Robot-Assisted Assessment in Neurological Rehabilitation of the Arm
Verena Klamroth-Marganska1,2, Sabine Schölch1,2, Urs Keller1,2,3, Armin Curt2, Robert Rienner1,2
1 Sensory-Motor Systems Lab, ETH Zurich, Switzerland
2 Balgrist University Hospital, University of Zurich, Switzerland
3 Pediatric Rehab Research Group, Rehabilitation Center, University Children's Hospital Zurich, Affoltern am Albis, Switzerland

Abstract
Background: ARMin is an exoskeletal robot for arm rehabilitation. Besides neurorehabilitative training it allows for automation of clinical assessments. A feasibility study on patients with spinal cord injury (SCI) was conducted in order i) to quantify the inter-rater-reliability of ARMin assessments and ii) to analyze correlations between clinical tests and automated assessments with the therapy robot ARMin.

Methods: Five patients with incomplete tetraplegia after SCI performed a battery of robotic assessments (ROM, WORKSPACE, FORCE, RESISTANCE TO PASSIVE MOVEMENT RPM, and QUALITY OF MOVEMENT QM) and clinical tests (active range of motion with goniometer, modified Ashworth Scale, Tardieu Scale, Van Lieshout Test, GRASSP, MMT of the upper limb) with both arms, on two occasions, by two different assessors. By calculating the Spearman’s rank correlation coefficient $r_s$, robotic and clinical test results were correlated and the inter-rater reliability was determined.

Results: Significant correlations with clinical tests were found for specific joints in ROM and most muscles in FORCE. Meaningful correlations with workspace were found in WORKSPACE. As there was hardly any joint stiffness detectable in the tested patients, RPM could not be evaluated. Significant results could be shown for most items in FORCE, ROM and WORKSPACE, in particular joints for RPM and particular movements in QM.

Conclusions: We identified a set of parameters that can be reliably measured with ARMin and used for standardized, reliable and repeatable assessment of the arm. A subset of the assessment routines have shown a reasonable validity and can be applied as surrogate markers regarding the clinical value of interventions.

1. Introduction
Robotic devices are effective tools for motor training in neurorehabilitation. Assessment routines integrated into the robotic training can assist in diagnosis and detect the therapy progress, identify the impairment level and adapt the therapy to optimize the therapy outcome, eventually contributing to an optimization of rehabilitation. There are several approaches for diagnostic tools either specifically developed for assessment or integrated into therapeutic robots, e.g. stiffness assessments [1,2] We developed robot-assisted assessment routines that were implemented into the ARMin robot for objective and quantitative evaluation of biomechanical arm functions in patients with neurological deficits, e.g. spinal cord injury (SCI).

2. Methods
a. ARMin Assessments
The assessments were implemented in the arm therapy robot ARMin which has been developed by the group of R. Rienner at ETH Zurich, and V. Dietz/A. Curt at the University Hospital Balgrist. [3] ARMin is an exoskeleton-based robot with 7 degrees of freedom allowing 3D shoulder rotation, elbow flexion/extension, pro/supination of the lower arm and wrist flexion/extension. A hand actuation module is added to support the opening and closing of the hand. ARMin is equipped with precise position sensors and backdrivable motors. Combined with virtual reality (VR) scenarios ARMin is interfaced with a virtual reality environment that allows performing games and activities of daily living. [4] ARMin is interfaced with a display that provides audiovisual feedback about the assessment tasks and patient’s performance.

The following assessment routines were developed and implemented into ARMin (see Fig.1):

- ROM assessment: measures active of the arm. ARMin moves the patient’s arm into a predefined position. The joint to be assessed is free in movement while the position of the remaining axes is fixed. The patient moves the limb in each single joint to the individual limit and ARMin quantifies the maximum angle in degrees ($^\circ$). The positions were defined by the “standardized neutral-0-method positions”. Tested were seven joint movements, namely shoulder flexion/-extension, abduction/-adduction, horizontal ab/-adduction and external/ internal rotation; elbow flexion/-extension; lower arm supination / pronation; and wrist flexion/-extension.

