Relationship between clinical assessments of function and measurements from an upper-limb robotic rehabilitation device in cervical spinal cord injury

José Zariffa, *Member, IEEE*, Naaz Kapadia, John L.K. Kramer, Philippa Taylor, Milad Alizadeh-Meghrazi, Vera Zivanovic, Urs Albisser, Rhonda Willms, Andrea Townson, Armin Curt, Milos R. Popovic, *Senior Member, IEEE*, and John D. Steeves

Abstract— **Upper limb robotic rehabilitation devices can collect quantitative data about the user's movements. Identifying relationships between robotic sensor data and manual clinical assessment scores would enable more precise tracking of the time course of recovery after injury and reduce the need for time-consuming manual assessments by skilled personnel. This study used measurements from robotic rehabilitation sessions to predict clinical scores in a traumatic cervical spinal cord injury (SCI) population. A retrospective analysis was conducted on data collected from subjects using the Armeo®Spring (Hocoma, AG) in three rehabilitation centres. 14 predictive variables were explored, relating to range of motion, movement smoothness, and grip ability. Regression models using up to 4 predictors were developed to describe the following clinical scores: the GRASSP (consisting of four sub-scores), the ARAT, and the SCIM. The resulting adjusted R^2 value was highest for the GRASSP "Quantitative Prehension" component (0.78), and lowest for the GRASSP "Sensibility" component (0.54). In contrast to comparable studies in stroke survivors, movement smoothness was least beneficial for predicting clinical scores in SCI. Prediction of upper-limb clinical scores in SCI is feasible using measurements from a robotic rehabilitation device, without the need for dedicated assessment procedures.**

***Index Terms*— ARAT, GRASSP, regression analysis, robotic rehabilitation, SCIM, spinal cord injury, upper extremity.**

Manuscript received XXX, 2011. This study was supported by the Rick Hansen Institute (grant #2009-34) and the NCCR Neural Repair, Zurich, Switzerland.

J. Zariffa (jose.zariffa@utoronto.ca) was with the International Collaboration On Repair Discoveries, Vancouver, Canada and is now with the Toronto Rehabilitation Institute, Toronto, Canada.

J.L.K. Kramer, P. Taylor and J.D. Steeves are with the International Collaboration On Repair Discoveries, Vancouver, Canada.

N. Kapadia, V. Zivanovic, M. Alizadeh-Meghrazi and M.R. Popovic are with the Toronto Rehabilitation Institute, Toronto, Canada. M.R. Popovic is also with the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada.

A. Townson and R. Willms are with the GF Strong Rehabilitation Centre, Vancouver, Canada, and the Division of Physical Medicine and Rehabilitation, Faculty of Medicine, University of British Columbia, Vancouver, Canada.

A. Curt and U. Albisser are with the Spinal Cord Injury Center, University Hospital Balgrist, University of Zurich, Switzerland.

I. Introduction

Rehabilitation following most neurological injuries consists largely of activity-dependent and goal directed training, where patients repeatedly move their limbs to produce functional patterns [1]. In many cases, the patient may be incapable of completing these movements unassisted, particularly in the early stages of recovery after injury. Therapists therefore help to support and move the limbs during these exercises, and continuously adjust the amount of assistance to the needs of the patient. In recent years, robotic rehabilitation devices have been proposed as a means to complement therapists' activities by assisting with the repetitive mechanical task of moving the patient's limbs. This robotic assistance increases the number of repetitions within a given time period and allows the therapists to focus on other aspects of training or simultaneously treat more than one individual at a time. When robotic devices are combined with engaging virtual reality training exercises, they also have the potential to increase the patient's enjoyment and compliance with rehabilitation. Examples of robotic rehabilitation devices focusing on the upper limb include the MIT-Manus [2, 3], the ARM guide [4], the MIME [5], the GENTLE/s [6], the Bi-Manu-Track [7], the T-WREX [8, 9] (the prototype for the device commercialized as the Armeo®Spring by Hocoma, AG, Switzerland), the ARMin [10, 11], and the ReJoyce [12].

An added benefit of robotic rehabilitation devices is that they can collect kinematic information about the user's movements as they are performed. If these measures could be converted into clinically meaningful information, they would have two benefits. First, frequent quantitative evaluation of the user's performance would open the door to more accurate tracking of the time course of recovery after SCI, which has application in the design of clinical trials and in developing responsive rehabilitation programs with data-driven session-to-session training adjustments. Second, automated and quantitative functional assessments would reduce the subjectivity inherent in many of the current rehabilitation trial outcome measures (functional assessment tests) [13].

A number of studies have examined the correlations between movement descriptors measured by robots (including smoothness, accuracy, speed and force produced) and clinical measures used for stroke survivors, such as the Fugl-Meyer test [14, 15, 16, 17, 18], the Motor Power Score [13, 16], the Motor Status Score [16, 17], and the Action Research Arm Test [18]. A number of moderate and statistically significant correlations have been identified [14, 16, 18]. The use of multiple regression analysis, to obtain predictions of clinical scores using combinations of robotic descriptors, has also been attempted in the past and has yielded stronger correlations [13, 15, 17].

