

## Original article

# Translation, and interobserver and test–retest reliability of the Brazilian Portuguese version of Children’s Hospital Oakland Hip Evaluation Scale for patients with sickle cell disease



Nágela Rayanne Barros de Freitas<sup>a</sup>, Natalia Duarte Pereira<sup>b</sup>, Maisa Santos Pessoa<sup>a</sup>, Fábio Galvão<sup>c</sup>, Sara Teresinha Olala Saad<sup>c</sup>, Camila Tatiana Zanoni<sup>c,\*</sup>

<sup>a</sup> Centro Universitário Nossa Senhora do Patrocínio (CEUNSP), Itu, SP, Brazil

<sup>b</sup> Universidade Federal de São Carlos (UFSCAR), São Carlos, SP, Brazil

<sup>c</sup> Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil

## ARTICLE INFO

## Article history:

Received 31 August 2016

Accepted 24 January 2018

Available online 24 March 2018

## Keywords:

Translation

Sickle cell disease

Scales

Hip

Physiotherapy

## ABSTRACT

**Background:** The Children’s Hospital Oakland Hip Evaluation Scale is a disease-specific tool for the clinical and functional assessment of the hip in sickle cell disease.

**Objectives:** To translate the tool into Brazilian Portuguese and evaluate the interobserver and test–retest reliability.

**Methods:** Eighteen patients diagnosed with sickle cell disease and a mean age of  $49 \pm 11.9$  years participated in the study. The scale was applied by two evaluators who did not speak to each other regarding their understanding of the tool and who had no prior training. Inter-observer and test–retest reliability of individual items and of the total score were evaluated using the intraclass correlation coefficient and the Bland–Altman method.

**Results:** When the overall score for each hip was considered, the test–retest intraclass correlation coefficient score for the right hip was 0.95 (0.89–0.98) and for the left hip it was 0.96 (0.91–0.98). Considering all assignments (total score), the score was 0.96 (0.90–0.98). The test–retest intraclass correlation coefficient varied from 0.76 to 1 for 18 of the 27 items (excellent) and from 0.53 to 0.75 for nine items (moderate). When the overall score for each hip was considered, the interobserver intraclass correlation coefficient for both hips was 0.94 (0.86–0.98). Considering all assignments, the total score was 0.94 (0.86–0.98). The inter-observer intraclass correlation coefficient varied from 0.48 to 0.75 for 18 out of 27 items (moderate) and varied from 0.77 to 1 for the remaining nine items (excellent).

**Conclusion:** The results demonstrate that the Brazilian version of the Children’s Hospital Oakland Hip Evaluation Scale presented adequate interobserver and test–retest reliability and that the version can be used to evaluate clinical function in sickle cell disease patients, producing consistent, standardized and reproducible results.

© 2018 Associação Brasileira de Hematologia, Hemoterapia e Terapia Celular. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Corresponding author at: Hematology and Blood Transfusion Center, Universidade Estadual de Campinas (UNICAMP), Rua Carlos Chagas, 480, CEP 13083-878, Campinas, SP, Brazil.

E-mail address: [czanoni@fcm.unicamp.br](mailto:czanoni@fcm.unicamp.br) (C.T. Zanoni).

<https://doi.org/10.1016/j.htct.2018.01.006>

2531-1379/© 2018 Associação Brasileira de Hematologia, Hemoterapia e Terapia Celular. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Sickle cell disease (SCD) is a autosomal recessive disorder caused by a physical–chemical change in the hemoglobin molecule, producing an abnormal type of hemoglobin called hemoglobin S (Hb S) instead of normal hemoglobin A (Hb A).<sup>1–3</sup> In cases of homozygotes or double heterozygotes for a mutation in trans in the other Beta-globin gene, this may result in hemolytic anemia and vaso-occlusion.<sup>2</sup>

It is estimated that approximately 7% of the world population is affected by hemoglobin disorders, mainly thalassemia and SCD. The delayed diagnosis of these disorders may lead to death during the first years of life.<sup>4</sup> Data from the Brazilian Health ministry estimates that more than 8000 patients and 700–1000 newborns/year are homozygous for Hb S.<sup>5</sup>

