# Kinematic Robot-Based Evaluation Scales and Clinical Counterparts to Measure Upper Limb Motor Performance in Patients With Chronic Stroke

**Caitlyn Bosecker**[1], **Laura Dipietro, PhD**[1], **Bruce Volpe, MD**[2], and **Hermano Igo Krebs, PhD**[1,2,3]

[1]Massachusetts Institute of Technology, Cambridge, MA, USA

[2]Weill Cornell Medical College, New York, NY, USA

[3]University of Maryland, Baltimore, MD, USA

## Abstract

**Background—**Human-administered clinical scales are the accepted standard for quantifying motor performance of stroke subjects. Though widely accepted, these measurement tools are limited by interrater and intrarater reliability and are time consuming to apply. In contrast, robot-based measures are highly repeatable, have high resolution, and could potentially reduce assessment time. Although robotic and other objective metrics have proliferated in the literature, they are not as well established as clinical scales and their relationship to clinical scales is mostly unknown.

**Objective—**Test the performance of linear regression models to estimate clinical scores for the upper extremity from systematic robot-based metrics.

**Methods—**Twenty kinematic and kinetic metrics derived from movement data recorded with the shoulder-and-elbow InMotion2 robot (Interactive Motion Technologies, Inc), a commercial version of the MIT-Manus. Kinematic metrics were aggregated into macro-metrics and micrometrics and collected from 111 chronic stroke subjects. Multiple linear regression models were developed to calculate Fugl-Meyer Assessment, Motor Status Score, Motor Power, and Modified Ashworth Scale from these robot-based metrics.

**Results—**Best performance–complexity trade-off was achieved by the Motor Status Score model with 8 kinematic macro-metrics ($R = .71$ for training; $R = .72$ for validation). Models including kinematic micro-metrics did not achieve significantly higher performance. Performances of the Modified Ashworth Scale models were consistently low ($R = .35–.42$ for training; $R = .08–.17$ for validation).

**Conclusions**—The authors identified a set of kinetic and kinematic macro-metrics that may be used for fast outcome evaluations. These metrics represent a first step toward the development of unified, automated measures of therapy outcome.

## Keywords

robotic therapy; rehabilitation; stroke assessment

## Introduction

Stroke is the leading cause of permanent disability in the United States.[1] Although occupational and physical therapy are widely accepted treatments for upper extremity dysfunction in stroke patients, they are labor intensive and therefore expensive. Changes in the health care system and budgetary constraints have provided pressure to reduce the cost of therapy. Until recently, that has been done by shortening inpatient rehabilitation hospitalizations, but once the practical limit of abbreviated inpatient stays is reached, further efficiencies will be attainable chiefly by addressing clinical practices themselves.

Our research has demonstrated that the efficiency of care among stroke patients could be increased by the application of robotics and information technology, and these results have been confirmed by others.[2–4] Labor-intensive procedures have traditionally been a primary field for the application of robotics. Over the past decade, we have shown that labor-intensive physical therapies for stroke patients can be delivered safely and effectively by robotic aids. Such an approach has several major advantages over traditional motor therapies. Robots can be used to generate objective measures of patients' impairment and therapy outcome, assist in diagnosis, customize therapies based on patients' motor abilities, and assure compliance with treatment regimens and maintain patients' records. For the upper extremity, the most employed and deployed therapeutic robot is the MIT-Manus.[5] Developed by our group, this robot controls patients' arm movement in the horizontal plane. A number of studies have shown that therapy with the MIT-Manus, and its commercial version InMotion2, are safe, tolerated, and clinically effective as measured by traditional clinical scales including the Fugl-Meyer Assessment (FMA) and the Motor Status Score (MSS).[6–15] Metrics extracted from movement data recorded from patients undergoing MIT-Manus therapy have been used for quantifying patients' motor abilities and their changes during motor recovery and have offered unprecedented insights into the process of motor recovery from stroke.[16–18] It remains unclear how these robot-based kinematic and kinetic metrics relate to traditional human-administered clinical scales for measuring outcome.

The goal of this study was to test whether linear regression models can estimate upper extremity clinical scores from robot-based metrics. Such models would facilitate objective outcome measurement, freeing the therapist to render additional therapy, and, ultimately, contribute to improving patients' care.

# Methods

## Persons With Chronic Impairment Due to Stroke

One-hundred and eleven community-dwelling volunteers who had suffered a stroke trained for 18 hours with the InMotion2 robot (Interactive Motion Technologies, Inc, Cambridge, MA). Table 1 shows these volunteers' demographics. Subjects were trained in point-to-point movements, which evoked significant improvement (as measured by clinical scales) by the end of the therapy.[11] The protocol included 5 clinical evaluation sessions, 3 before treatment and 2 more at mid-point and at discharge. Pretreatment evaluation sessions took place prior to admission, with their average serving as the admission score and ensuring the patient's condition was stable ("phase-in" procedure[19]). Kinematic and kinetic robot-based metrics were acquired at admission and at discharge. Therefore, each volunteer contributed 2 data points for the regression models. We constructed models for the admission data, for the discharge data, and for the combination of admission and discharge data. Because the data were not normally distributed, we employed a Wilcoxon signed rank test and determined that there were no significant differences between the predicted clinical score employing the combined model and the admission-only model or discharge-only models. In this article, we report on the combined model.

