**The A-test – reliability of functional recovery assessment during early rehabilitation of patients in an orthopedic ward**

**A-test – pouzdanost procene funkcionalnog oporavka bolesnika tokom rane rehabilitacije na ortopedskom odeljenju**

**Abstract**

**Background/Aim.** There are few tests for evaluation of functional abilities of patients surgically treated for hip fractures or osteoarthritis during early rehabilitation period. The aim of this study was to investigate reliability (interobserver reproducibility and internal consistency) of the A-test, an original test for functional recovery evaluation during early rehabilitation of patients in an orthopedic ward.

**Methods.** The investigation included 105 patients (55 patients with hip osteoarthritis that underwent arthroplasty and 50 surgically treated patients with hip fracture). It was conducted in an orthopedic ward during early inpatient rehabilitation (from 1st to 5th day). For their functional recovery evaluation during early rehabilitation we used the A-test, a performance-based test with 10 items for assessing basic activities by six level ordinal scale (0–5). For internal consistency of the test the Cronbach coefficient alpha was calculated for the A-test results collected during early rehabilitation for all patients (105 patients x 5 days = 525 measures) and separately for the results of patients with hip osteoarthritis (275 measures) and hip fracture (250 measures). Values of this coefficient > 0.7 imply good internal consistency of the test. Interobserver reproducibility was estimated as follows: two physiotherapists together conducted physical therapy with the patients, and then, separately, rated the performance of each activity from the test (78 measures). The agreement between their estimations was expressed by the linear weighted kappa coefficient (for very good agreement values of kappa coefficient have to be in the range 0.81–1).

**Results.** The Cronbach coefficient alpha was 0.98 (the results of all the patients and the results of the patients with hip osteoarthritis) and 0.97 (the results of the patients with hip fracture). The values of kappa coefficient were in the range 0.81–0.92 for all items.

**Conclusion.** The A-test is a reliable instrument for everyday evaluation of functional recovery during early rehabilitation of patients surgically treated in an orthopedic ward.

**Key words:** hip fractures; osteoarthritis, hip; hip prosthesis; orthopedic procedures; postoperative period; physical therapy; recovery of function; predictive value of the tests.
Introduction

Early rehabilitation period usually lasts only a few days, and it is particularly present in an orthopedic ward. In this period patients experience greater or lesser degree of functional disability followed by restoration efforts to return to premorbid activity level. Regardless of the short duration, reliable tests for monitoring and evaluating the functional recovery of patients and presenting the results of work are required in the period of early rehabilitation. However, there are few tests that have been created for the period of early rehabilitation.

The Cumulated Ambulation Score (CAS) with a simple three-level ordinal scale was designed to monitor functional recovery of older people who were surgically treated for hip fractures and in this population it has excellent reliability. The University of Iowa Level of Assistance Scale (ILAS) has a complex seven-level scale and demonstrated moderate reliability in a population of patients who underwent arthroplasty for osteoarthritis of the hip and knee. The existing tests are aimed to specific clinical entities and cannot be applied in another domain without any modification. All three tasks from the CAS are too simple for patients who underwent arthroplasty while some tasks from ILAS are too demanding for patients after surgically treated hip fracture. Thus, in an investigation of factors predictive of independence in transfer and ambulation of patients after hip fracture, only 3 from 5 tasks of the ILAS were used. In the absence of a single test that could be easily applied in a heterogeneous population of patients in an orthopedic department we established an original test that could be a solution to this problem. The test was called A-test (simple test and simple name “A” like Assessment or Activity).

We have 10 years of experience with the A-test. We use it in the orthopedic ward to assess the recovery of all patients surgically treated for diseases and injuries of the lower extremities. Unlike the CAS and the ILAS which were designed for estimation of only 3 and 5 activities, respectively, this performance based test consists of 10 items for assessing 9 basic activities and walking endurance by a six-level ordinal scale (0–5). In this way, a more reliable picture of a patient's physical ability is obtained. Total scores can range from 0 to 50 (inability to perform any activity despite the help of therapists until complete independence and safety in performing all activities). We designed the A-test in 2002 and used it initially for monitoring patients of interest for some studies. The study population consisted of patients with arthroplasty for osteoarthritis of the hip in the first study. In the second study, we observed functional recovery in patients surgically treated for hip fracture. Since 2007, the A-test has been used in routine practice of the rehabilitation team in our Orthopedic Ward. This test has proved to be a useful and practical measuring tool in routine and research work, but we did not have a solid evidence that this test is reliable.

