Development and Validation of a Culturally-Relevant Pain Scale for Kiswahili-Speaking Patients in a Tanzanian Emergency Department

by

Brian J. Meier

Duke Global Health Institute
Duke University

Date: May 24, 2017

Approved:

___________________________
Catherine Staton, Supervisor

___________________________
Nathan Thielman

___________________________
Charles Gerardo

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Duke Global Health Institute of Duke University
2017
ABSTRACT

Development and Validation of a Culturally-Relevant Pain Scale for Kiswahili-Speaking Patients in a Tanzanian Emergency Department

by

Brian J. Meier

Department of Global Health
Duke University

Date: May 24, 2017

Approved:

___________________________
Catherine Staton, Supervisor

___________________________
Nathan Thielman

___________________________
Charles Gerardo

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Duke Global Health Institute of Duke University

2017
Abstract

Background: Acutely painful conditions, responsible for a large proportion of Emergency Department patients around the world, are inadequately assessed and poorly treated. Routine use of scales to quantify pain is recommended to improve analgesic practice. Currently, no such scale has been validated for use in Kiswahili-speaking patients in Tanzania. The objective of this study was to develop and assess a culturally relevant pain intensity scale for use in injury patients at the Kilimanjaro Christian Medical Center in Moshi, Tanzania. Methods: This was a two-part study, with the initial phase using focus groups to develop a pain scale. The second phase used a convenience sample of injury patients to assess the scale for validity and reliability. Analysis of variance, intra-class correlation coefficients, and Bland-Altman Analysis were used to assess validity and reliability. We used focus groups and surveys to develop a pain scale, which was subsequently tested in injury patients. Results: A 100-point numeric pain scale was developed and tested among 98 injury patients. The intra-class correlation coefficient of scores was 0.97 (95% CI 0.96 - 0.98) and Bland-Altman analysis found that 95% of the differences were between -23.5 and +20.7. Conclusions: Our results suggest that a 100-point numeric rating scale is valid and reliable for use Tanzanian injury patients.
Table of Contents

Abstract ................................................................................................................................. iv

List of Tables ....................................................................................................................... vii

List of Figures ..................................................................................................................... viii

Acknowledgements ............................................................................................................ ix

1. Introduction ..................................................................................................................... 1

   1.1 Value of scales to quantify pain ............................................................................. 2

   1.2 Expressions and Measurements of Pain Between Cultures ............................... 3

2.1 Study Design ................................................................................................................. 5

2.2 Setting .......................................................................................................................... 5

2.3 Participants .................................................................................................................... 6

2.4 Procedures ..................................................................................................................... 7

   2.41 Development .......................................................................................................... 7

   2.42 Clinical Testing ....................................................................................................... 9

2.5 Analysis ......................................................................................................................... 10

3. Results ............................................................................................................................. 13

   3.1 Scale Development ................................................................................................. 13

   3.2 Validity/Reliability Testing .................................................................................... 14

   3.3 Validity ...................................................................................................................... 15

   3.4 Reliability .................................................................................................................. 16

4. Discussion ......................................................................................................................... 19
4.1 Limitations ................................................................. 20

5. Conclusion ........................................................................ 23

References ............................................................................ 24
List of Tables

Table 1: Patient Demographics .............................................................................................................. 15

Table 2: Comparison of Pain Intensity Score by Initial Severity ............................................................. 16

Table 3: Intra-class Correlation Coefficients for Pain Scores 1 Minute Apart................................. 17
List of Figures

Figure 1: Most Common Pain Scales .................................................................3

Figure 2: Examples of Pain Scales Tested .........................................................9

Figure 3: Correlation Between Ratings at Time 0 and Time 1 .......................17

Figure 4: Bland-Altman Plot of Paired Ratings ..............................................18
Acknowledgements

We would like to acknowledge our KCMC-Duke Research Team including Amina Mlay, Eliza Ngowi, Anna Mchaki, Evangeline Msanya, Msafiri Pesambili, and Joseph Kilisara. Without them, none of our research would be possible.
1. Introduction