These individuals are referred to physiotherapy due to osteonecrosis of the femoral head, a common complication of SCD, which results from the infarction of the articular surfaces of long bones including the humeral head joint, knees and the small joints in the hands and feet. Impairment of multiple joints is common and over 50% of the patients present with bilateral hip pain.<sup>6–8</sup>

Ischemic bone necrosis can lead to the destruction of the hip joint at a young age and can progress from an early to a late stage in just a few years.<sup>9,10</sup> Epiphyseal ischemic necrosis is common in sickle cell anemia, especially in the head of the femur and humerus, and often bilateral.<sup>11</sup> Symptomatic patients usually complain of joint pain and limited movement. About 50% of patients develop avascular necrosis near the age of 35 years.<sup>12</sup>

Recently more strict parameters have been defined to measure clinical changes. In addition to the physical examination and complementary tests, outcomes such as health-related quality of life, functional capacity, pain scales and satisfaction have been used to enable an analysis of the health situation and to evaluate manifestations of the disease in the life of the individual from their own perspective, thereby complementing the clinical data. For this reason instruments, questionnaires and scales that address this type of variable have been developed and published; they are classified as generic or specific. Generic instruments quantify the individual's perception of general health, whereas specific instruments target specific areas of the body, and can measure function more accurately.

The Children's Hospital Oakland Hip Evaluation Scale (CHOHES) is a specific scale for SCD.<sup>13</sup> Used to evaluate patient functionality, it is a modification of the internationally recognized Harris Hip Score, developed in 1969, which analyzes the efficiency of clinical and surgical treatments such as total hip arthroplasty.<sup>14</sup> CHOHES was published in 2005<sup>13</sup> when it was validated in 26 patients (mean age: 25 years) with avascular hip necrosis (AVN) and 14 patients without AVN (mean age: 16 years).

According to the Orthopedic Committee of the National Avascular Necrosis Trial in Sickle Cell Anemia, a modification of the score was necessary to evaluate early articular damage by a physical and functional evaluation of the hip.<sup>13</sup>

Therefore, the aim of this study was to evaluate the translation into Brazilian Portuguese of the CHOHES used to

evaluate AVN in SCD patients with the purpose of providing an additional evaluation tool for healthcare providers in Brazil, including physical therapists, which can be used as a resource for assessing the effectiveness of the clinical treatment of SCD patients. A second aim was to assess interobserver and test–retest reliability of the tool.

## Methods

This study is a measurement of the correlation and agreement rate. It was approved by the Human Research Ethics Committee of the Faculty of Medical Science of the Universidade Estadual de Campinas (FCM-UNICAMP).

Guidelines for Reporting Reliability and Agreement Studies (GRRAS) was used to establish the items that should be used to evaluate the reliability of this study and in order to minimize bias.<sup>15</sup>

### Translation of the Children's Hospital Oakland Hip Evaluation Scale

The items of the scale were translated to Portuguese by the first translator (T1) who is fluent in both Portuguese and English, thus generating version 1. The translation was approved by the author of this article.

Version 1 was back translated (version 2) by one of the authors, a physical therapist, who is also fluent in both languages, translator 2 (T2), who was unaware of the original version of the scale. This was sent to the main author who checked the version and found no incompatibility with the original version, validating the version as the definite version for the interobserver and test–retest reliability evaluation, with no need for substitutions or modifications.

### Application of the Children's Hospital Oakland Hip Evaluation Scale

The CHOHES was applied by two evaluators who exchanged no information whatsoever regarding their understanding or application of the scale, and who had access solely to the information of the original article.<sup>13</sup> The CHOHES was applied once by evaluator 1 (A1) and once again, after fifteen days by A1 (reliability test–retest). Evaluator 2 (A2) applied CHOHES after a period of 15–20 days from the last application by A1 (Interobserver reliability).

The scale ranges from 0 to 100 points for each hip (totalizing 200 points) and evaluates 27 items which are subdivided into three categories: pain, function and physical exam.<sup>13</sup> The first item is reported by the individual regarding pain in each hip during the four weeks prior to the evaluation and classifies pain according to different numbers. Scores range from 0 to 40, where 0 (zero) is disabling pain that limits activities, 10 is severe pain causing significant limitations, 20 is moderate pain with certain limitations to perform activities, 30 is mild pain with minimal or insignificant limitations and 40 absence of pain and limitations.

