Assessment of Neurological Impairment and Recovery Using Statistical Models of Neurologically Healthy Behavior

Stephen H. Scott, PhD[1,2], Catherine R. Lowrey, PhD[1] iD,
Ian E. Brown, PhD[3], and Sean P. Dukelow, MD, PhD[4,5],*

Abstract

While many areas of medicine have benefited from the development of objective assessment tools and biomarkers, there have been comparatively few improvements in techniques used to assess brain function and dysfunction. Brain functions such as perception, cognition, and motor control are commonly measured using criteria-based, ordinal scales which can be coarse, have floor/ceiling effects, and often lack the precision to detect change. There is growing recognition that kinematic and kinetic-based measures are needed to quantify impairments following neurological injury such as stroke, in particular for clinical research and clinical trials. This paper will first consider the challenges with using criteria-based ordinal scales to quantify impairment and recovery. We then describe how kinematic-based measures can overcome many of these challenges and highlight a statistical approach to quantify kinematic measures of behavior based on performance of neurologically healthy individuals. We illustrate this approach with a visually-guided reaching task to highlight measures of impairment for individuals following stroke. Finally, there has been considerable controversy about the calculation of motor recovery following stroke. Here, we highlight how our statistical-based approach can provide an effective estimate of impairment and recovery.

Keywords

stroke assessment, robots, kinematics, impairment, recovery

Introduction

In most areas of medicine, there has been rapid and continued improvement in clinical tools and biomarkers to quantify body function and dysfunction. Blood, urine, and saliva samples provide a wealth of information on the function of many organs, and imaging techniques provide detailed information on their structure. In contrast, there has been relatively limited improvement in assessment of brain function. Some indirect measures have been developed to look at patterns of activity in the brain (or muscle) such as electroencephalography (electromyography), positron emission tomography, and functional magnetic resonance imaging. However, techniques to assess perception, cognition, and motor impairments have changed very little over the years. For example, assessment of motor impairments continues to be largely based on visual and physical inspection by a clinician. These approaches have been honed over the years to focus on key problems for a given patient group, and commonly use criteria-based, ordinal scales to quantify impairments. Such techniques have minimal cost (beyond the clinician's time) and exploit the impressive capability of

our visual system to identify atypical performance such as slight asymmetry of gait, indicative of a stroke.

[1]Centre for Neuroscience Studies, Queen's University, Kingston, ON, Canada

[2]Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada

[3]Kinarm, BKIN Technologies Ltd. Kingston, ON, Canada

[4]Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada

[5]Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

*Co-Author Dukelow is a member of the American Society for Neurorehabilitation

Corresponding Author:
Stephen H. Scott, Queen's University, Botterell Hall, Room 219, Kingston, Ontario, K7L 3N6, Canada.
Email: steve.scott@queensu.ca

While criteria-based, ordinal scales may be useful in the clinic for making treatment decisions, their use is more problematic when used for clinical research and clinical trials (e.g., the modified Rankin Scale or Fugl-Meyer (FM) Assessment). The challenge is that our visual system may have evolved to identify atypical behavior, but we are only able to crudely quantify severity of atypical behavior from visual inspection,[1] limiting the levels of impairment that can be reliably defined. Thus, there is a clear need for approaches to assess objectively, reliably and with high precision, neurological impairments and changes in impairments due to stroke and other neurological injuries/diseases.

There is growing recognition that kinematic and kinetic-based measures can provide excellent quantification of impairments after stroke.[2-6] A primary challenge remains how to appropriately characterize these impairments compared to healthy control performance. Age, sex, and handedness can often impact kinematic performance. Many studies manage these effects by using age and sex-matched controls, permitting statistical comparison for differences between patient cohorts and healthy controls (group effects).[7-10] However, comparisons to a healthy mean do not allow one to identify if an individual with stroke is impaired (outside the range of performance expected for healthy individuals). Another approach is to compare individual patients to a distribution of healthy individuals. Cortes et al (2017)[11] compared reaching performance of individual participants with stroke to 12 healthy individuals of a similar age distribution making reaches with their dominant hand. They calculated the Mahalanobis distance to measure the distance of each individual patient's reaching performance from the distribution of healthy reaches in order to quantify impairment. This is a step in the right direction. However, using a very small number of age-matched individuals to estimate the performance of a healthy population is problematic as there can be substantial deviation between the estimated and actual distributions. This approach also does not account for the impact of age, sex, or handedness on performance.

In this article, we introduce a statistical approach for quantifying impairment based on measures of neurologically healthy individuals. We first review some of the challenges with using criteria-based ordinal scales for quantifying impairments and recovery in order to highlight how kinematic-based measures can overcome at least some of these challenges. We then describe our approach that develops a statistical model of healthy performance using a large number of healthy controls and then use this model to transform performance of individuals with stroke into standardized units. We illustrate our approach using a visually-guided reaching task. However, it is important to recognize that our approach can be used with any spatial or temporal features of motor performance, collected with wearable sensors, markered or markerless motion capture, or even as simple as the time to walk 10 m. Finally, there has been

considerable controversy and debate on how to quantify motor recovery following stroke.[12-18] Here we highlight how we can use our statistical-based approach to provide a simple and effective estimate of recovery for a cohort of individuals.

The Problem: Criteria-Based Scales

While there are many different criteria-based, ordinal scales to quantify sensory, motor, and cognitive impairments for different neurological injuries/diseases, we will focus attention on the properties of the arm motor component of the Fugl-Meyer Assessment[19] (FM), as this scale is commonly used to quantify upper limb motor impairments and recovery following stroke. Measures included in this scale are based on the pattern of recovery described by Twitchell[20] with 33 individual items related to arm function. For example, volitional movements within synergies. The examiner instructs the individual to move their hand from the contralateral knee to the ipsilateral ear. Performance on each task is quantified using a 3-point scoring system: "cannot perform" (0), "partially perform" (1), and "fully perform" (2).[19] While FM uses a rather coarse rating system, there is excellent test–retest and inter-rater reliability scores for FM for those trained in the implementation of this scale.[21-25]

Nonetheless, this three-point scoring system creates many clear problems when trying to quantify impairment. Most obvious is a large rounding error when using just three values or bins to define performance. Specifically, there is a considerable range in performance that can be defined as "partially perform" the task. Further, the bottom and top bin give us floor and ceiling effects, respectively. Floor effects reflect that two subjects "cannot perform" the task, but one individual is clearly more impaired than the other (i.e., one cannot move their arm at all, whereas another is able to initiate some movement but not enough to attain "partially perform"). Ceiling effects are the opposite, where two individuals pass the criteria for "fully perform," but one individual looks fluid and graceful, whereas the other is just a bit slower and uncoordinated. This coarseness also creates rounding error for measuring recovery as subjects may display substantial improvements in performance after therapy, but nevertheless, land in the same bin because they could not attain the next criteria. Again, this leads to excellent test–retest and inter-rater reliability at the expense of coarseness.

Another problem with criteria-based scales is converting these ordinal labels of motor performance into interval units. In other words, the transition from "cannot perform" to "partially perform" (0 to 1) is construed as numerically equivalent to "partially perform" to "fully perform" (1 to 2). The validity of this assumption hinges on the placement of the criteria in performance, or hurdles. If performance is viewed as a 100 m race, the scoring system numerically places the lower bounds of "partially" and "fully" at 33.3

and 66.7 m along the track. However, these criteria may better represent 20 and 75 m, respectively. This again is difficult to assess with criteria-based scales which are selected more because it is easier for raters to identify if a criteria was met than based on some equal space along some measure of performance.

