

Establishing intra- and inter-rater agreement of the Face, Legs, Activity, Cry, Consolability scale for evaluating pain in toddlers during immunization

Rebecca J Gomez BSc¹, Nick Barrowman PhD¹, Sonja Elia MAdvNursPrac², Elizabeth Manias RN MPharm MNursStud PhD³,
Jenny Royle MD², Denise Harrison RN PhD^{1,3,4,5}

RJ Gomez, N Barrowman, S Elia, E Manias, J Royle, D Harrison. Establishing intra- and inter-rater agreement of the Face, Legs, Activity, Cry, Consolability scale for evaluating pain in toddlers during immunization. *Pain Res Manag* 2013;18(6):e124-e128.

BACKGROUND: The Face, Legs, Activity, Cry, Consolability (FLACC) scale is a five-item tool that was developed to assess postoperative pain in young children. The tool is frequently used as an outcome measure in studies investigating acute procedural pain in young children; however, there are limited published psychometric data in this context.

OBJECTIVE: To establish inter-rater and intrarater agreement of the FLACC scale in toddlers during immunization.

METHODS: Participants comprised a convenience sample of toddlers recruited from an immunization drop-in service, who were part of a larger pilot randomized controlled trial. Toddlers were video- and audiotaped during immunization procedures. The first rater scored each video twice in random order over a period of three weeks (intrarater agreement), while the second rater scored each video once and was blinded to the first rater's scores (inter-rater agreement). The FLACC scale was scored at four time-points throughout the procedure. Intraclass correlation coefficients were used to assess agreement of the FLACC scale.

RESULTS: Thirty toddlers between 12 and 18 months of age were recruited, and video data were available for 29. Intrarater agreement coefficients were 0.88 at baseline, 0.97 at insertion of first needle, and 0.80 and 0.81 at 15 s and 30 s following the final injection, respectively. Inter-rater coefficients were 0.40 at baseline, 0.95 at insertion of first needle, and 0.81 and 0.78 at 15 s and 30 s following the final injection, respectively.

CONCLUSIONS: The FLACC scale has sufficient agreement in assessing pain in toddlers during immunizations, especially during the most painful periods of the procedure.

Key Words: Immunization; Pain measurement; Procedural pain; Toddler

Invasive, painful procedures, such as intravenous cannulation, blood draws and immunizations, are common and necessary in pediatric health care settings. Vaccine injections for immunization are the most frequent painful procedures performed in childhood (1). These injections cause severe distress in >90% of toddlers (2), prompting research aimed at ascertaining effective pain-management interventions.

Observational pain assessment tools based on behavioural indicators are necessary for measuring pain in this population of preverbal children who are too young to understand self-report scales (3-6). Many behavioural pain assessment tools have been developed to assess pain in preverbal children (3,7), including the Face, Legs, Activity, Cry, Consolability (FLACC) scale (8), the Children's Hospital of Eastern Ontario Pain Scale (9), the COMFORT scale (10), the Procedure Behavioral Rating Scale (11), the Parents' Postoperative Pain Measure (12) and the Toddler-Preschooler Postoperative Pain Scale (13). The

Établir la fiabilité intra-évaluateur et interévaluateur de l'échelle du visage, des jambes, de l'activité, des pleurs et de la consolabilité pour évaluer la douleur chez les tout-petits pendant la vaccination

HISTORIQUE : L'échelle FLACC du visage, des jambes, de l'activité, des pleurs et de la consolabilité est un outil à cinq éléments mis au point pour évaluer la douleur postopératoire chez les jeunes enfants. Elle est souvent utilisée comme mesure d'issue dans les études sur la douleur aiguë pendant les interventions chez les jeunes enfants, mais les données psychométriques publiées sont limitées dans ce contexte.

OBJECTIF : Établir la fiabilité interévaluateur et intra-évaluateur de l'échelle FLACC chez les tout-petits pendant la vaccination.