While the studies cited above have focused on clinical assessments in populations of stroke survivors, the study presented here used data from a robotic upper limb rehabilitation device to predict clinical assessment scores in a population of individuals with cervical spinal cord injury (SCI). We performed a retrospective analysis of data collected in several centres using an upper limb rehabilitation robotic device (Armeo®Spring) in a sample of in-patient subjects with traumatic cervical SCI. Our hypothesis is that useful diagnostic information can be derived from data recorded by the robot during therapeutic use, without the need for additional dedicated assessment procedures. Given that SCI typically involves a proximal-to-distal impairment gradient that is dependent on the level of injury, relationships may exist between clinical assessments of function and the joint-specific information obtained using a robotic exoskeleton.

II. Methods

We performed a retrospective study of subjects undergoing in-patient rehabilitation following acute cervical SCI and having various degrees of tetraplegia. Our approach consisted of the following 3 steps:

1. Identify robotic rehabilitation sessions that were performed close in time (within 2 weeks) to a clinical assessment.
2. Derive quantitative predictors from the kinematic information collected by the robot during the rehabilitation sessions.
3. Perform a multiple linear regression to identify combinations of the predictor variables that correlate well with the clinical scores.

The details of each of these steps are given in the sections below.

A. Data Collection

When using the Armeo®Spring, the user's arm is placed in an exoskeleton that helps to support the weight of the upper and lower arm through a system of springs (Figure 1). The

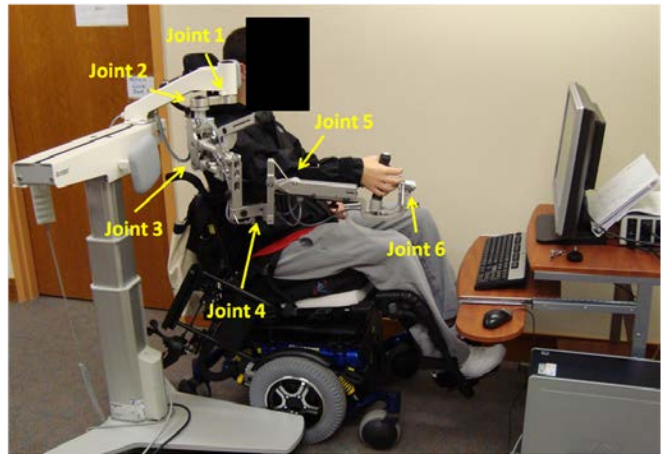

Figure 1: Picture of the Armeo®Spring. The 5 joints responsible for the position of the hand in space are identified (joint 6 is used for pronation/supination at the wrist).

device does not include actuators, but in this article we use to term "robotics" to refer to a general class of rehabilitation technology, rather than specifically to devices that can move on their own to accomplish a task. For the training sessions used in this study, the amount of weight support was chosen according to the rehabilitation goals of individual subjects; in the majority of cases, the support was chosen to approximately compensate for the combined weight of the arm and exoskeleton, and provide a neutral position (i.e. the arm was neither being pulled down of pushed up). In the version of the Armeo®Spring used in this study (version 1), the exoskeleton includes six joints: joints 1 and 2 are used for horizontal motion of the arm, joint 3 for vertical motion at the shoulder, joint 4 for internal shoulder rotation, joint 5 for flexion at the elbow, and joint 6 for pronation/supination at the wrist. Sensors detect the angle at each of these 6 joints, and a seventh sensor is used to measure hand grip pressure. The information collected from these sensors is transmitted to a computer, and used to control the cursor on a display during a variety of virtual reality (VR) tasks (e.g. putting fruit in a shopping cart, wiping a window, or catching moving targets on the screen). The Armeo®Spring can be adapted to the capabilities of an individual patient by adjusting the amount of anti-gravity arm support, the range of motion required to use the device, and the selection of VR tasks. While the exercises are conducted, the device includes the option to record the angle of each joint of the exoskeleton, the position of the hand in space (x, y, and z coordinates) and the grip pressure with a sampling rate of 100 Hz. It is therefore possible to investigate the path and timing of hand movements and the contributions of different joints, as well as grip function. The length of each VR task is adjustable, but typically falls in the range of 1 to 5 minutes.

Three clinical assessments commonly used internationally for the evaluation of upper limb impairment were considered: the Graded and Redefined Assessment of Strength, Sensibility and Prehension (GRASSP [19]), the Action Research Arm Test (ARAT [20]), and the Spinal Cord Independence

Measure III (SCIM III [21]). The GRASSP is designed specifically to evaluate hand function in subjects with tetraplegia, and consists of four components: (1) Manual muscle testing (MMT) of 10 muscles in the upper limb and hand; (2) Sensibility testing at the palmar and dorsal fingertips using monofilaments; (3) Qualitative prehension testing, which evaluates the ability of the subject to form different types of hand grips; (4) Quantitative prehension testing, which evaluates the ability of the subject to perform various functional tasks (e.g. pegboard, pouring water, turning a key, etc.). The ARAT is also a measure of the subject's ability to perform various functional tasks associated with activities of daily living (ADLs), and includes categories for grasp, grip, pinch, and gross movement. Lastly, the SCIM is designed to evaluate the degree of independence with which the subject can perform various ADLs (including self-care, mobility, and bowel and bladder care). The "self-care" sub-score of the SCIM has been shown to be most closely related to upper limb function [22], and we therefore examined both the total SCIM score and the self-care sub-score alone. GRASSP and ARAT are performed for each limb individually, so the assessments used in the analysis were those corresponding to the limb receiving robotic rehabilitation. The SCIM assessment provides a measure of independence and is not specific to one limb.

