

RESEARCH ARTICLE

# Robot-Assisted Arm Assessments in Spinal Cord Injured Patients: A Consideration of Concept Study

Urs Keller<sup>1,2</sup>\*, Sabine Schölch<sup>1,2</sup>, Urs Albisser<sup>2</sup>, Claudia Rudhe<sup>2</sup>, Armin Curt<sup>2</sup>, Robert Riener<sup>1,2</sup>, Verena Klamroth-Marganska<sup>1,2</sup>

**1** Sensory-Motor Systems Lab, Department of Health Sciences and Technology ETH Zurich, Zurich, Switzerland, **2** Balgrist University Hospital, University of Zurich, Zurich, Switzerland

© These authors contributed equally to this work.

\* [urs.keller@hest.ethz.ch](mailto:urs.keller@hest.ethz.ch)



**OPEN ACCESS**

**Citation:** Keller U, Schölch S, Albisser U, Rudhe C, Curt A, Riener R, et al. (2015) Robot-Assisted Arm Assessments in Spinal Cord Injured Patients: A Consideration of Concept Study. PLoS ONE 10(5): e0126948. doi:10.1371/journal.pone.0126948

**Academic Editor:** Hatem E Sabaawy, Rutgers-Robert wood Johnson Medical School, UNITED STATES

**Received:** December 29, 2014

**Accepted:** April 9, 2015

**Published:** May 21, 2015

**Copyright:** © 2015 Keller et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** We changed our online repository from sourceforge to GitHub (<https://github.com/kelleru/Robot-assisted-assessment-sci/tree/v1.0>), such that we can provide a DOI for the database. The DOI is [10.5281/zenodo.16933](https://doi.org/10.5281/zenodo.16933).

**Funding:** This study was funded by the Swiss National Science Foundation (SNF) in the framework of the National Centre of Competence in Research (NCCR, 51NF40-111381) transfer project RANA (Robot-Assisted Neurorehabilitation of the Arm). The funders had no role in study design, data collection

## Abstract

Robotic assistance is increasingly used in neurological rehabilitation for enhanced training. Furthermore, therapy robots have the potential for accurate assessment of motor function in order to diagnose the patient status, to measure therapy progress or to feedback the movement performance to the patient and therapist in real time. We investigated whether a set of robot-based assessments that encompasses kinematic, kinetic and timing metrics is applicable, safe, reliable and comparable to clinical metrics for measurement of arm motor function. Twenty-four healthy subjects and five patients after spinal cord injury underwent robot-based assessments using the exoskeleton robot ARMin. Five different tasks were performed with aid of a visual display. Ten kinematic, kinetic and timing assessment parameters were extracted on joint- and end-effector level (active and passive range of motion, cubic reaching volume, movement time, distance-path ratio, precision, smoothness, reaction time, joint torques and joint stiffness). For cubic volume, joint torques and the range of motion for most joints, good inter- and intra-rater reliability were found whereas precision, movement time, distance-path ratio and smoothness showed weak to moderate reliability. A comparison with clinical scores revealed good correlations between robot-based joint torques and the Manual Muscle Test. Reaction time and distance-path ratio showed good correlation with the “Graded and Redefined Assessment of Strength, Sensibility and Prehension” (GRASSP) and the Van Lieshout Test (VLT) for movements towards a predefined position in the center of the frontal plane. In conclusion, the therapy robot ARMin provides a comprehensive set of assessments that are applicable and safe. The first results with spinal cord injured patients and healthy subjects suggest that the measurements are widely reliable and comparable to clinical scales for arm motor function. The methods applied and results can serve as a basis for the future development of end-effector and exoskeleton-based robotic assessments.

and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Patients who suffer from a neurological disorder such as spinal cord injury (SCI) or stroke often face deficits in motor function. The global-incident rate for traumatic SCI is estimated to be 23 cases per million people (180'000 per year) [1]. Stroke has a prevalence of approximately 795'000 people in the US (Center of Disease Control and Prevention, 2010). These impairments due to stroke or SCI lead to a restriction of both independence and participation in daily life [2, 3]. An intensive rehabilitative intervention can help to improve motor function in stroke [4] and SCI patients [5] and, eventually, the patient's quality of life.

Plenty of clinical scores and assessments are available for different diseases, ages, movements and body parts to measure patient's motor functions. The assessments are often categorized using the international classification of functioning, disability and health (ICF) [6] to standardize the description of the health status. With this classification the scores can be grouped according to the disability they address, i.e., body functions and structure, activities and participation. Assessments covering these groups can be used for diagnosis of the patients' status, as measurement of therapy progress or as feedback about patients' performance. However, clinical assessments often show deficits in terms of reliability, validity, sensitivity and duration of execution [7].

Rehabilitation robots have the potential to provide an interface for objective, sensitive and reliable measurements. The prevalent use of robots for therapy and the positive findings of robot-assisted therapy contributed to an increased development of robot-assisted assessments in the last five years. Generally two fundamental approaches can be used to evaluate sensorimotor impairment using robot-assisted assessments: Using raw sensor data or feature extraction [7].

The first approach uses **raw sensor data** to directly extract information from sensors about body functions. Depending on the used sensors and parameters, different robot-assisted assessments have already been described. Several approaches focus on assessments of the upper extremity. In time-based assessments the duration is usually measured that is needed to finish a given point-to-point movement or position adjustment of the hand or a joint (e.g. using the MIT Manus [8], the Delta robot [9], the REAplan [10] or the HapticKnob [11]) or by measuring the time needed for a given task (e.g. using the PHANTOM [12] or the MIT Manus [8]). With sensors that measure kinematic or kinetic information, assessments can be performed such as measuring the joint range of motion (ROM) or the workspace (work area) of the hand that can be reached (e.g. using the Lokomat [13], the ACT<sup>3D</sup> [14], the ArmeoPower [15] or the Microsoft Kinect [16]) or the mean/peak/tangential speed (e.g. using the MIT Manus [8], the MEMOS [17], the IE2000 haptic joystick [18] or the REAplan [10]). An assessment device which can record forces or torques can be used to measure the active joint strength. This can be done recording the maximum voluntary isometric forces or torques (e.g. using the ARMin [19], its commercial version the ArmeoPower [15] or the Lokomat [20]) or isokinetic forces and torques (e.g. using the Kin-Com Dynamometer [21]).

In the **feature extraction** approach the sensor data is further processed, conditioned and characteristic properties are extracted. Using the time and position information during a movement, the quality of the corresponding joint or hand trajectory can be analyzed. Smoothness is a prominent metric to estimate the quality of a movement. Different metrics were used to calculate smoothness such as the ratio between mean speed and peak speed [10, 18, 22, 23], different jerk metrics [10, 11, 17, 22, 23, 24, 25], tent metric [23], mean arrest period [23], peak metric [17, 23, 25], number of submovements [26], comparison with an idealized normal speed profile [27], number of directional changes [22] or the spectral arc length metric [25]. Another feature is the hand-path-ratio [8, 10, 18, 28], sometimes referred to as straightness or trajectory error, which measures the deviation from a given (often straight) trajectory that has

to be followed. Moreover, the precision and accuracy of a given targeted movement [17, 27] or the shape accuracy [10, 29] are used as features to describe the quality of a movement. Patients with a neurological disorder often show abnormal synergy [30]. The assessment of abnormal synergies was previously performed using position information [29, 31] or isometric torque values during a motor task [32]. When kinetic information is available in an active device, the resistance to passive movement (RPM) can be extracted as a feature, which can be seen as a measure of spasticity in the measured joint. In this assessment the patient is passive and moved by the isokinetic robot, often at different speeds, while the resistive torque is recorded [33, 34, 35, 36]. Reaction time is a measure for the time the patient needs to initiate a movement. It is usually measured as the time needed between the moving instruction (visual cue, sound) and the movement onset, which is a predefined deviation from the starting position or a velocity threshold [28, 37].

A major challenge is that the robot-assisted assessments are clinically accepted. In clinical environments there already exists a multitude of established clinical assessments and scores that are regularly used and well known by the therapist. A possible approach is, therefore, to reconstruct clinical scores based on the robotic assessment to have an accepted measure understandable for therapists and physicians. Several publications evaluated the correlation between robotic assessments and accepted clinical scores, such as the Fugl-Meyer Assessment ([8], [11], [18], [38]), the Modified Ashworth Scale (MAS) ([8], [11], [13]), the Motor activity log [18], the Action Research Arm Test (ARAT) ([18], [22]), the Jebsen-Taylor Hand Function Test [18], the Graded and Redefined Assessment of Strength, Sensibility and Prehension (GRASSP) [22], Spinal Cord Independence Measure (SCIM) [22], Motor Status Score [8], Motor Power Scale [39], etc.

In this paper we evaluate a set of robot-assisted assessments for the upper extremity in healthy and SCI patients in a stage 1 consideration-of-concept study [40]. Five different assessment packages were implemented that include raw sensor data and feature extraction: ROM (active and passive joint range of motion), WORKSPACE (cubic arm reachable workspace), QOM (quality of arm reaching movements), STRENGTH (isometric joint torques) and RPM (resistance to passive joint movement). The assessments were tested in healthy subjects and SCI patients for applicability, safety, reliability and comparability with clinical scales. The assessments were implemented into the existing therapy robot ARMin III that, through its exoskeleton structure, allows not only for measurement of kinematic- and time-based parameters, but also for kinetic-based measures. The five different assessment packages were combined with a visual interface for an intuitive and standardized execution of the assessments. The implemented assessments are based on measurements used in other devices and are extended and adapted for the use of our robotic setup.

Up to now, most of the robot-assisted arm assessments were tested on stroke patients and the extension of the assessments to the SCI target group was rarely tested. An exception is Zariffa et al [22] who tested different kinematic measures on SCI patients with the passive arm exoskeleton robot ArmeoSpring and Perell et al [33] who investigated muscle tone in SCI patients with an isokinetic dynamometer. However, we are the first to offer a comprehensive measurement of patients' motor function within one single device.

