Robotic therapy for chronic stroke: general recovery of impairment or improved task-specific skill?

Authors: Tomoko Kitago¹, Jeff Goldsmith², Michelle Harran¹, Leslie Kane¹, Jessica Berard¹, Sylvia Huang¹, Sophia L. Ryan¹,³, Pietro Mazzoni¹, John W. Krakauer¹,⁴, Vincent S. Huang³

¹Department of Neurology, Columbia University College of Physicians and Surgeons, New York, NY, USA
²Department of Biostatistics, Mailman School of Public Health, New York, NY, USA
³Department of Neurology, Massachusetts General Hospital, Boston, NY, USA
⁴Departments of Neurology and Neuroscience, Johns Hopkins University, Baltimore, MD, USA

Corresponding author:
Tomoko Kitago
710 W. 168th St.
New York, NY 10032
Email: tk2229@columbia.edu

*These authors contributed equally to this work.
Abstract

There is a great need to develop new approaches for rehabilitation of the upper limb after stroke. Robotic therapy is a promising form of neurorehabilitation that can be delivered in more intensive regimens than conventional therapy. Here we sought to determine whether the reported effects of robotic therapy, which have been based on clinical measures of impairment and function, are accompanied by improved motor control. Patients with chronic hemiparesis were trained for three weeks, 3 days a week, with titrated assistive robotic therapy in two and three dimensions. Motor control improvements (i.e., skill) in both arms were assessed with a separate untrained visually guided reaching task. We devised a novel PCA-based analysis of arm trajectories that is sensitive to changes in the quality of entire movement trajectories without needing to pre-specify particular kinematic features. Robotic therapy led to skill improvements in the contralesional arm. These changes were not accompanied by changes in clinical measures of impairment or function. There are two possible interpretations of these results. One is that robotic therapy only leads to small task-specific improvements in motor control via normal skill learning mechanisms. The other is that kinematic assays are more sensitive than clinical measures to a small general improvement in motor control.

Keywords: stroke, neurorehabilitation, robotic therapy, kinematics, motor control
Introduction:

Robotic therapy has emerged as promising modality for stroke rehabilitation as it has the advantage of being able to deliver therapy at a much higher intensity and dosage than conventional therapy. The largest study of robotic therapy conducted to date is the VA ROBOTICS study, a randomized control trial (RCT) of robotic therapy of the upper limb for patients with chronic stroke using the MIT-Manus device (Lo et al. 2010). It was shown that after 12 weeks of therapy, the patients who received robotic training had slightly larger reductions in arm impairment compared to those who received usual care, but this difference in Fugl-Meyer Upper Extremity (FM) score was a clinically negligible 2.2 points. More recently, a study was conducted with a robot that allows 3-dimensional (3D) movements as compared to the 2-dimensional (2D) planar movements of the MIT-Manus. As in the VA ROBOTICS study, there were significant but negligible effect sizes on the FM score (Klamroth-Marganska et al. 2014). Several other smaller studies have also shown that chronic stroke patients undergoing robotic therapy have similar, if not better gains than conventional therapy (Housman et al. 2009; Liao et al. 2012; Lum et al. 2002; Page et al. 2012; Volpe et al. 2008), with mean FM changes ranging from 2-7 points. These results suggest that robotic therapy does have the potential to elicit improvements at the level of impairment, but the effect is consistently small (Kwakkel et al. 2008).

While the FM is a widely used measure of motor impairment after stroke, it requires a combination of strength and motor control. This makes it difficult to distinguish between these two different aspects of movement, which may dissociate with respect to recovery time course and respond differently to training (Noskin et al. 2007). Here we distinguish motor control from strength in that to be skillful, contractions of various muscle groups must be well coordinated.
Thus, we define motor control, or motor skill, as the ability to make accurate and precise, goal-directed movements without using compensatory movements (Kitago et al. 2013) or reducing movement speed (Reis et al. 2009; Shmuelof et al. 2012), and define reduced motor control as the loss of this ability.

In the present study, we sought to identify improvements in movements that occur with robotic therapy by performing kinematic analysis on visually guided gravity-supported planar reaching movements, which allowed us to isolate changes in motor control from recovery of strength (Kitago et al. 2013). Kinematic analyses also offer more objective measures of motor performance, compared to clinical assessments such as the FM, which is subject to considerable variance (See et al. 2013). To date, very few robotics rehabilitation studies have used kinematic analysis to look at the effects of robotic training. Krebs and colleagues (Krebs et al. 1998) showed that subacute stroke patients had improved ability to draw circles after robotic therapy. Kahn and colleagues (Kahn et al. 2006) reported that after 24 sessions of robotic therapy in chronic stroke, patients had straighter reaching trajectories and made fewer submovements. Another study (Lum et al. 2002) showed that after 2 months of robotic training, patients had longer reaching extent, along with significant improvements in FM scores.