Data was collected from in-patients with traumatic cervical SCI using the Armeo®Spring at three SCI rehabilitation centres. Two of the centres were located in Canada and were using the Armeo®Spring in the context of a pilot study studying the device's applicability to SCI [23], whereas the third centre was located in Switzerland and was using the Armeo®Spring as part of its standard of care. At each centre, the date of each training session was compared to the dates at which each of the three types of clinical assessments (GRASSP, ARAT, SCIM) was conducted. Sessions that were within two weeks of a clinical assessment were retained.

Our goal was to predict the clinical scores using the kinematic and grip data automatically collected from the robot during regular therapeutic use, rather than by using VR modules dedicated specifically to assessment (e.g., with targets at specific repeatable positions). The training programs were, however, tailored to the capabilities of each subject, such that each study participant used different VR tasks based on their level of function. The data used in our prediction models must therefore be general enough to be obtainable from different VR tasks. Taking this requirement into consideration, as well as the nature of the VR tasks available in the Armeo®Spring, predictor data was extracted from VR tasks that (1) required the subject to move the cursor in two dimensions within the vertical plane and (2) ensured that the movements covered a large portion of this plane. These choices were based on the fact that several of the available tasks met these criteria, and nearly all subjects had at least one of these tasks included in their rehabilitation program. In addition, we sought to incorporate data about grip capabilities into the predictive models. For subjects with sufficient function to use the Armeo®Spring's hand grip sensor, data was extracted from VR tasks requiring the subjects to grasp and release virtual objects at specific times and in specific locations. Again, several of the available tasks met this criterion. Subjects who had insufficient hand function to produce a detectable grip on the hand grip sensor were assigned a score of 0 in all of the grip-related predictive variables (see below).

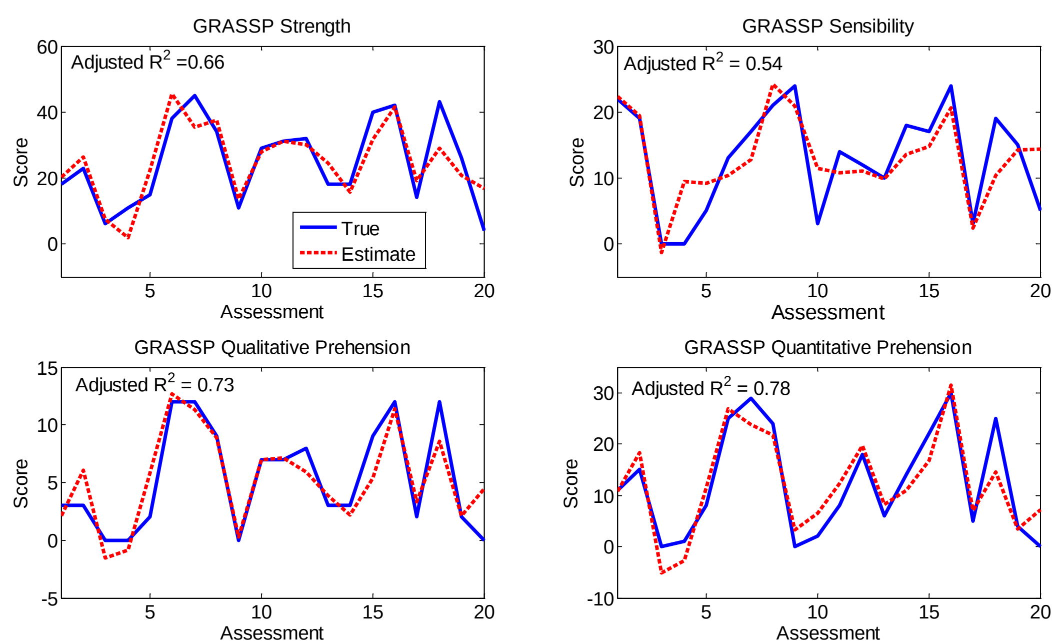

Figure 2: True clinical scores (blue solid line) and estimates from the regression model (red dashed line), for each of the four components of the GRASSP assessment. Each point on the x-axis represents one clinical assessment, and the y-axis provides the corresponding scores.

Three executions of a given VR task were associated with each clinical score, and each predictive variable was averaged over these three executions. One execution is defined as the subject playing through the entire VR task once, which typically corresponds to 1-5 minutes of data. The three executions most often corresponded to three consecutive training sessions, but depending on data availability, two executions were on occasion taken from a single training session where the task was performed several times. Thus, for each clinical score, up to three robotic training sessions were used, which were the ones closest in time to the date of the manual assessment and were all required to have taken place within two weeks of the assessment (either before or after).

As a result of these criteria, the number of available data points was 20 for the GRASSP and SCIM assessments (of which 11 had non-zero grip variables), and 18 for the ARAT (of which 11 had non-zero grip variables). The lower number of data points for the ARAT is due to the fact that only the two centres conducting the pilot study were performing this assessment. These data points were obtained from a total of 14 subjects for the GRASSP and SCIM, and 12 subjects for the ARAT. Some subjects provided more than one data point from the regression analysis, for example one baseline assessment and one discharge assessment, while in other cases only one clinical assessment was available or was sufficiently close (i.e. within 2 weeks) in time to three VR task executions to be included. The three executions used to obtain each data point were taken from sessions spanning an average of 4.3 ± 2.8 days.