Inclusion criteria were the following: (*a*) diagnosis of a single, unilateral stroke at least 6 months prior to enrollment verified by brain imaging; (*b*) sufficient cognitive and language abilities to understand and follow instructions (Mini-Mental Status Score of 22 and higher or interview for aphasic subjects); and (*c*) stroke-related impairments in muscle strength of the affected shoulder and elbow between grades 1/5 and 3/5 on the Motor Power (MP) scale.[20,21] Subjects were excluded from the study if they had a fixed contraction deformity in the affected limb, and also if they demonstrated improvement over 3 measurements made during the 4-week observation period prior to treatment. None of the subjects were engaged in conventional occupational or physical therapy programs or received pharmacological management of spasticity and tone (ie, Botox) during the experimental trial. The study was approved by all test site human research review boards, and all subjects volunteered for the study and gave their informed consent.

## Clinical Measures for the Upper Extremity

The FMA, MSS, the MP, and the Modified Ashworth Scale (MAS) were used to assess patients' abilities during the evaluation sessions. These scales have established reliability and validity as well as significant limitations, as described below.

The FMA has 5 evaluation domains (including sensory function, balance, and joint range of motion).[22–24] In this study, we only used the upper limb section of the FMA. Based on the levels of motor recovery after stroke, as proposed by Twitchell[25] and Brunnstrom,[26] such sections can be used to assess synergistic and voluntary motor abilities in about 30 minutes. A 3-point scale is used to measure performance (0 = *unable*; 1 = *partial*; 2 = *performs fully*) on 33 test items (total possible score = 66 points). The FMA is able to detect changes in persons with severe to moderate motor impairments after stroke, in part because of its inclusion of isolated joint movement rather than a sole emphasis on task-related actions.

Reported disadvantages of the FMA include a potential ceiling effect for higher-functioning patients and its limited number of test items to evaluate distal, fine motor function.[24] Our protocol included patients in the moderate to severe range so that the ceiling effect would not affect our results. The MSS was developed at Burke Rehabilitation Hospital due the perceived pitfalls of relying on the upper limb section of the FMA for subacute stroke patients.[27] The upper limb section of the FMA suffers from a decreased sensitivity because return of reflex and synergistic movements (a substantial component of the total score) often occur before admission and evaluation at the rehabilitation hospital. Instead, the MSS uses an end-of-scale value of 82, with a potential range of 40 points for isolated shoulder and elbow movements. Thus, the MSS aims at augmenting the FMA by further specifying the quality of voluntary movement in the hemiparetic upper limb.

The MSS elbow and shoulder section consists of a sum of scores (0–2) given to 10 isolated shoulder movements and 4 elbow/forearm movements. The MSS wrist and fingers section consists of a sum of grades for 3 isolated wrist movements and 12 hand movements.[27] To be an effective augmentation of the FMA, the MSS must preserve a linear relationship with the FMA. Results of a study with 56 subacute patients evaluated by the same "blinded" therapist showed that indeed the relationship between FMA and MSS was linear and that the MSS had good interrater as well as intrarater reliability.[27]

The Medical Research Council test of MP was used to assess strength. The MP measures strength in isolated muscle groups of the involved shoulder and elbow on an ordinal scale (scale range: 0 = *no muscle contraction*; 5 = *normal strength*).[20,21]

The MAS was used to measure hypertonia, and it rates the resistance to passive stretch in 14 different muscle groups of the upper limb.[28] It is an ordinal scale ([0, 1, 1+, 2, 3, 4] or [0, 1, 2, 3, 4, 5]). Evaluations are conducted by moving a limb about a joint at different speeds and by noting the muscular response throughout that limb range of motion (ie, both speed and position dependent).

### Kinematic Measures for the Upper Extremity

In this article, we will report on metrics obtained from 3 distinct sets of measurements: unconstrained reaching movement, unconstrained circle drawing, and shoulder strength.