In high-income countries (HICs) pain is the most common reason for presentation to the Emergency Department (ED).\textsuperscript{1-3} Although there is limited evidence regarding cause-specific presentations from low and middle-income countries (LMICs),\textsuperscript{4} pain is likely a leading cause of seeking emergency care in these settings as well. In fact, due to the disproportionate burden from injuries, there is perhaps an even greater number of pain-related presentations in LMICs.\textsuperscript{5,6} Unfortunately, the burden of pain is often not limited to the initial insult, as evidenced by high rates of post-operative pain, chronic pain, and injury-related disability in LMICs.\textsuperscript{7-9}

Despite the burden of pain and pain-related conditions, numerous studies in both HICs and LMICs has shown that pain is often undertreated most clinical settings, including Emergency Departments (EDs).\textsuperscript{2,6,10-15} In fact, according to the World Health Organization (WHO), globally up to 80% of people with moderate to severe pain, don’t have access to appropriate treatment.\textsuperscript{16} As studies have shown that inadequate analgesia leads to poorer surgical outcomes,\textsuperscript{17} increased in-hospital opioid use,\textsuperscript{18} and increased risk of developing of chronic pain,\textsuperscript{19} one can see how the increased incidence of injury combined with a shortage of analgesic treatments results in a double-burden of painful conditions in LMICs. As a result, numerous countries and international organizations,
including the WHO, have called for pain management to be considered a fundamental human right.20-22

1.1 Value of scales to quantify pain

One of the universal recommendations towards improving analgesic practice, is the routine use of scales to quantify pain.23-26 The use of these scales in U.S. EDs has been shown to increase the overall usage of analgesics and decrease to time to administration.15,27 These scales also provide a means of quantifying the effectiveness of a variety of pain-related interventions.28-33

The first scale to be widely used was the visual analog scale (VAS). Although originally validated for use with chronic and experimental pain,34 it has since been validated for wider use, most notably for ED patients with acute pain.35 The two other most commonly used pain intensity scales (figure 1) are the numeric rating scale (NRS), and the verbal rating scale (VRS), although additional adaptations exist with slight variations in anchors or descriptors of pain.33,36,37
Studies comparing these three scales show that they correlate well and all three are valid and reliable for clinical use.\textsuperscript{36-38} The verbal rating scale appears to be the simplest to administer, but is also the least sensitive given its limited number of responses.\textsuperscript{37} A recent systematic review found that the numeric rating scale is the most often recommended, and appears to be easier to understand and interpret than the visual analog scale.\textsuperscript{36}

### 1.2 Expressions and Measurements of Pain Between Cultures

The translation of these scales into different languages and cultures can be challenging as the expression of pain, and the language used to describe pain varies between cultures.\textsuperscript{39} For instance, a study comparing the expression of pain between college students in the United States and India found that the American students were more accepting of outward expressions of pain, and this correlated to a lower pain tolerance compared to Indian students.\textsuperscript{40} The difference in how different cultures express
and respond to pain can have dramatic effects on pain management. A survey of ED nurses in Rwanda, for example, found that many nurses believed taking pain medication was a sign of weakness and delayed healing.\textsuperscript{41}

Equally important is the language used to describe pain, as the verbal descriptors and orientation of pain scales can influence their interpretation. For instance, a Chinese study found better patient compliance with a vertical, rather than horizontal, pain scale.\textsuperscript{42} Previous work in Tanzania has used a translated verbal rating pain scale which categorizes pain based on how it affects a patient’s ability to sleep, these results are then converted into a number in order to quantify measurements.\textsuperscript{43}

It is critical therefore, to consider these differences in cultural expression and language when developing and utilizing pain measurement scales. To date, most work on translation and validation of pain scales has focused on longer, often condition-specific tools such as the McGill Pain Questionnaire,\textsuperscript{44-46} the Brief Pain Inventory,\textsuperscript{47-49} or the Quebec Back Pain Disability Scale.\textsuperscript{50-52} Although the VAS, VRS, or NRS are frequently used in non-English settings,\textsuperscript{53-55} including Tanzania,\textsuperscript{43,56,57} these scales have not previously been validated for use in these settings. The aim of this study, therefore, was to develop and validate a culturally relevant pain intensity scale for use in Kiswahili-speaking injury patients in Tanzania.
2. Methods