Reliability refers to error of an instrument. Reliability is best represented by reproducibility and internal consistency. The aim of this study was to examine the reliability of the A-test through the evaluation of the functional abilities of patients surgically treated for hip fractures and osteoarthritis.

Methods

Subjects

This prospective study was conducted at the Clinic for Orthopedic Surgery and Traumatology (COST), Military Medical Academy, Belgrade, and initially included 120 patients: 60 consecutive patients with acute hip fracture of both sexes who before the injury were able to walk with or without aids and up- and downstairs (help of another person was allowed for this activity); patients with dementia, pathological hip fracture, bilateral hip fractures, concurrent fracture in any other part of the body, and patients to whom surgical treatment were not included), and 60 consecutive patients who underwent hip arthroplasty due to osteoarthritis, without significant mental disability who were able to walk with or without aids before the operation and up- and downstairs (help of another person was allowed for this activity).

Exclusion criteria during the study were the presence of intraoperative or postoperative complications that prevented or delayed the beginning of rehabilitation, lethal outcome immediately after the surgery, and incomplete collected data for individual patient.

Procedure

All the patients were treated surgically. The modality of treatment depended on the type of fracture: osteosynthesis with dynamic hip screw was applied in patients with intertrochanteric fracture, and arthroplasty was performed in patients with fractures of the femoral neck (partial arthroplasty for older than 70 and total arthroplasty for younger than 70). All the patients admitted for arthritis of the hip underwent arthroplasty.

After the surgery, all the patients had the same rehabilitation treatment, which involved early mobilization of the patients at the bedside (from the first postoperative day, un-
less their general condition did not allow it) and activities such as getting out of bed (in accordance with the possibilities of the patients), walking with aids on the flat, as well as up-and downstairs, the use of walking aids, carrying out basic and instrumental activities, as well as socio-epidemiological data (marital status, housing conditions) were collected from all the patients. Mental status assessment was made using the Serbian version of shortened mental test score, while the functional status before injury for patients with hip fracture or on admission for patients with hip osteoarthritis (walking distance, the ability to walk up- and downstairs, the use of walking aids, carrying out basic and instrumental activities), as well as socio-epidemiological data (marital status, housing conditions) were collected from all the patients. Mental status assessment was made using the Serbian version of shortened mental test score, while the functional status before injury was assessed by the New mobility score (NMS).

In the postoperative period, functional abilities assessment of all the patients was performed by the A-test, the ILAS and the CAS from the first day until the fifth day of rehabilitation (this was the day of discharge for the majority of patients).

By the protocol, postoperative complications that slowed down the course of rehabilitation, the number of treatment days and duration of hospitalization after surgery were recorded.

We conducted this research with the approval of the Ethics Committee of our hospital.

**Measurement**

The A-test is a performance-based test that assesses 10 activities necessary for everyday life that the patient needs to achieve in the first days after the surgeries: (1) turning to side, (2) transition from supine to sitting position, (3) getting out of bed, (4) return back to bed, (5) standing, (6) walking with aids, use of (7) toilet and (8) dining room chairs, (9) walking up- and downstairs, (10) walking endurance.

Depending on the success of performance, the patient is evaluated from 0 to 5 for each activity: 0 – activity is not achieved; 1 – needs full assistance of the physiotherapist; 2 – requires adherence by the physiotherapist; 3 – activity performed with verbal suggestions of therapists; 4 – completely independent but insecure (while performing activities, a patient needs the presence of another person, for example a family member); 5 – fully independent and secure.