Another advantage of using kinematic analysis is that it is sensitive to potential changes in the “unaffected arm”, which would go undetected by measures like the FM. It is well documented that the ipsilesional arm of stroke patients also has deficits in motor control but without the confounding effect of weakness (Noskin et al. 2007; Schaefer et al. 2007; Yarosh et al. 2004). Recent studies in animals have shown that skill training is not limb-specific, and has effects on both the trained and untrained sides. Repetitive use of the impaired limb in rats following unilateral ischemic stroke triggered reorganization of the cortex in the intact
hemisphere (Barbay et al. 2013). Studies of motor learning in humans also show significant
degrees of improvement in the untrained arm and hand (Grafton et al. 2002; Parlow and
Kinsbourne 1989; Wiestler et al. 2014). Thus the unaffected arm can be viewed as a unique
opportunity to assay for a motor learning effect on the affected side.

To assay motor control, we use a gravity-supported, planar reaching task that minimizes
the contribution of anti-gravity strength and the use of compensatory strategies (Kitago et al.
2013). Unlike the robotic training protocol for which 3D-dimensional training was employed, the
2D planar assay does not have joint angle redundancy amenable to compensation. This makes it
an ideal choice for assessing motor control, and can be considered the proximal analogue of the
finger individuation task (Kitago et al. 2013; Lang and Schieber 2004). By using this planar assay,
we also sought to test motor control on a task separate from those performed during the training
sessions. We applied a novel kinematic analysis to the visually guided reaching trajectories and
compared this approach with standard functional and impairment assessments, the Action
Research Arm Test (ARAT) and the FM, respectively.

The first objective of the current study was to investigate whether motor control in the
affected arm changes in response to robotic training in patients with chronic stroke. Our second
objective was to investigate whether training the paretic limb caused changes in motor control
of the ipsilesional (non-paretic) limb, which has not been previously examined in studies of
robotic arm training for patients with chronic stroke.

Methods

Study participants
Nine patients with chronic stroke were recruited from the outpatient rehabilitation clinics at Columbia University Medical Center between June 2009 and May 2011. Subjects were included with the following criteria: (1) ischemic stroke at least 6 months prior to the start of the first assessment session, (2) motor deficit involving one arm with movements at least 20° wrist extension and 10° finger extension, (3) ability to sit and be active for an hour on a chair without cardiac, respiratory and/or pain disturbances. We chose to test subjects with moderate impairment of the upper extremity because we were specifically interested in motor control changes rather than weakness, and we reasoned that severely impaired subjects may have greater masking of motor control capacity by concomitant weakness. Subjects were excluded if they (1) were unable to understand and/or follow instructions, (2) had pain in shoulder or arm (visual analogue scale >= 4), (3) had other neurological or musculoskeletal affecting the upper limb, (4) were unable to give informed consent, or (5) were under 18 years of age. Kinematic data from fourteen neurologically healthy control subjects (9 women, 5 men, mean age 60.8±9.4 years) recruited from the local community were used as the reference group for the kinematic analysis.

**Ethics Statement**

The study was conducted in accordance with the principles expressed in the Declaration of Helsinki. All participants signed a written consent form that was approved by the Columbia University Medical Center Institutional Review Board. The study is registered with ClinicalTrials.gov (NCT02331407).
Intervention

Patients trained with their affected arm using a robotic device (ReoGo™) (Figure 1) that assisted them in various 3-dimensional, goal-directed reaching exercises. They received training three days a week for three consecutive weeks. On each day of training, there were two training sessions, each averaging 1.5 hours in duration, separated by a break. Each session consisted of the patient sitting next to the ReoGo™ system, a robotic guide that extended near the subject’s arm. The subject’s arm was then attached to the robotic guide through a brace with Velcro straps. On the first day of testing, subjects were fitted specifically to the device. During the therapy, the subjects controlled a cursor on a computer monitor by moving the robotic guide, and were instructed to move the cursor to targets that were highlighted.

The therapy was conducted according pre-determined protocols for low- and medium-functioning patients. Patients who were only able to abduct their arm less than 30° were considered ‘low-functioning’, and those who had between 30-60° of arm abduction were considered ‘medium-functioning’. The exercises included goal-directed tasks that required reaching to targets in 2-dimensional (controlling anterior/posterior and medio/lateral directions) or 3-dimensional (which had the addition of a vertical movement component) space. Nine different tasks were performed during the training sessions: forward thrust, horizontal reaching, forward reaching in 2D and 3D, horizontal abduction, making a circle, reaching in a star pattern, reaching in a zigzag pattern, and mimicking bringing a cup to the mouth.

The ReoGo™ has the capacity to operate in five modes of interaction: 1) Guided: the patient is 100% assisted by the robot but must be attentive to the passive movement, 2) Initiated: the patient correctly initiates the movement, and then is 100% assisted to complete the task, 3)
Step Initiated: the patient correctly initiates the movement, then is assisted for a short distance along a predefined trajectory, then is required to initiate the next segment of the trajectory, 4) Follow Assist: the patient is allowed to move along the correct predetermined trajectory, while the robot assists to prevent deviation from this trajectory, and 5) Free: The patient completely controls the movement.