The subjects had motor levels ranging from C4 to C6 on the side using the robotic device, and ASIA Impairment Scale (AIS) grades ranging from A to D (as defined by the International Standards for the Neurological Classification of Spinal Cord Injury (ISNCSCI), and evaluated by the clinical staff). The inclusion criterion for analysis in this study was to have sustained a traumatic cervical SCI with a motor level between C4 and C8. Subjects with a history of neuromuscular disease, upper limb spasticity too severe to effectively use the device, severe shoulder pain, unable to sit upright for 30 minutes, or unable to understand and follow instructions were excluded from the Armeo®Spring training. The time since injury for the assessments included in the regression varied from 21 to 227 days. 92.8% of the 14 subjects were male, and the average age was 43.6 ± 18.4 years.

B. Predictive Variables

A total of 14 movement descriptors were investigated: 12 descriptors extracted from the movement of the arm, and 2 descriptors extracted from the grip sensor information. Of the 12 descriptors related to the movement of the arm, 8 were concerned with range of motion, 4 with smoothness, and all were derived from the unfiltered joint angles and hand position data recorded by the machine. When computing the movement descriptors from the Armeo®Spring's kinematic and grip data, the first and last seconds of the VR task were ignored, because subjects were usually moving to or from their resting position during those periods. The descriptors used were:

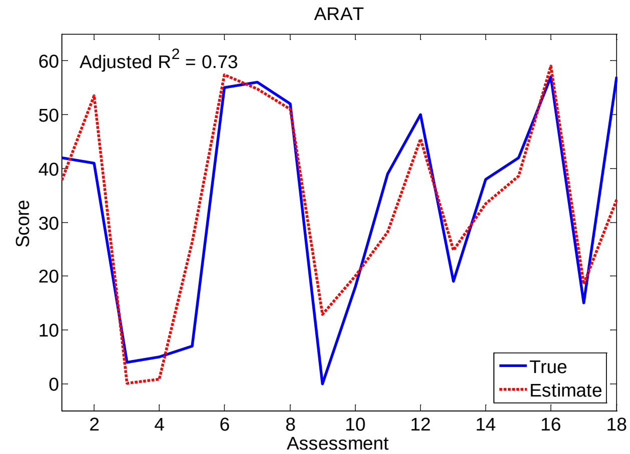

Figure 3: True clinical ARAT scores (blue solid line) and estimates from the regression model (red dashed line). Each point on the x-axis represents one clinical assessment, and the y-axis provides the corresponding scores.

Range of motion descriptors:

1-3. The range of motion in centimetres of the hand position in the x, y and z directions. The range of motion is defined here as the difference between the maximum and minimum position recorded for a given coordinate over the course of the VR task.

4-8. The range of motion in degrees for each of the 5 joints governing the position of the hand in space (joints 1 to 5 in Figure 1; the pronation/supination joint is not used here). The range of motion is defined here as the difference between the maximum and minimum angles recorded for a given joint over the course of the VR task.

Smoothness descriptors:

9. The number of directional changes in the hand's trajectory, normalized by the length of the task. A directional change is defined as a reversal in the sign of the first derivative of any of the three coordinates of the hand position. A smooth movement would have fewer directional changes than a jittery uneven movement.

10. The mean velocity of movement over the duration of the task. The VR tasks were based primarily on accuracy rather than speed, and subjects moved at their natural speed. Although not a true measure of smoothness, velocity is nonetheless indicative of ease of movement.

11. The ratio of the mean velocity to the maximum velocity recorded over the duration of the task [17].

12. The mean jerk (third derivative of the hand position) of the movement over the duration of the task [14, 16].

Grip descriptors:

TABLE I
Regression model for each clinical outcome: Outcome = Predictor1 * Coeff1 + Predictor2 * Coeff2 + Pred3 * Coeff3 + Predictor4 * Coeff4 + Constant.

Outcome	Predictor 1 *Coefficient 1*	Predictor 2 *Coefficient 2*	Predictor 3 *Coefficient 3*	Predictor 4 *Coefficient 4*	*Constant*
GRASSP (Strength)	Y range ***0.67***	Joint 1 range ***0.08***	Joint 4 range ***-0.34***	Grip Skewness ***6.83***	***15.48***
GRASSP (Sensibility)	Directional changes ***143.06***	Mean velocity/max velocity ***90.90***	Jerk ***-0.003***	Grip Skewness ***6.15***	***-19.98***
GRASSP (Qualitative Prehension)	X range ***0.12***	Joint 4 range ***-0.15***	Directional changes ***-63.01***	Grip Skewness ***2.90***	***13.74***
GRASSP (Quantitative Prehension)	Joint 3 range ***0.31***	Joint 4 range ***-0.21***	Mean velocity/max velocity ***102.35***	Grip Skewness ***11.51***	***-16.37***
ARAT	Z range ***-0.95***	Joint 3 range ***0.68***	Mean velocity/max velocity ***162.51***	Grip Skewness ***23.76***	***-17.83***
SCIM (Total)	X range ***1.63***	Z range ***-3.10***	Grip range ***254.81***	Grip Skewness ***22.95***	***19.27***
SCIM (Self-care)	X range ***0.37***	Z range ***-0.82***	Grip range ***78.24***	Grip Skewness ***4.96***	***5.04***

13. The range of the grip pressure, defined as the maximum minus minimum pressure observed over the course of the VR task. Note that the grip sensor of the Armeo®Spring returns measurements in volts and is not calibrated to provide units of pressure; it can therefore be thought of here as a dimensionless quantity.