FM sums together performance on 33 items, with some tasks being fairly easy and others harder. There has been some work using Rasch analysis to hierarchically categorize the items in order to determine the specific types of behaviors that patients can and can't perform at certain levels.[26] Overall, the inclusion of many items in the clinical tool may help smooth out the coarseness and asymmetries in the criteria inherent in each item, and also reduce floor and ceiling effects. However, previous work has shown that 20% or more of individuals with stroke will be impacted by floor or ceiling effects in commonly used clinical measures of motor impairment, including FM.[27]

Figure 1 A highlights the upper extremity FM scores extracted from several studies,[13,28-32] and initially published in Hawe and colleagues.[14] "Stroke Initial" denotes assessment completed within 2 weeks of stroke, whereas "Stroke Final" denotes assessment completed ~3–6 months following stroke. FM scores for Stroke Initial are fairly broadly distributed, however, there is a preponderance of individuals at the bottom of the range with fewer individuals receiving a score in the middle and higher range. The reverse problem is observed for Stroke Final where there is a preponderance of individuals near the very top of the scale. Approximately 70 (19%) of 373 participants received the maximum possible score of 66 meaning they could fully perform all 33 items of the scale and almost half received scores >60. However, it is likely that many of the individuals scoring 66 display atypical behaviors by visual inspection (see 37). In effect, criteria-based scales like FM have a substantive ceiling effect, ranging from the abilities necessary to 'fully perform' the FM tasks up to and through the range of motor skills of neurologically healthy individuals (Figure 1B). As also noted in Figure 1B, two individuals may have different levels of motor skill pre-stroke. The impact of stroke may initially lead to similar levels impairment for these two individuals. In the cases depicted in Figure 1B, a score of 66 may represent a performance closer to the pre-stroke motor skill for Subject 2, but Subject 1 was a highly skilled individual pre-stroke and 66 is a long way from their pre-stroke "normal."

The Solution: Kinematic-Based Assessment and Statistical Models of Healthy Performance

Human motor performance spans a spectrum, as emphasized and embraced at the Olympics. Even for a simple task like reaching, some individuals will move faster, straighter, and require minimal corrective movements compared to others. Thus, there is a distribution that characterizes healthy performance, with some better, and some worse.

For situations like stroke, it is not possible to know what was an individual's pre-stroke perfomance, and thus, it is not possible to measure an individual's true impairment from pre-stroke baseline. However, pre-stroke performance for a group of individuals should be equivalent to the performance of a group of healthy individuals (assuming no pre-existing neurological issues). Thus, expected mean and range of performance of individuals pre-stroke can be estimated by assessing the performance of a very large sample of healthy individuals.

Demographic factors can also impact performance. For example, reaction times increase with age[33] so what is atypical for a 40-year-old may be typical for an 80-year-old. Thus, factors such as age, sex, and handedness, need to be considered when estimating performance, just like growth charts estimate height and weight based on age and sex.

Thus, the key to our approach is to characterize the performance of a large cohort of neurologically healthy individuals. We develop mathematical models to capture healthy performance and then use these models to quantify performance of individuals with stroke providing quantitative measures of impairments relative to healthy performance. Reasonable estimates of the mean of a distribution may only require 30 to 50 individuals, but it requires many hundreds of individuals to estimate the shape of a distribution with reasonable precision (i.e., mean, variance and skew). Further, the model requires a good sample of all ages, both sexes, or any other factor that may impact performance. This requires hundreds of healthy individuals to be assessed. The benefit is that impairments can be defined for each individual. That is, if the patient is a 69-year-old, right-handed female, then we can assess their performance based on what is expected for a 69-year-old right-handed female. Once a normative model is established, there is no need to collect age- and sex-matched individuals for each study, as these factors are already considered in the model.

Here, we highlight our approach to quantify performance using a visually-guided reaching task, an upper limb task commonly assessed due to its importance for many daily activities.[34-38] For didactic purposes we use the Kinarm robotic platform, for which the basic paradigm has been described previously[39-43] and used to assess large cohorts of individuals, both neurologically health and those following stroke. Briefly, the Bimanual Kinarm Exoskeleton lab consists of 2 robotic exoskeletons that support the upper arms and forearm/hands to allow free movement of both arms in the horizontal plane. Participants are seated in a height-adjusted chair that provides truncal support, if required. The robot's shoulder and elbow joints are aligned with the subject's joints, and the robot and virtual reality systems are then calibrated for each participant. A virtual reality system
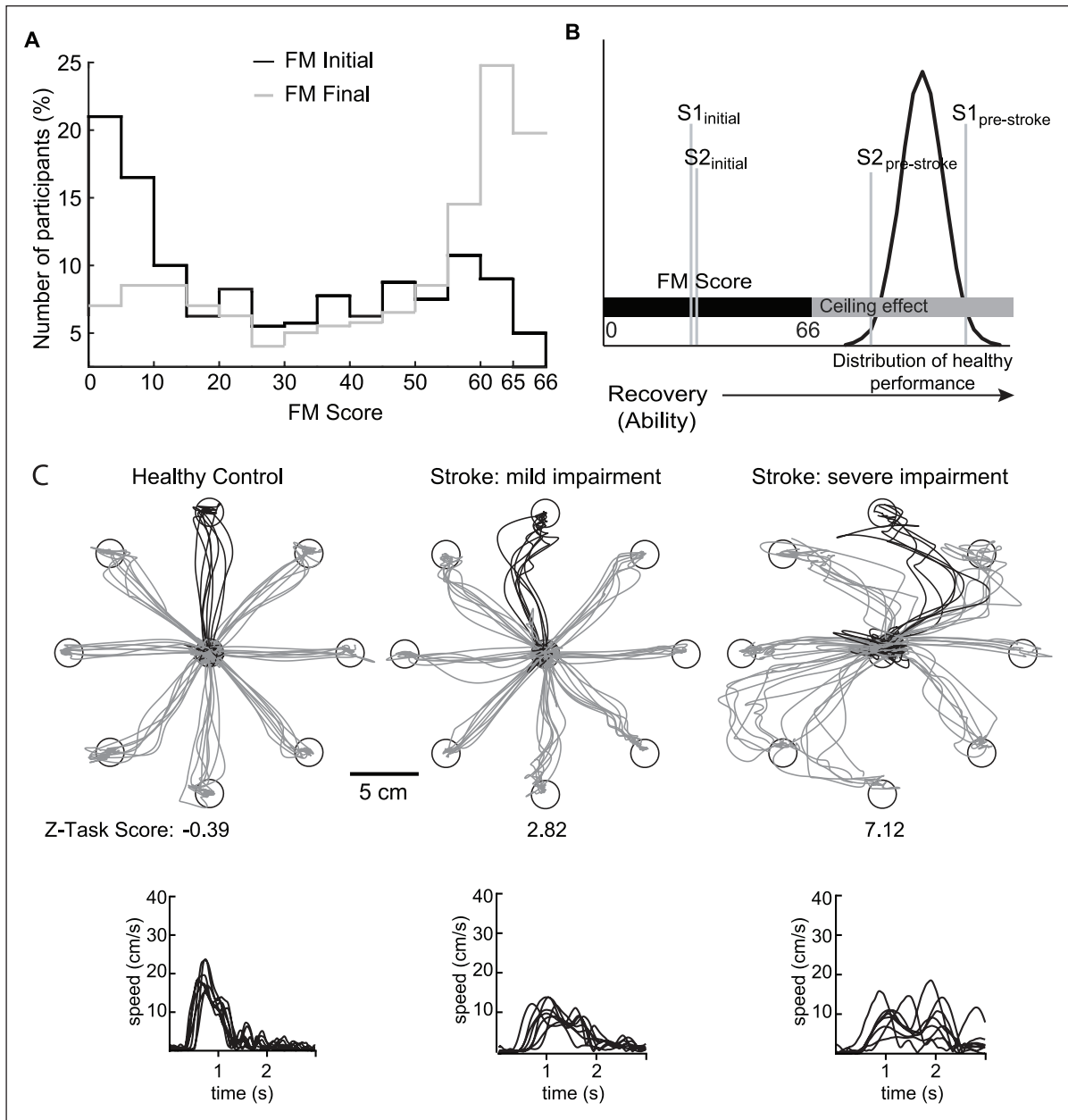
Figure 1. (A) Aggregate data for Fugl-Meyer scores for participants with stroke.[14] Black trace represents initial scores and grey trace represents final scores. (B) Schematic reflecting how the range of possible Fugl-Meyer scores falls below the range of abilities typically observed in healthy individuals. Gaussian distribution denotes variation of skills for healthy individuals. Vertical grey lines denote two individuals' (S1 and S2) abilities are different pre-stroke, but are similar initially after stroke. (C) Exemplar subject performance on the visually guided reaching (VGR) task. Top traces are hand paths for the most affected arm (or the non-dominant arm for the healthy control participant). The black traces are the trials that correspond to the hand speed shown below in the bottom traces. Participant on the left is a control participant (male, 70 years, right handed) who had no impairments on the task (Z-Task Score = −0.69). The middle participant (male, 71 years old, right affected) was assessed 2 days post-stroke and showed mild impairments on the task (Z-Task Score = 2.82). Participant on the right (male, 69 years, right affected) was assessed 28 days post-stroke and showed more severe impairments on the task (Z-Task Score = 7.12).