Separate protocols were followed for low-and-medium functioning patients, with pre-specified tasks, number of repetitions, and degree of assistance for each session. Each week, the therapy protocol emphasized progressively more challenging movements, with an increase in the number of arm movement repetitions, and a decrease in the amount of assistance from the robotic device.

Study design

Clinical and kinematic assessments were performed at four time points: 1) 3-weeks prior to therapy, 2) 1-week prior to therapy, 3) after completion of therapy, 4) 3-weeks post therapy. The two assessments performed prior to the robotic intervention were to confirm that the patients had a stable baseline and to examine the contribution of practice effects, particularly for the planar reaching task.

Clinical outcome measures

All clinical assessments were performed by a single occupational therapist blinded to the patients' performance during therapy. The primary clinical outcome measures were the Fugl-
Meyer Upper Extremity Scale (FM) (Fugl-Meyer et al. 1975) and the Action Research Arm Test (ARAT) (Lyle 1981). Both of these tests are frequently used and have been shown to have good reliability, validity, and responsiveness to motor change in patients with chronic stroke (Gladstone et al. 2002; Hsieh et al. 2009; Lang et al. 2006; Yozbatiran et al. 2008).

The Fugl-Meyer is a measure of impairment that considers movements arm, wrist, hand, and coordination. Each of the 22 items is scored on a 3-point ordinal scale for a maximum score of 66. The ARAT tests hand and arm function and consists of 19 items in 4 domains: grasp, grip, pinch, and gross movement. Each domain contains items arranged into hierarchical order of difficulty such that success at the most difficult item of a specific subclass assumes success for all items lower in the hierarchy of the same class. Each item is scored on a 4-point ordinal scale with a maximum score of 57.

**Kinematics: Motor Control:**

**Data Collection**

In order to assess the effect of robotic training on motor control of the upper extremity, patients were tested on an untrained planar reaching task, which has been previously described (Kitago et al. 2013). Subjects sat at a glass-surface table with their trunk securely belted to a high-backed chair. Table height was adjusted so that the shoulder and elbow were planar. The wrist, hand, and fingers were immobilized with a splint, which only allowed for movements of the shoulder and elbow. The forearm was supported using an air-sled system, which created a frictionless surface for movements. A mirror reflecting a computer display was placed just above the upper extremity so that the subjects were unable to see their arm. Hand position was tracked
real-time using Flock of Birds (Ascension Technology, Burlington, Vermont, United States) magnetic movement recording system at a frequency of 120 Hz and used to provide visual feedback.

Kinematic data from the hand, elbow, and shoulder were calculated and recorded using custom written routines in RealBasic (Real Software, Austin, TX). The target set consisted of eight radially-arrayed circles with a 1 cm radius, 45 degrees apart, 8 cm from a center start circle. Each trial began after the subject held the cursor inside the start circle for 750 ms. Patients were instructed to make straight, out-and-back movements with a sharp reversal within the target. To ensure that movements were made quickly and to minimize on-line corrections, cursor feedback stopped after 200 ms and the reversal point was indicated by a white square. Patients were given one or two practice runs of 88 movements for each arm to become familiar with the task. Patients then completed two experimental runs, each comprising 11 cycles of eight targets, for each arm.

Data Analysis

Hand position data were analyzed using custom routines in MATLAB (Mathworks, Inc., Natick, MA). Position time series were low-pass filtered (Butterworth filter) at 8 Hz. The first velocity peak above a threshold of 10 cm/s was identified for each trial. This threshold was chosen to exclude small movements made by some patients who had difficulty stabilizing their hand within the start circle. The start of the movement was defined as either the point at which the velocity crossed 1 cm/s or the first velocity minimum prior to the first velocity peak above 10 cm/s.
cm/s, whichever was later. The end point of the outward movement was defined as the reversal point—that is, the point where distance from the origin stopped increasing.

The following types of movements were excluded from analysis: movements that did not reach 30% of the distance to the target, movements without reversals, spatial outliers (in which the movement direction was >90° from the target direction).

Functional Principal Component Analysis

Traditional kinematic analysis of limb trajectories relies on examining specific kinematic variables, such as directional error, smoothness, and endpoint accuracy. The choice of variables is dictated by the study's hypotheses. In the case of recovery from stroke, the range of potential changes in limb trajectory during recovery is not known \textit{a priori}. Although some specific variables have been shown to change (e.g., number of submovements (Rohrer et al. 2002)), these were pre-selected based on specific questions about recovery. It would be desirable to examine trajectories at a global level with a method that is sensitive to any changes in overall movement quality, without preselecting a long list of variables and incurring the risk of unnecessary multiple comparisons. We therefore devised a novel method of trajectory analysis, based on functional principal components analysis (FPCA) (Goldsmith et al. 2013; Yao et al. 2005), to characterize reaching kinematics. This approach extends principal component analysis to time series data. Its main benefit is that it examines the entire trajectory and has sensitivity for changes undetected by conventional analyses that focus only on pre-selected measures such as end-point accuracy or peak velocity. It should be emphasized that FPCA will detect and incorporate all these kinematic features anyway.
We represent kinematic data as \((X_i(t), Y_i(t))\), where \(X_i(t)\) and \(Y_i(t)\) are the X- and Y-position of the hand at time \(t\). FPCA expresses each motion as the combination of population-level components, selected to capture the major features of the kinematic data, and motion-specific weights or scores:

\[
X_i(t) = \overline{X}(t) + \sum_{k=1}^{3} c_{ik}^X X_k(t) \quad \text{and} \quad Y_i(t) = \overline{Y}(t) + \sum_{k=1}^{3} c_{ik}^Y Y_k(t)
\]

(1)

Here \(\overline{X}(t)\) and \(\overline{Y}(t)\) are population mean functions, \(X_k(t)\) and \(Y_k(t)\) are shared components and the \(c_{ik}^X\) and \(c_{ik}^Y\) are the motion-specific scores. The mean and shared components are estimated using all curves and, given these, scores are estimated from individual trajectories. Interpretatively, \(X_k(t)\) and \(Y_k(t)\) give a data-driven summary of the directions that reaching motions differ in the population and the scores \(c_{ik}^X\) and \(c_{ik}^Y\) quantify how these directions appear in a particular motion. Focusing on the scores \(c_{ik}^X\) and \(c_{ik}^Y\) effectively reduces the dimension of the kinematic data: 3 scores for X and Y suffice to explain more than 99% of the observed variance in the kinematic data.

The distribution of reaching trajectories can be understood through the distribution of FPCA scores; differences comparing groups are apparent in shifts in the mean or changes in the variance of the scores. We compute the squared Mahalanobis distances

\[
(MD_i^2 = (c_i - \overline{c})^T S^{-1} (c_i - \overline{c}))
\]

for each score to measure the distance of each motion from the population average. \(MD^2\) is computed with respect to a reference population, which in this case is a collection of reaching trajectories from the dominant arm of a group of aged-matched healthy older adults using the same kinematic task. Intuitively, \(MD^2\) is a generalization of the squared Z-score and quantifies the difference between motions made by healthy controls and
subjects in this study. Subject-specific average squared Mahalanobis distances (AMD²) were computed for each subject at each target for each time point. FPCA analyses were carried out using R version 3.1.1 (R Development Core Team 2014).

Statistical Analysis

For each of the outcome variables (FM, ARAT, AMD²), we used repeated measures ANOVA to detect differences between adjacent time points assuming constant variance; separate analyses were conducted for the affected and unaffected arm. Our primary focus was comparing the difference between the first two pre-tests to the difference between the second pre-test and the post-test.

For the kinematic variable (AMD²), observations were made at each of eight targets at all time points. There were several analysis options: (i) to treat all targets as independent within subjects; (ii) to treat all targets as uniformly correlated within subjects; and (iii) to treat targets as uniformly correlated within subgroups of targets. The first analysis is the least restrictive, but may neglect any within-subject correlation. The second analysis is the most restrictive in the sense of the assumed correlation structure. The final analysis balances these by assuming a reasonable correlation structure within target subgroups, according to whether they required movements at single or multiple joints. Previous studies have shown that multi-joint reaching movements are particularly difficult for stroke patients because of the need to account for interaction torques at the shoulder and elbow (Beer et al. 2000; Cirstea et al. 2003). Targets where either shoulder or elbow joint excursion accounted for >70% of the combined joint excursion were considered single-joint targets, while the remainder were considered multi-joint
targets. Each analysis is conducted using a generalized estimating equation framework for the
assumed correlation structure and target subgroup.

An *a priori* power analysis based on effect size of 3.8 and SD of 4.3 in previously reported
Fugl-Meyer changes in chronic stroke (Kitago et al. 2013) was conducted and we determined that
9 subjects would be sufficient to yield an observed power of 76%. We also conducted a power
analysis on the trajectory variables to assess the probability of detecting a true alternative
hypothesis. Effect sizes, variances, and correlation across targets for the trajectory analysis were
determined from the comparison of healthy subjects to affected stroke patients (Kitago et al.
2012). Using these values and still assuming 9 subjects, our expected power to detect a training
effect on AMD$^2$ of size 13.5 in the affected arm is 98.8%, assuming targets are independent.

Results

Table 1 shows a summary of the demographic and lesion characteristics of the 9 subjects who
were enrolled in the study. All subjects were able to complete the robotic therapy with no
adverse events, and no patients were lost to follow up. Over the 3 weeks of therapy, both low and
medium-functioning patients performed an increasing number of repetitions with progressively
less assistance from the robotic device, and progressively more complex tasks, as seen in Figure 2.

