated (30)	Calculated using Fisher's Z-transformation to compare the magnitude of NRS-CAS correlations with the magnitude of NRS-FAS correlations
Interchangeability	The extent to which individual scores captured on one pain scale are transposable with the individual scores on another pain scale (ie, the quality of the intraindividual assessments)	Examined the absolute differences between individual participants' NRS and CAS pain ratings as well as by using the Bland-Altman method (31)

CAS Coloured Analogue Scale; FAS Facial Affective Scale; NRS Numerical Rating Scale

evaluation). Previous validation studies in children with acute pain have obtained moderate to strong correlations between the NRS and criterion pain measures (ranging from $r=0.78$ to $r=0.93$) (15-18). However, these correlations were obtained by administering study measures consecutively, by the same individual, in a nonblinded fashion (ie, children and examiners were aware of the study purpose). The present study examined the performance of the NRS in the context of a routine interdisciplinary clinic appointment during which the NRS and comparator pain measure were administered at different times, by a different individual, using a retrospective blinded design, so that neither the team nor patients were aware that pain ratings were being studied during the clinical encounter. Thus, lower correlations than found in previous studies were expected. Consistent with recommendations by Streiner and Norman (30), it was determined a priori that correlations ranging from $r>0.30$ to 0.50 (denoting a moderate effect) would be acceptable.

Interchangeability of the pain measures was evaluated by examining the absolute differences between individual participants' NRS and CAS pain ratings as well as by using the Bland-Altman method (31). Bland and Altman (31) argue that using a correlation to assess the relationship between two measures can be misleading because it is affected by the range of the sample and measures the relation, but not the agreement between measures. Instead, agreement can be assessed using graphical techniques and simple calculations. The Bland-Altman method involves plotting the difference between the two measures of interest (y-axis) against their mean (x-axis) and then superimposing upper and lower limits of agreement. The limits of agreement are calculated based on the SD of the mean difference between the measures. A common practice in the literature is to use the 95% CIs around the mean difference as the limits of agreement (ie, ± 1.96 SD). These limits of agreement are then compared with a priori limits of acceptability, which are generally set at $\pm 20\%$ of the scale range for pain self-report measures in the adult literature (18). Recently, von Baeyer (32) suggested that the 95% CIs may be too strict a criterion for assessing agreement among pediatric pain self-report scales, given the variability in children's development and recognizing that standard practice in pediatric pain assessment includes multiple sources of judgments about pain (32). Instead, von Baeyer recommends the 80% CI (32). Thus, in the present study, the 95% and 80% CIs were examined as limits of agreement, and the limits of acceptability were set as 20% of the scale range (ie, ± 2 points on the 0 to 10 scale).

The impact of sex, age, pain duration and pain diagnosis on NRS-CAS correlations were explored. All data were analyzed using SPSS version 17.0 (IBM Corporation, USA) for Windows (Microsoft Corporation, USA). Alpha values were set at 0.05 and adjusted for multiple comparisons using a Bonferroni adjustment.

RESULTS

During the study period, 168 children with chronic pain between eight and 17 years of age were assessed in the clinic. Twenty-five children

TABLE 2
Participant characteristics

Participants	Age, years	
Total, n	143	14.1 \pm 2.4
Female	103 (72)	14.0 \pm 2.3
Male	40 (28)	14.3 \pm 2.5
Age distribution, years		
8 to 12	47 (32.9)	
13 to 15	48 (33.6)	
16 to 17	48 (33.6)	
Pain diagnosis		
Musculoskeletal	55 (38.5)	
Neuropathic	51 (35.7)	
Headache	17 (11.9)	
Abdominal pain	15 (10.5)	
Other	5 (3.5)	
Pain duration	2.1 \pm 2.2 years	
Range	1 month* to 12 years	
≤ 6 months, %	22.4	13.1 \pm 2.5
7 to ≤ 12 months, %	18.9	13.5 \pm 2.5
13 to ≤ 24 months, %	25.2	14.6 \pm 2.1
25 to 36 months, %	13.3	14.2 \pm 2.2
> 3 years, %	20.3	15.0 \pm 2.1
Ethnicity		
White	84 (58.7)	
Asian	6 (4.2)	
Black	3 (2.1)	
Other	4 (9.8)	
Not answered	9 (6.3)	
Not assessed	27 (18.9)	