The items related to the evaluation of function comprised five questions with different scores for each item (Table 1). This category consists of a self-report on dressing, walking

**Table 1 – The Children’s Hospital Oakland Hip Evaluation Scale Scores.**

Item	Score
<b>Pain domain</b>	
Pain	0, 10, 20, 30, 40
<b>Function domain</b>	
Dressing	0, 2, 4
Using a walking aid	0, 8
Walking	
Using a walking aid	0, 3, 5, 7, 8
Without a walking aid	0, 3, 5, 7, 11
Sitting	0, 3, 5
Stairs	0, 1, 2, 4
<b>Physical exam</b>	
Hip JRoM (individual)	
Internal rotation	0, 1, 2, 3
External rotation	0, 1, 2, 3
Flexion	0, 1, 2, 3
Abduction	0, 1, 2, 3
Adduction	0, 1, 2, 3
Thomas test	0, 1
Hip muscle strength (individual)	
Flexion	0, 1, 2
Extension	0, 1, 2
Abduction	0, 1, 2
Adduction	0, 1, 2
Step height performance	0, 2, 6
Limp	0, 1

JRoM: joint range of movement.

with or without support, sitting and walking up stairs. The last item differs from the others, as instead of self-evaluation, this item was appraised by an assessor, who observed the patient walking up a flight of stairs and marked down the highest level reached (step by step, both feet per step, holding or not onto the handrail or incapacity to climb stairs). The overall score for this category ranges from 0 to 32 and is the same for both hips, therefore the result is multiplied by two at the end.

The physical exam is included in the third category and consists of an assessment of the patient’s walk, measuring range of movement (RoM) of the hips during flexion, abduction, internal and external rotation using a conventional goniometer. The Thomas test was used to rule out hip flexion contracture<sup>16</sup> and evaluate muscle strength of the hip (flexors, extensors, adductors, and abductors), graduated according to the Oxford Test.<sup>17</sup> The scores in this category range from 0 to 28 points for each hip and are different for each item (Table 1). The self-report also evaluates the best step height performance, where zero is incapacity to climb, 2 when the best performance occurs at step height (15–20 cm) and 6 when capable of reaching over 50 cm, exemplified by the height of public transportation steps. Each hip is rated individually and can have different scores. The classification of the patient’s walk receives a score of 1 when the individual does not limp and does not use a walking aid and 0 (zero) in cases where the individual limps or uses a walking aid.

The CHOHES additionally evaluates speed of movement over a distance of 25 m; however this item is not listed in the overall rating of the scale. This speed test has been assessed

for reliability in several papers; despite performing this test, the scores were not included.<sup>18</sup>

Thus, the subject who does not report pain, functional limitations and who presents a normal physical exam receives a one hundred point score for each hip.

### Participants

Eighteen adults (age: 49 ± 11.9 years) with a diagnosis of SCD followed at the Hematology and Transfusion Center (HEMO-CENTRO), UNICAMP participated in this study. After signing the written informed consent form, individuals who fulfilled the following selection criteria were enrolled: age >18 years and clinical and laboratory diagnosis for SCD. Individuals who suffered a worsening of the disease, who had neurological impairment and those who could not follow simple commands or who had orthopedic alterations that were not associated to the disease (resulting from trauma or not) or did not complete the three evaluations during the proposed period of time, were excluded from the study.

### Statistical analysis

The intraclass correlation coefficient (ICC) with a 95% confidence interval (CI) was calculated to evaluate the interobserver and test-retest reliability for the scores of the 27 individual items of the CHOHES and scores in general. Bland-Altman analysis was used for total interobserver and test-retest values.

The values adopted for the ICC reliability test were based on the classification by Fleiss et al.,<sup>19</sup> where values under 0.4 were classified as poor, between 0.4 and 0.75 moderate, and higher than 0.75 excellent.