2.1 Study Design

This was a two-part study designed to develop and assess a culturally relevant pain intensity scale for use in injury patients at the Kilimanjaro Christian Medical Center (KCMC). The first part of the study utilized focus group discussions and surveys in order to develop the scale including consensus verbal descriptors of pain, scale anchors, and pilot the use of potential scale designs. This phase also assessed the validity of the proposed scales using theoretical examples of painful conditions. The second phase of the study was a longitudinal study assessing the reliability and validity of the scale in injury patients in the KCMC ED. All study procedures were approved by the Institutional Review Boards of Kilimanjaro Christian Medical Center, Moshi, Tanzania, the Tanzanian National Institute of Medical Research, and Duke University.

2.2 Setting

This study took place at Kilimanjaro Christian Medical Center in Moshi, Tanzania. KCMC is one of four Zonal Consultant hospitals in Tanzania, serving as a referral center for over 15 million people in Northern Tanzania. The surrounding Moshi area is divided into urban Moshi town (population ~180,000) and Moshi rural (population ~466,000). The KCMC Emergency Department, sees upwards of 20,000 patients per year. The majority people in Moshi are members of the Chagga
ethnic group, with smaller proportions of Pare and Maasai. Kiswahili is the most common language, followed by English and Kichagga.69.

2.3 Participants

Participants consisted of patients presenting to the KCMC Emergency Department with the complaint of pain related to traumatic injury (musculoskeletal injuries, burns, lacerations, etc.). For the focus group portion of the study, there was one focus group consisting only of Emergency Department physicians and nurses (n = 10). The remaining focus groups consisted of patients who had recently presented to the Emergency Department with pain related to injury (n = 38). Kiswahili-speaking patients who were 18 years of age or older, presenting to the ED with pain related to traumatic injury, were medically stable, and able to provide consent were considered for participation in the second phase of the study. Patients were excluded if they were less than 18 years of age, known or suspected to be intoxicated, had an emergent medical condition necessitating treatment, non-Kiswahili speaking, or refused/unable to provide informed consent. A study investigator was present between the hours of 8am and 8pm, 6 days a week in order to recruit participants for participation in either the focus group or clinical portion of the study.
2.4 Procedures

2.41 Development

Patients or providers who met the inclusion and exclusion criteria were approached to participate in the study. For focus group discussions, additional consideration was given to age and gender in order to select a balanced sample of participants. Potential participants were provided with informed consent in a written and/or verbal form in Kiswahili. English translations of the informed consent were also available at patient request. After informed consent was administered, a copy of the signed consent was given to the participants.

Once participants were enrolled, the study investigator arranged to lead the focus group discussions at a time deemed convenient to each of the participants. Focus groups consisted of between 5-10 participants, with one focus group consisting of only providers. Discussions were led by trained, bilingual research nurses with previous experience in qualitative research methods. Focus groups lasted approximately 60 minutes, and participants were offered snacks and drinks, as well as compensated for their travel according to acceptable local standards ($2.50).

Focus group discussions were designed to develop an appropriate scale focused on topics relevant to understanding how pain is expressed and described in Tanzania.
This included verbal and nonverbal expressions of pain, descriptions of different types of pain, words used to categorize mild, moderate, and severe pain, as well as how this might differ based on age or gender. Participants were also asked to provide examples of painful conditions of varying intensity. Following this discussion, participants were asked to categorize several example conditions (e.g. severe burn, bruise, broken leg) by level of intensity (mild, moderate, or severe). Participants were then asked to provide responses to the same example conditions on several different pain intensity scales. Lastly, they were asked to rate these scales based on their preference and ease of use.