14. The skewness (third standardized moment) of the grip pressure, a measure of the asymmetry of the distribution, used here to detect imbalances between the subject's ability to grip and to release the sensor.

C. Construction and Evaluation of Predictive Models

Predictive models were sought for seven clinical outcomes: GRASSP (Strength), GRASSP (Sensibility), GRASSP (Qualitative Prehension), GRASSP (Quantitative Prehension), ARAT, SCIM (Total) and SCIM (Self-care). In each case, a multiple linear regression was performed.

Because the number of predictors being considered (14) is large compared to the number of available data points (18-20), using all predictors would result in overfitting of the model. We therefore limited the number of variables in a predictive model to 4, or approximately 20% of the number of data points. 4 variables was chosen as an upper limit due to sample size considerations, yet was high enough to produce regression models with good performance: an exploratory analysis found that 4 out of 7 of our clinical outcomes (GRASSP Strength, GRASSP Qualitative Prehension, GRASSP Quantitative Prehension and ARAT) were best modeled by combinations of 4 variables, even when no limit was placed on the number of independent variables (results not shown). In order to select the best predictive variables, a leave-one-out cross-validation approach was used. In other words, each data point in turn was left out, a model was fitted to the remaining data, and the error between the predicted value of the unused data point and the correct value was computed. In this way, a vector of error values with one entry per data point was obtained for each combination of 1, 2, 3 or 4 of the 14 variables. The combination that minimized the Euclidian norm of this error vector was selected, using an exhaustive search of all possible combinations.

Next, a regression model was computed for each outcome using the 4 (or less) variables selected in the cross-validation process, and all available data points. The quality of the prediction was evaluated using the following metrics:

- The adjusted R^2 value. This quantity is a modification to the coefficient of determination R^2, and decreases as the ratio between the number of predictors and the number of data points increases. Adjusted R^2, a more conservative measure than R^2, is used here to ensure that the results are not skewed by the relatively large number of predictors compared to the number of available data points.
- The mean absolute value of the prediction error, using the model based on all available data points.
- The mean absolute value of the prediction errors obtained during the cross-validation process. This value does not correspond exactly to the model obtained using all data points, but provides a more rigorous estimate of the error magnitude that might be expected on previously unseen data using the selected variables.

Note that a small difference between the prediction errors obtained using all data points and those obtained during the cross-validation process suggests that the model will be able to generalize to new data and therefore have prognostic utility. Conversely, a large discrepancy between the two quantities suggests that the model using all data points has overfitted the data.

Lastly, we sought to estimate the importance of each category of predictive variable (Range of Motion, Smoothness, or Grip). For each clinical outcome, each category of variable in turn was removed from the list of predictors, and a new predictive model was computed using the same methodology as before.

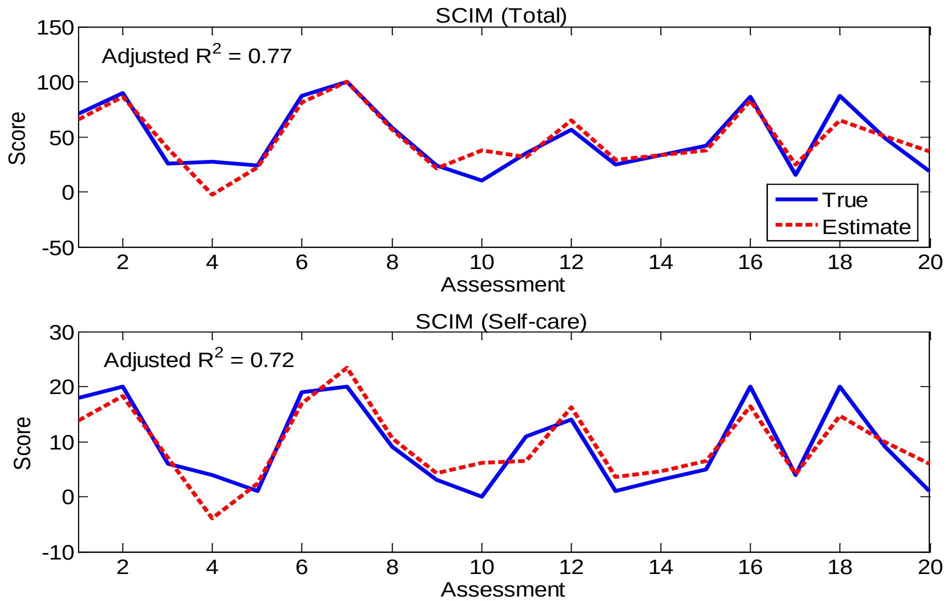

Figure 4: True clinical total SCIM scores and self-care SCIM sub-scores (blue solid line) and estimates from the regression model (red dashed line). Each point on the x-axis represents one clinical assessment, and the y-axis provides the corresponding scores.