projects visual targets in the horizontal workspace along with a visual representation of the location of the index fingertips. Details of the Visually-Guided Reaching (VGR)

task are described elsewhere.[39] In general, the VGR task required participants to keep their hand at a visual spatial goal in the center of the workspace. When another spatial

Table 1. Participant Demographics

	Healthy Controls (N = 321)	Participants with stroke (N = 112)
Age: mean (range)	48 (18–84)	61 (21–84) years
Sex	156 M/165 F	79 M/33 F
Handedness	300 RH/18 LH/3 mix	105 RH/6 LH/1 mix
Affected arm	–	61 Left/51 Right
Type of Stroke	–	93 ischemic/19 hemorrhagic
Lesion Location* [# of subjects]	–	[C/SC/C+SC/Cb/Br/Cb+Br/Sp/Uk] [24/42/26/2/12/1/1/4]

*C = Cortical, SC = Subcortical, C + SC = Cortical + Subcortical, Br = Brainstem, Cb = Cerebellum, Cb + Br = Cerebellar + Brainstem, Sp = Spinal, Uk = Unknown.

goal was presented, participants were asked to quickly and accurately reach to and then maintain their hand at this second spatial goal.

We have assessed 321 neurologically healthy individuals and 112 individuals with stroke. The associated demographic information is provided in Table 1. Each participant with stroke was assessed at four times points: Stroke Initial for our data set was collected on average 1–2 weeks post-stroke (0.25–0.5 months, range: 1–28 days). Subsequent time points were collected at ~6 weeks (1.5 months, 4–9 weeks), ~12 weeks (3 months, 10–20 weeks), and ~26 weeks after stroke (Stroke Final, 6 months, 21–38 weeks after stroke). Participants were recruited from the stroke unit at Foothills Medical Centre and the inpatient stroke rehabilitation unit at Dr. Vernon Fanning Care Centre in Calgary, Alberta, Canada. Healthy participants were recruited from the community in the cities of Kingston, Ontario, Canada, and Calgary, Alberta, Canada. All participants provided informed consent prior to participation in the study. This study was approved by the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (#ANAT042-05), and the University of Calgary's Conjoint Health Research Ethics Board (#22123).

Figure 1C highlights performance of one healthy control and two individuals with stroke, one with mild and one with more severe impairments. The healthy individual displays relatively straight reaches, and consistent reaction times and hand speeds. The individuals with stroke show varying amounts of directional errors, more corrective responses and slower reaction times, and peak hand speeds.

Task Parameters

While one could just identify whether or not a reaching movement attained a goal, kinematic-based measures allow us to assess the quality of that movement. This quality of performance can be captured using various spatial and temporal features of movement, such as reaction time, movement time, initial direction angle, number of corrections, etc.[34-36,39] Individuals may have impairments that

specifically impact one aspect of performance. For example, increases in reaction time are commonly observed in those with spatial neglect.[44] Similarly, impairments may be more pronounced in the initial phase of the reach, or in the final phase of the reach depending on which hemisphere is damaged.[45] For example, uncoordinated motion (limb ataxia) is commonly observed in those with cerebellar lesions.[46] Thus, it is important to track several measures of performance, defined here as task parameters, to capture the unique patterns of impairments that may be present for an individual. In VGR we collect fourteen different parameters within the domains of postural control of the arm, reaction time, initial movement, corrective response, total movement metrics, and target success. For a full description of each parameter please see Supplementary Material.

Figure 2 illustrates the basic steps in our process to quantify impairment. Over the years we have developed and refined our approach to develop best practices for this type of quantification, but there are other methods (see discussion). Our crucial first step is to develop a model of neurologically healthy performance that transforms values in raw units of time or distance into a standard, normal distribution, while also removing the influence of demographic factors on performance. Box-Cox equations are a common technique to transform distributions from an asymmetrically skewed distribution to one that is normal.[47] Linear regression models can then be used to adjust performance for factors such as age, sex, or handedness. These regressions can be weighted appropriately to account for heteroskedasticity in the original distribution. The normal distribution is useful as statistical tests commonly assume data are normally distributed. The standard normal distribution provides a simple approach to quantify performance with a mean and median both equal to zero and the standard deviation of the distribution equal to 1. Thus, performance of 68% of individuals is between 1 and −1, 95% between 1.96 and −1.96, and 99% between 2.58 and −2.58. Measures of healthy performance transformed to a standard normal distribution create a common language where differences and measures are scaled in units of standard deviation (Z-units). One can do the same
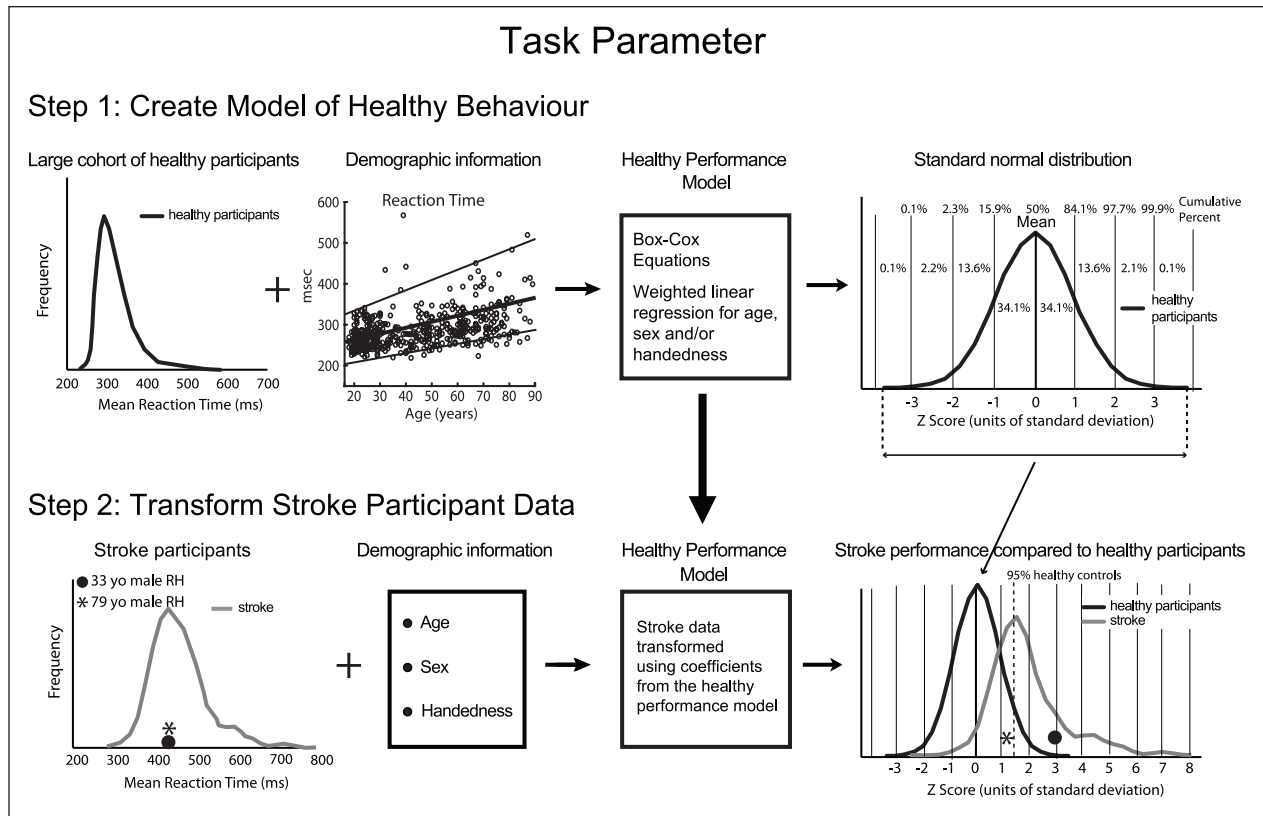
Task Parameter

Step 1: Create Model of Healthy Behaviour

Step 2: Transform Stroke Participant Data

Figure 2. Conversion of kinematic measures into standardized units. A) Calculations for task parameters. Step 1 creates a model that converts raw parameter scores to a standard normal distribution for a kinematic measure for a large healthy population. The distribution of values for a measure of performance (shown here is reaction time) is transformed to a standard normal distribution using Box-Cox equations and the effects of age, sex, and handedness are taken into account using weighted linear regression. Step 2 illustrates the calculations for the same task parameter for participants with stroke. For each participant, the outcome measure (i.e. reaction time) is transformed using the individual's demographic information and the coefficients from the healthy performance model (from Step 1) to generate a task parameter value in Z-units. Vertical dashed line in right panel indicates 95th percentile performance for healthy controls (1.64 Z-units).

for each spatial and temporal measure permitting direct comparisons across task parameters. For example, an individual with a reaction time of 300 ms may convert to −1.64 in Z-units (5th percentile), whereas movement time of 850 ms converts to 0 in Z-units (50th percentile).