## Results

### Sample characterization

The mean age of the 18 participants was 49 years (±11.9), 72% were women and 28% were men. Regarding clinical and laboratory diagnosis, 55.5% of the patients were homozygous for the Hb S mutation (Hb SS), 38.8% double heterozygotes for Hb S and Hb C (Hb SC) and 5.5% double heterozygotes for Hb S and Beta thalassemia (Hb S/β-thal). Table 2 presents patient characteristics.

### Test-retest reliability

Tables 3 and 4 present the means and standard deviation (DP) of the test-retest and interobserver ICC and the 95% CI for each item of the CHOHES.

When the overall score was considered for each hip, the test-retest ICC for the right hip was 0.95 (range: 0.89–0.98) and for the left hip it was 0.96 (range: 0.91–0.98). Considering all assignments (total score), the score was 0.96 (range: 0.90–0.98).

Test-retest ICC values ranged from 0.76 to 1 for 18 of the 27 items (excellent) and from 0.53 to 0.75 for the remaining nine items (moderate).<sup>19</sup>

**Table 2 – Characteristics of study participants.**

Participant	Age (years)	Gender	Diagnosis
1	32	M	Hb SS
2	49	F	Hb SS
3	62	F	Hb SS
4	44	M	Hb SS
5	52	F	Hb SS
6	31	F	Hb SS
7	49	M	Hb SC
8	43	M	Hb SS
9	69	M	Hb SC
10	27	F	Hb SS
11	55	F	Hb SC
12	43	F	Hb SS
13	49	F	Hb SC
14	72	F	Hb SS
15	57	F	Hb SC
16	52	F	Hb SC
17	46	F	Hb SS
18	53	F	Hb S/β-thal
Mean (SD)	49 (11.9)	13F/5M	10 SS/7SC/1SB

M: male; F: female; SD: standard deviation; Hb SS: homozygotes for Hb S mutation; Hb SC: double heterozygotes for Hb S and Hb C; Hb S/β-thal: double heterozygotes for Hb S and Beta thalassemia.

The Bland and Altman analysis (Figure 1) shows that the agreement of the test–retest did not significantly differ from zero; only one point was above the standard deviation (SD), thus, the difference between the two evaluations was minimal.

#### Interobserver reliability

When the overall scores for each hip were considered, interobserver ICC was 0.94 (range: 0.86–0.98) for both left and right hips. Taking all assignments into consideration (overall score), the score was 0.94 (range: 0.86–0.98).

Interobserver ICC values varied from 0.48 to 0.75 in 18 of the 27 items of the scale (moderate) and from 0.77 to 1 for the remaining nine items (excellent).<sup>18</sup>

When analyzing the Bland Altman interobserver plot (Figure 1), the mean difference between the two evaluations can be observed with only one point above the SD (not significantly different from zero).

## Discussion

This study translated the CHOHES into Brazilian Portuguese and assessed interobserver and test–retest reliability in a population of SCD patients. The overall results were adequate, with a similar ICC to that found by Aguilar et al.<sup>13</sup> in the original version. Moreover, the CHOHES was based on a sample of younger patients than in this study (25 vs. 49 years). It is important to point out that the frequency of AVN increases with age and the parameters evaluated in the CHOHES are applied to any age group as it takes into account common daily activities (sitting, walking, dressing and so on).

Analyzing the test–retest reliability individually, two out of nine items were classified as moderate (range: 0.53–0.75),