The unconstrained reaching movement required the patient to perform unconstrained movement toward targets presented in 8 positions equally spaced around a 14-cm radius circle and back to the center. Targets were presented in clockwise order similar to a robot-assisted therapy session. If the patient was unable to hit a target, the therapist cued the robot to stop recording, and the patient's hand was moved to the target where recording of patient's movement proceeded. A total of 80 movements were performed for each evaluation session. The following macro-metrics were extracted from the data: the deviation from the straight line connecting the targets, aiming, movement mean and peak speed, movement smoothness (ratio of mean to peak speed), and movement duration. Movement speed profiles were decomposed into support-bounded lognormal submovements as described in Rohrer et al.[17] The following micro-metrics were derived from submovement decompositions: number of submovements, submovement duration, overlap, peak, interpeak

interval, and shape metrics μ (skewness or asymmetry) and σ (kurtosis or "fatness"). The rationale for these micro-metrics was based on a century-old conjecture that human arm movement is controlled discretely, by combining elementary units of movements or submovements.[16]

During the unconstrained circle drawing, the patient's hand was initially positioned at 3 o'clock and at 9 o'clock (right or left to the workspace center). Patients were asked to draw clockwise and counterclockwise circles starting and ending at the same point. This task, which was not trained during therapy, required the coordination of shoulder and elbow joint movements. A total of 20 movements were performed for each evaluation session. The following macro-metrics were extracted from the data as described in Dipietro et al[18]: the axes ratio (ratio of the minor to major axes of the best-fitting ellipse) and the joint angle correlation (degree of independence of the shoulder and elbow movements).

### Kinetic Measures for the Upper Extremity

Shoulder extension/flexion and abduction/adduction measurements take advantage of the fact that the MIT-Manus is a SCARA-type device (selective compliance assembly robot arm), which is highly compliant in the horizontal plane and has very low compliance on the vertical plane (almost infinite impedance). This affords the unique potential of easily testing the shoulder during flexion/extension and abduction/adduction. The patient's arm was prepositioned at 90° of shoulder flexion/extension in the sagittal plane with elbow fully extended. The hand was supported in a plastic trough at the robot's tip with no support provided at the elbow. Once the patient was secured in the trough, the therapist manually stabilized the robot's arm in the required testing position. The patient was then instructed to attempt to lift and push-down his/her arm 5 consecutive times. During abduction/adduction testing, the patient was rotated in the chair 90° to left or right depending on the side of impairment, and his/her arm was prepositioned in 90° abduction. The elbow position and support procedure was the same as during flexion/extension. Once the arm was secured, the patient was instructed to lift and push-down the arm 5 times. A total of 20 attempts to lift or push-down the arm were completed for each evaluation session. Mean shoulder strength ($Z$ force) for flexion, extension, abduction, and adduction were calculated from the data.[29]

### Statistical Analysis

Least squares error multiple linear regression models were constructed in an attempt to estimate the FMA, the MSS, the MAS, and the MP from robot kinematic and kinetic metrics. Model performance was measured by the correlation between the therapist assigned and model calculated scores. Five types of models were investigated.

The first model calculated the FMA, MSS, MAS, and MP from all the kinematic and kinetic metrics. For the second model, prior to performing multiple regressions, a principal component analysis (PCA) was conducted on all kinematic and kinetic metrics in an attempt to reduce the total number of metrics. Fewer metrics would reduce the model complexity and perhaps the number of tasks performed by the patient, thereby reducing the overall time required for patient evaluation. Because the metrics had different units, the PCA was applied

to normalized kinematic and kinetic metrics. Normalization scaled the data so that each metric had 0 mean and standard deviation of 1.

Three additional models containing different combinations of the robot metrics were constructed: kinematic macrometrics and kinematic macro-metrics and micro-metrics (submovements) for the FMA, MSS, and MAS, and force metrics for the MP.

All models were developed by randomly separating 75% of the data to train the model with the remaining 25% set aside for model validation. A Mann–Whitney test on the clinical scores confirmed that the training and validation data sets were not significantly different. To verify that this result was not due to a type II error, the data set was randomly separated into 4 groups, and Mann–Whitney cross-validation verified that there was no significant difference ($\alpha = .05$) between any combination of these groups into training and validation sets.

Correlations among the clinical scales were also investigated to identify dependent scales. A strong correlation between 2 or more clinical scales would indicate measurement redundancy, meaning it may be possible to eliminate one, thus reducing the amount of time required for patient evaluation.

## Results

### Correlation Among Clinical Scales

Table 2 shows the correlation among clinical scales.

Correlation between FMA and MSS was strong, suggesting redundancy for this population of chronic stroke patients. This was not entirely surprising because the MSS was designed to be a finer grading scale of the upper extremity impairment than the FMA for subacute patients. This result is consistent with and extends the results of previous studies for subacute stroke (Ferraro et al[27]: 12 patients, $R = .981$, $P < .0001$; Krebs et al[29]: 56 patients, $R = .981$, $P < .0001$).

The correlation between the ability to move isolated joint movement (FMA and MSS) and strength (MP) was moderate, suggesting a common trait. In contrast, the correlation between MAS and the other scales was very low. If we assume that recovery can be characterized by improvement in the ability to move isolated joints and increased joint strength, then the observed low correlation suggests that tone might not be an important marker of recovery for this chronic stroke population.