Critically, the same model that is used to transform neurologically healthy performance into standard Z-units is then used to convert task parameters for each subject with stroke into Z-units (Figure 2, Step 2). With measures of performance now in Z-units, impairment can be defined as performance beyond some percentage of the healthy population, commonly the 95th percentile which is 1.64 Z-units. If poor performance could reflect either a large negative or positive value, then the line for impairment should be defined as ±1.96 Z-units (<2.5 and >97.5 percentiles). Correction for factors such as age means that two individuals with the same reaction time in milliseconds (star and filled circle in left panel of Figure 2, Step 2) receive different Z-scores when transformed by the model (right panel).

Figure 3A highlights distributions of reaction time with the affected arm for individuals 0.5 months (mean 11 days, range 1-28 days) after stroke (Stroke Initial) and 6 months (mean 26 weeks, range 21-38 weeks) after stroke (Stroke Final). The distribution of reaction times for Stroke Initial is shifted rightward and spread more broadly compared to the control distribution. Approximately 40% of individuals with stroke are identified as having impaired reaction times (i.e., $Z > 1.64$). The distribution of reaction times for Stroke Final is much less shifted compared to control and only 17% are identified as impaired.

Figure 3B displays the distributions for movement time during reaching. For Stroke Initial the distribution is shifted with an observable bimodal distribution in the scores (Hartigan's Dip statistic = 0.05; 95% likelihood of bimodality)[48] and 70% of individuals were identified as impaired. For Stroke Final, the distribution is shifted back towards the control distribution and is largely unimodal (Dip statistic = 0.023, 1% likelihood of bimodality), with a few
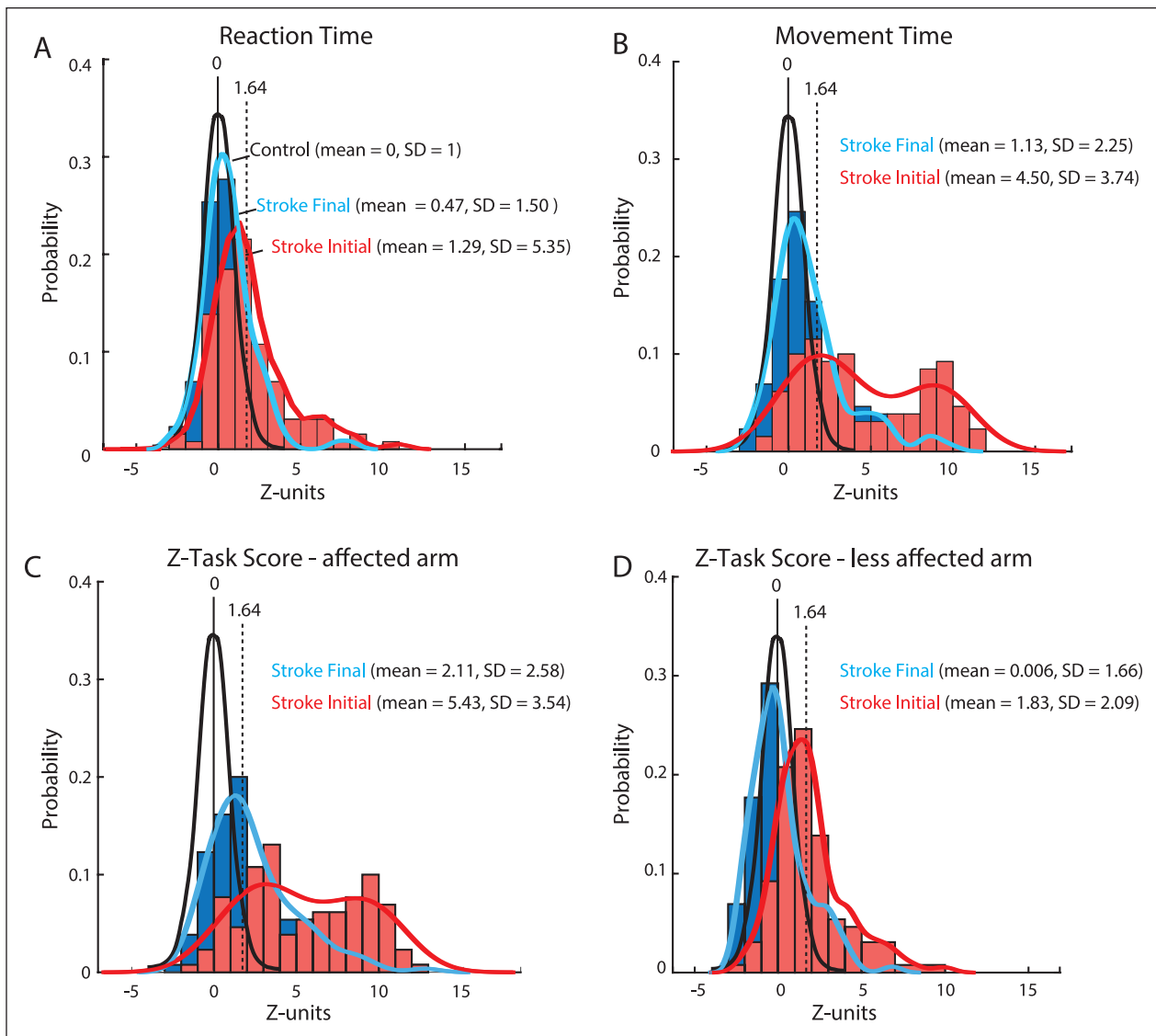
Figure 3. Histograms and probability distribution curves for (A) reaction time (B) movement time (C) Z-Task Scores (affected arm) and D) Z-Task Scores (less affected arm) for participants with stroke (N = 112) who completed the Visually Guided Reaching (VGR) task. Black traces are theoretical probability distribution curves for the healthy control subjects (mean/median = 0 and standard deviation = 1). Red curves and histograms represent data collected at the initial time point after stroke (Stroke Initial: ~0.5 months post-stroke). Blue curves and histograms represent the data collected at the final time point (Stroke Final: ~6 months post-stroke). Distribution curves calculated using a probability density estimate using a normal kernel function.

individuals that continue to have large values, and only 29% are impaired. The distributions for reaction and movement time clearly highlight that stroke leads to a greater impairment in the latter compared to the former at both time points, although this varies across individuals.

It is important to note that while the use of standardized units permits comparisons across parameters, some caution is necessary. In particular, comparison across parameters (or tasks) may be problematic when values are far from healthy performance. Near the healthy range, it is reasonable to assume that a reaction time of 2 in Z-units is better

than a movement time of 3, but a difference of 1 may not be meaningful when the two values are 9 and 10, respectively. Each parameter has limits on the range of potential values dictated by the distribution of performance for healthy individuals (mean, variance, and skew) as well as limits of possible values that can be attained. Reaction time and movement time are both limited by the maximal time for a given trial, whereas initial direction angle ranges from 0 to 180°. Floor effects can still remain in tasks such as reaching for individuals that cannot move the arm to the central target to initiate a trial. This is partially minimized by having

the robotic exoskeleton provide weight support so that some individuals with substantial weakness can move the limb to some degree.