**Table 3 – Results for test–retest reliability.**

Pain and functionality			
Task	Mean (SD)n = 18	ICC	95% CI
Pain R Hip	21.9 (±10.4)	0.92	0.80–0.97
Pain L Hip	21.9 (±10.4)	0.92	0.80–0.97
Dressing	2.2 (±1.7)	0.88	0.71–0.95
Using a walking aid	6.2 (±3.4)	1	1–1
Walking	6.8 (±2.0)	0.53	0.10–0.79
Sitting	2.2 (±2.5)	0.57	0.15–0.81
Stairs	2 (±1.3)	0.93	0.83–0.97
Physical exam			
Task	Mean (SD)n = 18	ICC	95% CI
<i>Hip JRoM</i>			
Internal Rotation R	1.6 (±0.7)	0.79	0.53–0.91
Internal Rotation L	1.7 (±1.1)	0.85	0.65–0.94
External Rotation R	1.8 (±0.8)	0.65	0.28–0.85
External Rotation L	1.7 (±0.8)	0.67	0.31–0.86
Flexion D	1.1 (±1.0)	0.75	0.45–0.90
Flexion E	1.1 (±1.2)	0.72	0.41–0.88
Abduction D	1.8 (±0.9)	0.82	0.58–0.92
Abduction E	1.6 (±1.0)	0.81	0.56–0.92
Thomas Test R	0.8 (±0.4)	0.86	0.66–0.94
Thomas Test L	0.7 (±0.5)	0.87	0.69–0.95
<i>Hip strength</i>			
Flexion R	1.8 (±0.4)	0.67	0.32–0.86
Flexion L	1.6 (±0.6)	1	1–1
Extension R	1.6 (±0.6)	0.82	0.59–0.93
Extension L	1.6 (±0.5)	0.89	0.73–0.95
Abduction R	1.7 (±0.6)	0.64	0.26–0.85
Abduction L	1.6 (±0.6)	0.73	0.42–0.89
Adduction R	1.4 (±0.8)	0.91	0.78–0.96
Adduction L	1.1 (±0.8)	0.88	0.72–0.95
Step height R Hip	5.9 (±0.7)	1	1
Step height L Hip	5.9 (±0.7)	1	1
Limp	0.2 (±0.4)	0.83	0.60–0.93
<i>Total</i>			
R Hip	61.1 (±18.3)	0.95	0.89–0.98
L Hip	60.1 (±19.4)	0.96	0.91–0.98
Overall	121.2 (±37.5)	0.96	0.90–0.98

JRoM: joint range of movement; R: right; L: left; SD: standard deviation; ICC: intraclass correlation coefficient; CI: confidence interval.

walking and sitting, as they presented low reliability levels (ICC <0.60). These results can be explained by the fact that these were self-reported items. The difficulty in evaluating the walking item was to obtain the distance covered comfortably without stopping ('6 blocks', '2–3 blocks' or '30–60 m') from the patients. The greatest difference occurred between the options '6 blocks' and '2–3 blocks'. This divergence could be minimized by standardizing the distance as there is no exact value defining the size of a block. A suggestion that could be adopted would be the time that the individual can walk without resting, as time is a unit that even individuals with no schooling and those with difficulties in understanding the concept of measurements, understand easily.

In contrast, the item sitting presented test–retest ICC of 0.57 and interobserver ICC of 0.56, thus, both presented low levels of reliability (ICC <0.60). One of the difficulties encountered was to do with the organization and scoring of items, one

**Table 4 – Results for interobserver reliability.**

Pain and functionality			
Task	Mean (SD)n= 18	ICC	95% CI
Pain R Hip	21.1 (±10.4)	0.91	0.79–0.96
Pain L Hip	21.1 (±10.4)	0.91	0.79–0.96
Dressing	2.1 (±1.8)	0.53	0.10–0.96
Using a walking aid	6.2 (±3.4)	1	1–1
Walking	6.8 (±1.9)	0.65	0.29–0.85
Sitting	1.9 (±2.3)	0.56	0.15–0.81
Stairs	2.0 (±1.2)	0.72	0.40–0.88
Physical exam – hip			
Task	Mean (SD)n= 18	ICC	95% CI
<i>JRoM</i>			
Internal Rotation R	1.8 (±0.9)	0.64	0.27–0.85
Internal Rotation L	1.8 (±1.1)	0.87	0.70–0.95
External Rotation R	1.8 (±0.7)	0.58	0.17–0.82
External Rotation L	1.7 (±0.9)	0.73	0.70–0.95
Flexion D	0.9 (±0.9)	0.72	0.39–0.88
Flexion E	0.9 (±1.0)	0.85	0.65–0.94
Abduction R	1.8 (±0.9)	0.6	0.39–0.88
Abduction L	1.8 (±1.1)	0.69	0.65–0.94
Thomas Test R	0.7 (±0.5)	0.66	0.20–0.83
Thomas Test L	0.6 (±0.5)	0.61	0.35–0.87
<i>Hip strength</i>			
Flexion R	1.6 (±0.5)	0.62	0.23–0.84
Flexion L	1.4 (±0.7)	0.71	0.38–0.88
Extension R	1.4 (±0.7)	0.77	0.49–0.90
Extension L	1.5 (±0.5)	0.48	0.37–0.77
Abduction R	1.6 (±0.7)	0.71	0.38–0.88
Abduction L	1.5 (±0.7)	0.83	0.60–0.93
Adduction R	1.5 (±0.7)	0.84	0.63–0.93
Adduction L	1.3 (±0.7)	0.75	0.46–0.90
Step height R Hip	5.9 (±0.7)	1	1
Step height L Hip	5.9 (±0.7)	1	1
Limp	0.2 (±0.4)	0.7	0.37–0.88
<i>Total</i>			
R Hip	59.2 (±17.9)	0.94	0.86–0.98
L Hip	58.5 (±19.0)	0.94	0.86–0.98
Overall	117.7 (±36.6)	0.94	0.86–0.98