### Estimating FMA, MSS, MP, and MAS From Kinematic and Kinetic Metrics

Training and validation results of the kinematic and kinetic model to estimate the FMA, MSS, MP, and MAS are shown in Table 3. We employed all the kinematic and kinetic metrics to build such models. All correlations except for the MAS validation group ($P = .266$) were significant at $P < .005$. The residual plots did not have any significant correlations or patterns, indicating no underlying trends missed in the data and the appropriateness of our linear model.

The PCA only reduced the original 20 metrics marginally to 17 variables for the FMA and 16 variables for the MSS, MP, and MAS. The residual plots did not have any significant correlations or patterns.

### Calculating FMA, MSS, and MAS From Kinematic Metrics

The FMA and MSS evaluate motor impairment including isolated and synergistic movements. Therefore, the kinematic macro-metrics extracted from the unconstrained reaching movements (point-to-point) and the circle drawing task (independent joint movement) were selected to develop a model. In addition, the benefits from including micro-metrics based on submovements in the model were explored.

Table 4 shows the training and validation results for the model. As with the previous models, no significant correlations or patterns were found in the residual plots, and all correlations except for the MAS validation groups ($P = .306$ and $.373$, respectively) were significant after Bonferroni correction for multiple measurements at $P < .007$. For the breakdown of the 8 kinematic macro-metrics and 7 kinematic micro-metrics (submovements), the training data performed better, namely, the macro-metrics and micro-metrics model yielded to the highest $R$ values. But for the validation data, the opposite was true—the model based on the kinematic macrometrics alone had larger $R$, indicating that this model generalized better to the data. In fact, it outperformed the model that included all the 20 robot-based metrics for the validation data (both kinematic and kinetic, see Table 3). The FMA showed the highest $R$ values for the training data but only moderate $R$ values for the validation data. The MSS performance remained remarkably consistent between training and validation data (small difference among the correlation values). Neither the macro-metric model nor the micro-metric model was able to produce more than a weak correlation for the MAS score.

For all 3 scales, including the micro-metrics based on submovements into the model (see validation data) produced no additional benefit.

Because this model was the most concise and outperformed the others, the macro-metric model is reported for the FMA and MSS:

$$\begin{aligned} \text{FMA} = &\ 4.58 - 11.63*[\text{Aim}] + 37.04*[\text{Deviation}] \\ &- 29.30*[\text{MeanSpeed}] + 62.55*[\text{PeakSpeed}] \\ &+ 83.96*[\text{Smoothness}] + 1.72*[\text{Duration}] \\ &+ 2.98*[\text{EllipseRatio}] - 17.28*[\text{JoinInd.}], \end{aligned}$$

$$\begin{aligned} \text{MSS} = &-29.64 - 2.96*[\text{Aim}] - 16.12*[\text{Deviation}] \\ &- 230.22*[\text{MeanSpeed}] + 161.99*[\text{PeakSpeed}] \\ &+ 184.74*[\text{Smoothness}] + 3.36*[\text{Duration}] \\ &- 16.55*[\text{EllipseRatio}] - 35.85*[\text{JoinInd.}]. \end{aligned}$$

Because the robot metrics are not independent of each other, these regression models have a high degree of multicollinearity. Multicollinearity does not affect the models' ability to estimate the clinical scores, but it does not afford any interpretation of the magnitude of the coefficients or any suggestion that a large coefficient indicates higher impact.

To determine the magnitude of the association of each metric to the model and eliminate multicollinearity, backwards regression analyses were conducted for the FMA and MSS models.[30] The model started with all 8 macro-metrics, each was removed separately, and the $R$ value for the actual and predicted scores were calculated. The metric that resulted in the smallest reduction in model performance was removed, and the process was repeated for the remaining 7 metrics and so on until only 1 metric remained. The resulting group of metrics that achieved a balance between reducing model complexity without reducing the $R$ value of the model by more than 5% included peak speed, smoothness, duration, and joint independence. The reduced FMA model had an $R$ value of .731 and .580 for training and validation, respectively, and the reduced MSS model had an $R$ value of .704 and .696 for training and validation, respectively. The residual plots did not have any significant correlations or patterns. The resulting reduced models are presented below, and the 95% confidence intervals of the metrics are listed in Table 5:

$$FMA = -22.39 + 44.04*(\text{PeakSpeed}) + 112.94*(\text{Smoothness}) + 2.15*(\text{Duration}) - 20.00*(\text{JointInd}),$$

$$MSS = -30.61 + 75.5*(\text{PeakSpeed}) + 141.17*(\text{Smoothness}) + 3.17*(\text{Duration}) - 27.83*(\text{JointInd}).$$

For the FMA model, smoothness provides the largest contribution, and for the MSS model, the upper range of the peak speed coefficient overlaps with the lower range of the smoothness coefficient. This is not surprising because we also constructed models employing a single robotic variable. The model employing smoothness was the best performer with an $R$ value of .616 and .400 for training and validation, respectively, for FMA, and .557 and .589 for training and validation, respectively, for MSS. Duration has the smallest contribution for both reduced models, but removing it causes the models' performance to decrease by more than 5%.