Robotic therapy had no demonstrable effect on clinical measures of impairment and function

There were no significant differences between the two baseline FM values (Pre-Test 1
=30.3, Pre-Test 2=32.8; p=0.07) nor were there any significant differences before and after
training (PreTest2=32.8, Post-Test 1 =36.0; p=0.07) (Table 2). Similarly, there were no significant
differences in the baseline ARAT measures (Pre-Test 1 =19.6, Pre-Test 2=21.3; p=0.38), nor
significant effects of training (PreTest2=21.3, Post-Test =23.7; p=0.29) (Table 2).

Movement times

Movement times in the affected arm decreased from Pre-Test 2 to Post-Test 1 (mean
difference -33 msec, p-value = 0.02, but there was no significant change in movement times
between Pre-Test 1 to Pre-Test 2 (mean difference -31 msec, p-value = 0.11), or between Post-Test
1 and Post-Test 2 (mean difference 8 msec, p-value = 0.55). In the unaffected arm, there was a
significant decrease in movement times from Pre-Test 1 to Pre-Test 2 (mean difference -26 msec,
p-value = 0.01), and between Post-Test 1 and Post-Test 2 (mean difference -15 msec, p-value =
0.00001) but there was no significant change in movement time with training between Pre-Test 2
and Post-Test 1 (mean difference -2.9 msec, p-value = 0.70).

Reaching trajectories

Figure 3 shows the reaching trajectories from a representative patient for both the
affected (trained) and unaffected (untrained) arms. Mean AMD^2 values for each testing session
are shown in Figure 4. Data for healthy control subjects is included in this figure for reference
but no statistical tests involving these values were conducted: the apparent heteroscedasticity
comparing patients to controls does not affect the validity of our tests, which consider only
within-group changes over time.
We analyzed the AMD² outcome using the three methods described above: (i) independence across targets; (ii) uniform correlation across all targets; and (iii) uniform correlation in target subgroups (Figure 5). The target subgroups we considered were single joint and multi-joint targets, defined by the degree of interjoint coordination used by healthy control subjects in reaching to these target directions. Decreases in AMD² indicate that trajectories are more similar to those of healthy controls and represent improvements in control.

A total of 251 movements (2% of total movements, 3.5% (range 0-10.7%) of affected arm movements and 0.5% (range 0-1%) of unaffected arm movements) were rejected by our pre-specified criteria. We also performed an analysis including all movements, without rejections, and obtained similar results (not shown).

Reaching skill improved in the robotically trained arm

In the contralesional arm, there was a significant decrease in AMD² with training (from Pre-Test 2 to Post-Test 1) assuming independence across targets (mean difference -16.31, p-value = 0.0073). There was a decrease assuming uniform correlation, although this failed to reach significance (mean difference -16.31, p-value = 0.077). Within the single-joint target subgroup, the decrease in AMD² was significant (mean difference -24.28, p-value = 0.030), although in the multi-joint target subgroup the decrease was not significant (mean difference -8.34, p-value = 0.348). Comparisons of the first and second pre-test time points and of the first and second post-test time points were not significant under any analysis strategy. The former indicates that testing itself was not the cause of the improvements, and the latter that the small robotic treatment effect on the control of reaching trajectories was sustained for 3 weeks. In summary,
there was a robust but small effect of robotic training on the control of visual-guided reaches in the affected arm, which was more apparent with single joint than multi-joint reaches.

Reaching skill improved in the untrained arm with initial practice but not with robotic training

In the ipsilesional untrained arm, there was a significant decrease in $\text{AMD}^2$ between the first and second pre-test assessments, assuming independence across targets (mean difference -3.05, p-value = 0.0001) and uniform correlation (mean difference -3.05, p-value = 0.042). Within the single-joint target subgroup, the decrease in $\text{AMD}^2$ was significant (mean difference -3.17, p-value = 0.015), although in the multi-joint target subgroup the decrease was not significant (mean difference -2.93, p-value = 0.11). This pattern of improvement in the control of single-joint movements but not multi-joint movements is similar to what we observed in the contralesional arm after robotic training. However, for the untrained arm, there was no significant change in $\text{AMD}^2$ with robotic training (from Pre-Test 2 to Post-Test 1) or between the first and second post-test time points under any analysis strategy. In summary, the untrained “unaffected” arm showed improvement with initial practice during testing of the kinematic task, but did not further benefit from robotic training or further practice with testing, despite not reaching the skill level of healthy controls (Fig. 4).

Discussion

We sought to detect an effect of robotic therapy on motor control in patients with chronic hemiparesis after stroke. We found that there were improvements in the control of visual-guided reaching trajectories in a planar task with the effects of gravity eliminated. There
were no concomitant improvements in measures of impairment (FM) or function (ARAT) in the contralesional arm. These results illustrate that robotic training can lead to improvements in a motor skill in the absence of changes in standard clinical assessments.