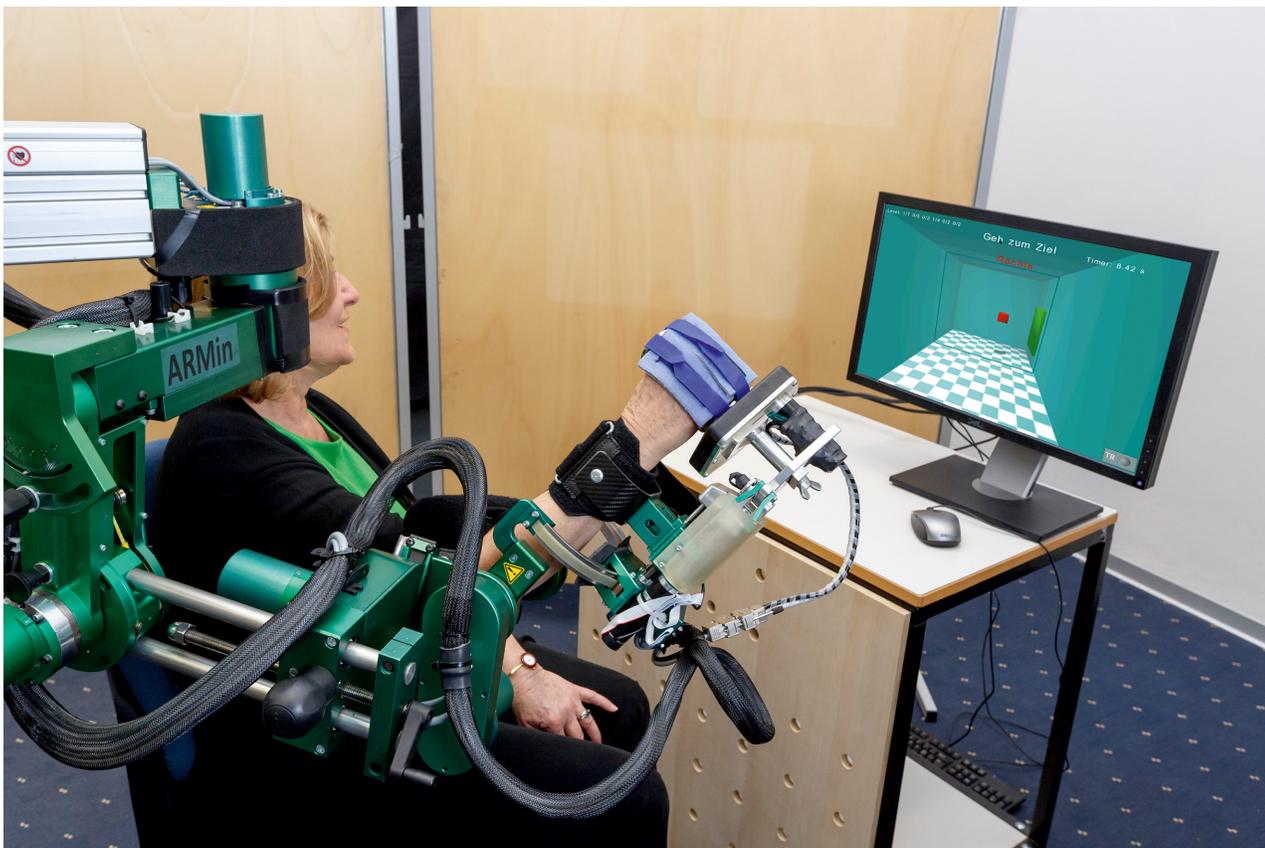
We hypothesized that the ARMin assessment packages provide an applicable, safe, reliable and comparable tool to measure arm motor functions in SCI patients. In order to evaluate the intra-rater reliability, data from ten healthy subjects was collected. To analyze correlations between clinical tests and ARMin assessments and to quantify the inter-rater reliability, a feasibility study on five SCI patients with two different testers was conducted. We believe that the robotic assessments are widely reliable and comparable to clinical scales for arm motor functions. Furthermore, they may offer a sensitive and objective measurement for more detailed insights in arm motor functions.

## Methods

To evaluate the five ARMin assessment packages—namely ROM, WORKSPACE, QOM, STRENGTH and RPM—healthy subjects and SCI patients participated in this study to investigate the four aspects intra-rater reliability, inter-rater reliability, comparison between healthy subjects and patients, and construct validity.

### ARMin arm rehabilitation robot

The arm therapy robot ARMin has been designed and evaluated by the groups of Riener and Dietz/Curt of ETH Zurich and University of Zurich [41, 42]. ARMin is used for neurorehabilitation of the arm and is characterized by an exoskeleton structure (Fig 1). The latest prototype has seven degrees of freedom allowing 3D shoulder rotation, elbow flexion/extension, pro/supination of the lower arm and wrist flexion/extension. A hand actuation module supports opening and closing of the hand. The patient is sitting on a chair and is fixed in the ARMin robot by the use of cuffs around the upper arm, forearm and around the fingers. The robot can be adjusted to the patient by changing the exoskeleton length settings for the upper arm, the forearm and the hand as well as the shoulder height. The same robot can be used for the training of the left and right arm by changing the hardware configuration. Mechanical end limits are provided for safety reasons to not overstretch joints or collide with the patient. ARMin is the research version of the ArmeoPower (Hocoma AG, Switzerland) and the kinematics, joint ranges and



**Fig 1. Subject performing assessments with the ARMin arm robot (Courtesy of Dietmar Heinz).** Published with written informed consent of the individual in the picture.

doi:10.1371/journal.pone.0126948.g001

actuation are comparable. The ARMin robot features different control modes covering the range from a tracking only setting, where the robot’s weight and friction are compensated, to complete guidance of the arm. In between, a path control approach can be used to assist the patient as needed. While the path controller is mainly used for the ADL training, the position controller is either used to move the passive patient arm during the RPM assessment or to fix the arm posture during single joint measurements such as ROM or STRENGTH assessment. The compensation mode is used for QOM and WORKSPACE assessments, where the robot should not interfere with the patient’s movement and only follow the arm to record kinematic data.

With its exoskeletal structure, precise position sensors, mechanical transparency and a visual display, ARMin is particularly useful to assess kinetic and kinematic arm functions on joint and end-effector level.

### Participants

Patients’ were recruited by contacting in- and out-patients of the University Hospital of Balgrist. This hospital located in Zurich, Switzerland, offers specialized treatment for SCI patients and treats about 230 inpatients and about 1’500 outpatients after SCI annually.

Eligibility criteria were i) cervical, complete or incomplete SCI with tetraparesis, ii) no severe subluxation of the shoulder, iii) no severe shoulder pain on the tested side and iv) no other illness or incapability that could compromise the assessments. Five patients (aged 19 to 49, mean 33.8, SD 13.8; 4 male) were eligible to perform the five robotic assessments as well as clinical tests. Three patients were in the subacute (i.e. <= 6 months) and two in the chronic (i.e. > 6 months) state post-SCI with different levels of severity (ASIA B to ASIA D) and arm dominance (four right-handed and one left handed) (Table 1). One out of the five patients reported severe shoulder pain on one side, so the corresponding arm was excluded. As the SCI was incomplete in the five patients and their two arms were affected to different degrees, nine unrelated data sets were sampled for evaluation.

Eleven healthy right-handed subjects (aged 21 to 64, mean 35.4, SD 15.4; 6 male) were assessed on their dominant side to determine the intra-rater reliability. Thirteen healthy right-handed subjects (aged 20 to 68 years, mean 34.9, SD 15.7; 6 male) performed the assessments with their non-dominant arm to compare the performance with the patient’s non-dominant arm. This data was also used to define norm values. None of the subjects had experience with ARMin prior to the study onset.

**Table 1. Characteristics of the patients.**

Patient ID	Status <sup>a</sup>	Level of lesion	Age in years	Sex	Sequence of arm side tested (r = right, l = left)	Hand dominance after accident <sup>b</sup> (r = right, l = left)	Sequence of tester
1	subacute	sub C5, ASIA B	20	m	r, l	r	T1, T2
2	subacute	sub C5, ASIA B	19	m	l, r	r	T2, T1
3	chronic	sub C4, ASIA C	36	f	r, l	l	T1, T2
4	chronic	sub C6, ASIA B	47	m	r, l	r	T2, T1
5	subacute	sub C1, ASIA D	47	m	l	r	T1, T2

<sup>a</sup>Chronic (> 6 months) vs. subacute (<= 6 months post SCI).

<sup>b</sup>Evaluated with questionnaire of Chapman [43].

### Ethics statement

All patients signed an informed consent; the study was approved by the responsible ethical committee (KEK, Zurich, Switzerland) and Swissmedic (Swiss Agency for Therapeutic Products, reference number: 2011-MD-0002). The individual in Fig 1 has given written informed consent (as outlined in PLOS consent form) to publish these case details.

### Study design

In this consideration-of-concept study (stage 1) we aimed to analyze the utility of the chosen assessment measurements and to evaluate how well they can be applied [40]. We tested for applicability, safety, reliability and validity. Forty minutes were scheduled for the performance of the assessment packages, including posing and instruction of the subject.

**Intra-rater reliability with healthy subjects.** The intra-rater reliability was tested with eleven right-handed healthy subjects. A set of norm values was determined. The subjects performed all ARMin assessments four times with their dominant right arm at intervals of one week. Hand dominance was evaluated with the questionnaire of Chapman [43]. No pretest was performed before starting the intra-rater measurement as only minor learning effects were expected. The settings of the robot (arm length settings, shoulder and chair height and position) and the sequence of the ARMin assessments were held constant. Three trained testers performed the measurements, where each subject was always assessed by the same tester.

**Comparison between healthy subjects and patients.** In order to complete a first set of norm values with data from non-dominant arms, 13 healthy subjects performed the assessments once with the non-dominant left hand. Again, three different testers performed the assessments. The acquired norm values for the non-dominant and the dominant arm (values from the first assessment session) were used to compare the assessment performance between the healthy subjects and the SCI patients.

**Inter-rater reliability and construct validity with patients.** For evaluation, inter-rater reliability and correlation between ARMin assessments and common clinical assessments were tested in a cross-sectional study. To assess the inter-rater reliability of the ARMin assessments, two testers performed the assessments. Both trained the ARMin assessments procedure at least ten times and with at least two patients to gain experience with the robot. Testers was given a manual that described the handling of the robot and the instructions to be given to the participants. The order of testers and the order of measured sides of each subject were randomized by lots which the patient drew on day 1. The sequence of ARMin assessments and clinical tests was held constant.

Each patient participated in six sessions (one pretest session, four robotic assessment sessions, one clinical assessments session) that were arranged over a period of twelve days (Table 2). While robotic assessments were conducted by both testers (T1 and T2), all clinical assessments were performed by the same one tester for all patients.

Table 2. Study protocol for the patients.

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7, Day 8 or Day 9	Day 10	Day 11	Day 12
Pretest	-	-	ARMin test 1	ARMin test 2	-	Clinical tests	-	ARMin test 3	ARMin test 4
robot settings, first test, define tester and arm by lot	-	-	arm 1, T1	arm 1, T2	-	aROM, pROM, MAS, MTS, GRASSP, SCIM, VLT, MMT	-	arm 2, T1	arm 2, T2

doi:10.1371/journal.pone.0126948.t002

**Table 3. Overview over implemented assessments and the measured parameters.**

Assessment name	Assessment description	Parameters
ROM	Active and passive ranges of motion of seven joint movements	Active ROM [°] ( <i>aROM</i> ); Passive ROM [°] ( <i>pROM</i> )
WORKSPACE	Actively achievable Cartesian workspace	<i>Workspace levels</i> [# reached levels]/ <i>Cubic volume</i> [dm <sup>3</sup> ]
QOM	Quality of hand movement while performing goal-directed reaching tasks	Distance-path ratio ( <i>D-P ratio to target</i> ) []; Distance-path ratio ( <i>D-P ratio to start</i> ) []; <i>Time to target</i> [ms]; <i>Time to start</i> [ms]; <i>Precision</i> (Deviation on target) [m]; <i>Number of peaks to target</i> []; <i>Number of peaks to start</i> []; <i>Reaction time to target</i> [ms]; <i>Reaction time to start</i> [ms]
STRENGTH	Isometric maximum torque for seven joints	<i>Joint torques</i> [Nm]
RPM	Resistance to passive movements for two different speeds in all seven joints	<i>Joint stiffness</i> [Nm/rad]

doi:10.1371/journal.pone.0126948.t003

### Outcome measures

**Robotic assessments.** Five assessment packages were implemented to evaluate various aspects of arm motor function. While the ROM and the STRENGTH assessment were purely based on raw sensor data, the WORKSPACE, QOM and RPM assessments calculated features that are extracted from the raw data (Table 3). For the QOM and the WORKSPACE assessment the robot is in compensation mode. While most of the gravitational and frictional effects are compensated there may still be forces disturbing the arm movement such as static friction or dynamical effects of the robot.