### Calculating MP From Kinetic Metrics

As the MP evaluates strength at each joint, the $Z$-force metrics were selected to develop a model. In contrast to the models for FMA, MSS, and MAS described above, this model was applied to logarithmic data. The logarithmic scale was introduced to account for Weber's law, which explains how actual forces (the $Z$-force measured by the robot force transducer) are perceived by humans (the MP score recorded by the therapist). According to Weber's law, the smallest perceivable difference in weight, $p$ (the least difference that the test person can still perceive as a difference), is proportional to the starting value of the weight:

$$p = k \ln \frac{S}{S_0},$$

where $S$ is the stimulus weight, $S_0$ is the stimulus threshold, and $k$ is a proportionality constant. This model was similar to the model proposed by Krebs et al[29] for calculating MP from shoulder strength in subacute stroke patients. Our model yielded $R = .564$ for training ($P < .0001$) and $R = .525$ for validation ($P < .0001$). The residual plots did not have any significant correlations or patterns. As with the kinematic models, these 4 kinetic metrics are not independent of each other and the relative impact of each metric cannot be determined, but the models remain highly significant.

The MP force metrics model is

$$\text{MP} = 6.77 + 0.79 * \ln(Z\text{Flexion}) - 0.07 * ln(Z\text{Extension}) + 2.01 * \ln(Z\text{Abduction}) + 2.91 * \ln(Z\text{Adduction}).$$

Or in $\log_{10}$ basis

$$\text{MP} = 6.77 + 1.82 * \log(Z\text{Flexion}) - 0.16 * \log(Z\text{Extension}) + 4.63 * \log(Z\text{Abduction}) + 6.70 * \log(Z\text{Adduction}).$$

## Discussion

First proposed over a decade ago, devices for robot-aided neurorehabilitation are increasingly being incorporated into stroke patients' care programs. In addition to delivering high-intensity, reproducible sensorimotor therapy, these devices are precise and reliable "measuring" tools that can be expanded with multiple sensors to record simultaneously kinematic and force data. These measurements are objective and repeatable and can be used to provide patients and therapists with immediate measures of motor performance. Reducing the time to evaluate improvement or deterioration may offer new opportunities for designing therapeutic programs and ultimately for increasing the efficiency of patients' care.

Across multiple regression models, we demonstrate that the FMA always performed the best in training, but the validation performance decreased. Correlations were higher than those reported in the study of Colombo et al,[31] which used simple regression models to calculate the FMA ($R = .53$–$.55$). The MSS performed the most consistently between training and validation, with training correlations similar to the FMA. The strong correlation between the MSS and the FMA suggests that the estimation of the MSS would allow for estimation of the FMA as well. In contrast to the positive results for the FMA and MSS, the MAS produced only weak nonsignificant correlations between the actual and calculated MAS scores and with the other scales. Tone and the variability of tone suggest that different clinical metrics should be developed to calculate this scale, a result supported by other clinical work.[32,33]

The extent to which micro-metrics based on submovements improved models' performance was also investigated. Submovement extraction can be a tedious, slow, computer-intensive task. In the past we employed 2 different strategies to extract submovements from patients' movement data: a global[34] or a local minimization scheme.[5] The efficiency of a global algorithm such as branch-and-bound is $O(N^M)$, where $M$ is the number of submovements. A greedy algorithm such as ISRBF (irregular sampling radial basis function)[5] is far more

efficient under this definition, $O$ ($MN$), but contrary to the branch-and-bound algorithm, ending close to a global minimum is not guaranteed. In this study, we found that the models that included submovement metrics performed only slightly better than those that did not (eg, compare performance of the macro-metric models with those of the macro-metric and micro-metric models). Because the impact of submovements was found to be low, the greedy-type algorithm may be employed for extracting submovements. Although limited by local minima, it would reduce the computation from hours to seconds and allow online computation. The low impact may also allow for the submovement metrics to be eliminated entirely from the model.

The MP force metrics model yielded relatively low correlation values, with $R = .564$ and $R = .525$ for training and validation, respectively. Using a similar model, Krebs et al[29] found a correlation of .853 between the MP score and the $\log^{10}$ of the Z-force metric. In contrast to our study that investigated chronic stroke subjects evaluated by 3 different evaluators, the study by Krebs et al investigated subjects in the subacute phase of recovery evaluated by the same therapist, which may account for the differences in results. Although our model achieved only a moderate correlation, it should be pointed out that the slope of the curves obtained in both models (the model used by Krebs et al and our model) are quite similar, giving additional credence to the proposed model to estimate the MP.