The effects of robotic therapy on motor control in the current study were small but robust, with sustained improvements 3 weeks later. Significant improvements were seen only for single-joint rather than multi-joint movements. The same pattern of improvement was seen in the ipsilesional arm after practice of the task, suggesting that a similar skill learning mechanism is occurring in both arms. This is interesting, as previous work, in healthy subjects and in patients with stroke, has shown that maintaining accuracy and path straightness is more challenging for multi-joint compared to single-joint movements because of the need to compensate for interaction torques (Beer et al. 2000; Cirstea et al. 2003). Thus it appears that after stroke, anisotropies due to limb dynamics are accentuated.

These improvements in motor control occurred without any significant increase in movement time. Improved trajectory accuracy for movements of the same or higher speed represents an improvement in the speed-accuracy trade-off. The observed improvements thus imply that a form of motor skill learning has taken place (Reis et al. 2009; Shmuelof et al. 2012). There was no ceiling effect for the planar task as healthy controls performed markedly better than the patients did even with their ipsilesional arm.

Previous studies of robotic therapy in patients with chronic hemiparesis have shown small effect sizes at the impairment level and none at the functional level (Kwakkel et al. 2008; Lo et al. 2010; Prange et al. 2006). Here we report significant but small effects only at the level of motor control. The magnitude in improvement in the FM in our study was comparable to what
was found in two larger studies of robotic arm training in chronic stroke patients, the VA ROBOTICS study (Lo et al. 2010) (3.9 points) and the recently published 3D robotics study (Klamroth-Marganska et al. 2014) (3.3 points). While the lack of significant improvements in our clinical measures may be related to our small sample size, the absolute changes with training were small (3.2 points on the FM, and 2.4 points on the ARAT) and unlikely to be clinically meaningful (Gladstone et al. 2002; Van der Lee et al. 2001). Reaching kinematics were not assessed in either of these aforementioned studies and so it is not possible to say whether the small improvements seen at the impairment level occurred through improvements in motor control.

The absence of significant effects with the ARAT and the FM could be interpreted to mean that the improvements in motor control were not large enough to generalize to either of these measures. That is to say, the motor control assay may be more sensitive to small effects of robotic training. Furthermore, the movements that are tested in ARAT and FM involve a combination of motor control, strength, and can incorporate the use of compensatory strategies. Thus, an improvement in motor control may not be sufficient to lead to improvements on these clinical assessments, if there is no improvement in the other domains.

Our study population encompassed a wide range of severity, with baseline scores at ranging from 9 to 50 for the FM-UE, and from 0-38 for the ARAT. We explored the relationship between baseline clinical scores and the amount of improvement with training for the affected arm. Subjects with lower FM-UE scores tended to have larger improvements with training, but for the ARAT, subjects with higher scores tended to improve more with training, though these correlations were non-significant. This study was not powered for subgroup analyses based on
the severity of the motor deficits, but we found little evidence that the heterogeneity in our population was masking an improvement in the lower or higher functioning groups.

For the ipsilesional, untrained arm, no further improvement was seen after the three weeks of training with the affected arm despite a persistent impairment when compared with healthy control subjects. This lack of generalization to the untrained arm does not necessarily signify that motor learning has not occurred for the trained arm. Previous studies that have demonstrated improvements in both trained and untrained limbs have not examined the quality of the movements as we did in this study, with the exception of one primate study in which the animals were trained and tested with a planar reaching task (Georgopoulos et al. 1981). It may be that the magnitude of motor skill improvement in our study was not large enough to generalize to the untrained side, or that in chronic stroke patients, the type of motor skill learning we tested is limb and task specific (Bavelier et al. 2012). Indeed, it is the assumption that rehabilitation is based on motor learning coupled with task-specificity that has led to the emphasis placed on task-oriented training after stroke (Bayona et al. 2005; French et al. 2010; Hubbard et al. 2009; Schaefer et al. 2013).

The critical question is whether task-oriented training after stroke is merely exploiting normal learning mechanisms to increase skill within the performance envelope available to the patient. This would be comparable to a healthy person getting better at handwriting with their non-dominant arm – it has a baseline level of performance that can be improved with practice. The core point is that healthy subjects and patients alike can augment performance to some degree on any task with practice. There is no need to invoke repair or reorganization. Thus the fundamental question raised by our results is whether the robotic training led to an increase in
task-specific skill or to a general improvement in motor control. One possibility is that there are 2D task elements in the robotic training that are similar to the movements in the planar assay, which would mean that improvements that we observed can be attributed to task-specific skill. Alternatively, as mentioned already, the training could have had a more general effect on motor control but it was too small to either generalize or be detected by the ARAT and FM. In the VA ROBOTICS study there was a small effect of planar training on the FM, which might be taken as evidence for generalization from 2D reaches with gravity eliminated to 3D multi-joint movements.