- WORKSPACE assessment: measures achieved workspace. The patient has to reach walls in a VR room. When accomplished the virtual room grows stepwise to a maximal volume that corresponds to 140dm$^3$ in real world.
FORCE assessment: measures maximum isometric force in the aforementioned seven joint directions. ARMin moves the patient’s arm in a fixed predefined position, all axes are blocked and the patient is asked to apply a maximal torque in the joint of interest. The force is measured for a period of 4 seconds and a moving average filter is applied to reduce the effect of single force peaks in the beginning of the measurement.

QM assessment: measures the quality of movement while reaching for targets by values such as distance-to-path ratio while moving, deviation on target position and number of speed peaks. Eight target positions are arranged in a circle around the starting position and appear successively. The distance of the target positions to the start position is based on the results of the WORKSPACE assessment. The patient is asked to move directly to the target position, rest on the target for three seconds and move directly back to the starting position. The trajectory of the movement and values for each target such as distance-to-path ratio are recorded. The number of speed peaks is calculated. Mean values for the four quadrants of frontal plane and over all targets were calculated.

RPM assessment: measures resistance to passive movement. The RPM assessment quantifies the torque necessary to move a single joint at two fixed velocities (30°/s, 60°/s) while the patient is instructed to keep the arm relaxed and passive. The torques are correlated to the corresponding angles of each joint. A linear function is fitted using the least squares method to calculate the slope. We interpreted a value > 0 as an indicator of stiffness.

Fig. 1 Visual display for a) WORKSPACE-, b) ROM- and c) QM-assessments

b. Clinical Assessments

In order to evaluate the validity the robotic assessments we correlated results to scores in established clinical tests, namely the manual ROM with a goniometer, the arm-reachable workspace [5], the van Lieshout test (VLT), The “Graded and Redefined Assessment of Strength, Sensibility and Prehension” (GRASSP), the modified Ashworth Scale (mAS), the Tardieu Scale, and the Manual Muscle Test (MMT) (see Table 1).

<table>
<thead>
<tr>
<th>ARMin assessment</th>
<th>Clinical assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROM assessment</td>
<td>manual ROM</td>
</tr>
<tr>
<td>WORKSPACE assessment</td>
<td>arm-reachable workspace, VLT, GRASSP</td>
</tr>
<tr>
<td>RPM assessment</td>
<td>modified Ashworth Scale, Tardieu Scale</td>
</tr>
<tr>
<td>FORCE assessment</td>
<td>MMT of corresponding muscles</td>
</tr>
<tr>
<td>QM assessment</td>
<td>VLT, GRASSP</td>
</tr>
</tbody>
</table>

Table 1 ARMin assessments and corresponding clinical tests

c. Statistics

For analysis of construct validity of the robotic assessments we calculated the Spearman’s rank correlation coefficient $r_s$ for single ARMin assessments and respective clinical tests (see Table 1). For evaluation of inter-rater reliability of the ARMin assessments, the Spearman’s rank correlation coefficient $r_s$ for the results of both testers was calculated.
3. Results

a. ROM assessment

Inter-rater reliability
For active shoulder abduction/adduction, horizontal abduction/adduction and internal rotation, pronation/supination and wrist flexion/extension the Spearman's rank correlation coefficient ($r_S$) between tester 1 and tester 2 ranged between 0.52* and 0.96**. Elbow data was not available. The reminding joints showed no significant results.

Validity
Spearman’s rank correlation coefficient $r_S$ was significant for the active range of motion in shoulder abduction (0.63*), shoulder flexion (0.59*), elbow flexion (0.81**), wrist extension (0.78**) and wrist flexion (0.67*).

b. WORKSPACE assessment

Inter-rater reliability
Spearman’s coefficient $r_S$ showed high correlations for the number of reached target levels and ranged between .75* and 1.0**.

Validity
Spearman’s coefficient $r_S$ between the calculated workspace of the ARMin assessments and the calculated „arm reachable workspace” was .69. The ARMin workspace correlates with the global test score of GRASSP (.53), and with the global score of VLT (.54).

c. RPM assessment

Inter-rater reliability
Spearman’s coefficient $r_S$ between testers was only significant for shoulder flexion at speed 30°/s and 60°/s and for shoulder internal rotation at 30°/s. Several other joints (shoulder extension at 60°/s; horizontal abduction at 60°/s, external rotation at 30°/s and internal rotation at 60°/s, elbow flexion for both speeds, and supination at 60°/s) showed moderate, but non-significant correlations between .52 and .77.