The proposed scales varied by inclusion of numbers, words, range (0-10 vs. 0-100), etc. Examples of the different types of scales used are shown in Figure 2. Discussions were audio-taped, transcribed, and translated in order to be analyzed. Survey data from focus group discussions were entered into an online database (REDCap) hosted Duke University. REDCap is a secure, web-based application designed to support data capture for research studies. Entered data was reviewed for quality by the project principle investigator (BM). The focus group results were then analyzed as described below, culminating in a consensus pain intensity scale to be used in the clinical portion of the study (Figure 2).
2.42 Clinical Testing

For the clinical portion of the study a convenience sampling method was used. After obtaining informed consent, study nurses administered a questionnaire to the patients, including questions regarding demographics, description of injury, treatments received, etc. The nurses then asked the patients to rate their pain intensity on the proposed scale (0-100, numeric rating scale). 1 to 2 minutes later, a second research nurse asked the patient to again rate their pain. 1 to 2 minutes was chosen based on previous pain scale validation studies as it is assumed that pain intensity would not markedly change during this time, nor would it allow for interval treatments to be administered. This process was repeated at 30 minutes, and a final measurement
was taken at the time of patient disposition from the ED (i.e. admit or discharge) for a total of 5 measurements. The first measurement taken at each time period is referred to as time 0 and the second referred to as time 1. Patients were asked to provide a categorical ranking of their pain at the first measurement. Additional information including ED treatment, overall satisfaction with analgesic treatment, diagnoses, etc. were also recorded at the time of disposition. Data from the questionnaires was entered into REDCap and reviewed for quality by the principal investigator (BM).

2.5 Analysis

As no objective “gold standard” exists for measuring pain, for both phases of the study, the standard comparator for pain was considered to be the patient’s ordinal, categorical description of their pain (i.e. mild, moderate, severe). This has been used in previous studies with the assumption that if a pain scale is valid, it should differentiate between these categories of pain, and change as a patient’s reporting pain increases or decreases. To assess this for the initial phase of the study, focus group participant responses to example conditions were recorded for each of the tested scales and compared to their individual categorical ranking of the condition. The mean scores were compared to the categorical ranking using an Analysis of Variance (ANOVA). Regressions analyses were performed to determine if responses differed based on age or gender. Participants preferences for a particular scale were also aggregated and compared directly.
A similar analysis was complete to assess validity in the clinical portion of the study. Patient’s numerical responses were compared to their individual categorical ranking of their pain and compares using an ANOVA. Again, our reasoning was that if the scale was valid, the mean scores of the numeric responses would be significantly different for each category. Summary statistics were used to describe presentation data and presented as mean ± SD, median, and percentages as appropriate. Variation in pain scores or severity rating based on age, gender, mechanism of injury, pain location, treatment prior to arrival, and time since injury was assessed with multiple regression analyses.

In order to assess reliability, scores taken at times 0 and 1 were correlated. Correlation was measured using Intraclass Correlation Coefficients (ICC) and a Bland-Altman graphical analysis. These have been recommended for use in reliability studies over other measures such as Pearson’s r due to the fact they are a better measure of observer variability as they require identical, rather than perfectly correlated scores. Scales should have an ICC of at least 0.6 to be considered “useful”, while values above 0.75 are considered “excellent.” ICC values may be falsely elevated depending on the range of possible scores. To account for this, and remain consistent with previous
validation studies,\textsuperscript{35,62,63} we have included additional ICCs for scores above and below the median possible score (50).

To assess the underlying basis of any variability between paired measurements. We separately analyzed all paired measurements which differed by at least 20. Although we anticipated that scores should not significantly differ at times 2 minutes apart, there exists the possibility for dynamic changes in a patient’s pain even in this short interval. Previous validation studies have reported that the expected variability between repeated scores can be up to 20\%.\textsuperscript{35,63,70} This was set, therefore, as our acceptable margin of variability between measurements.

Bland-Altman analysis provides an additional way to assess agreement between ratings. This is accomplished by plotting the difference between two paired measurements against the average of the two measurements. The resultant values are plotted, with a line indicating perfect agreement (i.e. zero difference), allowing a visual assessment of variation related to the size of the mean. Confidence intervals (i.e. limits of agreement) were also calculated based on the method recommended for non-normally distributed data.\textsuperscript{67,68}. 
3. Results

3.1 Scale Development

A total of 5 focus groups were conducted for a total of 48 participants, 10 of which were providers. The average age of participants was 33.7 years (range 18 - 77) and 20 (41.7%) were females.