The ambiguity in ascertaining precisely what robotic therapy accomplishes arises because of the inherent difficulties of using task A (planar assay) to assess the effect of training with task B (robot). Task A can be viewed both as an assay for general motor control or a specific task that one can become more skilled at through training. As a hypothetical example, imagine two stroke patients with hemiparesis. The first patient has a hemiparesis of moderate severity and is assessed on the planar reaching task before any rehabilitation is given. The second patient has more severe arm impairment and receives 2 weeks of robotic training. This second patient is then assessed post-training with the planar reaching task and is found to have exactly the same level of performance on it as the first patient. Thus the two patients now look phenotypically identical on the planar reaching task but one patient required robotic training and the other did not. Are they phenotypically identical because the robotic training has partially reversed the second patient’s more severe impairment, or has the robotic training only made the second patient a little more skilled at the planar task but is otherwise unchanged? This emphasizes the often unappreciated fact that pretraining correlation does not imply training covariation (Moreau and Conway 2014).
Ideally one would have a battery of pre-established compensation-proof tasks with some

*a priori* framework for how they differ from each other and from the training task. The question, however, of generalization of task-specific training is a vexed one because it pre-supposes that we have an *a priori* definition of task. If the degree of similarity between task A and task B is established via a transfer assay then that same assay cannot be taken as evidence for generalization in the extrapolative sense because transfer occurred only to the degree that there was *already* overlap between tasks A and B *within* the part of parameter space trained. Thus transfer of learning across tasks can only be considered generalization if there is a separate assay for task similarity. In a recent study that looked at training on three tasks for the arm and hand, kinematics of the proximal limb was used as the measure of similarity across tasks. Interestingly, kinematic similarity did not predict the degree of transfer. The authors postulated that goal similarity might be the feature that matters instead. Goal, however, is a fuzzy concept – at one level the goal is always to complete the task successfully, but successful task completion necessarily depends on the specifics of the task and its context. For example, in recent studies of split-treadmill adaptation after stroke, the poor generalization to over-ground walking has been taken as evidence for a context effect preventing transfer of what is a very similar locomotor pattern on and off the treadmill (Reisman et al. 2009; Torres-Oviedo and Bastian 2012). Finally it needs to be recognized that rehabilitative training can lead to increased cardiovascular fitness and muscular conditioning, in addition to neural changes. The very small amount of generalization to the FM scale seen in the VA ROBOTICS study (Lo et al. 2010) could be attributable to strength increase in the proximal limb rather than motor learning effects.

On balance we would tentatively conclude that the improvement in motor control in the planar reaching task after 3 weeks of robotic therapy is because of some skill improvement at the
task due to overlap with some of the planar components of robotic training. The lack of
significant improvement in the FM and ARAT further support our interpretation of a small
increment in task-specific skill.

We have recently shown that constraint-induced movement therapy (CIMT) seems to
promote functional recovery in chronic stroke mainly, if not exclusively, through compensation.
Specifically, we showed that there was no discernible improvement in motor control, using the
same planar reaching task as used in the current study, despite a clinically meaningful change in
the ARAT (Kitago et al. 2013). Thus for CIMT the results were the opposite of those obtained
here for robotic therapy, which suggests a dissociation for these two rehabilitation approaches
with respect to gains in motor control versus function, the latter presumably driven by
compensation. This conclusion is not inconsistent with previously reported results, which
overall suggest that robotic therapy exerts its small effects at the impairment level (Huang and
Krakauer 2009; Kwakkel et al. 2008; Lo et al. 2010; Prange et al. 2006) whereas CIMT exerts it
effects at the functional level (Kitago et al. 2013; Wolf et al. 2006).

Conclusion

Patients with chronic hemiparesis who underwent 3 weeks of 2D and 3D robotic training
showed small but robust improvements in the control of reaching trajectories, which were
assayed with an untrained planar task. The same pattern of skill improvement was seen on the
untrained ipsilesional side with practice. That patients can show small increases in skill on any
given task is hardly surprising, as to date there is no evidence that skill learning mechanisms are
impaired after stroke; hemiparesis is a motor control deficit (Krakauer 2006; Winstein et al.
We conclude that training can eke out small task-specific performance improvements in motor control in patients after stroke through normal motor learning mechanisms, which cannot reverse damage to the corticospinal tract or trigger reorganization. Generalization will only occur to the degree that the tasks share parts of command space. The much larger and much more general gains seen early after stroke are due to unique plasticity conditions in a limited time-window that interact with training to enable reorganization and repair processes that are qualitatively and quantitatively distinct from normal learning (Zeiler and Krakauer 2013). Robotic therapy initiated within this time-window might be able to augment such spontaneous biological recovery.

Acknowledgments

We would like to thank Motorika Medical Ltd. for supplying the ReoGo™ robotic device used in this study.

Grants
National Institutes of Health K02 NS048099 (JWK)
Orentreich Family Foundation (JWK)

Disclosures
The authors report no conflicts of interest.
References


Page SJ, Hill V, and White S. Portable upper extremity robotics is as efficacious as upper extremity rehabilitative therapy: a randomized controlled pilot trial. Clinical rehabilitation 2012.


Figure 1.
ReoGo robotic device. Picture provided courtesy of Motorika Medical Ltd.