Data presented as n (%) or mean \pm SD unless otherwise specified. *A child with complex regional pain syndrome-I who was fast-tracked into the clinic to receive urgent treatment. Thus, their pain duration is shorter than those of most new referrals

did not complete study measures due to developmental delay ($n=12$), scheduling difficulties ($n=8$), declined or too unwell ($n=3$), and pain-free status ($n=2$). Thus, both NRS and CAS pain scores were obtained for 143 participants (or 85% of the eligible patients assessed in clinic). Sixty-two percent of participants completed their CAS ratings before their appointment with the team, while 38% of participants completed the CAS afterwards. Demographics data and diagnostic classifications are presented in Table 2. Mean age did not differ significantly according to sex ($t=-0.04$, $P>0.05$). Pain duration did not differ significantly according to sex (25.8 ± 25.7 months for males and

TABLE 3
Numerical rating scale (NRS) and coloured analogue scale (CAS) pain values for current, lowest, usual and strongest pain levels

Pain level	NRS (0 to 10)					CAS (0 to 10)				
	n	Mean ± SD	Median	IQR	Range	n	Mean ± SD	Median	IQR	Range
Current	112	4.6±3.0	5.00	2.0–7.0	0–10	143	4.9±2.6	4.75	2.75–7.5	0–10
Lowest	138	2.7±2.8	2.00	0–5.0	0–9.0	143	3.3±2.5	3.00	1.25–5.25	0–8.75
Usual	122	5.8±2.4	6.00	4.5–7.5	0–10	143	6.0±2.4	5.50	4.75–7.75	0.25–10
Strongest	133	8.5±2.0	9.00	8.0–10	1–10	143	8.6±2.0	8.75	8.0–10.0	0–10

IQR Interquartile range

TABLE 4
Bivariate Pearson r correlations among numerical rating scale (NRS), coloured analogue scale (CAS) and Facial Affective Scale (FAS) scores for current, lowest, usual and strongest pain

	NRS			Current	CAS			FAS
	Lowest	Usual	Strongest		Lowest	Usual	Strongest	
NRS current	0.80**	0.37**	0.10	0.58**	0.50**	0.43**	0.26**	0.16
NRS lowest		0.44**	0.12	0.50**	0.68**	0.41**	0.22*	0.18
NRS usual			0.56**	0.56**	0.59**	0.58**	0.42**	0.41**
NRS strongest				0.40**	0.30**	0.47**	0.58**	0.43**
CAS current					0.71**	0.65**	0.48**	0.42**
CAS lowest						0.62**	0.30**	0.36**
CAS usual							0.68**	0.49**
CAS strongest								0.46**

Bolded values indicate NRS-CAS correlations to assess convergent validity. *P<0.05; **P<0.01

26.9±25.6 months for females; $t=-0.21$, $P>0.05$). χ^2 analyses showed that the frequency of diagnoses were similar across sexes ($\chi^2=1.25$, $P>0.87$).

Completion of pain ratings

Using the NRS, 91 children (63.6%) rated four levels of pain, 41 children (28.7%) rated three levels and 11 children (7.7%) rated only one or two levels. The most frequent ratings were for lowest pain (n=138), followed by strongest (n=133), usual (n=122) and current (n=112) pain levels. CAS ratings were obtained for all four pain levels for all children. The magnitude of correlations between the NRS and CAS were not significantly different for children who provided NRS ratings at all four pain levels compared with children who provided NRS ratings for <4 levels (all $Z<1.7$; $P>0.05$), except for lowest pain (all four levels $r=0.77$, <4 levels $r=0.53$, $Z=2.33$; $P=0.02$).

Approximately one-half of the children always used a whole number as their NRS rating (n=76), while 67 (46.9%) children used an intermediate or fractional rating for at least one pain rating. Five children (3.5%, mean age 14.9 years) used a range (eg, '6.0 to 7.0') to depict their pain level for the NRS. In these cases, the midpoint of their range in pain scores was used.

Distributions of pain scores

The NRS distributions for all four pain levels were skewed (Kolmogorov-Smirnov D test, all $D>0.129$; $P<0.01$). On the CAS, distributions for lowest and strongest pains were skewed (Kolmogorov-Smirnov D test, both $D>0.087$; $P<0.05$). The shape of NRS and CAS distributions (skewness and kurtosis) were similar at all pain levels. Spearman rank-order correlations were used in subsequent analyses to adjust for violations to normality.