Walking endurance is graded in a slightly different way: 0 – activity is not achieved; 1 – a patient walks across the room (up to 5 meters); 2 – a patient walks from 5 to 20 meters; 3 – a patient walks from 20 to 50 meters; 4 – a patient walks from 50 to 100 meters; 5 – a patient walks more than 100 meters.

For ease of walking endurance grading, we had a landmark in the hospital: 0 – activity is not achieved; 1 – a patient walks across the room; 2 – once crosses ward hallway; 3 – two times crosses ward hallway; 4 – once crosses hospital corridor; 5 – several times crosses hospital corridor.

The maximum sum is 50, which means that a patient is independent and secure in the performance of all activities envisaged in the early rehabilitation. The test is simple, convenient, does not take additional time and requires no additional equipment.

**Reliability**

We examined the reliability or the ability of the instrument to measure something twice or more in the same manner, through assessing internal consistency and reproducibility of the A-test.

**Internal consistency**

Internal consistency is a measure of how homogenous or consistent items in the scale are and it gives us information to what extent they measure the same thing. As an indicator of internal consistency, we calculated the Cronbach coefficient alpha for the A-test results collected during early rehabilitation for all the patients and separately for the results of the patients with hip osteoarthritis and hip fracture. Cronbach alpha coefficient greater than 0.7 was considered feature of good internal consistency.

**Reproducibility**

There are two forms of reproducibility: interobserver and test-retest.

Test-retest reproducibility or intratester reliability indicates the agreement in measurements over time. This approach assumes that there is no substantial change in the construct being measured between the two occasions. In our case, the A-test evaluates the patient's functional status which changes daily during early rehabilitation (usually improves day by day), so we could not examine this form of reproducibility.

Interobserver reproducibility or intertester reliability is the consistency of measurement when the measurement is performed independently by two or more examiners and indicates the agreement of measurements performed by different examiners. It was tested in the following way: two physiotherapists conducted together physical therapy with the patient, and then, separately, rated the performance of each activity from the test. The first physiotherapist in the team had extensive experience in rehabilitation of patients in the COST (29 years of work experience). Also, the first physiotherapist was involved in collecting data in our previous studies. The A-test form was known to this physiotherapist. Another physiotherapist had 12 years of work experience, but until this research did not work in orthopedic rehabilitation team. The second physiotherapist had no experience in completing the A-test form. We did not organize specific training for the use of the A-test for another therapist. We planned to examine the reliability in the whole population of respondents, one day during each patient's rehabilitation (105 measurements).

In order to compare the reproducibility of the A-test with the CAS and the ILAS, we also examined the reliability of the ILAS and the CAS. There is the same item in all three tests which assesses walking, but with different scales. We
considered that the reproducibility of this item, as representative of the test, could be adequately compared with 3 tests by the same methodology.

In addition to assessing the reliability of all tests, physiotherapists were doing their usual job. Thus, we created a situation which corresponds to real everyday work of physiotherapists.

Agreement of the results of each item between the examiners is expressed by kappa coefficient with corresponding 95% confidence interval that was calculated using the linear weighted kappa for ordinal scale 16. We considered that the discrepancy by one ordinal category was less than the discrepancy by two or more ordinal categories and that the discrepancy by one ordinal category was less than the basis for weighting. Kappa coefficient was evaluated according to widely accepted interpretation by Landis and Koch 17. The agreement between the examiners is good if the kappa coefficient is between 0.61 and 0.80, very good if the kappa coefficient is between 0.81 and 1.00. Linear weighted kappa was calculated according to the procedure given on the website: http://vassarstats.net/kappa.html.

We planned 105 measurements to test the interobserver reproducibility (one measurement for each patient during rehabilitation), but due to unplanned absence of the second examiner, 78 measurements were done. The agreement between examiners for each of the A-test items is 95% confidence interval was calculated.

We did not delay the beginning of early rehabilitation because of complications occurred in other patients like: confusion, gastric complaints, hypotension, urinary tract infection, short-term diarrhea, the occurrence of pressure ulcers in the sacral region and on the feet, vomiting.