There was unanimous agreement in all focus groups that the Kiswahili word “maumivu” was the best translation for the concept of pain. Similarly, there was near unanimous agreement the words or phrases used for severity category translation. These were “maumivu kidogo” - mild pain, “maumivu ya wastani” - moderate pain, and “maumivu makali sana” - severe pain. These words were therefore chosen as the categorical anchors for subsequent testing.

Overall, 7 different pain intensity scales were tested in focus groups. Scales without numbers (i.e. VAS) performed poorly, as nearly all ratings were 0mm or 100mm with few ratings in between. Subsequent ANOVA analysis showed no significant difference in pain ratings between categorical severity groups in these scales.

Conversely, all scales with numbers (with or without tick marks, verbal anchors, etc.) performed well. ANOVA of all numeric scales showed significant differences of the mean ratings between conditions in each pain category. The majority of patients
preferred a scale ranging from 0 - 100, rather than 0 - 10, and nearly all patients preferred the use of tick marks, rather than a flat line. Patients were relatively split, however, in their preference for whether or not to include verbal anchors on the scale. Based upon these results a 0 - 100 scale with tick marks and verbal anchors was chosen as the final scale to undergo clinical reliability and validity testing.

### 3.2 Validity/Reliability Testing

A total of 99 patients were enrolled for a total of 134 paired measurements (68 for time 0,1 and 66 for time 30,31). The mean age of the participants was 40.8 (range 18-85) and 54 (54%) patients were male. The majority of patients were injured in road-traffic crashes (53.5%), followed by falls (27.3%). Lower extremity pain was the most common (39.4%), with head/neck (20.2%), and upper extremities (19.2%) the other major contributors. Patients had a median time of 8 hours between injury and presentation, ranging from 30 minutes to 12 days. The mean(SD) pain score on arrival was 66.4 (29.8), and 77.8% of patients endorsed moderate(30.3%) or severe(47.5%) pain on arrival. Additional demographic information can be found in Table 1.
### Table 1: Patient Demographics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th># of Patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>40.81 (18-85)</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>54 (54.6%)</td>
</tr>
<tr>
<td>Mechanism of Injury</td>
<td></td>
</tr>
<tr>
<td>- RTC</td>
<td>53 (53.5%)</td>
</tr>
<tr>
<td>- Assault</td>
<td>9 (9.1%)</td>
</tr>
<tr>
<td>- Burn</td>
<td>1 (1.0%)</td>
</tr>
<tr>
<td>- Fall</td>
<td>27 (27.3%)</td>
</tr>
<tr>
<td>- Other</td>
<td>2 (2.0%)</td>
</tr>
<tr>
<td>- Not recorded</td>
<td>7 (7.1%)</td>
</tr>
<tr>
<td>Location of Pain</td>
<td></td>
</tr>
<tr>
<td>- Head/neck</td>
<td>20 (20.2%)</td>
</tr>
<tr>
<td>- Chest</td>
<td>6 (6.1%)</td>
</tr>
<tr>
<td>- Upper Extremity</td>
<td>19 (19.2%)</td>
</tr>
<tr>
<td>- Abdomen</td>
<td>2 (2.0%)</td>
</tr>
<tr>
<td>- Pelvis</td>
<td>11 (11.1%)</td>
</tr>
<tr>
<td>- Lower Extremity</td>
<td>39 (39.4%)</td>
</tr>
<tr>
<td>- Other</td>
<td>2 (2.0%)</td>
</tr>
<tr>
<td>Time since injury</td>
<td></td>
</tr>
<tr>
<td>(median/range)</td>
<td>Median (range)</td>
</tr>
<tr>
<td></td>
<td>8h (30m, 12 days)</td>
</tr>
<tr>
<td>Any treatment prior to</td>
<td>47 (47.5%)</td>
</tr>
<tr>
<td>arrival</td>
<td></td>
</tr>
<tr>
<td>Initial pain score</td>
<td>66.4 (29.8)</td>
</tr>
<tr>
<td>(mean/SD)</td>
<td></td>
</tr>
<tr>
<td>Initial severity of pain</td>
<td></td>
</tr>
<tr>
<td>- No Pain</td>
<td>2 (2.0%)</td>
</tr>
<tr>
<td>- Mild</td>
<td>20 (20.2%)</td>
</tr>
<tr>
<td>- Moderate</td>
<td>30 (30.3%)</td>
</tr>
<tr>
<td>- Severe</td>
<td>47 (47.5%)</td>
</tr>
</tbody>
</table>