Mean and median pain ratings

Mean NRS and CAS pain ratings with SDs and median values are presented in Table 3. As expected, values for the strongest pain levels were significantly higher than those for usual pain levels, which were significantly higher than those for the lowest pain levels for both the NRS and CAS (all $t>11.0$; $P<0.001$). Mean pain ratings did not differ according to sex (NRS ratings, all pain levels $t<0.73$, $P>0.05$; CAS pain ratings, all levels $t<1.02$; $P>0.05$) or according to order of administration (whether the CAS was administered before the NRS; NRS

ratings, all pain levels $t<1.4$, $P>0.05$; CAS ratings, all pain levels $t<1.05$, $P>0.05$).

VALIDITY ANALYSES

Convergent validity

NRS ratings correlated significantly with the CAS at all four pain levels: current ($r=0.58$, $n=112$), lowest ($r=0.68$, $n=138$), usual ($r=0.58$, $n=122$) and strongest ($r=0.58$, $n=133$), all $P<0.001$. Two-tailed Pearson correlations for NRS and CAS pain levels and FAS are presented in Table 4. Spearman rho (r_s) correlations (current: $r_s=0.61$, least: $r_s=0.64$, usual: $r_s=0.64$, strongest: $r_s=0.60$) were also computed given the skewed distributions for NRS and were also found to be significant (all $P<0.01$). Correlations did not differ significantly according to sex (Fisher's Z-transformation for independent samples, all $Z<1.62$; $P>0.05$). Multiple regression analyses were conducted to examine whether age (as a continuous variable) moderated the relationship between CAS and NRS for any of the four pain severity ratings. NRS was entered as the dependant variable. CAS pain rating and age were entered in step 1 (test of main effects) and product of CAS according to age entered in step 2 (for a test of the interaction). The interaction effect was not significant in any of the analyses (all R^2 change <0.06 , $P>0.05$; step 2), suggesting that the correlation between NRS and CAS did not vary according to age. Main effects for age were also not significant in any of the analyses (all Beta <0.07 ; $P>0.05$).

Discriminant validity

To examine the discriminant validity of NRS ratings from affective pain ratings, the magnitude of the correlations for NRS and CAS usual pain ratings ($r=0.63$; $P<0.001$, $n=113$) were compared with the correlation between NRS-usual and FAS-usual ($r=0.41$; $P<0.001$, $n=113$) using Fisher's Z-transformation test. The magnitude of the correlation between NRS-usual and FAS-usual was significantly lower than that between the NRS-usual and CAS-usual ($Z=2.84$; $P=0.005$).

Similarity of CAS and NRS pain ratings

Paired-samples t tests revealed that mean pain ratings on the NRS and CAS were similar at all pain levels. To further examine the similarity of the pain ratings on the NRS and CAS pain scales, the absolute differences between NRS and CAS pain ratings were calculated

TABLE 5
Bland-Altman analysis of agreement

Scales compared	Parameter	Limit of acceptability	Difference, mean \pm SD	Limits of agreement	
				95% CI	80% CI
NRS and CAS	Usual pain	± 2.0	-0.092 ± 2.1	-4.1 to 4.2	-2.8 to 2.6
	Lowest pain	± 2.0	-0.55 ± 2.1	-4.3 to 5.3	-3.2 to 2.1
	Strongest pain	± 2.0	-0.034 ± 1.6	-3.3 to 3.4	-2.0 to 2.1

CAS Coloured Analogue Scale; NRS Numerical Rating Scale

TABLE 6
Exploratory analyses according to pain duration and diagnosis

Pain duration	<6 months (n=32)			>3 years (n=43)			
	Pain level	NRS, mean \pm SD	CAS, mean \pm SD	Correlation	NRS, mean \pm SD	CAS, mean \pm SD	Correlation
Pain duration	Current	4.6 \pm 3.4	5.1 \pm 3.2	r=0.78; P=0.002	4.0 \pm 2.9	4.4 \pm 2.7	r=0.44; P=0.04
	Lowest	3.3 \pm 3.2	3.9 \pm 2.8	r=0.82; P<0.001	2.3 \pm 2.4	3.1 \pm 2.1	r=0.38; P=0.04
	Usual	5.3 \pm 2.7	6.0 \pm 2.5	r=0.60; P<0.001	6.4 \pm 1.7	6.2 \pm 2.1	r=0.59; P=0.002
	Strongest	8.4 \pm 1.6	8.1 \pm 2.3	r=0.74; P<0.001	9.2 \pm 0.8	8.9 \pm 1.2	r=0.48; P=0.009
Diagnosis	Musculoskeletal* (n=55)			Neuropathic† (n=51)			
	Pain level	NRS, mean \pm SD	CAS, mean \pm SD	Correlation	NRS, mean \pm SD	CAS, mean \pm SD	Correlation
Diagnosis	Current	4.9 \pm 3.1	5.2 \pm 2.8	r=0.54; P<0.001	4.3 \pm 2.8	4.4 \pm 2.8	r=0.65; P=0.01
	Lowest	3.0 \pm 2.6	3.4 \pm 2.6	r=0.72; P<0.001	2.4 \pm 2.7	3.0 \pm 2.4	r=0.76; P<0.001
	Usual	5.7 \pm 2.4	5.9 \pm 2.4	r=0.50; P<0.001	5.2 \pm 2.8	5.6 \pm 2.5	r=0.63; P<0.001
	Strongest	8.6 \pm 1.6	8.6 \pm 1.9	r=0.47; P=0.001	8.4 \pm 1.7	8.3 \pm 2.0	r=0.64; P<0.001