MÉTHODOLOGIE : Les participants se composaient d'un échantillon de commodité de tout-petits recrutés dans un service de vaccination sans rendez-vous, qui faisaient partie d'un essai pilote aléatoire et contrôlé plus vaste. Les tout-petits ont été enregistrés sur bande vidéo et audio pendant l'administration des vaccins. Le premier évaluateur a examiné chaque vidéo deux fois dans un ordre aléatoire, sur une période de trois semaines (fiabilité intraévaluateur), tandis que le deuxième évaluateur a examiné chaque vidéo une fois sans connaître les résultats du premier évaluateur (fiabilité interévaluateur). L'échelle FLACC a été évaluée à quatre moments de l'intervention. Les coefficients de corrélation intraclass ont permis d'évaluer la fiabilité de l'échelle FLACC.

RÉSULTATS : Trente tout-petits de 12 à 18 mois ont été recrutés, et il y avait des données vidéo sur 29 tout-petits. Les coefficients de fiabilité intra-évaluateur étaient de 0,88 au départ, de 0,97 à l'insertion de la première aiguille, et de 0,80 et 0,81 15 secondes et 30 secondes après l'injection finale, respectivement. Les coefficients de fiabilité interévaluateur étaient de 0,40 au départ, de 0,95 à l'insertion de la première aiguille et de 0,81 et 0,78, 15 secondes et 30 secondes après la dernière injection, respectivement.

CONCLUSIONS : L'échelle FLACC est suffisamment fiable pour évaluer la douleur chez les tout-petits pendant la vaccination, surtout pendant les périodes les plus douloureuses de l'intervention.

FLACC scale has been recommended for use in research in this age group (3) and is one of the tools used to assess pain at the study site (3,14). It is a five-item behavioural scale that measures facial expression, leg movement, activity, cry and consolability in young children (Table 1). Each item is scored on a scale of zero to 2, resulting in a total score ranging from zero to 10. The scale was developed in collaboration with clinicians to provide a tool that is reliable and simple to use in a busy clinical setting, and was originally validated by Merkel et al (8) to measure postoperative pain in children between two months and seven years of age. It has since undergone further psychometric testing and has been shown to be valid, reliable and feasible to use in a variety of settings, including minor noninvasive procedures and ear-nose-throat operations (3), pain from surgery, trauma, cancer or other disease processes (15), pain in critically ill patients (16) and postoperative pain in children with cognitive impairment (17). The

¹Children's Hospital of Eastern Ontario and the Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario; ²Royal Children's Hospital, Melbourne; ³University of Melbourne, Victoria, Australia; ⁴University of Ottawa, Faculty of Health Sciences, Ottawa, Ontario;

⁵Murdoch Children's Research Institute, Melbourne, Australia

Correspondence: Dr Denise Harrison, 401 Smyth Road, Ottawa, Ontario K1H 8L1. Telephone 613-737-7600 ext 4140, fax 613-773-7600, e-mail dharrison@cheo.on.ca

TABLE 1
Categories and scoring of the Face, Legs, Activity, Cry, Consolability scale

Category	Scoring		
	0	1	2
Face	No particular expression or smile	Occasional grimace or frown, withdrawn, disinterested	Frequent to constant quivering chin, clenched jaw
Legs	Normal position or relaxed	Uneasy, restless, tense	Kicking, or legs drawn up
Activity	Lying quietly, normal position, moves easily	Squirming, shifting back and forth, tense	Arched, rigid or jerking
Cry	No cry (awake or asleep)	Moans or whimpers; occasional complaint	Crying steadily, screams or sobs, frequent complaints
Consolability	Content, relaxed	Reassured by occasional touching, hugging or being talked to, distractible	Difficult to console or comfort