Demographic characteristics, comorbidity, mental and functional status before admission (for the patients with hip osteoarthritis) or injury (for the patients with hip fracture), hospital stay and duration of early rehabilitation are shown in Table 1. Due to the large influx of patients in the Orthopedic Ward, patients were discharged relatively quickly, so most patients in both populations had only 5 days for early rehabilitation.

Analyzing all the A-test results collected from the first to the fifth day of rehabilitation, Cronbach alpha coefficient was 0.98, indicating excellent internal consistency (Table 2). Table 2 presents the results of correlation between all items, as well as between each item and total score. A strong correlation exists between all the items and total score, and the removal of any of the items does not contribute to increasing alpha.

A similar result was obtained when the A-test was used for observation of the patients with hip fracture (n of cases = 250, n of variables = 10, alpha = 0.97) and the patients with osteoarthritis of the hip (n of cases = 275, n of variables = 10, alpha = 0.98).

We planned 105 measurements to test the interobserver reproducibility (one measurement for each patient during rehabilitation), but due to unplanned absence of the second examiner, 78 measurements were done. The agreement between examiners for each of the A-test items is shown in Table 3. The kappa coefficient was 0.81 and

**Table 1**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>The group of patients with osteoarthritis of hip (n = 55) [mean ± SD, median (range) or number (percent)]</th>
<th>The group of patients with hip fracture (n = 50) [mean ± SD, median (range) or number (percent)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>65 ± 12; 53 (32–85)</td>
<td>75 ± 10; 76 (47–89)</td>
</tr>
<tr>
<td>Female</td>
<td>32 (58%)</td>
<td>37 (74%)</td>
</tr>
<tr>
<td>Number of comorbid diseases</td>
<td>1 ± 1; 1 (0–4)</td>
<td>2 ± 1; 2 (0–4)</td>
</tr>
<tr>
<td>Number of used drugs</td>
<td>2 ± 2; 2 (0–8)</td>
<td>3 ± 2; 3 (0–9)</td>
</tr>
<tr>
<td>Shortened mental test score (Siberian version)</td>
<td>10 ± 0; 10 (10–10)</td>
<td>9.84 ± 0.51; 10 (8–10)</td>
</tr>
<tr>
<td>New Mobility Score</td>
<td>7 ± 2; 6 (2–9)</td>
<td>7 ± 2; 9 (1–9)</td>
</tr>
<tr>
<td>Limited walking distance</td>
<td>41 (74.5%)</td>
<td>26 (52%)</td>
</tr>
<tr>
<td>Aids when walking</td>
<td>28 (51%)</td>
<td>16 (32%)</td>
</tr>
<tr>
<td>Up and down stairs with difficulty:</td>
<td>51 (93%)</td>
<td>32 (64%)</td>
</tr>
<tr>
<td>Lives in the flat without elevator</td>
<td>18 (33%)</td>
<td>14 (28%)</td>
</tr>
<tr>
<td>Lives alone</td>
<td>7 (13%)</td>
<td>10 (20%)</td>
</tr>
<tr>
<td>Hospital stay (day)</td>
<td>7.44 ± 1.08; 7 (7–12)</td>
<td>8.52 ± 3.40; 7 (7–24)</td>
</tr>
<tr>
<td>Rehabilitation (day)</td>
<td>5.25 ± 0.78; 5 (5–10)</td>
<td>6.20 ± 2.28; 5 (5–16)</td>
</tr>
<tr>
<td>5 days of rehabilitation</td>
<td>46 (84%)</td>
<td>33 (66%)</td>
</tr>
</tbody>
</table>

**Results**

Out of a total of 120 patients included in the study, 15 patients (10 with hip fracture and 5 with osteoarthritis of the hip) were excluded during the study: 2 patients with intertrochanteric fracture due to poor operative stabilization of the fracture and orthopedic surgeon recommendations to rest after surgery, 2 patients with hip fracture due to cardiac disorders and recommendations of cardiologists to delay mobilization, 3 patients (2 with hip fracture and one with osteoarthritis) because of debilitating diarrhea, severe electrolyte imbalances and extreme hypotension, so the physiatrist recommended postponing initiation of early rehabilitation, in 1 patient with hip fracture and with symptoms of pulmonary embolism, early rehabilitation was interrupted in the first days after surgery as recommended by pulmonologists, 4 patients died in the first days after the surgery (3 patients with hip fracture and one with osteoarthritis of the hip), 3 patients with osteoarthritis had no completely collected data (hospital discharge was performed before the seventh day after surgery).
higher for all items, indicating very good inter-observer reproducibility.