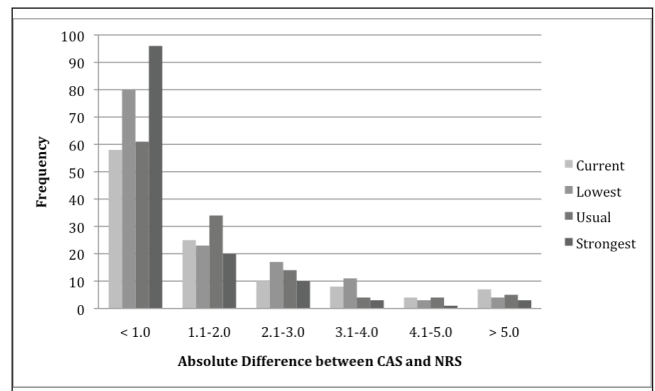
*Duration = 30.4 \pm 23.2 months; †Duration = 14.2 \pm 17.0 months. CAS Coloured analogue scale; NRS Numerical rating scale

for each of the four pain levels (Figure 1). Mean pain ratings on the CAS and NRS ratings were most similar at the strongest level (0.98 \pm 1.39), followed by usual (1.52 \pm 1.48), lowest (1.61 \pm 1.88) and current pain levels (1.75 \pm 1.93). Whether larger differences between scores were related to age was also evaluated. Pearson correlation coefficients for age and absolute difference values were significant or approached significance for current (r=-0.26; P=0.005), lowest (r=-0.19, P=0.03) and strongest (r=-0.17; P=0.04) pain ratings, suggesting that younger children tended to have higher discrepancies between NRS and CAS ratings. Independent-samples *t* test comparing absolute difference values showed a significant difference for the mean absolute difference for NRS-CAS current ratings only, such that younger children (defined as eight to 12 years of age; n=41) had higher absolute difference values (2.24 \pm 2.12) than older children (defined as 16 to 18 years of age; n=45) (1.24 \pm 1.73) (*t*=2.28; P<0.05).

The Bland-Altman method was used to assess agreement between NRS and CAS scores for usual, lowest and strongest pain (Table 5). Current pain ratings were excluded because children's pain may vary naturally over the approximately 1.5 h assessment period. The calculated limits of agreement (95% and 80% CIs) surpassed the a priori limits of acceptability for all measured parameters, indicating that the scales are not interchangeable.

Pain diagnosis and duration

NRS and CAS mean pain ratings did not vary between the two largest diagnostic groups within the sample – musculoskeletal pain and neuropathic pain (all *t*<-0.63; P>0.05). Mean values and correlations for the two diagnostic groups are outlined in Table 6. NRS-CAS correlations were not significantly different across diagnostic groups (all *Z*<-0.43; P>0.05, two-tailed). To determine whether correlations between NRS and CAS varied according to pain duration, patients with <6 months' pain duration (n=32, mean age 13.1 \pm 2.45 years, 71.9% female) were compared with those with >3 years' pain (n=43, mean age 14.8 \pm 2.03 years, 63% female). Groups significantly differed with regard to age (*t*=3.21; P=0.002), and NRS-CAS correlations differed only for children's lowest pain ratings (<6 months [r=0.82] versus >3 years [r=0.38]; *Z*=2.73; P=0.006).