### 3.3 Validity

The mean (SD) initial numeric rating was 66.4 (29.8). As shown in Table 2, the mean rating increased as verbal categorical description of pain increased. ANOVA
analysis showed a significant difference between the mean rating for each category (F = 135.1, p < 0.001). There were no significant differences in categorical ranking or initial pain rating by age, gender, mechanism of injury, pain location, treatment prior to arrival, or time since injury.

Table 2: Comparison of Pain Intensity Score by Initial Severity

<table>
<thead>
<tr>
<th>Severity</th>
<th>#</th>
<th>Mean (SD)</th>
<th>Median</th>
<th>Range</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Pain</td>
<td>2</td>
<td>0 (0)</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mild</td>
<td>20</td>
<td>30 (19.74)</td>
<td>30</td>
<td>10 - 90</td>
<td>20.76 - 39.24</td>
</tr>
<tr>
<td>Moderate</td>
<td>30</td>
<td>54.67 (12.24)</td>
<td>50</td>
<td>40 - 80</td>
<td>50.10 - 59.24</td>
</tr>
<tr>
<td>Severe</td>
<td>47</td>
<td>92.13 (10.20)</td>
<td>100</td>
<td>70 - 100</td>
<td>89.13 - 95.12</td>
</tr>
</tbody>
</table>

3.4 Reliability

Overall the correlation between scores at time 0 and 1 was excellent with an ICC 0.97 (95% CI 0.96 - 0.98). ICCs between ratings at time 0,1 and 30,31 were 0.96 and 0.98, respectively, with confidence intervals suggesting excellent correlation in scores (Table 3).
Table 3: Intra-class Correlation Coefficients for Pain Scores 1 Minute Apart

<table>
<thead>
<tr>
<th>Time</th>
<th>ICC</th>
<th>Score 0 - 50</th>
<th>ICC</th>
<th>Score 51 - 100</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>r</td>
<td>95% CI</td>
<td>n</td>
<td>r</td>
</tr>
<tr>
<td>0, 1</td>
<td>68</td>
<td>0.96</td>
<td>0.93 - 0.97</td>
<td>26</td>
<td>0.88</td>
</tr>
<tr>
<td>30, 31</td>
<td>66</td>
<td>0.98</td>
<td>0.96 - 0.99</td>
<td>46</td>
<td>0.96</td>
</tr>
<tr>
<td>Total</td>
<td>134</td>
<td>0.97</td>
<td>0.96 - 0.98</td>
<td>72</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Similarly, ICCs in subgroups (above and below 50), although slightly lower, were still above the suggested threshold for “excellent.”

A visual representation of correlation between all scores at times “0” and “1” is shown in Figure 3.

Figure 3: Correlation Between Ratings at Time 0 and Time 1

Rating differences between measurements 2 minutes apart ranged from -60 to 70, with a median difference of 0. Bland-Altman analysis found that 95% of the differences
between paired ratings were between -23.5 and +20.7, with 50% of the differences between -6.0 and +8.9 (Figure 4). Overall there were 5 sets of measures (3.7%) with differences of at least 20. Two participants (2.0%) had both sets of paired measurements differ by at least 20.