Adapted from reference 8

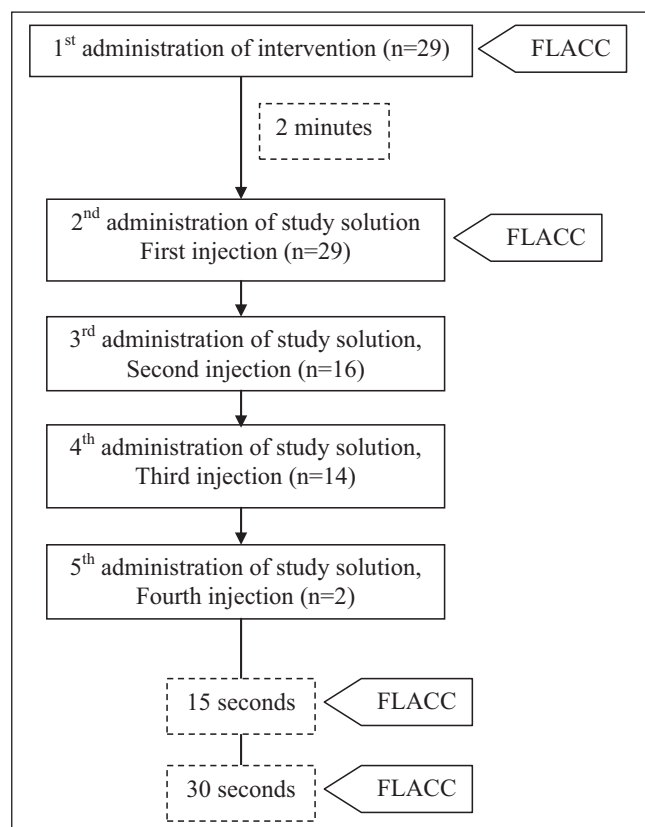


Figure 1 Pain assessment protocol. FLACC Face, Legs, Activity, Cry, Consolability scale

FLACC is one of six scales recommended by the Pediatric Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (Ped-IMPACT) for behavioural assessment in pediatric pain, and is classified as well established for hospital measurement of postoperative pain (18). However, at the time of commencement of the study, we identified no published reports establishing agreement of the FLACC for assessment of acute procedural pain in toddlers, although the tool had been both recommended (3) and widely used as an outcome measure in this context. For instance, the FLACC was used as a primary outcome measure by Vaughan et al (19) in 2005 to evaluate the use of lidocaine gel during bladder catheterization in children younger than two years of age in the emergency department, and also by Babl et al (20) in 2009 to evaluate an intervention to reduce pain and distress during nasogastric tube insertion in children between one and five years of age. Recently, Taddio et al (21) compared inter-rater reliability of three measures of acute pain, including the FLACC; however, the subjects were infants between two and six months of age. To the best of our knowledge, no agreement testing of the FLACC scale in toddlers during acute painful procedures has been published (22).

The objective of the present study was to establish intrarater and inter-rater agreement of the FLACC scale for measuring pain during immunization in toddlers 12 to 18 months of age.

METHODS

The present study is part of a larger study – a pilot randomized controlled trial (RCT) of sucrose compared with placebo (registry # ACTRN12610000355077, <http://www.anzctr.org.au/>) in toddlers 12 to 18 months of age during immunization, which was approved by the Royal Children's Hospital (RCH) Human Research Ethics Committee in Melbourne, Australia (HREC No. 30101 A). Informed consent was obtained from parents/guardians because all study participants were below the legal age of consent. The pilot RCT was performed at the RCH, while the data analysis for the establishment of intrarater and inter-rater agreement was performed at the Children's Hospital of Eastern Ontario (CHEO) Research Institute, Ottawa, Ontario.

Participants and design

Video files (including audio) from the pilot RCT of sucrose versus placebo during immunization were used for establishing agreement of FLACC scores. The participants were 30 toddlers between 12 and 18 months of age who were recruited at the Immunization Service Drop-in Centre at the RCH. All participants attended the centre for scheduled childhood immunization and received one to four injections. Video data were missing for one participant (the video recorder was inadvertently not turned on), leaving 29 video recordings for analysis.

The goal of the larger study – the pilot RCT – was to pilot the methods and inform sample size for the future full-scale RCT; in the present study, the goal was to describe the observed agreement of the FLACC scale within and between raters. The methods of Donner and Eliasziw (23) were used to inform sample size requirements. With two raters, a sample size of as few as 15 may provide sufficient evidence to reject the hypothesis that the reliability is only 0.4 if the population reliability is above 0.8 ('substantial', in the terminology of Landis and Koch [24]). A larger sample size of at least 40 is required to reject the hypothesis that the reliability is only 0.6 when the population reliability is above 0.8. A sample size of 30 was deemed to be sufficient for the present study.