Task Scores

The use of many different spatial and temporal parameters provides considerable richness to understand how stroke impacts performance, and potentially, specify targets for a rehabilitation plan. However, many measures of performance can also lead to information overload and difficulties in assessing overall impairment in a given task. Thus, there is also a need for a more general measure of performance. Our tentative strategy is to aggregate task parameters to generate a single Task Score (see Supplemental Figure 1). Distributions that characterize performance in a task parameter are either one-sided or two-sided. One-sided distributions either have poor performance indicated by larger values (e.g., reaction time) or smaller values (e.g., hand speed). In contrast, two-sided distributions have poor performance indicated by both larger and smaller than normal scores (e.g., arm performance asymmetries in bimanual tasks).[40] In order to aggregate different task parameters equally, one-sided task parameters are transformed such that best performance is 0 and poor performance is a large positive value. For two-sided parameters, (where poor performance could be either a large positive or a large negative) the absolute value is used. The end result is that all parameters are now aligned such that 0 is best performance and higher values reflect poorer performance. We chose to aggregate the values using root-sum-squares (RSS) and then to transform to standard normal space using Box-Cox equations. This model, developed from the performance of healthy controls, is then used to transform the aggregate measures for each individual with stroke, again into Z-units (termed Z-Task Score).

Figure 3C displays Z-Task Scores for the affected arm of individuals with stroke. The distribution for Stroke Initial is shifted compared to the standard control distribution with 85% of individuals identified as impaired. Task Scores tend to decrease for Stroke Final and shift towards the control distribution, but 49% of individuals are still identified as impaired. Figure 3D displays the Z-Task Scores for the less affected arm. The distribution for Stroke Initial is shifted compared to the standard control distribution and 46% of individuals were identified as impaired. In contrast, the distribution for Stroke Final is similar to the healthy control distribution and only 13% of individuals are identified as impaired.

Estimates of Recovery

Measures of recovery are important as they provide essential information to assess the efficacy of novel therapeutic interventions. Figure 4A displays hand traces of a participant with stroke assessed at four time points: 0.5, 1.5, 3, and 6 months following stroke. Their performance improves at each time point which is reflected by the decreasing Task Scores: 10.68 reflects a severe impairment at Stroke Initial, 4.36 and 2.36 reflect moderate impairments at 1.5 and 3 months post-stroke, respectively. Task Score of 2.31 at Stroke Final, 6 months post-stroke, shows that the participant has improved but still has some mild impairments. Figure 4B displays measures of impairment at the four time points for the stroke cohort. Z-Task Score means are plotted against their Standard Deviations (SD). The affected arm displays both a very large mean (5.43) and SD (3.54) reflecting the broad range in performance for Stroke Initial, as also displayed in Figure 3C. Mean and SD values reduce in magnitude with time, signifying a reduction in impairment and a reduction in the range of performance in the stroke cohort, respectively. The less affected arm also shows an improvement over time reaching a mean near zero and an SD of about 1.5 within 3 months. Individuals with stroke showed statistically significant improvements from initial measures to 1.5 months, and from 1.5 months to 3 months for reaching with both the affected and less affected arms. While there is a small reduction in the means from 3 to 6 months, it is not statistically significant (paired t-test, $P > .05$).

As the distribution of healthy performance in Z-units has a mean value of zero, recovery as a population can be estimated simply as the difference between mean initial impairment and mean impairment at a given time point, divided by the mean initial impairment (Stroke Initial). Figure 4C displays recovery (mean and confidence interval) and the percentage of individuals impaired at different time points following stroke. The 95% confidence intervals are calculated using repeat samples with replacement of individuals from the stroke cohort. These estimates suggest that individuals with stroke display ~60% recovery in performing VGR in the affected arm at 6 months (confidence interval 48 to 66%). We can also track the percentage of individuals impaired which decreases over time to approximately 44% at 6 months. The less affected arm displays almost 100% recovery at 6 months, with a confidence interval of 80 to 112%. The percentage of individuals who are impaired at 6 months with the less affected arm approaches 10%, which is just above the 5% of individuals who would be expected to show impairments in a healthy population, based on using a 95% cutoff for defining impairment.

It is important to recognize that the calculation of recovery is for the population and not individual participants. For example, in some cases individuals with stroke can attain Z-Task Scores below zero, denoting performance above average for healthy individuals. However, this does not mean that this individual had greater than 100% recovery.
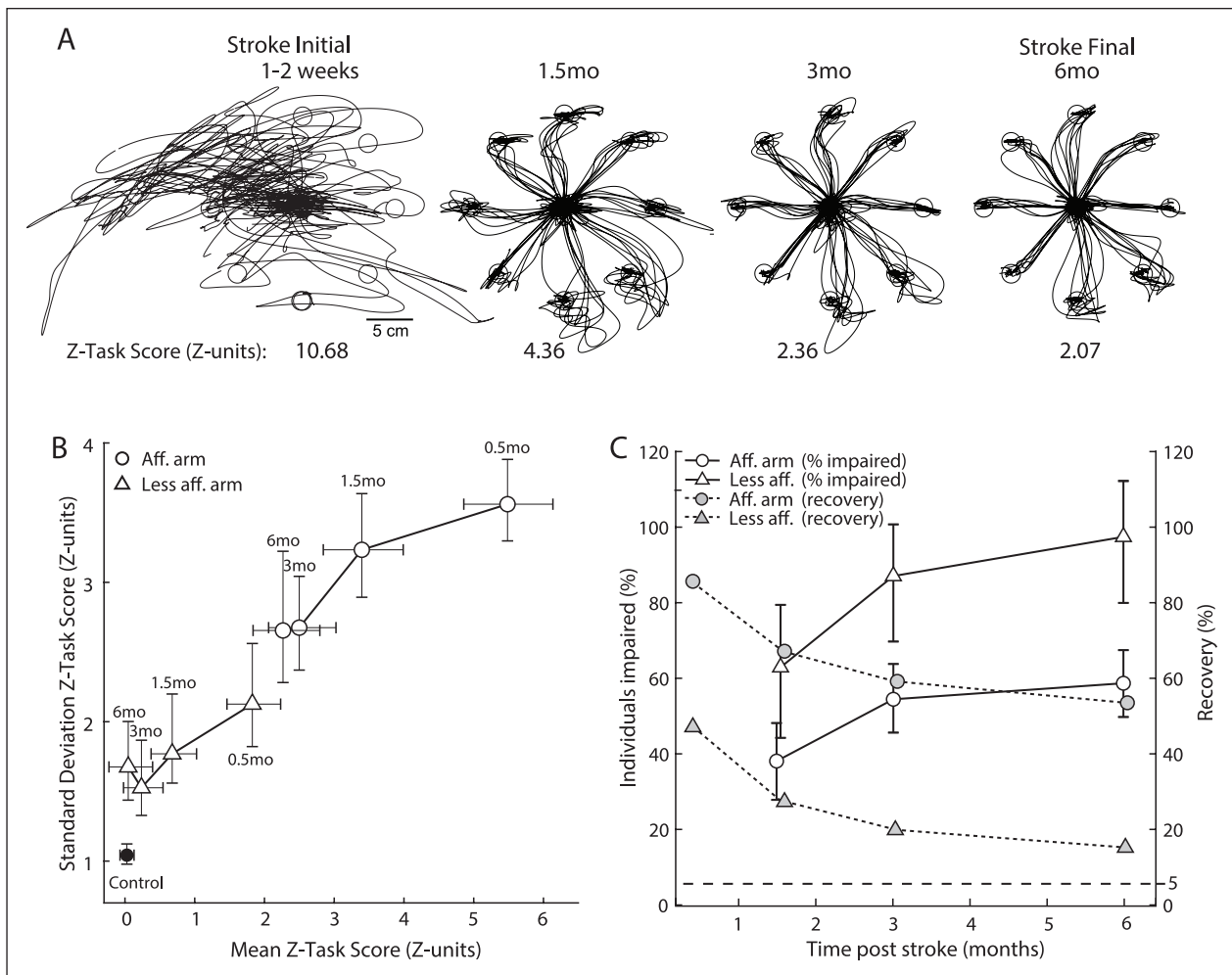
Figure 4. (A) Hand traces for the VGR task plotted for an exemplar participant assessed at 0.5, 1.5, 3, and 6 months post-stroke. (B) Mean Z-Task Score plotted vs the Standard Deviation of Z-Task Scores. Black circle represents healthy controls, white circles represent the affected arm of participants with stroke and white triangles represent the less affected arm of participants with stroke. Error bars denote 95% confidence interval for the mean and SD estimated by measuring the mean and SD from repeat samples with replacement of Z-Task Scores from the original distribution and repeating this process 1000 times. (C) Left axis: the percentage of individuals classified as impaired based on Z-Task Score (i.e., Task Score higher than 95% of healthy controls). Right axis: % Recovery plotted across time points. Recovery is calculated as the mean of ((Z-Task Score initial—Z-Task Score T(i))/Z-Task Score initial)*100. Circles represent data collected for reaches with the most affected arm and triangles represent data collected for the less affected arm. Errors bars denote 95% confidence interval for recovery based on repeat samples with replacement of individuals from the stroke cohort.).