JRoM: joint range of movement; R: right; L: left; SD: standard deviation; ICC: intraclass correlation coefficient; CI: confidence interval.

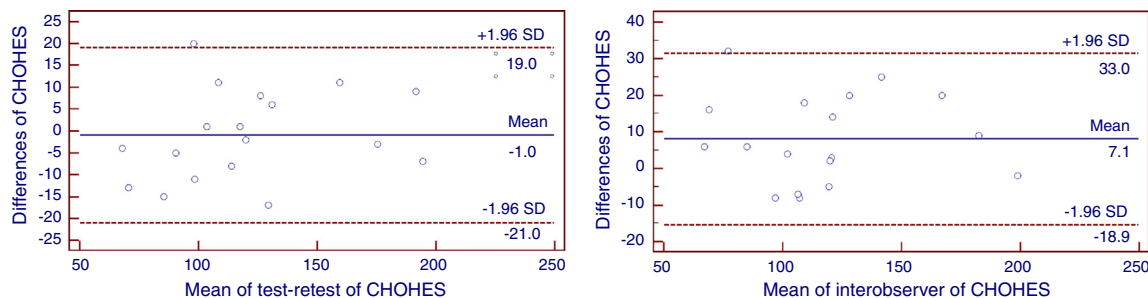
was scored on performance and the rest were self-reported. ‘Can sit comfortably in any position (i.e. have patient demonstrate ring sitting – seated with an erect spine, hips and knees flexed, and with the soles of the feet pressed together)’ was

scored on performance. Whereas ‘Can sit comfortably at a table or in the movies but not able to tolerate other sitting positions’ and ‘unable to sit comfortably for more than a few minutes without changing positions’ were self-reported. Not one of the participants was capable of ring sitting. Concerning sitting, the option of staying in the same position for longer than a few minutes without changing position is very imprecise, as the item is related to time and not to whether the position is comfortable or not. To reduce this divergence, the individual is recommended to sit in each position in the presence of an assessor and report whether the position is comfortable or not. This enables the possibility of timing each position, rendering the item more reliable and standardized, facilitating the visualization of any type of complication which would justify intolerance to a specific position, in addition to simplifying the statistical analyses.

The test-retest ICC for total scores was 0.96, similar to the value reported in the original version, where 0.96 was the value for affected hips and 0.95 for healthy hips. Regarding the methodology applied for test-retest reliability in the original version, the only specified fact was the application of the scale to eight different hips by seven different evaluators. The number of affected or unaffected hips was not specified nor was the side, i.e., right or left hip. Furthermore, the report did not mention whether the score was applied on both hips,<sup>13</sup> thus complicating a more detailed comparison of the results.

Regarding interobserver reliability, when analyzing the items of the scale individually, there was a variation ranging from 0.48–0.75 for 18 of the 27 items, classified as moderate. The task of getting dressed presented a lower level of reliability in this classification (ICC <0.60) and is also a self-reported item. This item evaluates a complex task and requires an objective measurement, advocating real time information, minimizing inconsistencies and errors in the information provided where past events must be recorded.<sup>20</sup> Thus, the use of a more objective measuring unit is suggested to evaluate each task, with the presence of an assessor giving more reliability to the results.