Before viewing the video recordings, training of the raters was conducted by the principal investigator (DH) at CHEO through discussion of the items, observation of sample videos (used for training of research nurses in a Canadian Institutes of Health Research Team in Children's Pain study [Stevens 2006-2011]) and subsequent discussion of scoring techniques. Raters of the FLACC were the first author (RG) and an experienced neonatal and pediatric nurse familiar with using the FLACC. Neither of the raters was present during the initial immunization procedures.

FLACC scores were obtained at four timepoints, representing the key periods for pain during immunization: on first administration of the study solution; on insertion of the first needle; and at 15 s and 30 s following completion of the final injection (Figure 1). Although the original publication describing the FLACC provided no instructions on the observation time required to score the FLACC (8), subsequent

instructions for scoring the FLACC in the postoperative period have specified that the child should be observed for 1 min to 5 min before scoring. Because our goal was to evaluate acute procedural pain, consolability was scored in the 15 s interval following each timepoint. As per the FLACC instructions (Table 1), a score of 0 was scored if the child was relaxed, a score of 1 was assigned if the toddler was able to be consoled by verbal or physical means, or able to be distracted (ie, by bubbles and toys), and a score of 2 was assigned if the toddler was difficult to console, or unable to be consoled or comforted.

Intra- and inter-rater agreement testing

The study was performed in two phases: intrarater agreement (first rater) and inter-rater agreement (first and second raters). Raters could view each video recording repeatedly until they were comfortable with the accuracy of the pain scores.

In the first phase, to establish intrarater agreement, the first rater scored pain in all 29 toddlers twice over a period of three weeks using the FLACC scale. The videos were copied, then scrambled in random order using the =RAND() function in Excel 2003 (Microsoft Corporation, USA) so that the videos were no longer in consecutive order. The second viewing of each video was not undertaken within eight videos or within 24 h of the first viewing. On average, nine to 10 videos were viewed on each day of scoring. This methodology was determined to be sufficient to minimize potential observer bias, along with the fact that other data (ie, duration of the procedure, timing of procedure segments, distraction methods used during procedure, crying time, etc) were being recorded simultaneously and, therefore, recollection of specific FLACC scores was highly unlikely.

For the second phase of the study, the second rater was trained as described above; although the second rater was experienced in the use of the FLACC, the education served as a revision and explanation of the study methods. The second rater scored the same 29 toddler video recordings in the same order as the first rater's first viewing of the videos. The same four timepoints for FLACC measurement were used (Figure 1). The second rater was blinded to the first rater's scores.

Data analysis

All analyses were performed using SPSS version 19 (IBM Corporation, USA). Intraclass correlation coefficients (ICCs) for intrarater and inter-rater agreement were computed using the ICC (2,1) model described by Shrout and Fleiss (25), together with 95% CIs.

In some cases, not all of the five FLACC items were able to be rated. For example, sometimes the video was filmed too close and the legs were out of view, or the toddler was picked up by the parent so that the face was outside of the video frame. Because such events occurred at random, when only one item of five was missing for a given timepoint, the mean value of the other items at that timepoint was imputed to allow a scale score on the full 10 points to be computed (26). If more than one item was missing, the FLACC score for that timepoint was not included in the analysis. Finally, to determine which of the five items most and least strongly influenced the total score, corrected item-total correlations were calculated.

RESULTS

FLACC scores were obtained from video recordings of 29 toddlers receiving scheduled childhood immunizations. Two of 29 videos had no baseline data scores because the video camera was not turned on until the first injection. The number of imputed data points ranged from nine to 12 at any timepoint. In the majority of cases, it was the 'legs' item that was missing and, therefore, imputed. In eight cases, the 'face' item was imputed; four of these occurred at the 15 s time point.

The mean (\pm SD) age of the group was 15.92 ± 3.02 months, comprising 16 males (55%) and 13 females (45%). Vaccinations administered were: varicella zoster (Varilrix; GlaxoSmithKline [GSK], Belgium); measles, mumps and rubella (Priorix; GSK, Belgium); meningitis C vaccine (NeisVac-C; Baxter, USA); influenza B vaccine (Hiberix; GSK, Belgium); combined diphtheria-tetanus-acellular