ROM (range of motion) assessment measured active and passive ranges of motion (*aROM* and *pROM*) of the arm. During this assessment ARMin held the patient’s arm in a predefined posture, while the assessed joint was free to move. The patient (*aROM*) or the therapist (*pROM*) moved the free joint in both directions (e.g. flexion and extension), while the software recorded both achievable extreme positions. The order of joints measured was fixed. The predefined robotic postures depended on the joint measured and were chosen to be as similar as possible to the “standardized neutral-0-method positions” [44] (postures described in S1 Table). Seven different joint movements were assessed: shoulder flexion/extension, lateral shoulder ab-/adduction, horizontal shoulder ab-/adduction, shoulder external/internal rotation, elbow flexion/extension, forearm supination/pronation and wrist flexion/extension.

WORKSPACE assessment aimed to measure the reachable cubic workspace of the end effector (i.e., the hand). The starting position of the hand was 30 cm in front of the breast (i.e., the xiphoid process of sternum). On the screen, a small cubic room (corresponding to an initial size of 20 cm x 20 cm x 20 cm in the real world) was presented (Fig 2). Each wall in the room indicated a direction to move to (top, bottom, left, right, towards the body, away from the body). For simplification, more directions such as diagonals were not assessed. The aimed movement direction was indicated by a green cube on the wall in randomized order. After the patient had reached the indicated wall, the room grew 5 cm in this direction. If the subject missed an indicated wall the room did not grow in this direction and the direction was shown once more later. The number of discrete increases in a certain direction refers to as a level and was used as an outcome parameter (*workspace level*, in numbers). Furthermore, the achieved room size (*cubic volume*, in dm<sup>3</sup>) was calculated from the workspace levels reached. The maximal volume of the room was 140 dm<sup>3</sup>, the initial size was 8 dm<sup>3</sup> and the maximal distances to the given room walls were 35 cm for the left and right (five movements in each direction to reach maximum expansion), 30 cm for the top (four movements to reach maximum expansion), and 20 cm for the bottom direction, towards the body and away from the body (two



**Fig 2. Visual representation of the WORKSPACE assessment package.** Screenshot of the WORKSPACE assessment. A room was presented on the screen. The patient looked directly into this room. The end effector of the robot (the position of the patient's hand) was represented as a small red cube. In the shown situation the patient had to move to the green target position to the right.

doi:10.1371/journal.pone.0126948.g002

movements in each direction to reach maximum expansion). This results in a total of 20 movements to discretely increase the room size to its maximum.

**QOM (Quality of movement) assessment** measured accuracy and smoothness of a point-to-point movement. Eight target positions appeared successively around a starting position in the frontal plane. The starting position was a circle that allowed some end-effector position variations in the range of 50 mm (rated as “on the start position”). The location of the target positions depended on the volume reached in WORKSPACE, i.e., the targets were at 80% of the reachable distance. The patient was asked to move directly from the start to the target position as soon as it appeared, to rest on the target position until it disappeared (after 3 seconds) and then move directly back to the starting position. Output parameters were the distance-path ratio on the way to the target (*D-P ratio to target*), the standard deviation on the target position (*precision*), and the D-P ratio back to start (*D-P ratio to start*). Furthermore, the *reaction time* to initiate a movement (i.e., the time to leave a circle that was chosen 20% wider than the starting position) and the *time to target* were calculated for each movement. For each movement the *number of peaks* of the end-effector speed (i.e., the zero transition of the end-effector acceleration) was counted as a measure of smoothness (using a peak detection threshold of 7.5% of the maximum speed).

STRENGTH assessment measured the maximum isometric torque of the arm (*joint torques*). ARMin moved the patient's arm to a predefined measuring posture similar to the one in ROM, but with the measured joint in the midrange (postures are described in [S2 Table](#)). The posture was fixed during the joint torque measurement. For a baseline measurement the patient stayed passive for 5 s and then applied the maximum possible torque in the measured joint for 5 s. A simple visual display showed the joint of interest and the timing. This procedure was repeated in both directions of each joint. The applied torque was estimated from the motor current. A moving average filter (window size 1 s) was applied to reduce the effect of single force peaks and the maximal *joint torque* was recorded.

RPM (**Resistance to passive movement**) assessment quantified resistance of a single joint to passive movement within the prior measured *pROM*. ARMin moved the arm to a predefined start posture (identical to the postures in STRENGTH except for the measured joint which moves in the *pROM*, [S2 Table](#)). The patient was instructed to keep the arm relaxed and passive while ARMin moved the joint of interest with two different speeds (30°/s, 60°/s). A calibration routine was used to identify the torque contributions that are caused by the robot arm mass. This routine performed the same movements as the RPM assessment but without the patient. A 5<sup>th</sup> order Butterworth filter with 3 Hz cutoff was used to filter the calculated interaction torques. Knowing the torques applied by the patient the characteristic angle-torque relation could be used to calculate the resistance during the movement. It was assumed that this resistance is a combination of a constant offset torque and a stiffness contribution that changes when the joint angle increases or decreases. Therefore, a linear function was fitted into the data using the least squares method. Further analyzed for this paper was the *joint stiffness* using the slope of the linear fit.

**Clinical assessments.** To evaluate the validity of the robot-assisted assessments the results were compared to clinical outcomes. The active and passive ranges of motion were measured in a sitting position with a handheld goniometer and according to the neutral-0-method [\[44\]](#). For better comparison with the measurement of WORKSPACE, the arm reachable workspace (ARW) was calculated with a dedicated program as proposed by Klopčar et al [\[45\]](#). The ARW is based on the subject's body height, the maximal flexion/extension, abduction/adduction and internal/external rotation of the shoulder and the elbow flexion angle in standing position.

The modified Ashworth Scale (MAS) [\[46\]](#) and the modified Tardieu Scale (MTS) [\[47\]](#) rate the degree of spasticity and were measured for all relevant joints. For the MAS the joint is passively moved with a moderate velocity, for the MTS with high velocity. The tests use an ordinal scale in the range of 0 to 5 for the MAS and 0 to 4 for the MTS, where 0 is equivalent to "no spasticity".

The GRASSP [\[48\]](#) is an upper extremity assessment for patients after SCI and combines a muscle test for relevant upper extremity muscles, a Semmes-Weinstein-monofilament-test [\[49\]](#) and qualitative and quantitative grasping tasks. In addition to the muscle tests performed within the GRASSP, the Manual Muscle Test (MMT) [\[50\]](#) was performed on muscles that were not covered by the GRASSP but assessed with ARMin. The MMT was conducted in a sitting position. Arm positions were chosen according to the instructions of Daniels and Worthingham [\[51\]](#) and if necessary adapted to the sitting position.

Furthermore, the SCIM [\[52\]](#), a questionnaire for SCI patients that measures the independence in activities of daily living, was conducted to characterize the patients.

To measure functionality of the arm, the scientific or short version of the Van Lieshout Test (VLT) [\[53\]](#) and specific optional items for proximal function of the arm from the clinical or longer version of the VLT (wheelchair propulsion, transfers, push-ups while seated, stabilization of the arms, reaching low, reaching high) were conducted.

**Table 4. Statistical methods for the analysis of the intra-rater reliability, inter-rater reliability and construct validity.**

Analyzed aspect	Analyzed data		Statistical method
Intra-rater reliability	Parametric and nonparametric data		Friedman test with pairwise multiple comparison
Inter-rater reliability	Parametric data		ICC (two-way mixed model, single measure)
	Nonparametric data		Spearman's rank correlation coefficient
	Parametric and nonparametric data		Bland-Altman plot with one-sample Student's t-test
Construct validity	ROM	manual ROM	Spearman's rank correlation coefficient, Bland-Altman plot
	WORKSPACE	ARW, VLT, GRASSP, SCIM	Spearman's rank correlation coefficient
	QOM	VLT, GRASSP, SCIM	Spearman's rank correlation coefficient
	STRENGTH	MMT	Spearman's rank correlation coefficient
	RPM	MAS, MTS	Spearman's rank correlation coefficient

doi:10.1371/journal.pone.0126948.t004

### Statistical analysis

The feature extraction for the robotic assessments was performed using MATLAB (Mathworks, R2010b). The resulting parameters were exported to IBM SPSS Statistics for further statistical analysis. All the robotic measures were tested for normal distribution with histograms and Q-Q-Plots. The statistical methods used are summarized in (Table 4).

**Intra-rater reliability.** Results of the four repeated assessments in healthy subjects were analyzed with the Friedman test against the null hypothesis that there are no differences. A significant difference ( $p < 0.05$ ) indicates a poor intra-rater reliability. Pairwise multiple comparisons were done by post hoc analysis with Wilcoxon signed-rank tests to determine where the differences between the repeated assessment measures occurred. To analyze the amount of variability between the four measurements the mean difference between the maximum and the minimum value measured was calculated for the aROM, pROM, WORKSPACE, STRENGTH and RPM assessments. For QOM the standard deviation was calculated.

**Comparison between healthy subjects and patients.** To analyze the differences between healthy subjects and patients, a qualitative approach was used calculating the ratio between mean values of the patients and mean values of healthy subjects. The hypothesis was that patients would show a lower performance compared to healthy subjects. Accordingly, for the assessments that produce ascending scores for better performance, the ratio was expected to be lower than 1 in patients. The dominant and non-dominant arms were analyzed independently. For the ROM, rather than comparing the minimal and maximal values of the joint range (e.g.,  $-5^\circ$  and  $130^\circ$ ) the full ranges (for this example,  $135^\circ$ ) were calculated and compared to healthy subjects.

**Inter-rater reliability.** For evaluation of the inter-rater reliability nine patient arms were measured. The Spearman's rank correlation coefficient was calculated for those ARMin assessments which were nonparametric. For parametric data the intraclass correlation coefficient (ICC) was used (two-way mixed model, single measure). A significant correlation in either of these tests indicates a good inter-rater reliability for the assessment.

The differences between the two testers were further analyzed with Bland-Altman plots. The difference of two measurements was plotted against the mean of the two measurements. The one-sample Student's t-test was used with the null hypothesis that the mean difference between the testers is zero. A significant Student's t-test ( $p < 0.05$ ) indicates a significant difference between results of testers 1 and 2. The Bland-Altman plots were calculated without outliers (rejected by visual inspection; number of outliers: ROM: 20/252, WORKSPACE: 0/63, QOM: 31/514, STRENGTH: 7/128, RPM: 4/184).

**Construct validity.** The construct validity was analyzed by comparing ARMin assessment parameters of the patient arms with clinical measurements, using the Spearman's rank correlation coefficient. The ARMin assessments and corresponding clinical tests are listed in [Table 4](#). Mean values of the ARMin assessment parameters of testers 1 and 2 were used. Values which were rated as outliers by visual inspection in the reliability analysis were excluded from validity analysis as well. A correlation of 0.0–0.5 was considered as weak, 0.5–0.75 as moderate and 0.75–1.0 as strong correlation. Correlations with a significance level  $p < 0.05$  are marked with \*, correlations with significance level  $< 0.01$  with \*\* in the text.

## Results

Only STRENGTH values were normally distributed. Nonparametric methods were used for analysis in the other assessments.

### Intra-rater reliability

Intra-rater reliability was calculated from four complete assessment sessions performed in eleven healthy subjects. The results are summarized in [Table 5](#). The amount of variability between the four tests can be seen in [S4 Table](#).