As mentioned above, we do not know an individual's pre-stroke abilities so we cannot accurately assess impairment or recovery at the individual level. In situations where baseline tests are available, recovery can be estimated for each individual. In contrast, the distribution functions do allow us to estimate the probability that someone will attain some level of recovery. For example, reaching performance with the affected arm will fall within a typical healthy range for ~50% of individuals 5 months after stroke. Further, recovery distributions could be quantified for a sub-set of individuals

with a specific level of initial impairment (i.e., Initial Task Score between 7 and 10)."

While we and others have recommended against this approach,[12,14,15,49,50] we also estimated recovery based on commonly used techniques of regressing initial impairment versus change in impairment for comparison purposes (Figure 5A and B). For the affected arm, the regression slope is 0.55 which is similar to the 60% recovery calculated for the affected arm in Figure 4C. However, for the less affected arm, the slope is 0.63 which is far lower than
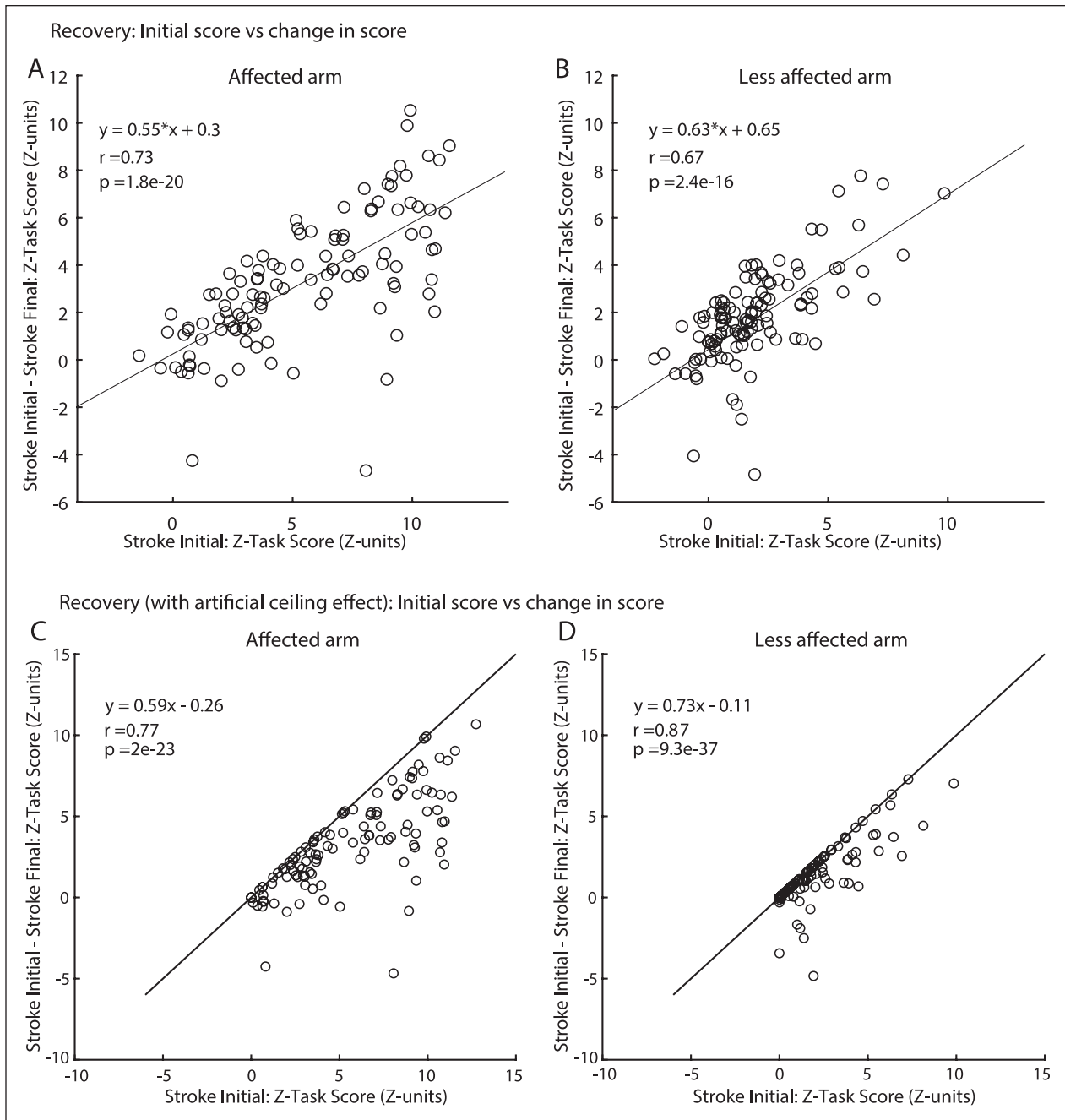
Figure 5. (A) Recovery plotted for VGR Z-Task Scores for the affected arm. Black lines are the linear regression lines represented by equation in the top left corner and *r*-values are the Pearson correlation coefficient. (B) Same as (A) but for the less affected arm. (C) The same data plotted for VGR Z-Task Scores for the affected arm but with an artificial ceiling added to the data of Z-Task Score = 0 (median of healthy performance). (D) Same as (C) but for the less affected arm.

the ~100% recovery in the less affected arm estimated in Figure 4C. These plots of recovery look different than observed in previous studies that plot initial impairment versus change in impairment based on FM because our Z-Task Scores do not have a ceiling effect. When we add an artificial ceiling to our Z-Task Score data and cap the possible values at 0 (Figure 5C and D) we can create a plot that looks more like the regression plots generated for the FM scale.[14,17,51,52] The calculated slope for the affected arm is now 0.59 and for the less affected arm is 0.73. For the
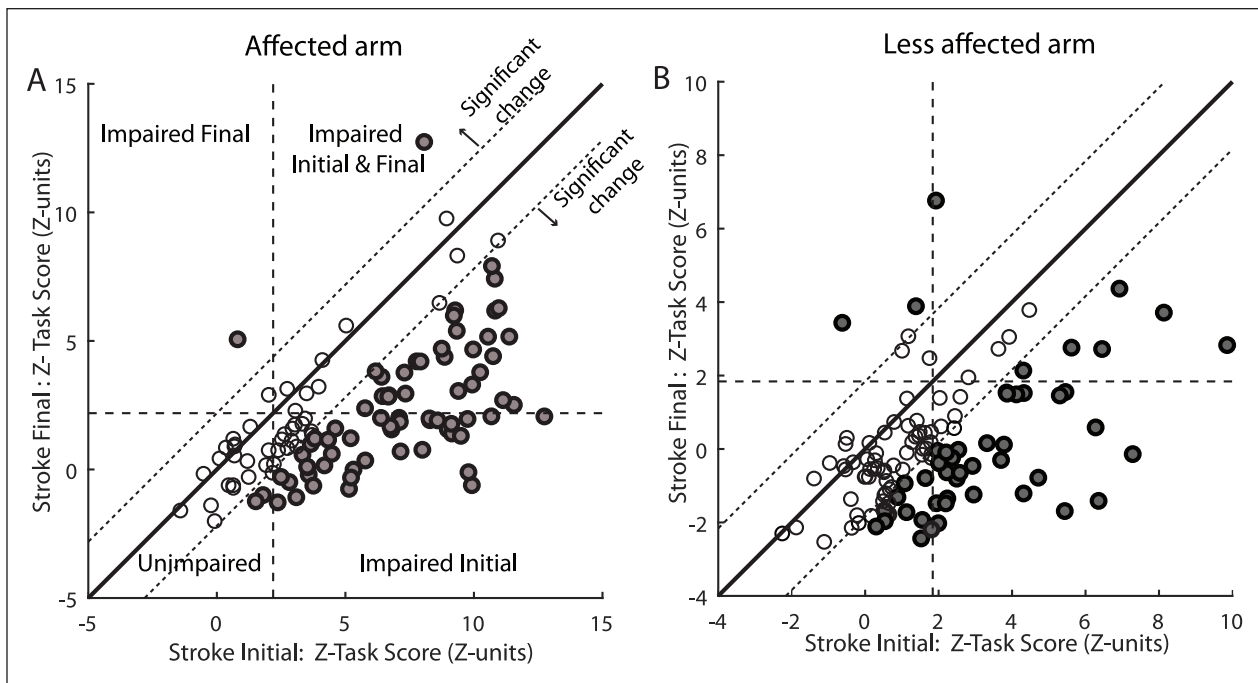
Figure 6. VGR Z-Task Score in Z-Units plotted for individual participants with stroke for the affected (A) and less affected (B) arms at Stroke Initial (0.5 months post-stroke) vs Stroke Final (6 months post-stroke). Vertical and horizontal dashed lines represent the cutoff line for impairment compared to 95% of healthy controls. Solid black diagonal lines represent the unity line and dashed diagonal lines represent significant change values calculated from repeat testing in a healthy normal population. Circles are filled grey if they surpassed thresholds for significant change from the initial exam to the final exam.

less affected arm regressions still predict far lower than the 100% recovery, as predicted in Figure 4C.