In the physical exam, the RoM items for external rotation of the right hip and extension strength of the left hip presented low levels of reliability (moderate). As these issues are related to the individual’s physical aptitude, they are directly dependent upon the individual’s disposition and are related to movement, pain and clothes that directly influence RoM measurements.



**Figure 1 – Differences and means of assessments using the Bland–Altman plot. SD: standard deviation. The horizontal lines in blue are the means of differences; the dotted red lines indicate the upper and lower 95% confidence intervals.**

Variations between 2 and 7 degrees are acceptable in particular in respect to interobserver reliability.<sup>21</sup>

Another difficulty encountered is the determination of specific anatomical landmarks of the lower limbs, as an error of only a few millimeters in the delimitation of these points can affect angular values obtained due to compensation.<sup>22</sup> According to Van Trijffel et al.,<sup>23</sup> when using a conventional universal goniometer, there is a lower reliability of RoM values of the legs compared to the arms due to the greater difficulty of locating anatomical landmarks precisely thus complicating the correct alignment of the goniometer.

Extension strength ICC of the left hip was moderate (0.48), showing that in some evaluations the individuals presented more pain thereby complicating the assessment of muscle strength and maximum RoM. This limitation can be correlated with lumbar curvature.<sup>24</sup> The most common manifestations of SCD patients are vaso-occlusive and pain crises, which are closely related to tissue ischemia secondary to the sickling of red blood cells with the most common site of these vaso-occlusive crises being in the lumbar region.<sup>25</sup>

The interobserver ICC value for the overall score was 0.94, similar to the values reported in the original version, where 0.95 was the value for affected hips and 0.94 for healthy hips. The ICC was not verified individually for all items in the original article as validated herein, making a more detailed comparison of the statistical data complicated.<sup>13</sup>

The function domain has performance and self-report items, where the self-reported items are subjective. Thus, we suggest that as the self-reported items are related to complex tasks, these should be transformed into an objective evaluation by a quantitative evaluation of the assessor.<sup>20</sup>

The use of the Thomas test for scoring the physical exam is questionable, as this test does not take into consideration the standard of normality of ten degrees used to evaluate the presence of hip flexion contracture,<sup>26</sup> therefore both individuals who had contractures below this standard, as well as the individuals who had contractures above this standard, scored zero. Another questionable item related to the physical exam was speed of movement; this item is present in the scale but is not scored. The actual need for this item in the scale should be raised and should the item really be required, a score should be assigned.

There is no doubt that in current scientific studies the use of protocols for evaluations and specific scales are mandatory in order to compare results, evaluate any intervention efficiently and obtain increasingly reliable results.

## Conclusion

Results demonstrated that the Brazilian version of the CHOHEs presented adequate test-retest and interobserver reliability when scores were considered, and can be used to evaluate clinical function of SCD patients, with the advantage of rendering uniform, standardized and reproducible results.

## Conflicts of interest

The authors declare no conflicts of interest.

## Acknowledgements

The authors thank Raquel Foglio, Translator 1 (T1), for the editing and English revision of the manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.htct.2018.01.006](https://doi.org/10.1016/j.htct.2018.01.006).