### Comparison between values of healthy subject and patients

The different assessment parameters were compared in the dominant ( $n = 5$ ) and non-dominant ( $n = 4$ ) arms. The results are listed in [Table 6](#). The results for age- and age/gender-matched (only male patients and subjects) comparisons are similar. For better visualization of the differences between patients and healthy subjects in the RPM assessment, the average *joint stiffness* parameters for all joint movements are plotted in [Fig 3](#).

### Inter-rater reliability

The inter-rater reliability was analyzed using the Spearman's rank correlation coefficient and the Bland-Altman plot. In [Fig 4](#) an exemplary Bland-Altman plot for the wrist flexion is

**Table 5. Summary of the significant differences found from the Friedman test of the intra-rater reliability.**

Assessment	Significant differences between the parameters from the four assessment tests
ROM <sup>a</sup>	<i>p</i> ROM in pronation ( $p = 0.03$ )
WORKSPACE <sup>b</sup>	No significant differences
QOM	<i>D-P ratio to start</i> for target 1 ( $p = 0.003$ ) (test 1 → test 3/4); <i>D-P ratio to target</i> for target 2 ( $p = 0.033$ ) (test 1 → test 2); <i>Precision</i> on target 1 ( $p = 0.008$ ) (test 1 → test 3/4); <i>Reaction time to target</i> for target 2 ( $p = 0.041$ ) (not sign. post hoc); <i>Reaction time to target</i> for target 7 ( $p = 0.045$ ) (not sign. post hoc); <i>Reaction time to start</i> for target 6 ( $p = 0.048$ ) (test 2 → test 4)
STRENGTH	Hand opening ( $p = 0.013$ ) (test 1 → test 3)
RPM	Shoulder external rotation 60°/s ( $p = 0.001$ ) (test 2 → test 4); Elbow flexion 60°/s ( $p = 0.018$ ) (not sign. post hoc)

<sup>a</sup>As healthy subjects almost exclusively reached the mechanical limits ROM values correspond in most cases to the mechanical end limit.

<sup>b</sup>The maximal *workspace levels* and therefore the maximal *cubic volume* were reached for the evaluated subjects for all the directions ( $p = 1.00$ ).

doi:10.1371/journal.pone.0126948.t005

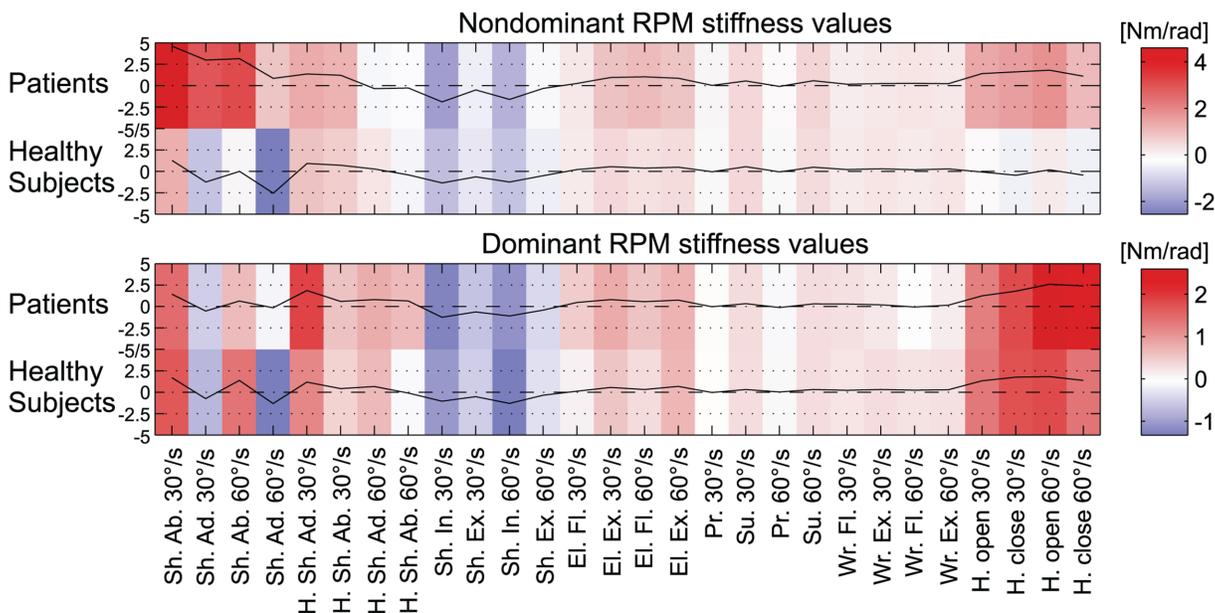
**Table 6. Summary of the comparison between data of patients and healthy subjects.**

Assessment	Patients in comparison with healthy subjects
ROM	<b>Dominant arm:</b> <i>aROM</i> : 81%; <i>pROM</i> : 95%.
	<b>Non-dominant arm:</b> <i>aROM</i> : 67%; <i>pROM</i> : 90%.
WORKSPACE	<b>Dominant arm:</b> <i>cubic volume</i> : 79%.
	<b>Non-dominant arm:</b> <i>cubic volume</i> : 73%.
QOM	<b>Dominant arm:</b> <i>D-P ratio to target</i> : 102%; <i>D-P ratio to start</i> : 101%; <i>time to target</i> : 129%; <i>time to start</i> : 125%; <i>number of peaks to target</i> : 115%; <i>number of peaks to start</i> : 168%; <i>precision</i> : 85%.
	<b>Non-dominant arm:</b> <i>D-P ratio to target</i> : 93%; <i>D-P ratio to start</i> : 99%; <i>time to target</i> : 84%; <i>time to start</i> : 107%; <i>number of peaks to target</i> : 69%; <i>number of peaks to start</i> : 72%; <i>precision</i> : 88%.
STRENGTH	<b>Dominant arm:</b> <i>Joint torques</i> between 23% (hand opening) and 97% (elbow flexion).
	<b>Non-dominant arm:</b> <i>Joint torques</i> between 8% (hand closing) and 97% (supination).
RPM	<b>Dominant arm:</b> The difference between <i>joint stiffness</i> in the patients' arms (mean: 0.49 Nm/rad) and the healthy subjects (mean: 0.34 Nm/rad) is 0.15 Nm/rad.
	<b>Non-dominant arm:</b> The difference between <i>joint stiffness</i> in the patients' arms (mean: 0.72 Nm/rad) and the healthy subjects (mean: -0.08 Nm/rad) is 0.80 Nm/rad.

For *D-P ratio*, *time to start/target*, *number of peaks* and *precision* a value below 100% indicates a better performance of the patients while for *ROM*, *cubic volume* and *joint torque* a value below 100% indicates a better performance of the healthy subjects.

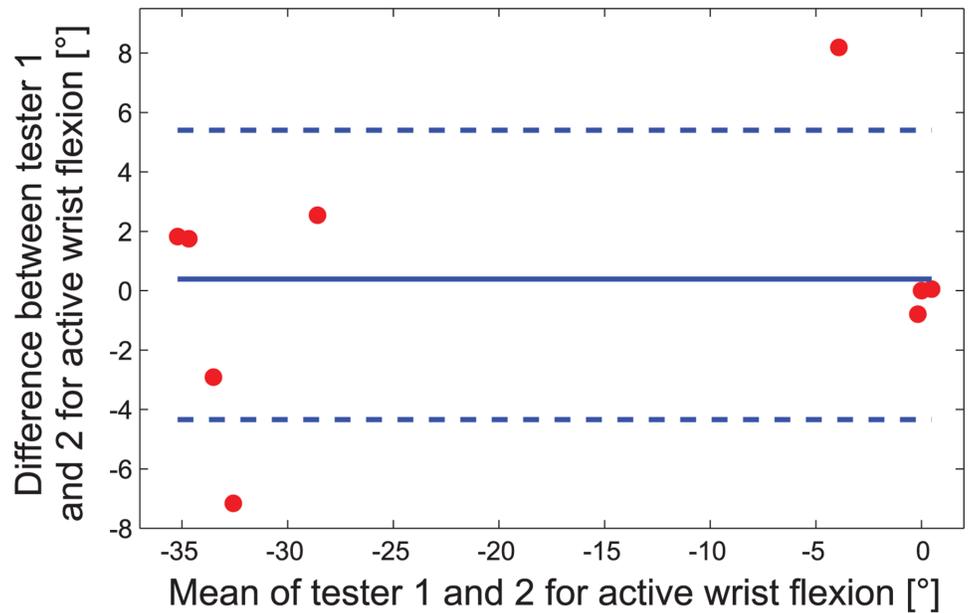
doi:10.1371/journal.pone.0126948.t006

shown. The analysis of the WORKSPACE and QOM parameters is shown in more detail in [S3 Table](#). The mean *cubic volume* was 107.5 dm<sup>3</sup>. [Fig 5](#) is an example of a patient's hand movements to the eight different targets in the QOM assessment package. The summarized results for the inter-rater reliability between testers 1 and 2 are summarized in [Table 7](#).



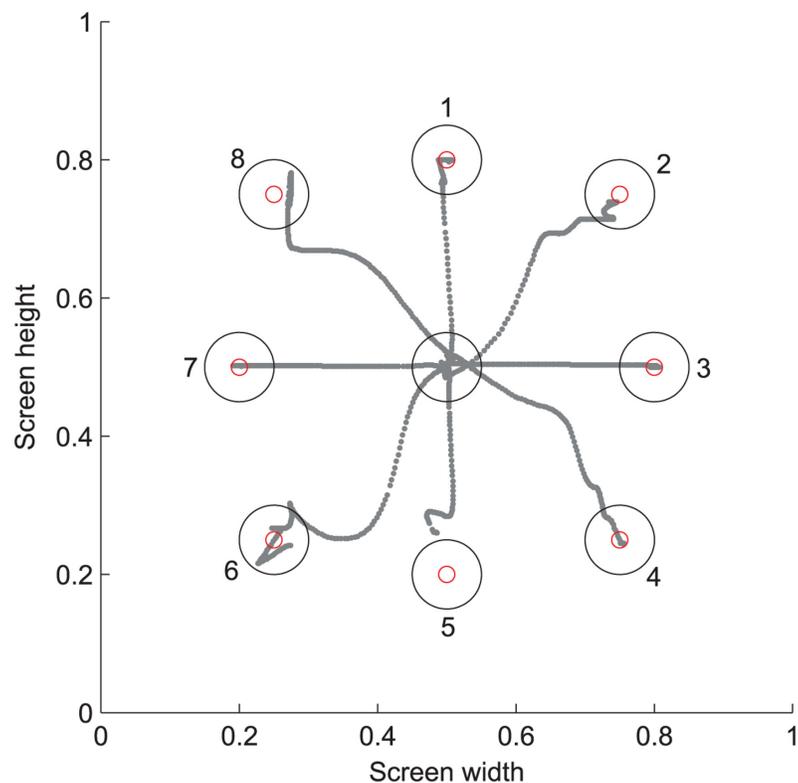
**Fig 3. Visualization of joint stiffness measurements within the RPM assessment for the non-dominant and dominant arms of healthy subjects and patients.** The more a joint counteracted the robot movement over the angle, the higher and more intense in color (red) the value. The color gradient ranges from minimal (blue) to maximal values (red) and is different for the non-dominant (-2 to 4) and dominant arms (-1 to 2).