For the item that assesses walking, kappa coefficient was very high and almost equalized in all three tests (Table 4). This study investigated the reliability of the A-test in the assessment of functional recovery of patients treated surgically due to hip fracture and osteoarthritis in an orthopedic department. Internal consistency analysis is an integral part of estimating the test reliability. According to the Cronbach alpha values, the A-test has excellent internal consistency (alpha = 0.98). Although all the A-test items strongly correlated with the total score, it is evident that the correlation magnitude of the item that estimates walking up- and downstairs with a total score is slightly lower than the others. This is the most difficult activity in early rehabilita-

Table 2

Reproducibility of the A-test’s 10 items

<table>
<thead>
<tr>
<th>Items</th>
<th>Observed kappa</th>
<th>Std. error</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>From supine to side lying</td>
<td>0.86</td>
<td>0.04</td>
<td>0.79–0.94</td>
</tr>
<tr>
<td>From supine to sitting</td>
<td>0.81</td>
<td>0.04</td>
<td>0.73–0.90</td>
</tr>
<tr>
<td>From sitting to standing</td>
<td>0.89</td>
<td>0.03</td>
<td>0.84–0.94</td>
</tr>
<tr>
<td>Standing</td>
<td>0.88</td>
<td>0.03</td>
<td>0.81–0.93</td>
</tr>
<tr>
<td>Back to bed</td>
<td>0.87</td>
<td>0.03</td>
<td>0.81–0.93</td>
</tr>
<tr>
<td>Walking</td>
<td>0.86</td>
<td>0.03</td>
<td>0.81–0.93</td>
</tr>
<tr>
<td>Use of toilet</td>
<td>0.91</td>
<td>0.03</td>
<td>0.85–0.97</td>
</tr>
<tr>
<td>Sit on and get up a chair</td>
<td>0.84</td>
<td>0.04</td>
<td>0.77–0.92</td>
</tr>
<tr>
<td>Up and down stairs</td>
<td>0.92</td>
<td>0.04</td>
<td>0.84–1.00</td>
</tr>
<tr>
<td>Walking endurance</td>
<td>0.85</td>
<td>0.04</td>
<td>0.78–0.92</td>
</tr>
</tbody>
</table>

Table 3

Reproducibility of items that assess walking with three different scales of the A-test, the University of Iowa Level of Assistance Scale (ILAS) and the Cumulated Ambulation Score (CAS)

<table>
<thead>
<tr>
<th>Tests (score range)</th>
<th>Observed kappa</th>
<th>Std. error</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A-test (0–5)</td>
<td>0.86</td>
<td>0.03</td>
<td>0.81–0.93</td>
</tr>
<tr>
<td>2. ILAS (0–6)</td>
<td>0.86</td>
<td>0.03</td>
<td>0.80–0.91</td>
</tr>
<tr>
<td>3. CAS (0–2)</td>
<td>0.88</td>
<td>0.04</td>
<td>0.79–0.97</td>
</tr>
</tbody>
</table>

Discussion

This study investigated the reliability of the A-test in the assessment of functional recovery of patients treated surgically due to hip fracture and osteoarthritis in an orthopedic department. Internal consistency analysis is an integral part of estimating the test reliability. According to the Cronbach alpha values, the A-test has excellent internal consistency (alpha = 0.98). Although all the A-test items strongly correlated with the total score, it is evident that the correlation magnitude of the item that estimates walking up- and downstairs with a total score is slightly lower than the others. This is the most difficult activity in early rehabilita-

cient was very high for all three items (0.92–0.97) 5. On the intertester reliability study of the CAS, kappa coefficient was slightly more complex than the scale of the CAS which could partly explain the differences in reliability obtained in these three studies. However, a disadvantage of this study is that we used the Pearson's correlation coefficient for statistical analysis and presentation of the results. In addition, we did not examine another form of reliability – internal consistency.