Patient-Specific Impairments

As stated earlier, we cannot quantify true impairment or true recovery for each individual following stroke as we do not know their pre-stroke abilities. However, the use of neurologically healthy individuals provides a basis to quantify patient-specific impairments.[53] This is a similar concept to diagnostic blood tests or many other clinical tests, where individual samples are compared to a normalized range based on healthy individuals. Figure 6 displays Z-Task Scores for Stroke Initial and Stroke Final for each individual (A: affected arm, B: less affected arm). The horizontal and vertical dashed lines denote the 95th percentile performance for healthy controls, and thus, values greater than this cutoff signify impairment. The solid diagonal unity line denotes identical performance for both initial and final assessments. The dashed diagonal lines denote the region of equivalence, where there is no significant change in performance between the two time points (open circles). Measures of significant change were identified by repeat testing of healthy individuals to assess performance consistency.[54] Below the lower diagonal dashed line denotes significant

improvement in the second assessment, whereas above the upper diagonal dashed line denotes significant degradation in performance (both indicated by filled circles). Not surprisingly, many individuals significantly improved reaching performance. In some cases, some residual impairment remained (above horizontal dashed line), whereas in other cases, performance returned to within the range for healthy individuals (below horizontal dashed line). A few individuals that were already within the healthy range for the initial assessment statistically improved performance. Z-Task Scores below zero denote performance better than the average for healthy individuals.

Discussion

Our objective here is to describe how kinematic-based measures can deal with many of the key drawbacks of criteria-based, ordinal scales. Here we propose using the natural distribution of performance observed in healthy individuals as a foundation to assess impairment and recovery. Standardized units create a common language for comparing performance. Age and other factors impact performance, obfuscating interpretation of performance in native units, but can be corrected when transforming into standardized units. We used a visual-guided reaching task to demonstrate

our approach, but the basic approach can be applied to any behavioural task or technology (i.e., wearable sensors, motion capture, force plates) where performance can be quantified with some level of resolution, including even simple measures like the 10m walk test.

It is recognized that quantifying the quality of tasks like reaching is important when assessing impairments,[36] although it is less clear which parameters or measures of performance should be considered when assessing this quality. In the VGR task we use 14 parameters related to the motion and position of the hand to quantify performance of postural stability before movement, the speed, and accuracy of the initial motor response as well as measures of secondary movements to attain the spatial goal (for simplicity, here we only presented reaction time and movement time). However, over 150 different kinematic metrics have been used previously to characterize upper limb movements across studies.[5] Some of this variety reflects differences in technologies (motion capture versus interactive robotics) or behavioural task (2D or 3D movements). In other cases, studies focus on more detailed measures related to the joints to explore issues related to joint coordination.[55,56] While the plethora of scientific questions examined dictates the need for many different measures to be assessed, some standardization would clearly be useful to facilitate ease of interpretation, and for clinical trials and meta-analyses.

While individual metrics allow us to observe the quality of movement and identify potential therapeutic targets, there is also a need for a singular measure of performance to characterize overall impairment and recovery. This is particularly true for clinical trials, where a global measure of performance is desirable as an outcome measure to determine efficacy of treatment. How many and which measures of performance should be included in an overall performance metric is an open question. Inter-rater reliabilty analysis may be useful in identifying which parameters best separate performance of different individuals, and thus, would be advantageous to include in an overall score. Other ways to reduce the data could include principal component analysis techniques based on extracted parameters[11,57] or raw kinematic data,[58] regression models that weight parameters based on some criteria,[59] or machine learning techniques.[2,60]

Similarly, once parameters are identified, there are many possible ways to aggregate different measures of performance. We chose root-sum-squares which equally weights each parameter, and thus values with larger deviations from healthy performance are weighted more. An alternate approach to the Euclidean distance is the Mahalanobis distance (MD) that removes inter-correlations between parameters.[61] However, measures of MD can have unexpected outcomes. For example, reaction and movement times tend to co-vary together both in healthy individuals and individuals following stroke. However, MD distance will increase more for an individual if movement time increases without a corresponding increase in reaction time. That is, an individual with only an impairment in movement time and not reaction time can be scored as more impaired as compared to someone that was impaired in both movement and reaction time. It is possible that one can receive an MD that is beyond the healthy range, and yet all task parameters are within the healthy range but the pattern of values is atypical. This is not to say that using the MD is not worthwhile. Where Task Scores using Euclidean distances are useful for describing overall impairment, Mahalanobis distances may be useful to identify when individuals are impaired in unexpected ways. Regardless of the approach, we suggest that transforming the data based on performance of a large cohort of healthy individuals should still be performed to aid in the interpretation of the level of impairment and the amount of recovery.

There has been considerable debate on how to assess recovery following stroke. A popular approach with the FM Scale has been to regress the amount of impairment observed from the first time point (66-FM Initial) versus the difference between the two assessments (FM Final-FM Initial). The resulting slope of the regression tends to be near 0.7 and explains a substantial amount of variance. This has been interpreted as evidence that most strokes recover 70% of function.[13,29] Subsequent studies have found that this 70% proportional recovery rule holds for many other clinical scales, including lower extremity FM,[52,62] aphasia,[63] hemispatial neglect,[28,64] and even the Functional Independence Measure.[14] The ubiquitous nature of these observations across multiple stroke impairments and scales has drawn criticism. Several recent studies have challenged this regression approach due to mathematical concerns, related to mathematical coupling and variance reduction between time points.[12,14,15,49,50] It is important to recognize that these mathematical concerns do not mean proportional recovery may not exist. Simply that the use of regression models between initial impairment and change in impairment is flawed and poor evidence, at best, to support a 70% proportional recovery rule.

Our approach for quantifying recovery highlights the clear problem with regressing the amount of initial impairment versus change in impairment. For the impaired arm, our measures are comparable although slightly lower to that predicted with regression, but there are glaring differences when quantifying recovery of reaching with the less affected arm. The regression approach found recovery was 63%, whereas our comparison to the healthy cohort suggests virtually complete recovery of function (Figure 5E). This is obvious when simply looking at the distribution of performance for healthy controls and Stroke Final in Figure 3D, as they are almost identical. This highlights the clear benefits of quantifying impairment as well as recovery based on

comparisons to kinematic-based measures of performance observed for healthy neurologically-intact individuals.

ORCID iD

Catherine R. Lowrey https://orcid.org/0000-0002-8379-9924

References

1. Williams G, Morris ME, Schache A, McCrory P. Observational gait analysis in traumatic brain injury: accuracy of clinical judgment. *Gait Posture*. 2009;29(3):454-459. doi:10.1016/j.gaitpost.2008.11.005

2. Krebs HI, Krams M, Agrafiotis DK, et al. Robotic measurement of arm movements after stroke establishes biomarkers of motor recovery. *Stroke*. 2014;45(1):200-204. doi:10.1161/STROKEAHA.113.002296

3. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. *Int J Stroke*. 2017;12(5):451–461. doi:10.1177/1747493017711813

4. Lambercy O, Maggioni S, Luenenburger L, Gassert R, Bolliger M. Robots and wearable sensor technologies for measurement/clinical assessment. In: Reinkensmeyer D, Dietz V, eds. *Neurorehabilitation Technology*. 2nd ed. Springer Verlag; 2016:183-207.