doi:10.1371/journal.pone.0126948.g003



**Fig 4. Example of an aROM Bland-Altman plot regarding wrist flexion for the nine patient arms.** The limits of agreement (dashed lines for lower  $-4.3^\circ$  and upper  $5.4^\circ$  limit) and the mean difference (solid line at  $0.5^\circ$ ) are shown. The x-axis shows the mean values of the two measurements of tester 1 and 2 (negative values indicate flexion, positive values stand for extension), while the y-axis shows the measurement difference between tester 1 and 2.

doi:10.1371/journal.pone.0126948.g004



**Fig 5. QOM: The hand paths of a patient moving to the 8 different targets.**

doi:10.1371/journal.pone.0126948.g005

**Table 7. Summary for the inter-rater reliability analysis.**

Assessment	Spearman correlation, ICC, Bland-Altman analysis
ROM	Spearman: <i>aROM</i> : Shoulder flexion/extension (0.95/0.85); lateral shoulder ab-/adduction (0.83/0.96); horizontal shoulder ab-/adduction (0.78/0.83); shoulder internal rotation (0.87); supination (0.92); wrist flexion/extension (0.72/0.85). <i>pROM</i> : Lateral shoulder adduction (0.98); shoulder extension (0.92); horizontal shoulder ab-/adduction (0.77/0.90); internal rotation (0.80); wrist flexion (0.67). Bland-Altman: <i>aROM</i> : No significant differences; <i>pROM</i> : Shoulder extension ( $p = 0.047$ ), internal-/external rotation ( $p = 0.033/p = 0.028$ ).
WORKSPACE	Spearman: <i>Workspace levels</i> (between 0.75* and 1.0**); <i>cubic volume</i> (0.77*). Bland-Altman: <i>Workspace levels</i> downwards ( $p = 0.035$ ); <i>cubic volume</i> ( $p = 0.025$ ).
QOM	Spearman: <i>Number of peaks to start</i> : 0.71*. Bland-Altman: No significant difference between testers.
STRENGTH <sup>a</sup>	ICC: Between 0.80** and 0.98** (except for supination <i>joint torque</i> : 0.38). Bland-Altman: Wrist extension ( $p = 0.01$ ).
RPM	Spearman: Shoulder flexion at speed 30°/s and 60°/s; Shoulder internal rotation at 30°/s. Bland-Altman: Elbow flexion at 60°/s ( $p = 0.019$ ).

The numbers show the significant Spearman correlation coefficients, the significant ICC values and the significant differences from 0 in the t-test for the Bland-Altman analysis.

\* =  $p < 0.05$ .

\*\* =  $p < 0.01$ .

<sup>a</sup>Data of testers 1 and 2 were normally distributed for STRENGTH.

doi:10.1371/journal.pone.0126948.t007

### Construct validity

**ROM.** Spearman’s rank correlation coefficient between the clinical ROM measurement and the robot’s ROM was significant for *aROM* in lateral shoulder abduction, elbow flexion and wrist flexion/extension, and showed a moderate but not significant correlation for shoulder flexion (0.59,  $p = 0.06$ ). The other correlations were not significant ( $p = 0.1$  to  $p = 0.43$ ). Regarding *pROM*, only lateral shoulder abduction and horizontal abduction were significant. The results are shown in [Table 8](#).

**Table 8. Summary of the results for the construct validity.**

Robotic assessment	Clinical assessment	Spearman correlation
ROM	manual ROM	<i>aROM</i> : Lateral shoulder abduction (0.63, $p = 0.034^*$ ), elbow flexion (0.81, $p = 0.004^{**}$ ) and wrist flexion/extension (0.78, $p = 0.007^{**}$ /0.67, $p = 0.035^*$ ); <i>pROM</i> : Lateral shoulder abduction (0.78, $p = 0.006^{**}$ ) and horizontal abduction (0.64, $p = 0.031^*$ ).
WORKSPACE	ARW	<i>Cubic volume</i> significantly correlates with ARW (0.69*).
	VLT, GRASSP, SCIM	<i>Cubic volume</i> correlates moderately with the test score of GRASSP (0.53), VLT (0.54) and SCIM (0.51). The <i>workspace levels</i> correlate on a 0.7* level with the VLT and GRASSP items.
QOM	VLT, GRASSP, SCIM	Particularly <i>D-P ratio to start</i> (GRASSP: -0.700*, VLT: -0.644*) and <i>reaction time to start</i> (GRASSP: -0.667*, VLT: -0.636*) show moderate correlations with the clinical scores.
STRENGTH	MMT	The correlations between STRENGTH <i>joint torque</i> and MMT range between 0.69* and 0.91** (except for shoulder extension).
RPM	MAS, MTS	No analysis due to lack of data.

The table shows the significant correlations between the robotic and clinical assessments.

\* =  $p < 0.05$ .

\*\* =  $p < 0.01$ .

doi:10.1371/journal.pone.0126948.t008

**Table 9. Correlations between the QOM assessment and the clinical GRASSP, VLT and SCIM scores.**

QOM metric (Mean values)	GRASSP	VLT	SCIM
<i>D-P ratio to target</i>	-0.450	-0.326	-0.441
<i>D-P ratio to start</i>	-0.700*	-0.644*	-0.653*
<i>Precision</i>	0.217	0.209	0.314
<i>Number of peaks to target</i>	0.233	0.142	0.144
<i>Number of peaks to start</i>	0.000	-0.050	-0.059
<i>Reaction time to target</i>	-0.350	-0.293	-0.398
<i>Reaction time to start</i>	-0.667*	-0.636*	-0.687*
<i>Time to target</i>	0.400	0.427	0.373
<i>Time to start</i>	-0.017	0.000	0.008

Negative values result when scales have diametric changes with improvement.

\* =  $p < 0.05$ .

doi:10.1371/journal.pone.0126948.t009

Bland-Altman plots were generated to evaluate the degree of agreement between clinical and robotic measurements. The mean difference between the two measurement methods were significantly different regarding *aROM* for 6 out of 12 measurements (lateral shoulder adduction, shoulder extension, internal and external rotation, supination and wrist flexion) and regarding *pROM* for 10 out of 12 measurements (lateral shoulder adduction, shoulder extension, horizontal ab-/adduction, internal and external rotation, elbow flexion, pronation and wrist flexion/extension). This difference results from predominantly higher clinical ROM values compared to the ranges measured in the robot. This was also reflected in correlations between the assessment differences and the mean values in the evaluation of the Bland-Altman plots, i.e. the difference was higher, for higher mean values.

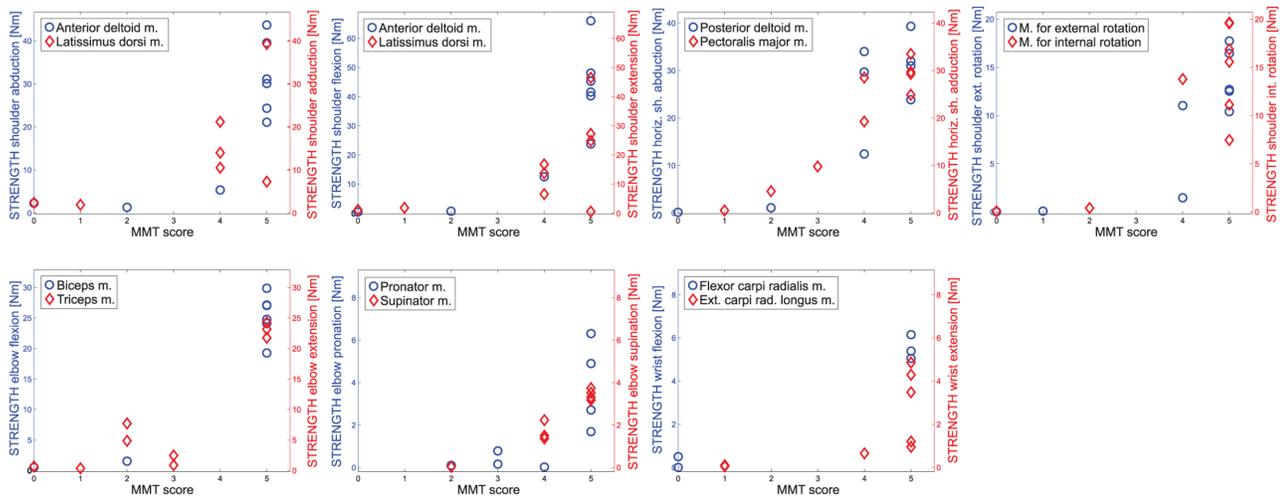
**WORKSPACE.** The arm reachable workspace (ARW) was calculated from ROM values measured with a goniometer. Spearman’s correlation coefficient between *cubic volume* and the ARW was 0.69\*. The *cubic volume* was approximately half the size of the ARW.

*Cubic volume* correlates on a 0.5–0.6 level with VLT and the single prehension items of the GRASSP but not with the sensibility items of GRASSP. The correlation of *cubic volume* was 0.53 for the GRASSP, 0.54 for the VLT and 0.51 for the SCIM. The *workspace levels* in the different ARMin directions correlated on a 0.7\* level with VLT and GRASSP items, and on a 0.8\* level with the targets in the “up” direction, but not for the sensibility items.

**QOM.** Overall, results regarding correlations of QOM with GRASSP, VLT and SCIM were inconsistent but there was a tendency that patients with higher values in GRASSP, VLT and SCIM got better results in QOM values (Table 9). Particularly, *D-P ratio to start* and *reaction time to start* showed moderate correlations with the clinical scores.

**STRENGTH.** The Spearman’s correlation coefficients between STRENGTH and the manual muscle test of corresponding joints were very high and ranged between 0.69\* and 0.91\*\*. The corresponding data for all joints is plotted in Fig 6. Only for shoulder extension there was no significant correlation (0.54,  $p = 0.066$ ).

**RPM.** Only one single patient showed a Tardieu and/or Ashworth Scale value higher than 0 (for shoulder extension, internal rotation, elbow flexion, supination and wrist extension), all other patients had no clinically detectable spasticity. Therefore, no further statistical analysis was possible to calculate the correlations. An example for RPM is shown in Fig 7. It represents both the elbow flexion data of a patient with a Tardieu score of 2 and a patient with a Tardieu score of 0.



**Fig 6. Plots of the patient joint torques from STRENGTH vs. MMT scores for all joints and in both directions.** (x-axis = MMT score, y-axis = joint torque from STRENGTH).

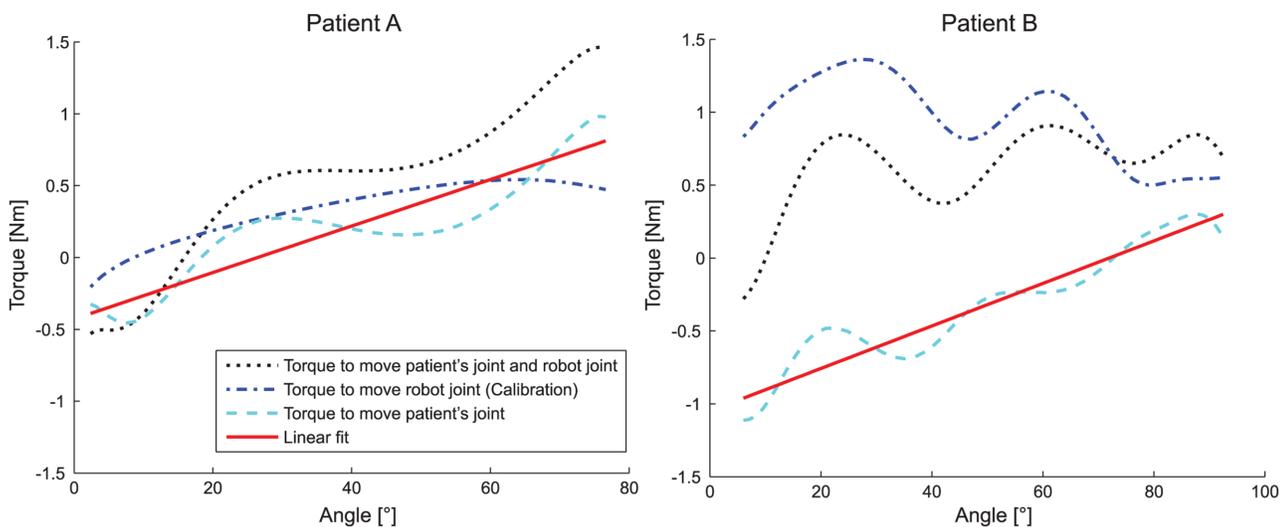
doi:10.1371/journal.pone.0126948.g006

## Discussion

Despite the small sample size, first conclusions could be drawn regarding applicability, reliability, validity and limitations of the single assessment packages. All subjects and patients were able to understand and perform the different assessment packages.