TABLE 2
Demographic characteristics

Characteristic	
Age, months, mean \pm SD	15.9 \pm 3.0
Sex	
Female	13 (45)
Male	16 (55)
Number of injections	
1	13 (45)
2	2 (7)
3	12 (41)
4	2 (7)
Duration of procedure, min, mean \pm SD	3.5 \pm 0.67
Vaccinations administered	
Varilrix (Glaxo Smith Kline, Belgium)	15 (20)
Priorix (Glaxo Smith Kline, Belgium)	15 (20)
NeisVac-C (Baxter, USA)	14 (18)
Hiberix (Glaxo Smith Kline, Belgium)	14 (18)
Infanrix IPV (Glaxo Smith Kline, Belgium)	14 (18)
Vaxigrip (Sanofi Pasteur, France)	3 (4)
Prevnam 7 (Pfizer, USA)	1 (2)

Data presented as n (%) unless otherwise indicated. Composition of vaccines: Varicella zoster, chicken pox vaccine; NeisVac-C, meningitis C vaccine; Hiberix, influenza type B vaccine; Infanrix IPV, combined diphtheria-tetanus-acellular pertussis and inactivated poliovirus vaccine; Vaxigrip, inactivated influenza vaccine trivalent types A and B (split virion); Prevnar Streptococcus pneumoniae vaccine; Priorix, measles, mumps and rubella vaccine; Varilrix, varicella zoster vaccine

pertussis and inactivated poliovirus vaccine (Infanrix IPV; GSK, Belgium); inactivated influenza vaccine trivalent types A and B (split virion) (Vaxigrip; Sanofi Pasteur, France); and *Streptococcus pneumoniae* vaccine (Prevnam; Pfizer, USA). Thirteen toddlers received only one injection, two received two injections, 12 received three injections, and two received four injections. The mean duration of the entire procedure, from first administration of the study solution until 30 s after the final injection, was 3.52 ± 0.67 min. Table 2 summarizes demographic information of the participants. All toddlers were held during the procedure, and distraction with toys or blowing bubbles was attempted for all toddlers with the exception of one, who sucked on a pacifier throughout her single injection.

Intra- and inter-rater agreement

ICCs for intrarater agreement were 0.88 for baseline scores, 0.97 for the time of insertion of the first needle and 0.81 for 30 s after the final injection, which indicate 'almost perfect' agreement according to the standards for strength of agreement for kappa coefficients by Landis and Koch (24). The ICC for 15 s after the final injection was 0.80, which indicates substantial agreement between first and second ratings (Table 3) (24).

Inter-rater agreement was 0.40, classified as a fair ICC, for the total scores of the FLACC at the first timepoint (baseline). The ICC for the time of insertion of the first needle was 0.95, followed by 0.81 and 0.78 for 15 s and 30 s after the final injection, respectively, indicating substantial to almost-perfect agreement between raters (Table 3) (24). The highest level of agreement between raters was on needle insertion, with an ICC of 0.95 (Figure 2). At baseline, or time of first administration of intervention, the inter-rater agreement was lowest (Figure 2). Specifically, at baseline, the lowest individual item kappas were 0.12 for activity, 0.29 for face and 0.46 for legs, indicating that those items were inconsistently scored between the two raters at that timepoint. A complete set of item kappas for each data point is presented in Table 4. Kappas were calculated based on actual values only, excluding missing data (Table 4).

Corrected item-total correlation was calculated for rater 1 to determine which of the five items most and least strongly influenced total

TABLE 3
Intra- and inter-rater reliability for the Face, Legs, Activity, Cry, Consolability scale scores across all timepoints

Time point	Intrarater		Inter-rater	
	ICC	95% CI	ICC	95% CI
Baseline (first administration of study solution), n=26	0.88	(0.76–0.95)	0.40	(0.10–0.68)
Insertion of first needle, n=29	0.97	(0.94–0.99)	0.95	(0.90–0.98)
15 s following final injection, n=28	0.80	(0.61–0.90)	0.81	(0.60–0.91)
30 s following final injection, n=28	0.81	(0.62–0.91)	0.78	(0.58–0.90)