Of all the models investigated in this study, the kinematic macro-metric model for MSS appeared to offer the best trade-off between performance and complexity and strong, similar correlation values for training ($R = .711$) and validation ($R = .721$). The smoothness metric had the best individual performance when calculating both the FMA (.731 and .580 for training and validation, respectively) and MSS (.704 and .696 for training and validation, respectively) scores, and the combination of smoothness, peak speed, duration, and joint independence balance the compromise of model complexity and model performance (FMA .616 and .400, and MSS .557 and .589 for training and validation, respectively).

Robot measurements can potentially outperform human-administered clinical scales and are only limited by the performance of the robot sensors. For example, MIT-Manus can measure positions with a resolution of 0.1 mm. The reliability of human-administered clinical scales has often been questioned; for example, Sanford et al[35] reported an interrater variability of $\pm$ 18 points on a 95% confidence interval for the total FMA scale, pointing out that small patient improvements will not be able to be identified by the score. Krebs et al[29] found up to a 15% discrepancy between therapists when evaluating the same patient for the upper extremity FMA scale. Gregson et al[20] estimated an interrater agreement of 59% for the MAS. The MAS is considered a reliable clinical scale by some[20] but totally unreliable by others.[36] Besides having questionable reliability, human-administered clinical scales are also time consuming. In contrast, robot measurements can potentially provide therapists and patients with immediate feedback. Real-time scoring can not only greatly reduce the amount of time required for evaluations of patients' motor improvements, but it is also becoming a key need for the new robot-aided neurorehabilitation scenarios. These include systems that continuously adapt the amount and type of delivered therapy based on patient's motor abilities.[37,38]

Despite their potential advantages, robot-mediated therapy is presently not well established as human-administered modality of therapy, and robot-mediated measurements have even smaller dissemination as compared with clinical scales, which are still the "gold standard" for measuring outcomes. In addition, this study was limited to chronic stroke a population and hence it is unclear if the present correlations extrapolate to other phases of stroke recovery. Future studies must include the acute and subacute stroke population. Furthermore, some of these clinical scales have been used to evaluate other populations and hence these populations must be included in future correlation studies.[39] Finally, we were not successful in establishing any meaningful correlation between kinematic or kinetic metrics with the MAS. To address this limitation, we are developing robot-mediated measurement techniques to estimate stiffness, both passive and global stiffness, and will test the correlation between these metrics in our next cohort of patients.[40,41] Nevertheless, this study is one of the very few studies that attempted to unify different sets of metrics for outcome measurement, potentially establishing a basic taxonomy between very distinct practices. We demonstrated the feasibility of developing models to calculate the well-established clinical scales FMA, MSS, and MP from the robot-derived metrics. The ideal model would have as few, easy-to-calculate metrics as possible (to reduce the evaluation time) while maintaining good training and validation model performance.

## Acknowledgments

## References

1. American Stroke Association. Web site. http://www.strokeassociation.org.

2. Prange G, Jannink M, Groothuis-Oudshoorn C, Hermens HJ, Ijzerman MJ. Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke. J Rehabil Res Dev. 2006; 43:171–184. [PubMed: 16847784]

3. Kwakkel G, Kollen BJ, Krebs HI. Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review. Neurorehabil Neural Repair. 2008; 22:111–121. [PubMed: 17876068]

4. Mehrholz J, Werner C, Kugler J, Pohl M. Electromechanical-assisted training for walking after stroke. Cochrane Database Syst Rev. 2007; (4):CD006185. [PubMed: 17943893]

5. Krebs HI, Hogan N, Aisen ML, Volpe BT. Robot-aided neurorehabilitation. IEEE Trans Rehabil Eng. 1998; 6:75–87. [PubMed: 9535526]

6. Aisen ML, Krebs HI, Hogan N, McDowell F, Volpe BT. The effect of robot assisted therapy and rehabilitative training on motor recovery following a stroke. Arch Neurol. 1997; 54:443–446. [PubMed: 9109746]

7. Krebs HI, Volpe BT, Aisen ML, Hogan N. Increasing productivity and quality of care: robot-aided neuro-rehabilitation. J Rehabil Res Dev. 2000; 37:639–652. [PubMed: 11321000]

8. Volpe BT, Krebs HI, Hogan N, Edelstein L, Diels C, Aisen ML. Robot training enhanced motor outcome in patients with stroke maintained over 3 years. Neurology. 1999; 53:1874–1876. [PubMed: 10563646]

9. Volpe BT, Krebs HI, Hogan N, Edelstein L, Diels C, Aisen M. A novel approach to stroke rehabilitation: robot aided sensorimotor stimulation. Neurology. 2000; 54:1938–1944. [PubMed: 10822433]