## ROM

**Assessment evaluation.** The mechanical joint limits of the robot were reached by the healthy subjects and sometimes by the patients, leading to a saturation effect of the range



**Fig 7. Example RPM plot for elbow flexion (60°/s) of two patients (A and B).** The black dotted line shows the torque progression measured during the assessment. The blue dash-dotted line is the data from the calibration routine without the patient. The cyan dashed line is the estimated torque progression of the patient's joint. The red line is the linear least square fit for the patient's joint torques. The movement in the left picture was rated a 2 on the Tardieu scale. The movement on the right example was rated a 0 on the Tardieu scale. The measured stiffness by the robot assigned 0.93 Nm/rad to the left and 0.84 Nm/rad to the right movement.

doi:10.1371/journal.pone.0126948.g007

values. Therefore, ROM showed good intra-rater (except for *pROM* pronation) and inter-rater (except for shoulder extension and internal-/external shoulder rotation) reliabilities. The *aROM* and *pROM* values of the patients were both smaller compared to healthy subjects. Furthermore, the comparison of the *aROM* and *pROM* values with the manual ROM test was good for several joints (active lateral shoulder abduction, elbow flexion, wrist extension, shoulder flexion and passive lateral shoulder abduction). We assume that this correlation is better for severely and moderately affected patients, who do not reach the robot joint ranges. The results support our use of the ROM values as a basis for exercises or assessments (such as the RPM assessment) to define the usable and safe region.

**Limitations.** The mechanical limits of the robot are not only due to kinematic constraints but also chosen for safety reasons and, therefore, do not cover the whole range that can maximally be reached by a healthy subject (e.g. in [54]). This was reflected by the significant differences between tester 1 and 2 found in the Bland-Altman plot and the strong correlation between the differences and the mean of clinical and robotic results. As most patients in the sample got higher ROM values in the clinical measurements with the goniometer, a goniometer has to be preferred to robots (with similar ROM) for the measurement of joint range.

**Applicability.** The chosen robot postures for measuring the range were applicable for all patients. Even patients sitting in a wheelchair could be measured in the predefined positions. The discrepancy between clinical and robotic ROM in wrist extension was mainly due to an offset between the zero position of the robot (i.e. grasping the hand module leads to an initial dorsal extension) and the zero position measured by the therapist (straight hand with stretched fingers and no dorsal extension).

## WORKSPACE

**Assessment evaluation.** The healthy subjects reached the maximum cubic workspace, therefore, intra-rater reliability was good regarding *cubic volume* and *workspace levels*. The comparison with the patients showed that their *cubic volume* was around 70–80% of healthy subjects (Table 6). Furthermore, a good inter-rater reliability was found for *workspace level* and *cubic volume*. However, the Bland-Altman plot showed significant differences between the two testers for *cubic volume*. A closer look at the values revealed that the volume assessed by tester 2 was always equal (4 out of 9 tests) or higher (5 out of 9 tests) as compared to tester 1. The actual source for this difference could not be identified. Possibly, the difference came from discrepancies in how the patient was motivated to reach in the different directions. The *cubic volume* correlated well with the ARW and moderately with the VLT/ GRASSP items. As WORKSPACE is reliable, it can be used for calibration of exercises (e.g. to place objects in the virtual environment, so that the patient can reach them) or as a basis for further assessments (e.g. as we did in QOM). More parameters might be extracted from this assessment to analyze the movements in more detail, such as the interjoint coordination. However, changing distances to the room walls have to be taken into account.

**Limitations.** WORKSPACE should only be used for assessment of severely affected patients, who do not reach the maximum cubic volume. The inter-rater reliability was poor in the downward direction for *workspace level*. The reason may be that downward movements were limited by the legs and depending on the robot forearm posture the robot could be moved more or less in this direction. A solution would be to fix the pro-/supination in a predefined position during this measurement such that the conditions stay the same for each patient.

Another limitation related to the clinical validity was that the absolute values of the workspace volume were much higher for the clinical assessment than in WORKSPACE, the reason being that WORKSPACE did not calculate the real workspace volume (e.g. as shown in [16]) but a cubic volume in front of the patient.

**Applicability.** No problems with the applicability of the assessment. The workspace was assessed in several steps reaching walls of discretely increasing distance. It could be simplified by continuously moving the walls away from the center position.

## QOM

**Assessment evaluation.** The QOM assessment had largely good intra-rater reliability. The only values that changed significantly over the course of the four ARMin sessions were almost exclusively due to different results for the first targets of the first session pointing to a learning effect. The comparison between patients and healthy subjects revealed that the patients performed similar or better than healthy subjects with the non-dominant arm. Currently, we have no explanation for this trend. In the dominant arm the healthy subjects gained better results (as expected, [Table 6](#)). The inter-rater analysis showed neither significant correlation between the testers (except for *number of peaks to start*) nor could differences be detected in the Bland-Altman analysis. The reason is in our opinion the small sample size. The *precision* metric showed a tendency to moderately correlate between the testers. The construct validity analysis revealed moderate but significant correlations comparing *D-P ratio to start* and *reaction time to start* with GRASSP and VLT.

**Limitations.** The fact that specifically the movements to the first target in the first test showed significant differences to later tests indicates that the first couple of movements should not be included in the assessment as it is likely influenced by learning factors or inattention in the beginning.

The results for the comparison and the inter-rater reliability showed clear differences between movements to the target and start position. This effect may be caused by a clear knowledge about the direction in which the start position will appear and, therefore, the movement could be planned before the start position showed up. For the target positions the direction was not known beforehand and the reaction and the movement may have been influenced by where the target appeared.

The results of the peak metric from the comparison between patients and healthy subjects were very ambiguous. It is known from literature that the peak metric performs fairly well for stroke patients but is insensitive, nonrobust and unsuitable for healthy subjects [25] and insensitive to brief resting periods [24]. Although our assessment was performed in the frontal and not in the transverse plane as described therein, conclusions from the *number of peaks* metric should be drawn carefully. Further metrics such as the spectral arc-length metric for smoothness [25] should be considered for evaluation in a next assessment version.

*Reaction time* would be more reliable, when measuring the time to overcome a certain distance from the real starting position, rather than measuring the time to leave a starting circle, where the real starting position inside this circle was unknown. Therefore, the distance that had to be moved until the *reaction time* was measured changed in each trial.

The location of the targets in the frontal plane without a restriction for the distance to the body may have been a source for higher variability in the assessed parameters, as the patient could choose slightly closer or more distant targets in different trials. An additional assessment package covering the quality of movement could be a tracking task of a figure e.g., a circle [10] or a Lissajous figure [55].

**Applicability.** No problems with the applicability of the assessment.

## STRENGTH

**Assessment evaluation.** The STRENGTH assessment had a good intra-rater and inter-rater reliability (except for wrist extension which showed a significant difference between the

two testers in the Bland-Altman analysis). All the six valid comparisons for wrist extension showed slightly higher values in the assessments of one tester. The reason for this effect is unclear and we assume that it comes from the small number of valid samples in this joint and the normal variability in the force data.

Furthermore, the construct validity showed a very good correlation with the MMT. The STRENGTH measurement detected even small changes in torques applied which cannot be detected with the MMT score. Moreover, the MMT scores showed a clear ceiling effect (Fig 6) which was not present in the STRENGTH assessment (i.e. a maximum MMT score of 5 can still be continuously graduated with the STRENGTH measurement). The MMT is the standard assessment to measure muscle forces. Cuthbert et al [56] reported an inter-rater-reliability from 0.63 to 0.98 with very well trained testers. As the STRENGTH assessment reached ICC scores for the inter-rater reliability from 0.80\*\* to 0.98\*\* the robotic assessment is comparable with the MMT score. Further aspects that could potentially be investigated in a later version are endurance or fatigue.

**Limitations.** The measurement of the mid-hand closing forces may not be reproducible in other robots without knowledge of the exact design of the specific hand module.

**Applicability.** The interface was used to show the currently assessed joint and in which direction the torque has to be applied. However, difficulties in understanding the indicated direction and the correct timing were reported sporadically.

## RPM

**Assessment evaluation.** The RPM assessment aimed to measure the torque that was needed to move single joints of the arm. The stiffness portion of the arm resistance may be used as an indicator for arm stiffness or spasticity. The patients participating in this assessment study had almost no clinical relevant spasticity and, therefore, the interpretability of the results is very vague. Accordingly, the *joint stiffness* values were rather noisy with a trend for an increased resistance in patients. Mainly for shoulder ab-/adduction, elbow flexion/extension and hand opening/closing (Fig 3).

If further investigations proof the validity of RPM assessments this would be a very good alternative for clinical scales such as MTS and the MAS which are neither reliable measurements for the upper extremity [57] nor have a rational scale with an appropriate resolution to measure subtle changes in the arm stiffness.

Similar robot-assisted RPM measurements were already performed by other groups with the therapist moving the arm and measuring force and angle with sensors achieving promising results [58]. Therefore, we hold to this approach and plan more tests with patients with spasticity or biomechanical stiffness in the arm.

**Limitations.** The variation between ARMin tests on different days were high, this could be due to a real variation of the spasticity over time. Furthermore, clinical tests and the ARMin assessments were not performed on the same day or during the same time, which also influences the measured stiffness. For further studies the clinical investigation of spasticity has to be shortly before or after the RPM measurement.

Higher joint speeds would be desirable for measuring aspects of speed-dependent stiffness (as in [33] or [59]). However, the ARMin robot is not powerful enough to reach higher speeds than roughly 60°/s over a sufficient joint range (after subtracting the joint region for acceleration and deceleration of the joint).

**Applicability.** For the RPM measurements in this paper there was a calibration routine of the robot necessary in order to know the torques without patient. An accurate dynamic model of the robot would make the calibration redundant and it could be skipped.

## General remarks

In this consideration of concept study only mean values for all directional movements were used (QOM and WORKSPACE). However, in a later study when there is more patient data available, specific directions, sectors or quadrants of the workspace should be grouped for a more detailed direction-dependent analysis.

Although the presented assessment packages were implemented in the ARMin robot, the methods can be applied to other exoskeleton robots. However, for STRENGTH the actuators and the robot structure have to be strong and stiff enough to counteract the torques applied by the patient. WORKSPACE and QOM can also be applied to end-effector robots, as both assessments only need the hand position as an input.