ICC Intraclass correlation coefficient

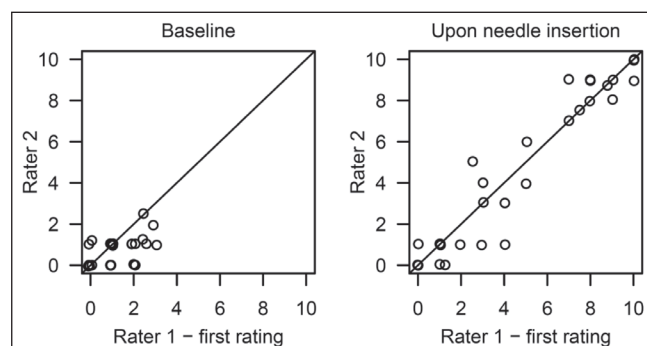


Figure 2 Scatter plots showing inter-rater agreement of Face, Legs, Activity, Cry, Consolability scale scores at baseline (first administration of intervention) and on needle insertion. A one-to-one line is also displayed in each panel to show the ideal agreement between raters. Perfect agreement is when the data points are directly on the line

score. During insertion of the first needle and at 15 s and 30 s after the final injection, the legs item showed the lowest corrected item-total correlation (0.667, 0.640, 0.899) and, therefore, did not contribute to the total score as much as the other items. The consolability item achieved the highest corrected item-total correlation at insertion of the first needle (0.903), while at 15 s after the final injection it was the cry item (0.957), and at 30 s after the final injection it was shared by the cry and consolability items (0.988). Baseline corrected item-total correlations were all below 0.3 except for the legs item (0.341), which indicates poor correlation to the total score, and is not surprising due to the poor agreement between raters at this timepoint.

DISCUSSION

Our study indicates that the FLACC scale has acceptable intra- and inter-rater agreement for use with toddlers during immunization procedures, with the highest agreement between raters occurring at higher FLACC scores. At the time of commencement of the study, there was no published evidence of reliability or validity testing of the FLACC scale during acute minor painful procedures (22), despite being recommended for use in this context (3). Since the completion of the present study, Taddio et al (21) have reported on the reliability, validity and practicality of three measures of acute pain (the FLACC, the Modified Behavioural Pain Scale [MBPS], and the Neonatal Infant Pain Scale [NIPS]) in infants. Five raters scored pain in 120 infants two to six months of age during immunization from video recordings. Similar to our results, intrarater agreement was very high for the FLACC, as well as the other two scales. Inter-rater ICC scores for the FLACC were lowest at baseline (0.85) compared with the MBPS (0.94) and the NIPS (0.90), while they were highest for the FLACC during vaccine injection (0.94) compared with the MBPS (0.90) and the NIPS (0.92) (21). In the current study, both the intra- and inter-rater agreement was demonstrated to be substantial to almost perfect, with the exception of the baseline inter-rater scores, which showed only fair agreement. The highest amount of agreement between raters occurred at the time of needle insertion and injection. Similar to our findings, Taddio et al (21) reported that inter-rater agreement for the FLACC was lower at baseline compared with during the injection. This is not surprising because maximal pain is

TABLE 4
Inter-rater agreement

Time	Face	Legs	Activity	Cry	Consolability
0	0.29 (25)	0.46 (21)	0.12 (26)	* (26)	* (26)
1	0.42 (29)	0.68 (24)	0.61 (29)	0.95 (29)	0.53 (29)
2	0.42 (26)	0.29 (23)	0.33 (28)	0.58 (29)	0.53 (29)
3	0.51 (25)	0.78 (22)	0.78 (26)	0.61 (27)	0.34 (27)

Data presented as kappa (n). *Could not be computed because all ratings were 0

experienced during the injection, which makes it easier to score (27). Intuitively, scoring becomes more challenging when the behaviour changes drastically within a defined interval of time, displaying moderate pain intensities, which occurs as time progresses after the final needle injection. This observation is reflected in the results that show decreasing ICCs from the time of needle insertion (0.95) to 15 s (0.81) and 30 s (0.78) after the final injection.