Figure 2.
Average number of repetitions per session, divided into A) the 5 modes of interaction and B) the different exercises performed. Separate pre-specified protocols were followed by low- and medium-functioning patients. Each week, there were progressively more challenging movements, with an increase in the number of repetitions, a decrease in the amount of assistance from the robotic device, and incorporation of more complex movements. Init = initiated; S.Init = step initiated; F. assist = follow assist; For. = forward, Horz. = horizontal; Abd. = abduction.

Figure 3.
Reaching trajectories from a representative subject, before and after training for the affected (top) and unaffected (bottom) arms.

Figure 4.
Average squared Mahalanobis distances (± confidence intervals) at each time point, for the affected and unaffected arms. Values from a reference population of healthy control subjects are presented for comparison. Even with the unaffected arm, patients do not reach the level of performance of the healthy control subjects.

Figure 5.
Changes in Average squared Mahalanobis distances (± confidence intervals) across time points for three methods: assuming independence across targets (Ind.), assuming uniform correlation across targets (Unif.), and assuming correlations within subgroups for single-joint (Sing.) and multi-joint (Mult.) targets. Negative values represent an improvement in reaching kinematics. The only changes that reach statistical significance (*) occur between Pre-Test 2 and Post-Test 1, for the affected / trained arm.
<table>
<thead>
<tr>
<th>Week</th>
<th>Avg Repititions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>400</td>
</tr>
</tbody>
</table>

**Exercises: Low Functioning Patients, n=3**
- For Reach 2D
- For Reach 3D
- Horz Abd
- Star
- Horz Reach
- For Thrust

**Exercises: Medium Functioning Patients, n=6**
- For Reach 2D
- For Reach 3D
- Horz Abd
- Star
- Horz Reach
- For Thrust

**Modes: Low Functioning Patients, n=3**
- Guided
- Init
- S. Init
- Free

**Modes: Medium Functioning Patients, n=6**
- Guided
- Init
- S. Init
- Free

1 trial of Circle
1 trial of ZigZag
Affected

Single-joint

Multi-joint

(cm)

Pre Test 2

Post Test 1

Affected

Unaffected

(cm)

(cm)

Single-joint

Multi-joint
PreTest1 v PreTest2
PreTest2 v PostTest1
PostTest1 v PostTest2

AMD: Estimate and CI

<table>
<thead>
<tr>
<th>Model</th>
<th>PreTest1 v PreTest2</th>
<th>PreTest2 v PostTest1</th>
<th>PostTest1 v PostTest2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unif.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sing.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mult.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05
Table 1: Description of subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Months since stroke</th>
<th>Stroke location</th>
<th>Stroke Side (Brain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>60</td>
<td>14</td>
<td>pons</td>
<td>R</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>45</td>
<td>12</td>
<td>MCA (fronto-temporal-parietal)</td>
<td>L</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>79</td>
<td>8</td>
<td>internal capsule and corona radiate</td>
<td>L</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>35</td>
<td>29</td>
<td>MCA</td>
<td>L</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>63</td>
<td>49</td>
<td>pons</td>
<td>L</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>75</td>
<td>15</td>
<td>internal capsule</td>
<td>L</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>73</td>
<td>7</td>
<td>MCA (parietal)</td>
<td>L</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>54</td>
<td>15</td>
<td>internal capsule</td>
<td>L</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>47</td>
<td>9</td>
<td>internal capsule</td>
<td>R</td>
</tr>
</tbody>
</table>
Table 2. Clinical scores.

| Subject | FM-UE Scores | | | | ARAT Scores | | | |
|---------|--------------|---|---|---|---|---|---|---|---|
|         | Pre Test 1  | Pre Test 2 | Post Test 1 | Post Test 2 | Pre Test 1 | Pre Test 2 | Post Test 1 | Post Test 2 |
| 1       | 33          | 37          | 42          | 45          | 36          | 37          | 36          | 36          |
| 2       | 50          | 57          | 55          | 59          | 37          | 41          | 42          | 39          |
| 3       | 9           | 9           | 13          | 13          | 0           | 0           | 1           | 2           |
| 4       | 13          | 16          | 19          | 20          | 4           | 6           | 5           | 8           |
| 5       | 45          | 46          | 44          | 46          | 36          | 37          | 34          | 36          |
| 6       | 21          | 18          | 30          | 31          | 0           | 10          | 12          | 19          |
| 7       | 37          | 45          | 44          | 50          | 25          | 33          | 36          | 39          |
| 8       | 48          | 48          | 51          | 51          | 38          | 28          | 46          | 42          |
| 9       | 17          | 19          | 26          | 30          | 0           | 0           | 1           | 4           |
| Mean    | 30.3        | 32.8        | 36.0        | 38.3        | 19.6        | 21.3        | 23.7        | 25.0        |
| SD      | 15.8        | 17.4        | 14.6        | 15.5        | 18.0        | 17.1        | 18.6        | 16.7        |