## REFERENCES

- Burch-Sims GP, Matlock VR. Hearing loss and auditory function in sickle cell disease. *J Commun Disord.* 2005;38(4):321–9.
- Beutler E. The sickle cell diseases and related disorders. In: Beutler E, Kipps TJ, Lichtman MA, Seligsohn U, Coller BS, editors. *Williams hematology*. 6th ed. New York: McGraw-Hill; 2001. p. 581–605.
- Mack AK, Kato GJ. Sickle cell disease and nitric oxide: a paradigm shift? *Int J Biochem Cell Biol.* 2006;38(8):1237–43.
- Loureiro MM, Rozenfeld S. Epidemiologia de internações por sickle cell disorder no Brasil. *Rev Saúde Pública.* 2005;39(6):943–9.
- Brasil. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. *Manual de Diagnóstico e Tratamento de Doenças Falciformes*. Brasília: Ministério da Saúde; 2002.
- Renaudier P. Sickle cell pathophysiology. *Transfus Clin Biol.* 2014;21(4–5):178–81.
- Almeida A, Roberts I. Bone involvement in sickle cell disease. *Br J Haematol.* 2005;129(4):482–90.
- Zanoni CT, Galvão F, Cliquet Junior A, Saad ST. Pilot randomized controlled trial to evaluate the effect of aquatic and land physical therapy on musculoskeletal dysfunction of sickle cell disease patients. *Rev Bras Hematol Hemoter.* 2015;37(2):82–9.
- Drescher W, Pufe T, Smeets R, Eisenhart-Rothe RV, Jager M, Tingart M. Avascular necrosis of the hip – diagnosis and treatment. *Z Orthop Unfall.* 2011;149(2):231–40.
- Al-Mousawi FR, Malki AA. Managing femoral head osteonecrosis in patients with sickle cell disease. *Surgeon.* 2007;5(5):282–9.
- Malizos KN, Siafakas MS, Fotiadis DI, Karachalios TS, Soucacos PN. An MRI-based semiautomated volumetric quantification of hip osteonecrosis. *Skelet Radiol.* 2001;30(12):686–93.
- Styles LA, Vichinsky EP. Core decompression in avascular necrosis of the hip in sickle-cell disease. *Am J Hematol.* 1996;52(2):103–7.
- Aguilar CM, Neumayr LD, Eggleston BE, Earles AN, Robertson SM, Jergesen HE, et al. Clinical evaluation of avascular necrosis in patients with sickle cell disease: Children's Hospital Oakland Hip Evaluation Scale – a modification of the Harris Hip Score. *Arch Phys Med Rehabil.* 2005;86(7):1369–75.
- Guimarães RP, Alves DP, Silva GB, Bittar ST, Ono NK, Honda E, et al. Tradução e adaptação transcultural do instrumento de avaliação do hip “Harris Hip Score”. *Acta Ortop Bras.* 2010;18(3):142–7.
- Kottner J, Audigé L, Brorson S, Donner A, Gajwski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96–106.

16. Peeler JD, Anderson JE. Reliability limits of the modified Thomas Test for assessing rectus femoris muscle flexibility about the knee joint. *J Athl Train.* 2008;43(5):470-6.
17. Hoppenfeld S, Hutton R. *Physical examination of the spine and extremities.* 1st ed. Pearson Education Limited; 2013. p. 143-70.
18. Vasconcelos KS, Dias JM, Dias RC. Relação entre intensidade de dor e capacidade funcional em indivíduos obesos com osteoartrite de joelho. *Rev Bras Fisioter.* 2006;10(2):213-8.
19. Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions.* 3rd ed. New York: John Wiley & Sons; 2003.
20. Marszalek J, Adamowicz NM, Rutkowska I, Kosmol A. Using ecological momentary assessment to evaluate current physical activity. *BioMed Res Int.* 2014:1-9.
21. Gouveia VH, Araújo AG, Maciel SS, Ferreira JJ, Santos HH. Reliability das medidas inter e intra-avaliadores com goniômetro universal e flexímetro. *Fisioter Pesq.* 2014;21(3):229-35.
22. Santos CM, Malacco PL, Sabino GS, Sabino GS, Moraes GF, Felício DC. Intra and inter examiner reliability and measurement error of goniometer and digital inclinometer use. *Rev Bras Med Esporte.* 2012;18(1):38-41.
23. van Trijffel E, van de Pol RJ, Oostendorp RA, Lucas C. Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: a systematic review. *J Physiother.* 2010;56(4):223-35.
24. Clapis PA, Davis SM, Davis RO. Reliability of inclinometer and goniometric measurements of hip extension flexibility using the modified Thomas test. *Physiother Theory Pract.* 2008;24(2):135-41.
25. Soderman P, Malchau H. Is the Harris Hip Score system useful to study the outcome of total hip replacement? *Clin Orthop Relat Res.* 2001;384:189-97.
26. Kendall FP, McCreary EK, Provance PG, Rodgers MM, Romani WA. Lower extremity. In: *Muscles: testing and function with posture and pain.* 5th ed. Lippincott Williams & Wilkins; 2005. p. 359-464.