The robot as a measuring tool may produce unwanted interaction forces with the patient. These were reduced by active compensation of both gravity and friction of the robot. However, the arm dynamics may still be influenced, e.g., by additional inertia or non-modeled gravitational and frictional effects. This makes it more difficult to compare the assessment outcomes with other platforms or measurements (e.g., a free arm movement outside the robot) or with a platform with different inertial properties. This is mainly an issue for the assessments with free end-effector movements (QOM and WORKSPACE). Furthermore, we cannot exclude that small imprecisions in the modeling of the nonlinear spring compensation may have led to a systematic bias for a certain direction in the shoulder flexion/extension joint. However, this effect could not be observed from the recorded assessment data.

Besides these modeling factors of the robot, the therapist as well has some influence on the outcome. Even though the recordings are performed by the robot (except for the pROM) the positioning, instruction and motivation of the patient—though unintended—can influence the performance during the assessment and is reflected in the observed inter-rater reliability. This influence could further be reduced by using standardized audio-instructions or self-aligning exoskeleton axes [60]. In pROM the range measurement is directly affected by the therapist's (and also patient's) rating of how strong and far the joint can be moved before the resistance gets too high or the movement causes pain.

The different assessments can be used independently from each other. Therefore, single assessment packages can be used to assess changes on daily basis, i.e. before or after the robotic arm therapy session. Considering the duration to perform all assessments successively, the complete assessment package may only be used for admission and discharge measurements or for outcome tests in research studies.

## Conclusions

We evaluated five robot-assisted assessment packages that are integrated into the arm therapy device ARMin and measured different aspects of the human arm motor capability in SCI patients. We found very good reliability and validity for the *joint torque* measurements despite the small number of patients. Furthermore, the measurements for the reachable cubic workspace and different quality of movement metrics showed tendencies for good intra-rater and inter-rater reliability as well as validity.

To our knowledge this is the first study integrating a comprehensive set of assessment packages in a single actuated robotic platform to measure kinematic, kinetic and timing parameters on joint and end-effector level. Furthermore, this study presents for the first time the application of an actuated arm robot for assessment of SCI patients. The actuation allows, on one hand, free arm movements (QOM and WORKSPACE) but on the other hand precise performance of desired movements (RPM) and complete resistance against the arm (STRENGTH).

Moreover, the backdrivable and transparent design of the robot combined with the sensitive sensors allows for continuous and accurate measurements of arm movements and torques.

We believe that all the described methods and results can be generalized and transferred to other robotic exoskeleton platforms (e.g. the commercial version of the ARMin robot, the ArmeoPower) or other patient groups, such as stroke patients. WORKSPACE and QOM assessments can also be applied to end-effector robots, such as the ACT<sup>3D</sup> [14], the GENTLE/G [61] or the PASCAL robot [62]. By giving detailed precise instructions for conducting the assessments we offer a reproducible guidance for future assessment implementations into rehabilitation robots.

With this consideration of concept study we could show the feasibility of our assessment packages and their applicability to SCI patients. We started extending and improving the assessments based on the knowledge acquired during this study. The next generation of assessments will have to be tested with more patients to investigate the different trends found and aspects discussed in this paper and to work out the potential and possible application of the robot-assisted assessments in a clinical setting. Currently an extended set of assessments is being implemented in our new robotic platform ChARMin [63] for children suffering from cerebral palsy and other motor deficits.

We conclude that these robotic assessments are a new opportunity to assess patients and cannot only replace clinical assessments but offer a sensitive, objective and reliable measurement for more detailed insights in arm motor functions.

## Supporting Information

**S1 Table. ARMin postures for the different joints measured during the ROM assessment.**  
(PDF)

**S2 Table. ARMin postures for the different joints measured during the STRENGTH assessment.**  
(PDF)

**S3 Table. Summary of the inter-rater analysis of WORKSPACE and QOM.**  
(PDF)

**S4 Table. Variability in the assessment parameters measured.** For aROM, pROM, WORKSPACE, STRENGTH and RPM the mean differences are shown between the maximum and minimum values measured during the four recordings of the intra-rater reliability. For the QOM assessment the standard deviation is used as an indicator for the variability of the recorded parameters.  
(PDF)

## Acknowledgments

The authors thank Marco Guidali for his help with the ARMin robot and Peter Schenk, Silvia Rohner and Alexander Duschau-Wicke from Hocoma AG, Switzerland for their input on the implementation and visual representation of the assessments, as well as all the subjects and patients who kindly participated in the study.

## Author Contributions

Conceived and designed the experiments: UK SS UA CR AC RR VK. Performed the experiments: SS UA CR. Analyzed the data: UK SS VK. Contributed reagents/materials/analysis tools: UK AC RR. Wrote the paper: UK SS UA CR AC RR VK. Implementation of robot

control, visual interface and evaluation algorithms: UK. Coordination of the study, performing the clinical assessments: SS.

## References

1. Lee B, Cripps R, Fitzharris M, Wing P. The global map for traumatic spinal cord injury epidemiology: update 2011, global incidence rate. *Spinal cord*. 2013; 52(2):110–116. doi: [10.1038/sc.2012.158](https://doi.org/10.1038/sc.2012.158) PMID: [23439068](https://pubmed.ncbi.nlm.nih.gov/23439068/)
2. Mayo NE, Wood-Dauphinee S, Côté R, Durcan L, Carlton J. Activity, participation, and quality of life 6 months poststroke. *Arch Phys Med Rehabil*. 2002; 83(8):1035–1042. PMID: [12161823](https://pubmed.ncbi.nlm.nih.gov/12161823/)
3. Kennedy P, Lude P, Taylor N. Quality of life, social participation, appraisals and coping post spinal cord injury: a review of four community samples. *Spinal Cord*. 2005; 44(2):95–105.
4. Schaechter JD. Motor rehabilitation and brain plasticity after hemiparetic stroke. *Prog Neurobiol*. 2004; 73(1):61–72. PMID: [15193779](https://pubmed.ncbi.nlm.nih.gov/15193779/)
5. Kloosterman M, Snoek G, Jannink M. Systematic review of the effects of exercise therapy on the upper extremity of patients with spinal-cord injury. *Spinal Cord*. 2008; 47(3):196–203. doi: [10.1038/sc.2008.113](https://doi.org/10.1038/sc.2008.113) PMID: [18825160](https://pubmed.ncbi.nlm.nih.gov/18825160/)
6. World Health Organization and others. Towards a common language for functioning, disability and health: ICF. World Health Organ; 2002.
7. Lambercy O, Lünenburger L, Gassert R, Bolliger M. Robots for Measurement/Clinical Assessment. In: *Neurorehabilitation Technology*. Springer; 2012. p. 443–456.
8. Bosecker C, Dipietro L, Volpe B, Krebs HI. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. *Neurorehabil Neural Repair*. 2010; 24(1):62–69. doi: [10.1177/1545968309343214](https://doi.org/10.1177/1545968309343214) PMID: [19684304](https://pubmed.ncbi.nlm.nih.gov/19684304/)
9. Dovat L, Lambercy O, Salman B, Johnson V, Gassert R, Burdet E, et al. Post-stroke training of a pick and place activity in a virtual environment. In: *Virtual Rehabilitation, 2008*. IEEE; 2008. p. 28–34.
10. Gilliaux M, Lejeune TM, Detrembleur C, Sapin P, Dehez B, Selves C, et al. Using the robotic device REAplan as a valid, reliable, and sensitive tool to quantify upper limb impairments in stroke patients. *J Rehabil Med*. 2014; 46(2):117–125. doi: [10.2340/16501977-1245](https://doi.org/10.2340/16501977-1245) PMID: [24213596](https://pubmed.ncbi.nlm.nih.gov/24213596/)
11. Lambercy O, Dovat L, Yun H, Wee SK, Kuah C, Chua K, et al. Robotic assessment of hand function with the HapticKnob. In: *Proceedings of the 4th International Convention on Rehabilitation Engineering & Assistive Technology*. Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre; 2010. p. 33.
12. Fluet MC, Lambercy O, Gassert R. Upper limb assessment using a virtual peg insertion test. In: *IEEE Int Conf Rehabil Robot*. IEEE; 2011. p. 1–6.
13. Lünenburger L, Colombo G, Rienen R, Dietz V. Clinical assessments performed during robotic rehabilitation by the gait training robot Lokomat. In: *IEEE Int Conf Rehabil Robot*. IEEE; 2005. p. 345–348.
14. Ellis MD, Sukal-Moulton T, Dewald JP. Progressive shoulder abduction loading is a crucial element of arm rehabilitation in chronic stroke. *Neurorehabil Neural Repair*. 2009; 23(8):862–869. doi: [10.1177/1545968309332927](https://doi.org/10.1177/1545968309332927) PMID: [19454622](https://pubmed.ncbi.nlm.nih.gov/19454622/)
15. <http://www.hocoma.com/products/armeo/armeopower/>;
16. Kurillo G, Chen A, Bajcsy R, Han JJ. Evaluation of upper extremity reachable workspace using Kinect camera. *Technol Health Care*. 2013; 21(6):641–656. doi: [10.3233/THC-130764](https://doi.org/10.3233/THC-130764) PMID: [24284552](https://pubmed.ncbi.nlm.nih.gov/24284552/)
17. Colombo R, Pisano F, Micera S, Mazzone A, Delconte C, Carrozza M, et al. Assessing mechanisms of recovery during robot-aided neurorehabilitation of the upper limb. *Neurorehabil Neural Repair*. 2008; 22(1):50–63. PMID: [17626223](https://pubmed.ncbi.nlm.nih.gov/17626223/)
18. Celik O, O'Malley MK, Boake C, Levin HS, Yozbatiran N, Reistetter TA. Normalized movement quality measures for therapeutic robots strongly correlate with clinical motor impairment measures. *IEEE Trans Neural Syst Rehabil Eng*. 2010; 18(4):433–444. doi: [10.1109/TNSRE.2010.2047600](https://doi.org/10.1109/TNSRE.2010.2047600) PMID: [20388607](https://pubmed.ncbi.nlm.nih.gov/20388607/)
19. Klamroth-Marganska V, Blanco J, Campen K, Curt A, Dietz V, Ettl T, et al. Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomised trial. *Lancet Neurol*. 2014; 13(2):159–166. doi: [10.1016/S1474-4422\(13\)70305-3](https://doi.org/10.1016/S1474-4422(13)70305-3) PMID: [24382580](https://pubmed.ncbi.nlm.nih.gov/24382580/)
20. Bolliger M, Banz R, Dietz V, Lünenburger L. Standardized voluntary force measurement in a lower extremity rehabilitation robot. *J Neuroeng Rehabil*. 2008; 5:23. doi: [10.1186/1743-0003-5-23](https://doi.org/10.1186/1743-0003-5-23) PMID: [18957092](https://pubmed.ncbi.nlm.nih.gov/18957092/)
21. Tredinnick TJ, Duncan PW. Reliability of measurements of concentric and eccentric isokinetic loading. *Phys Ther*. 1988; 68(5):656–659. PMID: [3362978](https://pubmed.ncbi.nlm.nih.gov/3362978/)