10. Volpe BT, Krebs HI, Hogan N. Is robot-aided sensorimotor training in stroke rehabilitation a realistic option? Curr Opin Neurol. 2001; 14:745–752. [PubMed: 11723383]

11. Ferraro M, Palazzolo JJ, Krol J, Krebs HI, Hogan N, Volpe BT. Robot-aided sensorimotor arm training improves outcome in patients with chronic stroke. Neurology. 2003; 61:1604–1607. [PubMed: 14663051]

12. Fasoli SE, Krebs HI, Stein J, Frontera WR, Hogan N. Effects of robotic therapy on motor impairment and recovery in chronic stroke. Arch Phys Med Rehabil. 2003; 84:477–482. [PubMed: 12690583]

13. Fasoli SE, Krebs HI, Stein J, Frontera WR, Hughes R, Hogan N. Robotic therapy for chronic motor impairments after stroke: follow-up results. Arch Phys Med Rehabil. 2004; 85:1106–1111. [PubMed: 15241758]

14. Stein J, Krebs HI, Frontera WR, Fasoli SE, Hughes R, Hogan N. Comparison of two techniques of robot-aided upper limb exercise training after stroke. Am J Phys Med Rehabil. 2004; 83:720–728. [PubMed: 15314537]

15. Daly JJ, Hogan N, Perepezko EM, et al. Response to upper-limb robotics and functional neuromuscular stimulation following stroke. J Rehabil Res Dev. 2005; 42:723–736. [PubMed: 16680610]

16. Krebs HI, Aisen ML, Volpe BT, Hogan N. Quantization of continuous arm movements in humans with brain injury. Proc Natl Acad Sci U S A. 1999; 96:4645–4649. [PubMed: 10200316]

17. Rohrer B, Fasoli S, Krebs HI, et al. Movement smoothness changes during stroke recovery. J Neurosci. 2002; 22:8297–8304. [PubMed: 12223584]

18. Dipietro L, Krebs HI, Fasoli SE, et al. Changing motor synergies in chronic stroke. J Neurophysiol. 2007; 98:757–768. [PubMed: 17553941]

19. Dobkin BH. Progressive staging of pilot studies to improve phase iii trials for motor interventions. Neurorehabil Neural Repair. 2009; 23:197–206. [PubMed: 19240197]

20. Gregson JM, Leathley MJ, Moore AP, Smith TL, Sharma AK, Watkins CL. Reliability of measurements of muscle tone and muscle power in stroke patients. Age Ageing. 2000; 29:223–228. [PubMed: 10855904]

21. Medical Research Council/Guarantors of Brain. Aids to the Examination of the Peripheral Nervous System. London, UK: Bailliere Tindall; 1986.

22. Hsueh IP, Hsu MJ, Sheu CF, Lee S, Hsieh CL, Lin JH. Psychometric comparisons of 2 versions of the Fugl-Meyer motor scale and 2 versions of the Stroke Rehabilitation Assessment of Movement. Neurorehabil Neural Repair. 2008; 22:737–744. [PubMed: 18645189]

23. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. Phys Ther. 1983; 63:1606–1610. [PubMed: 6622535]

24. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery and stroke: a critical review of its measurement properties. Neurorehabil Neural Repair. 2002; 16:232–240. [PubMed: 12234086]

25. Twitchell T. The restoration of motor function following hemiplegia in man. Brain. 1951; 74:443–480. [PubMed: 14895765]

26. Brunnstrom, S. Movement Therapy in Hemiplegia. New York, NY: Harper & Row; 1970.

27. Ferraro M, Demaio JH, Krol J, et al. Assessing the motor status score: a scale for the evaluation of upper limb motor outcomes in patients after stroke. Neurorehabil Neural Repair. 2002; 16:283–289. [PubMed: 12234090]

28. Bohannon RW, Smith MB. Interrater reliability of a Modified Ashworth Scale of muscle spasticity. Phys Ther. 1987; 67:206–207. [PubMed: 3809245]

29. Krebs HI, Volpe BT, Ferraro M, et al. Robot-aided neurorehabilitation: from evidence-based to science-based rehabilitation. Top Stroke Rehabil. 2002; 8:54–70. [PubMed: 14523730]

30. Montgomery, DC.; Peck, EA. Introduction to Linear Regression Analysis. New York, NY: Wiley; 1992.

31. Colombo R, Pisano F, Micera S, et al. Robotic techniques for upper limb evaluation and rehabilitation of stroke patients. IEEE Trans Neural Syst Rehabil Eng. 2005; 13:311–324. [PubMed: 16200755]

32. Sommerfeld DK, Eek EU, Svensson AK, Holmqvist LW, von Arbin MH. Spasticity after stroke: its occurrence and association with motor impairments and activity limitations. Stroke. 2004; 35:134–139. [PubMed: 14684785]