Now we find that interobserver reproducibility of the A-test is very good (kappa coefficient was 0.81–0.92). By interpretation of Landis and Koch, the A-test is found in the same gradation of reproducibility as the CAS, which has the simplest scale and strong evidence of highly reliable test for recovery assessment of elderly patients with hip fracture 3. In their intertester reliability study of the CAS, kappa coefficient was very high for all three items (0.92–0.97) 5. On the other hand, the results of reliability for the A-test in this study seem to have the advantage over moderate interobserver reliability of the ILAS which was demonstrated in patients after hip and knee arthroplasty (0.48–0.78) 7. The scale of the A-test is simpler than the scale of the ILAS and slightly more complex than the scale of the CAS which could partly explain the differences in reliability obtained in these three studies. However, it is known that Cohen's kappa, weighted for ordinal data does not allow comparability between studies and scales 21. That is why we have selected an item that assessed walking. This activity is assessed in all three tests, but with different scales. Interestingly, the agreement between the two examiners was not much affected by the complexity of the scale and kappa coefficient was almost the same for all three tests.

There are methodological limitations in this paper that we emphasize on this occasion. In analyzing the data, we calculated Cronbach's alpha. Due to the nature of the data obtained by the A-test measuring, this is not the most appropriate statistical method 21. However, this indicator of reliability is requested in estimates used to determine whether a test was examined adequately from all aspects 11, 18, 19. This was why we showed it in this paper. Also, in examining interobserver reliability we limited our study to the assessment of only one pair of examiners. We were not able to avoid this limitation of the study in the situation when the research was adjusted with the possibilities of routine work in the department. However, the results showed that estimates agreement of an experienced therapist and other therapists with no special preparation for the use of the A-test was very good. This situation corresponds to a real everyday practice for which we recommend the A-test.

Patients with various injuries and diseases of the lower extremities are treated surgically in an orthopedic ward of a general hospital. But, after surgery, they have a similar form of physical disability by International classification of functioning, disability and health (ICF). Therefore, there is a need for a single test that would facilitate functional recovery monitoring of patients in an orthopedic ward. Two large groups of patients with surgical treatment distinguish in the heterogeneous population of patients in our Orthopedic Department: patients with hip fracture and patients with hip osteoarthritis. We chose this mixed study population to test reliability of the A-test because we wanted to show that the A-test could be a reliable tool in this situation. Instead of using separate tests for different clinical entities, one can use a single test that is reliable in both cases.

Each test should be evaluated from several aspects 2 and we are preparing the results of validity, diagnostic test accuracy and practical applicability of the A-test in this same study population. However, we believe that future research should focus on other clinical entities present in an orthopedic ward. Moreover, the recovery of every patient who experienced sudden functional disability due to illness or injury could be monitored by the A-test during early rehabilitation. This year, we started to use the A-test outside the orthopedic department. Future research could be focused on the usefulness of this test in early rehabilitation of patients in departments of neurology, neurosurgery, cardiology and cardiac surgery, plastic and vascular surgery.

Conclusion

This study showed that the A-test could be a reliable instrument for monitoring functional recovery of patients surgically treated for hip fractures and osteoarthritis of the hip during early rehabilitation in an orthopedic ward.

Acknowledgements

The authors thank Miodrag Ćolić, Ph.D., Academician, the Dean of the Faculty of Medicine of the Military Medical Academy, University of Defence, Belgrade for making comments on the manuscript and Ljiljana Mihajlovic, PT, and Sladana Zec, PT, for their assistance in data collection.

REFERENCES


Received on January 18, 2013.
Revised on April 17, 2013.
Accepted on April 22, 2013.