22. Zariffa J, Kapadia N, Kramer JL, Taylor P, Alizadeh-Meghbrazi M, Zivanovic V, et al. Relationship between clinical assessments of function and measurements from an upper-limb robotic rehabilitation device in cervical spinal cord injury. *IEEE Trans Neural Syst Rehabil Eng*. 2012; 20(3):341–350. doi: [10.1109/TNSRE.2011.2181537](https://doi.org/10.1109/TNSRE.2011.2181537) PMID: [22203726](https://pubmed.ncbi.nlm.nih.gov/22203726/)
23. Rohrer B, Fasoli S, Krebs HI, Hughes R, Volpe B, Frontera WR, et al. Movement smoothness changes during stroke recovery. *J Neurosci*. 2002; 22(18):8297–8304. PMID: [12223584](https://pubmed.ncbi.nlm.nih.gov/12223584/)
24. Hogan N, Sternad D. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *J Mot Behav*. 2009; 41(6):529–534. doi: [10.3200/35-09-004-RC](https://doi.org/10.3200/35-09-004-RC) PMID: [19892658](https://pubmed.ncbi.nlm.nih.gov/19892658/)
25. Balasubramanian S, Melendez-Calderon A, Burdet E. A robust and sensitive metric for quantifying movement smoothness. *IEEE Trans Biomed Eng*. 2012; 59(8):2126–2136. doi: [10.1109/TBME.2011.2179545](https://doi.org/10.1109/TBME.2011.2179545) PMID: [22180502](https://pubmed.ncbi.nlm.nih.gov/22180502/)
26. Milner TE. A model for the generation of movements requiring endpoint precision. *Neuroscience*. 1992; 49(2):487–496. PMID: [1436478](https://pubmed.ncbi.nlm.nih.gov/1436478/)
27. Daly JJ, Hogan N, Perepezko EM, Krebs HI, Rogers JM, Goyal KS, et al. Response to upper-limb robotics and functional neuromuscular stimulation following stroke. *J Rehabil Res Dev*. 2005; 42(6):723. PMID: [16680610](https://pubmed.ncbi.nlm.nih.gov/16680610/)
28. Zimmerli L, Krewer C, Gassert R, Müller F, Riener R, Lünenburger L. Validation of a mechanism to balance exercise difficulty in robot-assisted upper-extremity rehabilitation after stroke. *J Neuroeng Rehabil*. 2012; 9(1):6.
29. Dipietro L, Krebs HI, Fasoli SE, Volpe BT, Stein J, Bever C, et al. Changing motor synergies in chronic stroke. *J Neurophysiol*. 2007; 98(2):757–768. PMID: [17553941](https://pubmed.ncbi.nlm.nih.gov/17553941/)
30. Cheung VC, Turolla A, Agostini M, Silvoni S, Bennis C, Kasi P, et al. Muscle synergy patterns as physiological markers of motor cortical damage. *Proc Natl Acad Sci U S A*. 2012; 109(36):14652–14656. doi: [10.1073/pnas.1212056109](https://doi.org/10.1073/pnas.1212056109) PMID: [22908288](https://pubmed.ncbi.nlm.nih.gov/22908288/)
31. Krabben T, Prange GB, Molier BI, Stienen AH, Jannink MJ, Buurke JH, et al. Influence of gravity compensation training on synergistic movement patterns of the upper extremity after stroke, a pilot study. *J Neuroeng Rehabil*. 2012; 9(1):44.
32. Ellis MD, Holubar BG, Acosta AM, Beer RF, Dewald J. Modifiability of abnormal isometric elbow and shoulder joint torque coupling after stroke. *Muscle Nerve*. 2005; 32(2):170–178. PMID: [15880629](https://pubmed.ncbi.nlm.nih.gov/15880629/)
33. Perell K, Scremin A, Scremin O, Kunkel C. Quantifying muscle tone in spinal cord injury patients using isokinetic dynamometric techniques. *Spinal Cord*. 1996; 34(1):46–53. PMID: [8848323](https://pubmed.ncbi.nlm.nih.gov/8848323/)
34. Reinkensmeyer DJ, Schmit BD, Rymer WZ. Assessment of active and passive restraint during guided reaching after chronic brain injury. *Ann Biomed Eng*. 1999; 27(6):805–814. PMID: [10625152](https://pubmed.ncbi.nlm.nih.gov/10625152/)
35. Wood DE, Burridge J, Wijck FV, McFadden C, Hitchcock R, Pandyan A, et al. Biomechanical approaches applied to the lower and upper limb for the measurement of spasticity: a systematic review of the literature. *Disabil Rehabil*. 2005; 27(1–2):19–33.
36. Kumar RT, Pandyan AD, Sharma AK. Biomechanical measurement of post-stroke spasticity. *Age Ageing*. 2006; 35(4):371–375. PMID: [16675479](https://pubmed.ncbi.nlm.nih.gov/16675479/)
37. Coderre AM, Zeid AA, Dukelow SP, Demmer MJ, Moore KD, Demers MJ, et al. Assessment of upper-limb sensorimotor function of subacute stroke patients using visually guided reaching. *Neurorehabil Neural Repair*. 2010; 24(6):528–541. doi: [10.1177/1545968309356091](https://doi.org/10.1177/1545968309356091) PMID: [20233965](https://pubmed.ncbi.nlm.nih.gov/20233965/)
38. Del Din S, Patel S, Cobelli C, Bonato P. Estimating Fugl-Meyer clinical scores in stroke survivors using wearable sensors. In: *Conf Proc IEEE Eng Med Biol Soc*. IEEE; 2011. p. 5839–5842. doi: [10.1109/IEMBS.2011.6091444](https://doi.org/10.1109/IEMBS.2011.6091444) PMID: [22255667](https://pubmed.ncbi.nlm.nih.gov/22255667/)
39. Krebs H, Volpe B, Ferraro M, Fasoli S, Palazzolo J, Rohrer B, et al. Robot-aided neurorehabilitation: from evidence-based to science-based rehabilitation. *Top Stroke Rehabil*. 2002; 8(4):54–70. PMID: [14523730](https://pubmed.ncbi.nlm.nih.gov/14523730/)
40. Dobkin BH. Progressive staging of pilot studies to improve phase III trials for motor interventions. *Neurorehabil Neural Repair*. 2009; 23(3):197–206. doi: [10.1177/1545968309331863](https://doi.org/10.1177/1545968309331863) PMID: [19240197](https://pubmed.ncbi.nlm.nih.gov/19240197/)
41. Nef T, Mihelji M, Riener R. ARMin: a robot for patient-cooperative arm therapy. *Med Biol Eng Comput*. 2007; 45(9):887–900. PMID: [17674069](https://pubmed.ncbi.nlm.nih.gov/17674069/)
42. Guidali M, Duschau-Wicke A, Broggi S, Klamroth-Marganska V, Nef T, Riener R. A robotic system to train activities of daily living in a virtual environment. *Med Biol Eng Comput*. 2011;p. 1–11. doi: [10.1007/s11517-011-0837-9](https://doi.org/10.1007/s11517-011-0837-9) PMID: [22033662](https://pubmed.ncbi.nlm.nih.gov/22033662/)
43. Chapman LJ, Chapman JP. The measurement of handedness. *Brain Cogn*. 1987; 6(2):175–183. PMID: [3593557](https://pubmed.ncbi.nlm.nih.gov/3593557/)
44. Hepp WR, Locher HA. *Orthopädisches Diagnostikum*. Georg Thieme Verlag; 2014.

45. Klopčar N, Tomšič M, Lenarcic J. A kinematic model of the shoulder complex to evaluate the arm-reachable workspace. *J Biomech.* 2007; 40(1):86–91. PMID: [16387308](#)
46. Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther.* 1987; 67(2):206–207. PMID: [3809245](#)
47. Haugh A, Pandyan A, Johnson G. A systematic review of the Tardieu Scale for the measurement of spasticity. *Disabil Rehabil.* 2006; 28(15):899–907. PMID: [16861197](#)
48. Kalsi-Ryan S, Beaton D, Curt A, Duff S, Popovic MR, Rudhe C, et al. The Graded Redefined Assessment of Strength Sensibility and Prehension: reliability and validity. *J Neurotrauma.* 2012; 29(5):905–914. doi: [10.1089/neu.2010.1504](#) PMID: [21568688](#)
49. Mackin E, Callahan A, Skiver T, Schneider L, Osterman A. *Rehabilitation of the hand and upper extremity.* vol. 1:240–262. Mosby; 2002.
50. Clarkson HM. *Musculoskeletal assessment: joint range of motion and manual muscle strength.* Lippincott Williams & Wilkins; 2000.
51. Hislop HJ, Montgomery J. *Manuelle Muskeltests: Untersuchungstechniken nach Daniels und Worthingham.* Elsevier, Urban&FischerVerlag; 2007.
52. Itzkovich M, Gelernter I, Biering-Sorensen F, Weeks C, Laramée M, Craven B, et al. The Spinal Cord Independence Measure (SCIM) version III: reliability and validity in a multi-center international study. *Disabil Rehabil.* 2007; 29(24):1926–1933. PMID: [17852230](#)
53. Post M, Van Lieshout G, Seelen H, Snoek G, Ijzerman M, Pons C. Measurement properties of the short version of the Van Lieshout test for arm/hand function of persons with tetraplegia after spinal cord injury. *Spinal cord.* 2006; 44(12):763–771. PMID: [16773040](#)
54. Soucie J, Wang C, Forsyth A, Funk S, Denny M, Roach K, et al. Range of motion measurements: reference values and a database for comparison studies. *Haemophilia.* 2011; 17(3):500–507. doi: [10.1111/j.1365-2516.2010.02399.x](#) PMID: [21070485](#)
55. Squeri V, Masia L, Casadio M, Morasso P, Vergaro E. Force-field compensation in a manual tracking task. *PLoS ONE.* 2010; 5(6):e11189. doi: [10.1371/journal.pone.0011189](#) PMID: [20567516](#)
56. Cuthbert SC, Goodheart GJ. On the reliability and validity of manual muscle testing: a literature review. *Chiropr Man Therap.* 2007; 15(1):4.
57. Gregson JM, Leathley M, Moore AP, Sharma AK, Smith TL, Watkins CL. Reliability of the Tone Assessment Scale and the modified Ashworth scale as clinical tools for assessing poststroke spasticity. *Arch Phys Med Rehabil.* 1999; 80(9):1013–1016. PMID: [10489001](#)
58. Pandyan A, Price C, Rodgers H, Barnes M, Johnson G. Biomechanical examination of a commonly used measure of spasticity. *Clin Biomech.* 2001; 16(10):859–865. PMID: [11733123](#)
59. Calota A, Feldman AG, Levin MF. Spasticity measurement based on tonic stretch reflex threshold in stroke using a portable device. *Clin Neurophysiol.* 2008; 119(10):2329–2337. doi: [10.1016/j.clinph.2008.07.215](#) PMID: [18762451](#)
60. Stienen AHA, Hekman EEG, Van Der Helm FCT, Van Der Kooij H. Self-aligning exoskeleton axes through decoupling of joint rotations and translations. *IEEE Trans Robot.* 2009; 25(3):628–633.
61. Loureiro RC, Harwin WS. Reach & grasp therapy: design and control of a 9-DOF robotic neuro-rehabilitation system. In: *IEEE Int Conf Rehabil Robot.* IEEE; 2007. p. 757–763.
62. Koenig A, Keller U, Pfluger K, Meyer-Heim A, Riener R. PASCAL: Pediatric arm support robot for combined arm and leg training. In: *Proc IEEE RAS EMBS Int Conf Biomed Robot Biomechatron.* IEEE; 2012. p. 1862–1868.
63. Keller U, Klamroth V, van Hedel HJ, Riener R. ChARMin: A robot for pediatric arm rehabilitation. In: *IEEE Int Conf Robot Autom.* IEEE; 2013. p. 3908–3913.