Validity
As we had only few patients with clinical relevant spasticity, only shoulder extension, internal rotation, elbow flexion and wrist extension reached values higher than “0” in the Tardieu and/or modified Ashworth scales. For these joints there were low to moderate correlations to the ARMin measurements (see Table 1). Lower speeds showed higher correlation coefficients than higher speeds.

<table>
<thead>
<tr>
<th>Movement</th>
<th>RPM at 30°/s</th>
<th>RPM at 60°/s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tardieu</td>
<td>Ashworth</td>
</tr>
<tr>
<td>shoulder extension</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>internal rotation</td>
<td>.76*</td>
<td>.76*</td>
</tr>
<tr>
<td>Elbow flexion</td>
<td>.63*</td>
<td>.66*</td>
</tr>
<tr>
<td>Wrist extension</td>
<td>.20</td>
<td>.20</td>
</tr>
</tbody>
</table>

Table 2 Spearman's correlation coefficient for RPM and clinical spasticity scales

d. FORCE assessment

Inter-rater reliability
Spearman’s coefficient $r_S$ showed high correlations with values between .6 and .97** for all movements but supination.

Validity
Spearman’s coefficient $r_S$ with MMT of corresponding muscles was high and ranged between .69* and .91** for all movements but shoulder extension which was not significant.

e. QM assessment

Inter-rater reliability
Spearman’s coefficient $r_S$ was low to moderate for the mean distance to path ratio, the mean deviation on target and the mean number of peaks.
<table>
<thead>
<tr>
<th>QM assessment items</th>
<th>Spearman's correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean distance to path ratio</td>
<td>.286</td>
</tr>
<tr>
<td>Mean deviation on target</td>
<td>.679</td>
</tr>
<tr>
<td>Mean number of peaks</td>
<td>.464</td>
</tr>
</tbody>
</table>

Table 3 Inter-rater reliability ($r_S$) for mean values in QM assessment

Validity
The eight targets in QM were split into quadrants. The Distance-to-Path-Ratio for quadrants 1 and 2 showed little to moderate correlations with the clinical tests (VLT and GRASSP). The targets in quadrant 3 correlated negatively with the clinical tests. The deviation on target in quadrants 2 and 3 correlated moderately to high with the clinical tests. The other values were not informative.

4. Discussion

From a battery of assessments we determined the relevant tests that show reliable inter-rater reliability and clinical validity.
The FORCE assessment proofs useful for the measurement of muscle force. It detects even small changes which are not reflected in the manual muscle score.
WORKSPACE and ROM are valuable tools to measure the ROM and workspace achievable in ARMin. In moderately to severely affected patients it can give clinically useful results, while in less severely affected patients the mechanical limits of the robot prevent a reliable measurement. For calibration issues for the subsequent assessments, and for therapeutic robotic games these assessments are reliable and feasible.
Concerning QM, differences in the correlation to clinical tests were dependent on the location of targets. It is questionable whether the clinical scores (GRASSP and VLT) are precise enough to describe the characteristics measured within the QM. Although the QM assessments are not significant with the low number of patients, we observed a strong tendency.
The RPM assessment may be used as an indicator for arm stiffness or spasticity of the arm. We could not reliably validate the tool as our patients showed little spasticity.

5. Conclusion

The clinical evaluation of arm functions in patients suffering from SCI can be complemented and improved by robot-assisted assessments. We identified a set of parameters that can be reliably measured with ARMin and used for standardized reliable and repeatable assessment of the arm. We identified complex measures of ROM, WORKSPACE and FORCE that provide a meaningful inter-rater reliability. The identified assessment routines have shown a reasonable validity and can be applied as surrogate markers regarding the clinical value of interventions. Finally, the seamless integration of such measures into VR augmented training provides the ability of an instant monitoring of progress during rehabilitation.

6. Funding

The study was supported by the National Center of Competence in Research Neural Plasticity and Repair (NCCR), Swiss National Fonds (SNF), Switzerland.

7. Literature


Contact: Dr. med. Verena Klamroth. Verena.klamroth@hest.ethz.ch, Tel.: +41 44 632 28 40