The lowest agreement between raters was during the first administration of the study solution (baseline). One possible explanation for that finding may be that restricting the range of values scores reduces the correlation coefficient (28). To further clarify the low ICC scores at baseline, after data were collected and analyzed, the raters reviewed the videos that showed fair agreement at baseline and discussed the scoring. Most of the discrepancies in scores were in the face and activity items. The first rater, who was a research student and naive to the FLACC scale before the present study, rated the face and activity items higher than the second rater, who was an experienced pediatric nurse. The second rater admittedly rated many baseline scores as zero (instead of 1) because any reaction from the toddlers when given the study solution was considered to be a normal reaction and not painful. This assumption of no pain based on the context is a common finding in health care workers who score patient pain and is a possible explanation for the lower baseline scores of the second rater compared with the first rater. For example, in an emergency department study of acute pain measurement and in a study rating cancer pain, health care professionals rated pain significantly lower than the patients themselves (29,30). This finding highlights the importance of establishing exact parameters during training for the FLACC scale, and especially when using raters with differing levels of clinical experience. Clear scoring guidelines are necessary to facilitate consistency in pain measurement and communication among health care providers (5). This includes timing of observations, particularly the consolability item. Although no instructions for the period of observation were included in the original publication of the FLACC (28), subsequent instructions for using the FLACC for scoring postoperative pain state that the child should be observed for 1 min to 5 min. This time period is not possible when using the FLACC for pain assessment during short-lasting acute painful procedures. However, ensuring consistency in the period of timing for observing the consoling effect of caregivers is important. Although we used a 15 s interval to observe consolability, other studies that used the FLACC to assess acute procedural pain did not specify the observation time (21), making true comparisons and interpretation of the scores difficult. Another important point is that the FLACC facial actions are not

indicative of acute pain, and differ from the facial actions based on the Facial Action Coding System and the Neonatal Facial Coding System (31,32), which form the basis for nearly all infant and pediatric pain assessment tools. Interpretation in the acute procedural pain context of 'withdrawn, disinterested' when allocating a Face score of one, or 'Frequent to constant quivering chin, clenched jaw' when allocating a score of two, was not discussed by Crellin et al (22) in their analysis of the validation of behavioural pain scales, or by Taddio et al (21) in their evaluation of the reliability and validity of the FLACC for immunization pain.

It is, therefore, not known how these descriptors are changed or interpreted by coders of video recordings and bedside scorers. Implications for further use of the FLACC as an outcome measure in intervention studies as well as further psychometric testing studies would be to clearly explain the methods of FLACC scoring, including the duration of observation before scoring and how the facial actions are scored.

Strengths and limitations

Our study was limited by focusing on evaluating intra- and inter-rater agreement of the FLACC when used by video coders to assess immunization pain and distress in toddlers, and did not include other validation testing or a conceptual critique of the tool. In addition, our sample size was small, although sufficiently large to fulfill the purpose of the present study.

CONCLUSION

Our study demonstrated that the FLACC has acceptable intra- and inter-rater agreement in assessing pain in toddlers 12 to 18 months of age during the acute painful procedure of immunization, especially at the highest pain scores during the most painful part of the procedure. We can, therefore, be confident that the FLACC score has acceptable reliability, based on intra- and inter-rater agreement, to warrant use as an outcome measure in future intervention studies of pain management during short-lasting acute procedural pain in toddlers.

ACKNOWLEDGEMENTS: The authors thank the nurses at the Immunization Service Drop-in Centre, Royal Children's Hospital, Melbourne, Australia for their collection of data (video files), and Cynthia Joly, at the Children's Hospital of Eastern Ontario, for data collection (inter-rater scores).

FUNDING: The Children's Hospital of Eastern Ontario Research Institute funded this study through its summer studentship program, and the Nurses Board of Victoria Major Research Grant supported the larger pilot RCT.