III. Results

Figure 2 shows the true and estimated values of the clinical scores for each of the four components of the GRASSP test, Figure 3 shows the prediction results for the ARAT, and Figure 4 shows the prediction results for the SCIM. Table 1 provides the movement descriptors used for the predictions and the corresponding coefficients. The adjusted R^2 value was 0.66 for GRASSP (Strength), 0.54 for GRASSP (Sensibility), 0.73 for GRASSP (Qualitative Prehension), 0.78 for GRASSP (Quantitative Prehension), 0.73 for the ARAT, 0.77 for the SCIM (Total), and 0.72 for the SCIM (Self-care). All of these values have square roots (R-values) greater than 0.7, a common rule of thumb for denoting strong correlations [24]. In all cases, the predicted and true values showed statistically significant correlations with $p < 0.01$.

Table 2 provides the prediction errors, both for the models derived from all data points and for the cross-validation process. The results are also expressed as a percentage of the maximum possible score for each clinical assessment. Using this scale, the mean prediction error of the regression model built using all data points was 10.4% for GRASSP (Strength), 14.8% for GRASSP (Sensibility), 12.2% for GRASSP (Qualitative Prehension), 11.8% for GRASSP (Quantitative Prehension), 11.8% for ARAT, 8.3% for SCIM (Total), and 14.5% for SCIM (Self-care).

Table 3 shows the adjusted R^2 values when each category of variables (Range of Motion, Smoothness, or Grip) is removed in turn. Note that in some cases the adjusted R^2 value with some of the predictors removed may be higher than the value when all predictors are included. This is because the selection of the predictors during the cross-validation process is based on minimizing the norm of the prediction error vector, not on maximizing the correlation. Removing the range of motion variables decreased the adjusted R^2 by 30.9 ± 20.7% over the seven clinical outcomes considered, removing the smoothness variables resulted in a decrease of 6.8 ± 15.3%, and removing the grip variables resulted in a decrease of 35.0 ± 23.1%. Removing smoothness was the least harmful to performance for all clinical outcomes except GRASSP (Sensibility), for which range of motion was the least useful category of predictor.

IV. Discussion

We investigated the prediction of clinical assessment scores using the data collected by an upper limb robotic rehabilitation device in an SCI population. The regression models obtained had adjusted R^2 values greater than 0.7 for all functional measures: ARAT, SCIM (Total and Self-care), GRASSP (Qualitative Prehension), and GRASSP (Quantitative Prehension), the latter corresponding to the highest adjusted R^2 value found, 0.78. Somewhat lower prediction quality was found for the measure of strength (0.66 for GRASSP (Strength)), and the lowest value found was for the measure of sensation (0.53 for GRASSP (Sensibility)).The finding that prediction quality was lowest for GRASSP (Sensibility) is in accordance with expectations, because the prediction is based on motor function characteristics. Unlike measures of function, GRASSP (Sensibility) scores do not directly depend on ROM, grip ability or movement smoothness, resulting in a weaker relationship in the regression model.. Overall, our results are in line with the R^2 values previously reported in similar studies involving stroke survivors [13, 15, 17]. Our regression models proved general enough to accommodate wide variations in the functional abilities of the study subjects,

TABLE II
Prediction error during the cross-validation process, and using the selected regression model with all data points. Results are also shown as a percentage of the maximum possible score for each assessment. The maximum possible scores are as follows. GRASSP (Strength): 50; GRASSP (Sensibility): 24; GRASSP (Qualitative Prehension): 12; GRASSP (Quantitative Prehension): 30; ARAT: 57; SCIM: 100.

Outcome	Mean Absolute Prediction Error (Cross-Validation)	As % of Maximum Possible Score	Mean Absolute Prediction Error (All Data)	As % of Maximum Possible Score
GRASSP (Strength)	6.9	13.8%	5.2	10.4%
GRASSP (Sensibility)	4.6	19.3%	3.6	14.8%
GRASSP (Qualitative Prehension)	1.9	15.9%	1.5	12.2%
GRASSP (Quantitative Prehension)	4.7	15.6%	3.5	11.8%
ARAT	8.9	15.6%	6.7	11.8%
SCIM (Total)	11.8	11.8%	8.3	8.3%
SCIM (Self-care)	4.2	20.8%	2.9	14.5%

which are reflected in the range of functional test scores in Figures 2 to 4.

The high level of performance was achieved despite the fact that the movement descriptors used in the regressions do not always correspond to what the clinical assessments are explicitly evaluating. For instance, many components of the test involve specific functional tasks that primarily require dextrous hand movements. Nonetheless, the robotic device evaluated a combination of arm (through ROM and trajectory tracking) and hand function (through the grip sensor), and these two components proved sufficiently representative of the underlying level of function to provide useful clinical predictions. In this respect, it is not surprising that the models obtained involved both ROM measures and grip measures (Table 1), for all outcomes except GRASSP (Sensibility).

Our regression models were built using data collected during the course of therapeutic use of the Armeo®Spring, without the need for any VR modules specifically dedicated to assessment. In addition, the predictive variables were not specific to a particular VR task (as opposed to, for example, measuring the time and trajectory smoothness between targets at known locations). As a result, the regression models can be applied to data from virtually any patient using the Armeo®Spring, without any modifications needed to the subject’s rehabilitation program. Although more standardized robotic assessment methods might have yielder even more predictive regression models, our demonstration that robotic rehabilitation sessions can provide useful diagnostic information without the need for additional time or personnel resources is highly relevant given the overburdened state of current clinical environments in many countries.