**Figure 1** Distribution of absolute differences for current, lowest, usual and strongest coloured analogue scale (CAS) and numerical rating scale (NRS) pain ratings

DISCUSSION

Although the NRS is used widely in clinical practice to assess pain intensity in children with chronic pain and is included as an outcome measure in studies of children's chronic pain (22-24,33,34), the present study is the first to examine aspects of its validity as it is routinely used within a clinical setting in this population. Correlations between the NRS and CAS met a priori limits of acceptability for convergent validity (r's across all pain levels >0.50) and correlations did not vary according to sex or children's pain diagnosis (musculoskeletal versus neuropathic pain). With respect to age differences, younger children exhibited greater discrepancies between their NRS and CAS ratings. This finding is consistent with the notion that young children, although able to count, may have yet to develop a full understanding of the quantitative significance of numbers (35,36). Thus, lack of consistency between NRS and CAS ratings among younger children may be related to developmental limitations rather than a limitation of the NRS. The NRS possesses face validity for use with children with chronic pain, such that children's ratings for strongest pain were greater than those for usual pain, which were greater than lowest pain ratings. The NRS also demonstrated discriminant validity when compared with a pain affect measure.

Only speculative explanations can be provided for the finding that, for lowest pain ratings, children with shorter pain durations (<6 months) have greater correspondence between NRS and CAS ratings than children with longer pain durations (≥ 3 years). It is possible that children with shorter pain durations are at the beginning of their pain trajectories and have just completed or are still in the process of undergoing multiple medical investigations to rule out a 'cause' for their pain. Thus, these children may be highly attentive to, and repeatedly reminded of, their pain intensity given frequent medical appointments and inquiries from worried parents, leading to greater consistency in their pain ratings. In contrast, our clinical experience suggests that children who have endured pain for >3 years report being 'accustomed' to their discomfort and, thus, may be less attached to a specific number. Indeed, in our study, children with longer pain durations generally demonstrated lower NRS-CAS correlations across all four pain levels compared with children with shorter pain durations.

Although correlations in the present study met a priori criteria for validity ($r=0.58$ to $r=0.68$), as expected, correlations were lower than those obtained in studies of the NRS for pediatric acute pain, which ranged from $r=0.78$ to $r=0.93$ (15,17,18). The lower correlations obtained in the present study may reflect aspects of our methodology rather than limitations of the NRS. Our study methodology differed from previous NRS studies in several significant ways. In the current study, children's pain ratings were obtained as part of routine clinic practice. Thus, children were not informed ahead of time that we would evaluate their pain ratings against another measure, information that is usually required as part of the informed consent process for a prospective clinical study. Consequently, children would not have focused on the rating task itself and perhaps provided more similar pain ratings on the measures. Because children with chronic pain report variations in pain intensity depending on their activity, time of day and medication use, we assessed four different levels of pain. Previous NRS validation studies have focused on assessing a particular level or episode of acute pain, such as the strongest pain experienced in the previous three months, pain on the second day postoperatively or pain on admission to the emergency department (15-19). In the current study, the NRS and CAS were administered by different individuals (members of the clinic team and psychosocial test administrator) who were blinded to children's pain ratings on the other measure, unlike previous studies in which the same individual administers two to three pain scales to determine whether they are correlated (15-19). Furthermore, the NRS and CAS were administered separately with a 45 min to 90 min period between ratings, rather than consecutively as in other studies. Each of these factors minimized biases that can artificially inflate correlation coefficients used to assess validity (30).

Despite these factors, in the current study, children provided fairly similar numerical scores when using a 0 to 10 NRS to those obtained on the 0 to 10 CAS. Ratings were most similar for children's 'strongest' pain, for which the average difference between NRS and CAS ratings was <1 point. In contrast, ratings were least similar for 'current' pain, for which children's ratings differed by an average of slightly less than two points. The explanation for these findings likely rests in anchor effects along with the nature of chronic pain. When asked to rate their strongest pain level, greater consistency in ratings may have occurred because children were drawn toward the end points of the scales, where there is meaning and salience (on the NRS 10 = 'the worst pain you can imagine'; on the CAS 10 = 'most pain'). Indeed, children provided high median ratings for their strongest pain on both the NRS and CAS of 9.00 and 8.75 of 10, respectively. Children were less consistent in rating their current pain, which may be expected to naturally vary during the 45 min to 90 min that separated their pain ratings.