![Bland-Altman Plot of Paired Ratings](image)

**Figure 4: Bland-Altman Plot of Paired Ratings**
4. Discussion

This is the first study to develop and validate a pain scale for use in Kiswahili speaking patients presenting to the Emergency Department. Although various pain scales have been used in Tanzania previously,\textsuperscript{43,56,57} no scale has previously been validated for use in Kiswahili.

This study was unique in that we did not chose a scale a priori, instead choosing to iteratively develop a scale using focus groups of patients and providers. Through this process, we found that understanding of a visual analog scale, even among providers, was low. This is consistent with previous reports which suggest that the utility of a VAS may be influenced by patient factors such as age, visual acuity, verbal fluency, etc.\textsuperscript{36,71,72} Therefore we do not recommend the use of the VAS in this setting, and feel that a numeric scale will potentially be more inclusive and easier to administer than other self-reporting scoring systems.

The results of our reliability analysis (ICC 0.97) are similar to those found in pain scale validation studies in other settings. Bijur and colleagues had ICCs between 0.95 and 0.99 across several ED pain scale validation studies,\textsuperscript{35,62,63} while a comparison of a NRS with a VAS by Daoust, et al, reported ICCs between 0.88 and 0.92.\textsuperscript{73} Similarly, our Bland-Altman analysis is consistent with the existing literature. Our 95\% limits of
agreement were approximately +/- 20, while previous reports have used Bland-Altman analysis to suggest imprecision of between 10 and 20% when using a pain intensity scales in various settings.\textsuperscript{35,63,70}

As there is difficulty in finding a “gold standard” for a subjective condition such as pain, previous studies used verbal descriptors of pain in order to assess validity. Our analysis during the development phase, found that there was an appropriate change in pain scores for all scales, other than the VAS, when comparing between categorical descriptions of pain. This was also the case for clinical testing of the resultant numeric rating scale. Our means rating for each category in clinical testing was similar to those found elsewhere in the literature,\textsuperscript{74-76} providing further evidence for the validity of our scale.

4.1 Limitations

Our findings came from a subset of injury patients presenting to the Emergency Department. KCMC is a referral hospital, and as a result, many patients are seen at local clinics or smaller hospitals, prior to being sent to KCMC. This could explain the substantial delay in presentation (median 8h) from the time of their injury, as well as the finding that nearly 50% of patients had received some analgesic medication prior to
arrival. This did not seem to have a significant impact on their clinical presentation, however, as the majority still rated their pain as severe.

Additionally, although our reliability findings are similar to the other validation studies in the literature, there was a small portion of patients whose ratings on paired measurements was >20. This is notable for two reasons. First, previous studies report the minimum clinically relevant difference in pain ratings between 9 and 20%. This would mean that a small percentage of patients (3.7%) in our study had a clinically relevant difference in their pain during a 2 minute interval. Secondly, 2 patients had both sets of paired measurements differ by more than 20. These findings may suggest difficulty in understanding the scale in a small subset of patients, or may simply represent the dynamic nature of pain.

Lastly, there has been debate in the literature regarding the clinical benefit of routine use of pain scales. Critics argue that adverse medication events have increased since the institution of policies mandating pain assessment. Proponents, however, state that pain scales lead to decreased time to administration and increased use of analgesic medications. Pain scales are often the basis for analgesic protocols, which have been especially useful in trauma patients. Additionally, a quantifiable measure is needed in order to provide a mechanism for comparing analgesic
interventions, as well as assessing the overall effectiveness of analgesic practices. Their benefits, therefore, outweigh the potential risks, especially when combined with proper clinical assessment.
5. Conclusion

In conclusion, this study shows the validity and reliability of a 100-point numeric rating scale for use in Kiswahili-speaking patients with acute pain for injuries in Moshi, Tanzania. Our study also demonstrated the difficulty of using a visual analog scale in this setting.
References


34. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983;17:45-56.


57. Masigati HG, Chilonga KS. Postoperative pain management outcomes among adults treated at a tertiary hospital in Moshi, Tanzania. Tanzania journal of health research 2014;16.


78. Kelly A. The minimum clinically significant difference in visual analogue scale pain score does not differ with severity of pain. Emergency Medicine Journal 2001;18:205-7.