The adjusted R^2 values (Figures 2 to 4) and prediction errors (Table 2) found for the ARAT, SCIM and GRASSP Prehension measures show that the regression models are sufficient to provide meaningful information, though not to replace manual clinical assessments altogether. Assessments meant to evaluate the effects of an intervention (e.g. baseline and discharge) in clinical trial are critical factors in guiding best practices and the adoption of new technology, and in that context it is crucial that any error be minimized. Before robotic assessments can be used for this purpose, the correlations between clinical and robotic data must be further increased. To do so, further studies like the one presented here must be conducted, in which a variety of predictive variables (which will depend on the capabilities of different robots) are studied with sufficiently large sample sizes to obtain highly accurate device- and population-specific regression models. On the other hand, with the current level of performance, quantitative data from robotic devices could potentially be used for interim assessments (assuming that GRASSP, ARAT, or SCIM were judged to be appropriate outcomes for the study). A crucial benefit of using robotic data for assessment is to enable a more detailed characterization of the time course of recovery after SCI, both spontaneously and as a result of rehabilitation interventions. Precise temporal profiles of recovery would typically be prohibitively time-consuming to collect using manual assessments, but are needed in the design of clinical trials [25], and might also be useful for guiding session-by-session training adjustements and thus improving the effectiveness of the rehabilitation process.

Although the coefficients in Table 1 are specific to the Armeo®Spring, generalization of our methods and qualitative results to other devices will depend on the design of the machines and their measurement capabilities. The variables derived from the position of the hand in space can be generalized easily to any other device allowing and measuring 3-dimensional movement. Likewise, the grip variables can be generalized to any device with a grip sensor, though the exact coefficients would change depending on the calibration of the sensor. The variables hardest to generalize are those derived from the joint angles of the exoskeleton, but we expect that alternative regression models with similar performance could be developed for other devices that can provide joint-specific angle information (e.g. other exoskeletons). Improved and more generalizable regression models might also be obtained by incorporating variables that describe the settings of the robotic device (e.g., amount of weight support, difficulty settings). This was not explored in the current study, partly due to the limited information provided by the Armeo®Spring. For instance, the amount of weight support is described by a coarse unitless scale, rather than standard units of weight.

TABLE III
Adjusted R^2 values of the regression models for each outcome. The first column shows the results when all predictive variables are available. The second, third, and fourth columns show the results when range of motion-, smoothness- and grip-related variables are excluded, respectively. The values in parentheses show the percent change when comparing with the values in the first column.

Outcome	All Predictors	No ROM Predictors	No Smoothness Predictors	No Grip Predictors
GRASSP (Strength)	0.66	0.52 (-21.8%)	0.66 (-0%)	0.59 (-11.1%)
GRASSP (Sensibility)	0.54	0.54 (-0%)	0.35 (-34.3%)	0.31 (-42.0%)
GRASSP (Qualitative Prehension)	0.73	0.61 (-17.4%)	0.74 (+1.0%)	0.65 (-10.8%)
GRASSP (Quantitative Prehension)	0.78	0.48 (-38.1%)	0.61 (-21.9%)	0.48 (-38.3%)
ARAT	0.73	0.53 (-27.1%)	0.79 (7.8%)	0.61 (-16.1%)
SCIM (Total)	0.77	0.37 (-51.9%)	0.77 (-0%)	0.31 (-59.7%)
SCIM (Self-care)	0.72	0.29 (-59.7%)	0.72 (-0%)	0.24 (-66.7%)

The number of available data points is the main limitation of our study, but the use of the adjusted R^2 value and of a cross-validation process demonstrate that our conclusions hold true despite the small sample size. The prediction errors obtained during cross-validation were found to be in the range of 2-5% higher (as a percentage of the maximum possible score) than the prediction errors obtained using the final regression models with all data points (Table 2). These differences are small enough to provide confidence in the qualitative accuracy of our results; however, they also suggest that the regression model coefficients could be further refined if a larger data set were available.

The regression models shown in Table 1 employ a combination of range-of-motion, smoothness, and grip variables. When examining the performance degradation that would result from eliminating each of these categories in turn, movement smoothness was found to generally be the least crucial for predicting functional abilities (Table 3). This observation may be partly due to the fact that smoothness has little impact on the scores of the GRASSP, ARAT, or SCIM, whereas active range of motion and grip ability will directly impact the ability to complete various components of these tests successfully. Nonetheless, our results create an interesting contrast with similar studies performed on populations of stroke survivors. In that context, movement smoothness is often cited as an important predictor of function [14, 15, 17, 18]. It is therefore important to frame any relationship between robot-derived data and clinical scores within the context of a specific injury. Which predictive variables are found to be most useful should be reflective of the nature of the injury. After mild or moderate stroke, the damaged regions of the brain often result in poor control of movements, which can be reflected in hand trajectory smoothness. In contrast, after SCI, the impairment is due to more fundamental poverty of movement (i.e. muscle paresis or paralysis), and may therefore be better predicted by measures of range of motion and/or muscle strength (e.g. grip). This argument is supported by a previous study by Wirth *et al.*, which demonstrated that in the ankle joint impairment after stroke involves reductions in both dexterity and muscle strength, whereas impairment in SCI involves reductions primarily in muscle strength, not dexterity [26]. Further studies using the automated and quantitative data collection of robotic rehabilitation devices may therefore result in improved understanding of the underlying recovery processes in each type of injury, as well as inform and more accurately guide rehabilitation strategies.