While children's pain ratings on the NRS and CAS were generally similar, the scales were not interchangeable or equivalent. The literature is largely inconsistent on equivalency among self-report pain scales for assessing children's acute pain. While some authors report a lack of agreement across scales (16,37), others find them to be equivalent (17,38). von Baeyer (32) recently argued cogently for adoption of

more liberal criteria for determining agreement between self report of children's pain – advocating for use of the 80% CI versus the 95% CI. On application of the 80% CI to our data, NRS and CAS scores attained near-acceptable agreement (± 2 points) for strongest pain ratings (-2.0, 2.1). Nevertheless, given the general lack of agreement between the NRS and the CAS in the current study at even the more liberal 80% CI, it is advisable for those using the NRS as an outcome measure in pediatric chronic pain research to also use the NRS to evaluate any pre-post treatment effects. Additionally, if a study using the NRS is to be replicated, the same scale should be used to compare results.

The NRS has been recommended in the consensus Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (Ped-IMPACT) statement (39) as a core outcome measure for clinical trials on pain in adults due to its versatility, ease of administration, and concurrent validity with VASs (40,41) – although no single self-report scale has consistently shown greater responsiveness than other scales in detecting improvement with pain treatment (12,42,43). We chose to evaluate the NRS in clinical practice, recognizing that there may be more variability in how the NRS was administered and that there would be a longer time interval between administrations of the two pain scales. Thus, it is possible that the lower correlation coefficients obtained in the present study compared with values obtained in studies evaluating the NRS for acute pain in children (15,17,18) reflect these methodological differences rather than a limitation of the NRS as a pain scale for children with chronic pain. Data from clinical studies indicate that the NRS is responsive to treatment changes for children with chronic pain (44) and that children may prefer the NRS to a VAS (18).

Future studies

Future studies on the NRS for children with chronic pain should formally evaluate its sensitivity to detecting change in pain after treatment. While the NRS is often used as a core outcome measure that reflects a child's reduced pain in response to treatment (45), no study in pediatric chronic pain has yet formally documented its responsiveness. Additionally, studies should evaluate whether composite measures (ie, global index of pain versus four levels of pain) may provide more sensitivity to detecting pain changes in clinical settings, as has been shown for adults (46).

Future studies should also examine potential biases associated with the different study designs used to validate pediatric pain measures. Almost all validation studies of pediatric pain scales have used within-subject designs to determine the concurrent validity of the pain scale by comparing children's pain scores on the proposed pain scale with their ratings on a validated scale (2,6,47). The same individual typically administers all measures and knows children's ratings on each scale. Children are informed about the purpose of the study and then may use a pain scale differently when the emphasis is on the pain scale (and reliability in using them) versus using a scale to answer a clinician's question about their pain level. While concurrent administration within a standardized research protocol should lead to the highest possible correlation coefficients, it is important to understand the inherent biases associated with different study designs. Finally, future studies may wish to explore the research and clinical implication regarding the present study's finding that almost one-half of the sample provided a fraction for their pain rating.

Key recommendations

Our findings indicate that the NRS should be considered for measuring chronic pain in children ≥ 8 years of age, which corroborates existing age-related guidelines (1). When used clinically, we recommend that health care providers assess children's general ability to use the NRS by asking them to rate a few types of pain (typical injuries that most children experience and that vary in intensity) and that they use consistent language in administering the NRS (ie, 'how much, how big, how strong') and the same end point descriptors such as 'no pain' to 'strongest possible pain'. For a brief description of an

assessment to determine children's ability to use the NRS, see McGrath and Gillespie (47). Although the NRS and CAS ratings were correlated, our results indicate that these scores are not interchangeable. Thus, if the NRS is used to obtain pain ratings at a child's initial chronic pain appointment, the NRS should also be used to track the child's pain intensity over time.

Limitations

Our study has certain limitations that must be acknowledged. NRS pain ratings were not obtained for all participants at all four pain levels. However, CAS ratings were consistently obtained for all four pain levels. While the magnitude of NRS-CAS correlations did not differ between children who provided NRS ratings at all four levels and children who provided NRS ratings at <4 levels, it would have been helpful to record reasons why children did not provide a rating to determine the extent that child characteristics (eg, child distress level or developmental limitations), or situational characteristics (eg, limited time during the clinical interview) could account for this variability. In addition, we did not consecutively administer the two pain scales to children. While this may have resulted in some variation in pain values due to the time period between ratings rather than due to a difference between the scales, we believe that this methodological choice improves the generalizability of our results to clinical practice. Finally, the NRS ratings obtained during the interdisciplinary team appointment were not elicited using a strictly standardized approach

but instead were obtained as part of our regular clinical assessment, which may have resulted in lower correlations between the NRS and CAS. Again, this may improve generalizability of our findings to clinical practice.