33. Landau WM. Spasticity after stroke: why bother? Stroke. 2004; 35:1787–1788. [PubMed: 15232114]

34. Rohrer B, Hogan N. Avoiding spurious submovement decompositions: a globally optimal algorithm. Biol Cybern. 2003; 89:190–199. [PubMed: 14504938]

35. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. Phys Ther. 1993; 73:447–454. [PubMed: 8316578]

36. Pomeroy VM, Dean D, Sykes L, et al. The unreliability of clinical measures of muscle tone: implications for stroke therapy. Age Ageing. 2000; 29:229–233. [PubMed: 10855905]

37. Krebs H, Palazzolo J, Dipietro L, et al. Rehabilitation robotics: performance-based progressive robot-assisted therapy. Auton Robots. 2003; 15:7–20.

38. Dipietro L, Ferraro M, Palazzolo JJ, Krebs HI, Volpe BT, Hogan N. Customized interactive robotic treatment for stroke: EMG-triggered therapy. IEEE Trans Neural Syst Rehabil Eng. 2005; 13:325–334. [PubMed: 16200756]

39. Fasoli SE, Fragala-Pinkham M, Hughes R, Hogan N, Krebs HI, Stein J. Upper limb robotic therapy for children with hemiplegia. Am J Phys Med Rehabil. 2008; 87:929–936. [PubMed: 18936558]

40. Palazzolo JJ, Ferraro M, Krebs HI, Lynch D, Volpe BT, Hogan N. Stochastic estimation of arm mechanical impedance during robotic stroke rehabilitation. IEEE Trans Neural Sys Rehabil Eng. 2007; 15:94–103.

41. Roy A, Krebs H, Williams D, et al. Robot-aided neurorehabilitation: a novel robot for ankle rehabilitation. IEEE Trans Robotics. 2009; 25:569–582.

**Table 1**

Subject Characteristics

| | |
|---|---|
| Number of Subjects | 111 |
| Male/Female | 70/41 |
| Age (years; average ± standard deviation) | 59.9 ± 12.6 |
| Days since stroke (average ± standard deviation) | 1150.8 ± 944.3 |
| Left-sided/right-sided lesion | 52/59 |
| Left hand/right hand dominant | 16/95 |

**Table 2**

Correlation of Clinical Scores for Chronic Stroke Patients (*R* Values)

|  | FMA | MSS | MP | MAS |
|---|---|---|---|---|
| FMA | 1 | | | |
| MSS | .948 | 1 | | |
| MP | .785 | .774 | 1 | |
| MAS | −.23 | −.213 | −.013 | 1 |

Abbreviations: FMA, Fugl-Meyer Assessment; MSS, Motor Status Score; MP, Motor Power scale; MAS, Modified Ashworth Scale.

**Table 3**

R Values for Linear Fit Between Measured and Scores Calculated With All (Kinematic and Kinetic) Metrics

|  | Training | Validation |
|---|---|---|
| FMA | .802[a] | .427[a] |
| MSS | .788[a] | .696[a] |
| MP | .797[a] | .449[a] |
| MAS | .428[a] | .171 |

Abbreviations: FMA, Fugl-Meyer Assessment; MSS, Motor Status Score; MP, Motor Power scale; MAS, Modified Ashworth Scale.

[a] $P < .005$.

**Table 4**

R Values for Linear Fit Between Measured Scores and Scores Calculated With Kinematic Metrics

| | Training | | Validation | |
|---|---|---|---|---|
| | **Macro-Metrics** | **Macro- and Micro-Metrics** | **Macro-Metrics** | **Macro- and Micro-Metrics** |
| FMA | .740[a] | .779[b] | .425[a] | .410[a] |
| MSS | .711[a] | .772[a] | .721[a] | .688[a] |
| MAS | .357[a] | .399[a] | .158 | .137 |

Abbreviations: FMA, Fugl-Meyer Assessment; MSS, Motor Status Score; MAS, Modified Ashworth Scale.

[a] $P < .0001$.

[b] $P < .007$.

**Table 5**

95% Confidence Intervals for the Backwards Regression Model for Both FMA and MSS

| | FMA | | MSS | |
|---|---|---|---|---|
| | **Min** | **Max** | **Min** | **Max** |
| Constant | −42.26 | −2.53 | −56.58 | −4.63 |
| Peak speed | 11.67 | 76.42 | 37.02 | 113.99 |
| Smoothness | 80.20 | 145.68 | 98.83 | 183.52 |
| Duration | 1.21 | 3.09 | 1.83 | 4.51 |
| Joint independence | −30.22 | −9.79 | −40.11 | −15.56 |

Abbreviations: FMA, Fugl-Meyer Assessment; MSS, Motor Status Score.