References

[1] Sadowsky CL, and McDonald JW. Activity-based restorative therapies: Concepts and applications in spinal cord injury-related neurorehabilitation. Developmental Disabilities Research Reviews 2009; 15: 112-6.

[2] Krebs HI, Hogan N, Aisen ML, Volpe BT. Robot-aided neurorehabilitation. IEEE Transactions on Rehabilitation Engineering 1998; 6: 75-87.

[3] Volpe BT, Krebs HI, Hogan N, Edelstein L, Diels C, Aisen M. A novel approach to stroke rehabilitation: Robot-aided sensorimotor stimulation. Neurology 2000; 54: 1938-44.

[4] Reinkensmeyer DJ, Kahn LE, Averbuch M, McKenna-Cole A, Schmit BD, Rymer WZ. Understanding and treating arm movement impairment after chronic brain injury: Progress with the ARM guide. Journal of Rehabilitation Research and Development 2000; 37: 653-62.

[5] Lum PS, Burgar CG, Shor PC, Majmundar M, Van der Loos M. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. Archives of Physical Medicine and Rehabilitation 2002; 83: 952-9.

[6] Loureiro R, Amirabdollahian F, Topping M, Driesse B, Harwin W. Upper limb robot mediated stroke therapy - GENTLE/s approach. Autonomous Robots 2003; 15: 35-51.

[7] Hesse S, Schulte-Tigges G, Konrad M, Bardeleben A, Werner C. Robot-assisted arm trainer for the passive and active practice of bilateral forearm and wrist movements in hemiparetic subjects. Archives of Physical Medicine and Rehabilitation 2003; 84: 915-20.

[8] Sanchez RJ, Liu J, Rao S, et al. Automating arm movement training following severe stroke: Functional exercises with quantitative feedback in a gravity-reduced environment. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2006; 14: 378-89.

[9] Housman SJ, Scott KM, Reinkensmeyer DJ. A randomized controlled trial of gravity-supported, computer-enhanced arm exercise for individuals with severe hemiparesis. Neurorehabilitation and Neural Repair 2009; 23: 505-14.

[10] Nef T, Mihelj M, Riener R. ARMin: A robot for patient-cooperative arm therapy. Medical & Biological Engineering & Computing 2007; 45: 887-900.

[11] Staubli P, Nef T, Klamroth-Marganska V, Riener R. Effects of intensive arm training with the rehabilitation robot ARMin II in chronic stroke patients: Four single-cases. Journal of Neuroengineering and Rehabilitation 2009; 6: 46.

[12] Kowalczewski J, Chong SL, Galea M, Prochazka A. In-home tele-rehabilitation improves tetraplegic hand function. Neurorehabilitation and Neural Repair, E-pub ahead of print.

[13] Krebs HI, Volpe BT, Ferraro M, et al. Robot-aided neurorehabilitation: From evidence-based to science-based rehabilitation. Topics in Stroke Rehabilitation 2002; 8: 54-70.

[14] Rohrer B, Fasoli S, Krebs HI, et al. Movement smoothness changes during stroke recovery. The Journal of Neuroscience 2002; 22: 8297-304.

[15] Chang JJ, Yang YS, Wu WL, Guo LY, Su FC. The constructs of kinematic measures for reaching performance in stroke patients. Journal of Medical and Biological Engineering 2008; 28: 65-70.
[16] Colombo R, Pisano F, Micera S, et al. Assessing mechanisms of recovery during robot-aided neurorehabilitation of the upper limb. Neurorehabilitation and Neural Repair 2008; 22: 50-63.
[17] 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. Neurorehabilitation and Neural Repair 2010; 24: 62-9.
[18] Celik O, O'Malley MK, Boake C, Levin H, Yozbatiran N, Reistetter T. Normalized movement quality measures for therapeutic robots strongly correlate with clinical motor impairment measures. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2010; 18: 433-444.
[19] Kalsi-Ryan S, Beaton D, Curt A et al. The Graded Redefined Assessment of Strength Sensibility and Prehension (GRASSP) - Reliability and Validity. J Neurotrauma. 2011; Epub ahead of print.
[20] Carroll D. A quantitative test of upper extremity function. J Chronic Diseases 1965; 18:479-491.
[21] Catz A, Itzkovich M, Tesio L, et al. A multicenter international study on the spinal cord independence measure, version III: Rasch psychometric validation. Spinal Cord 2007; 45: 275-91.
[22] Rudhe C, van Hedel HJ. Upper extremity function in persons with tetraplegia: Relationships between strength, capacity, and the spinal cord independence measure. Neurorehabilitation and Neural Repair 2009; 23: 413-21.
[23] Zariffa J, Kapadia N, Kramer JLK, et al., Feasibility and efficacy of upper limb robotic rehabilitation in a subacute cervical SCI population, Spinal Cord, 2011, in press.
[24] Dawson-Saunders B, Trapp RG. Basic and Clinical Biostatistics 2nd ed. East Norwalk CT: Appleton & Lange 1994.
[25] Fawcett JW, Curt A, Steeves JD, Coleman WP, Tuszynski MH, Lammertse D, et al. Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP panel: spontaneous recovery after spinal cord injury and statistical power needed for therapeutic clinical trials. Spinal Cord 2007;45:190-205.
[26] Wirth B, van Hedel HJ, Curt A. Ankle dexterity remains intact in patients with incomplete spinal cord injury in contrast to stroke patients. Experimental Brain Research. 2008; 191: 353-61