CONCLUSIONS

The present study provides evidence of convergent and discriminant validity for the 0 to 10 NRS among children with chronic pain eight to 17 years of age in the clinical setting of an interdisciplinary pain clinic. Children's pain ratings did not vary according to sex or pain diagnosis, and younger children showed the greatest discrepancies between NRS and CAS ratings. While children's ratings on the NRS and CAS were similar, the scales were not interchangeable. Thus, if the NRS is used to obtain pain ratings at a child's initial chronic pain appointment, then the NRS should also be used to track the child's pain intensity over time.

FUNDING: This research was funded by a Canadian Institutes of Health Research (CIHR) New Emerging Team Grant (GHL – 63209), which was awarded to Dr Patricia McGrath. No conflicts of interest are declared by study authors.

INSTITUTION: This research was performed at The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8.

REFERENCES

1. von Baeyer CL. Children's self-reports of pain intensity: Scale selection, limitations and interpretation. *Pain Res Manag* 2006;11:157-62.
2. Stinson JN, Kavanagh T, Yamada J, Gill N, Stevens B. Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. *Pain* 2006;125:143-57.
3. von Baeyer CL, Spagrud LJ. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. *Pain* 2007;127:140-50.
4. Champion GD, Goodenough B, von Baeyer CL, Thomas W. Measurement of pain by self-report. In: Finley GA, McGrath PJ, eds. *Measurement Of Pain In Infants And Children*. Seattle: IASP Press, 1998:123-60.
5. Cohen LL, Lemanek K, Blount RL, et al. Evidence-based assessment of pediatric pain. *J Pediatr Psychol* 2008;33:939-55.
6. Ruskin DA, Amaria K, Warnock F, McGrath PA. Pain assessment: Infants, children and adolescents. In: Turk DC, Melzack R, eds. *Handbook of Pain Assessment*, 3rd edn. New York: The Guilford Press, 2011:213-41.
7. Tomlinson D, von Baeyer CL, Stinson JN, Sung L. A systematic review of faces scales for the self-report of pain intensity in children. *Pediatrics* 2010;126:e1168-98.
8. Jacob E, Puntillo KA. A survey of nursing practice in the assessment and management of pain in children. *Pediatr Nurs* 1999;25:278-86.
9. von Baeyer CL, Hicks CL. Support for a common metric for pediatric pain intensity scales. *Pain Res Manag* 2000;4:157-60.
10. von Baeyer CL. Children's self-report of pain intensity: What we know, where we are headed. *Pain Res Manag* 2009;14:39-45.
11. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: A comparison of six methods. *Pain* 1986;27:117-26.
12. Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, eds. *Handbook Of Pain Assessment*. New York: The Guilford Press, 2001:15.
13. Jensen MP, Karoly P, O'Riordan EF, Bland F, Burns RS. The subjective experience of acute pain: An assessment of the utility of 10 indices. *Clin J Pain* 1989;5:153-9.
14. Paice JA, Cohen FL. Validity of a verbally administered numeric rating scale to measure cancer pain intensity. *Cancer Nurs* 1997;20:88-93.
15. Miró J, Castarlenas E, Huguet A. Evidence for the use of a numerical rating scale to assess the intensity of pediatric pain. *Eur J Pain* 2009;13:1089-95.
16. Sánchez-Rodríguez E, Miró J, Castarlenas E. A comparison of four self-report scales of pain intensity in 6- to 8-year-old children. *Pain* 2012;153:1715-9.
17. von Baeyer CL, Spagrud LJ, McCormick JC, Choo E, Neville K, Connelly MA. Three new datasets supporting use of the Numerical Rating Scale (NRS-11) for children's self-reports of pain intensity. *Pain* 2009;143:223-7.
18. Bailey B, Daoust R, Doyon-Trottier E, Dauphin-Pierre S, Gravel J. Validation and properties of the verbal numeric scale in children with acute pain. *Pain* 2010;149:216-21.
19. Pagé GM, Katz J, Stinson JN, Isaac L, Martin-Pichora AL, Campbell F. Validation of the numerical rating scale for pain intensity and unpleasantness in pediatric acute postoperative pain: Sensitivity to change over time. *J Pain* 2012;13:359-69.
20. Connelly M, Neville K. Comparative prospective evaluation of the responsiveness of single-item pediatric pain-intensity self-report scales and their uniqueness from negative affect in a hospital setting. *J Pain* 2010;11:1451-60.
21. Merskey H, Bogduk N. *Classification of Chronic Pain: Description of Chronic Pain Syndromes and Definitions of Pain Terms*. Seattle: IASP Press, 2002.
22. Hechler T, Dobe M, Kosfelder J, et al. Effectiveness of a 3-week multimodal inpatient pain treatment for adolescents suffering from chronic pain: Statistical and clinical significance. *Clin J Pain* 2009;25:156.
23. Scharff L, Langan N, Rotter N, et al. Psychological, behavioral, and family characteristics of pediatric patients with chronic pain: A 1-year retrospective study and cluster analysis. *Clin J Pain* 2005;21:432.
24. Palermo TM, Wilson AC, Peters M, Lewandowski A, Somhegyi H. Randomized controlled trial of an Internet-delivered family cognitive-behavioral therapy intervention for children and adolescents with chronic pain. *Pain* 2009;146:205-13.
25. Connelly M. The verbal numeric rating scale in the pediatric emergency department: What do the numbers really mean? *Pain* 2010;149:167-8.
26. McGrath PA, Seifert CE, Speechley KN, Booth JC, Stitt L, Gibson MC. A new analogue scale for assessing children's pain: An initial validation study. *Pain* 1996;64:435-43.
27. Bulloch B, Garcia-Filion P, Notricia D, Bryson M, McConahay T. Reliability of the Color Analog Scale: Repeatability of scores in traumatic and nontraumatic injuries. *Acad Emerg Med* 2009;16:465-9.