REFERENCES

- Schechter NL, Zempsky WT, Cohen LL, McGrath PJ, McMurtry CM, Bright NS. Pain reduction during pediatric immunizations: Evidence-based review and recommendations. *Pediatrics* 2007;119:e1184-98.
- Jacobson RM, Swan A, Adegbenro A, et al. Making vaccines more acceptable – methods to prevent and minimize pain and other common adverse events associated with vaccines. *Vaccine* 2001;19:2418-27.
- 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.
- Merkel S, Malviya S. Pediatric pain, tools, and assessment. *J PeriAnes Nurs* 2000;15:408-14.
- Herr K, Coyne PJ, Key T, et al. Pain assessment in the nonverbal patient: Position statement with clinical practice recommendations. *Pain Manag Nurs* 2006;7:44-52.
- 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.
- Duhn LJ, Medves JM. A systematic integrative review of infant pain assessment tools. *Adv Neonatal Care* 2004;4:126-40.
- Merkel SI, Voepel-Lewis T, Shayevitz JR, Malviya S. Practice applications of research. The FLACC: A behavioral scale for scoring postoperative pain in young children. *Ped Nurs* 1997;23:293-7.
- McGrath P, Johnson G, Goodman J, Shillinger J, Dunn J, Chapman J. CHEOPS: A behavioral scale for rating postoperative pain in children. *Adv Pain Res Ther* 1985;9:395-402.
- Ambuel B, Hamlett K, Marx C, Blumer J. Assessing distress in pediatric intensive care environments: The COMFORT scale. *J Pediatr Psychol* 1992;17:95-109.
- Katz E, Kellerman J, Siegel S. Behavioural distress in children with cancer undergoing medical procedures: Development considerations. *J Pediatr Psychol* 1980;48:356-65.
- Chambers CT, Reid GJ, McGrath PJ, Finley GA. Development and preliminary validation of a postoperative pain measure for parents. *Pain* 1996;68:307-13.
- Tarbell SE, Cohen T, Marsh JL. The Toddler-Preschooler Postoperative Pain Scale: An observational scale measuring postoperative pain in children aged 1-5. *Pain* 1992;50:273-80.
- Royal Children's Hospital. Assessing Pain in Children, Children's Pain Management Service. Melbourne, 2011. <www.rch.org.au/emphlibrary/anaes/Pain_assessment.pdf> (Accessed November 20, 2011)
- Manworren R, Hynan L. Clinical validation of FLACC: Preverbal patient pain scale. *Ped Nurs* 2003;29:140-6.
- Voepel-Lewis T, Zanoliti J, Dammeyer JA, Merkel SI. Reliability and validity of the Face, Legs, Activity, Cry, Consolability behavioral tool in assessing acute pain in critically ill patients. *AJCC* 2010;19:55-62.
- Voepel-Lewis T, Merkel SI, Tait AR, Trzcinka A, Malviya S. The reliability and validity of the face, legs, activity, cry, consolability observational tool as a measure of pain in children with cognitive impairment. *Anesth Analg* 2002;95:1224-9.
- Blount RL, Loiselle KA. Behavioural assessment of pediatric pain. *Pain Res Manag* 2009;14:47-52.
- Vaughan M, Paton EA, Bush A, et al. Does lidocaine gel alleviate the pain of bladder catheterization in young children? A randomized, controlled trial. *Pediatrics* 2005;116:917-20.
- Babl FE, Goldfinch C, Mandrawa C, et al. Does nebulized lidocaine reduce the pain and distress of nasogastric tube insertion in young children? A randomized, double-blind, placebo-controlled trial. *Pediatrics* 2009;123:1548-55.
- Taddio A, Hogan ME, Moyer P, et al. Evaluation of the reliability, validity and practicality of 3 measures of acute pain in infants undergoing immunization injections. *Vaccine* 2011;29:1390-4.
- Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. *Paed Anaes* 2007;17:720-33.
- Donner A, Eliasziw E. Sample size requirements for reliability studies. *Statist Med* 1987;6:441-8.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979;86:420-27.
- Polit DF, Hungler BP. *Nursing Research: Principles and Methods*. 6th edn. Philadelphia: Lippincott, 1999.
- Bijur PE, Silver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. *Acad Emerg Med* 2001;8:1153-7.
- Bland JM, Altman DG. Correlation in restricted ranges of data. *BMJ* 2011;342:d556.
- Guru V, Dubinsky I. The patient vs. caregiver perception of acute pain in the emergency department. *J Emerg Med* 2000;18:7-12.
- Grossman SA, Sheidler VR, Swedeen K, Mucenski J, Piantadosi S. Correlation of patient and caregiver ratings of cancer pain. *J Pain Symptom Manage* 1991;6:53-7.
- Ekman P, Friesen W. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press, 1978.
- Grunau RE, Craig K. Pain expression in neonates: Facial action and cry. *Pain* 1987;28:395-410.