28. Bulloch B, Tenenbein M. Validation of 2 pain scales for use in the pediatric emergency department. *Pediatrics* 2002;110:e33.
 29. McGrath PA. *Pain in Children: Nature, Assessment, and Treatment*. New York: Guilford Press, 1990.
 30. Streiner D, Norman G. *Health measurement scales: A practical guide to their development and use*, 4th edn. New York: Oxford University Press, 2008:170-3.
 31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.
 32. von Baeyer CL. Reported lack of agreement between self-report pain scores in children may be due to a too strict criterion for agreement. *Pain* 2012;153:2152-3.
 33. Lewandowski AS, Palermo TM, De la Motte S, Fu R. Temporal daily associations between pain and sleep in adolescents with chronic pain versus healthy adolescents. *Pain* 2010;151:220-5.
 34. Nilsson IM, Drangsholt M, List T. Impact of temporomandibular disorder pain in adolescents: Differences by age and gender. *J Orofac Pain* 2009;23:115-22.
 35. Gaffney A, McGrath P, Dick B. Measuring pain in children: Developmental and instrument issues. In: Schechter N, Berde CB, Yaster M, eds. *Pain in Infants, Children and Adolescents*, 2nd edn. Philadelphia: Lippincott Williams & Wilkins, 2003:128-41.
 36. Gelman R, Meck E. Preschoolers' counting: Principles before skill. *Cognition* 1983;13:343-59.
 37. Bailey B, Bergeron S, Gravel J, Daoust R. Comparison of four pain scales in children with acute abdominal pain in a pediatric emergency department. *Ann Emerg Med* 2007;50:379-83.
 38. Newman CJ, Lolekha R, Limkittikul K, Luangxay K, Chotpitayasunondh T, Chanthavanich P. A comparison of pain scales in Thai children. *Arch Dis Child* 2005;90:269-70.
 39. Turk DC, Dworkin RH, Allen RR, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2003;106:337-45.
 40. Jensen MP. The validity and reliability of pain measures in adults with cancer. *J Pain* 2003;4:2-21.
 41. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113:9-19.
 42. McQuay H. Consensus on outcome measures for chronic pain trials. *Pain* 2005;113:1-2.
 43. McGrath PJ, Walco GA, Turk DC, et al. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. *J Pain* 2008;9:771-83.
 44. Hechler T, Dobe M, Damschen U, et al. The pain provocation technique for adolescents with chronic pain: Preliminary evidence for its effectiveness. *Pain Med* 2010;11:897-910.
 45. Ayling Campos A, Amaria K, Campbell F, McGrath P. Clinical impact and evidence base for physiotherapy in treating childhood chronic pain. *Physiother Can* 2011;63:21-33.
 46. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain* 1999;83:157-62.
 47. McGrath PA, Gillespie JM. Pain assessment in children and adolescents. In: Turk DC, Melzack R, eds. *Handbook of Pain Assessment*. New York: Guilford Press, 2001:97-117.
-