5. Schwarz A, Kanzler CM, Lambercy O, Luft AR, Veerbeek JM. Systematic review on kinematic assessments of upper limb movements after stroke. *Stroke*. 2019;50(3):718-727. doi:10.1161/STROKEAHA.118.023531

6. Scott SH, Dukelow SP. Potential of robots as next-generation technology for clinical assessment of neurological disorders and upper-limb therapy. *J Rehabil Res Dev*. 2011;48(4):335. doi:10.1682/JRRD.2010.04.0057

7. van Dokkum L, Hauret I, Mottet D, Froger J, Métrot J, Laffont I. The contribution of kinematics in the assessment of upper limb motor recovery early after stroke. *Neurorehabil Neural Repair*. 2014;28(1):4-12. doi:10.1177/1545968313498514

8. Hussain N, Alt Murphy M, Sunnerhagen KS. Upper Limb Kinematics in Stroke and Healthy Controls Using Target-to-Target Task in Virtual Reality. *Front Neurol*. 2018;9:300. doi:10.3389/fneur.2018.00300

9. Saes M, Mohamed Refai MI, van Kordelaar J, et al. Smoothness metric during reach-to-grasp after stroke: part 2. longitudinal association with motor impairment. *J NeuroEngineering Rehabil*. 2021;18(1):144. doi:10.1186/s12984-021-00937-w

10. Tomita Y, Mullick AA, Levin MF. Reduced kinematic redundancy and motor equivalence during whole-body reaching in individuals with chronic stroke. *Neurorehabil Neural Repair*. 2018;32(2):175-186. doi:10.1177/1545968318760725

11. Cortes JC, Goldsmith J, Harran MD, et al. A short and distinct time window for recovery of arm motor control early after stroke revealed with a global measure of trajectory kinematics. *Neurorehabil Neural Repair*. 2017;31(6):552-560. doi:10.1177/1545968317697034

12. Bonkhoff AK, Hope T, Bzdok D, et al. Bringing proportional recovery into proportion: Bayesian modelling of post-stroke motor impairment. *Brain*. 2020;143(7):2189-2206. doi:10.1093/brain/awaa146

13. Byblow WD, Stinear CM, Barber PA, Petoe MA, Ackerley SJ. Proportional recovery after stroke depends on corticomotor integrity: proportional recovery after stroke. *Ann Neurol*. 2015;78(6):848-859. doi:10.1002/ana.24472

14. Hawe RL, Scott SH, Dukelow SP. Taking proportional out of stroke recovery. *Stroke*. 2019;50(1):204-211. doi:10.1161/STROKEAHA.118.023006

15. Hope TMH, Friston K, Price CJ, Leff AP, Rotshtein P, Bowman H. Recovery after stroke: not so proportional after all? *Brain*. 2019;142(1):15-22. doi:10.1093/brain/awy302

16. Krakauer J, Marshall R. The proportional recovery rule for stroke revisited: the proportional recovery rule for stroke revisited. *Ann Neurol*. 2015;78(6):845-847. doi:10.1002/ana.24537

17. Kundert R, Goldsmith J, Veerbeek JM, Krakauer JW, Luft AR. What the proportional recovery rule is (and is not): methodological and statistical considerations. *Neurorehabil Neural Repair*. 2019;33(11):876-887. doi:10.1177/1545968319872996

18. Senesh MR, Reinkensmeyer DJ. Breaking proportional recovery after stroke. *Neurorehabil Neural Repair*. 2019;33(11):888-901. doi:10.1177/1545968319868718

19. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7(1):13-31.

20. Twitchell TE. The restoration of motor function following hemiplegia in man. *Brain*. 1951;74(4):443-480. doi:10.1093/brain/74.4.443

21. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Phys Ther*. 1983;63(10):1606-1610. doi:10.1093/ptj/63.10.1606

22. Beckerman N, Vogelaar TW, Lankhorst GJ, Verbeek AL. A criterion for stability of the motor function of the lower extremity in stroke patients using the Fugl-Meyer assessment scale. *Scand J Rehabil Med*. 1996;28(1):3-7.

23. Woodbury ML, Velozo CA, Richards LG, Duncan PW, Studenski S, Lai SM. Longitudinal stability of the Fugl-Meyer assessment of the upper extremity. *Arch Phys Med Rehabil*. 2008;89(8):1563-1569. doi:10.1016/j.apmr.2007.12.041

24. 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(7):447-454. doi:10.1093/ptj/73.7.447

25. Sullivan KJ, Tilson JK, Cen SY, et al. Fugl-Meyer Assessment of sensorimotor function after stroke: standardized training procedure for clinical practice and clinical trials. *Stroke*. 2011; 42(2):427-432. doi:10.1161/STROKEAHA.110.592766

26. Woodbury ML, Velozo CA, Richards LG, Duncan PW. Rasch analysis staging methodology to classify upper extremity movement impairment after stroke. *Arch Phys Med Rehabil*. 2013;94(8):1527-1533. doi:10.1016/j.apmr.2013.03.007

27. Lin JH, Hsu MJ, Sheu CF, et al. Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Phys Ther*. 2009;89(8):840-850. doi:10.2522/ptj .20080285

28. Winters C, van Wegen EEH, Daffertshofer A, Kwakkel G. Generalizability of the proportional recovery model for the upper extremity after an ischemic stroke. *Neurorehabil Neural Repair*. 2015;29(7):614-622. doi:10.1177/1545968314562115

29. Zarahn E, Alon L, Ryan SL, et al. Prediction of motor recovery using initial impairment and fMRI 48 h poststroke. *Cereb Cortex*. 2011;21(12):2712-2721. doi:10.1093/cercor/bhr047

30. Feng W, Wang J, Chhatbar PY, et al. Corticospinal tract lesion load: an imaging biomarker for stroke motor outcomes: CST lesion load predicts stroke motor outcomes. *Ann Neurol*. 2015;78(6):860-870. doi:10.1002/ana.24510

31. Guggisberg AG, Nicolo P, Cohen LG, Schnider A, Buch ER. Longitudinal structural and functional differences between proportional and poor motor recovery after stroke. *Neurorehabil Neural Repair*. 2017;31(12):1029-1041. doi:10 .1177/1545968317740634

32. Buch ER, Rizk S, Nicolo P, Cohen LG, Schnider A, Guggisberg AG. Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging. *Neurology*. 2016;86(20):1924-1925. doi:10.1212/WNL.000000000000 2675

33. Gottsdanker R. Age and simple reaction time. *J Gerontol*. 1982;37(3):342-348. doi:10.1093/geronj/37.3.342

34. Cirstea MC, Levin MF. Compensatory strategies for reaching in stroke. *Brain*. 2000;123(5):940-953. doi:10.1093/brain/123 .5.940

35. Kamper DG, McKenna-Cole AN, Kahn LE, Reinkensmeyer DJ. Alterations in reaching after stroke and their relation to movement direction and impairment severity. *Arch Phys Med Rehabil*. 2002;83(5):702-707. doi:10.1053/apmr.2002.32446

36. Kwakkel G, van Wegen EEH, Burridge JH, et al. standardized measurement of quality of upper limb movement after stroke: consensus-based core recommendations from the second stroke recovery and rehabilitation roundtable. *Neurorehabil Neural Repair*. 2019;33(11):951-958. doi:10.1177/1545968319886477

37. Otaka E, Otaka Y, Kasuga S, et al. Clinical usefulness and validity of robotic measures of reaching movement in hemiparetic stroke patients. *J NeuroEngineering Rehabil*. 2015;12(1):66. doi:10.1186/s12984-015-0059-8

38. Rohrer B, Fasoli S, Krebs HI, et al. Movement smoothness changes during stroke recovery. *J Neurosci*. 2002;22(18): 8297-8304. doi:10.1523/JNEUROSCI.22-18-08297.2002

39. Coderre AM, Amr Abou Zeid, Dukelow SP, et al. Assessment of upper-limb sensorimotor function of subacute stroke patients using visually guided reaching. *Neurorehabil Neural Repair*. 2010;24(6):528-541. doi:10.1177/1545968309356091

40. Tyryshkin K, Coderre AM, Glasgow JI, et al. A robotic object hitting task to quantify sensorimotor impairments in participants with stroke. *J NeuroEngineering Rehabil*. 2014; 11(1):47. doi:10.1186/1743-0003-11-47

41. Dukelow SP, Herter TM, Moore KD, et al. Quantitative assessment of limb position sense following stroke. *Neurorehabil Neural Repair*. 2010;24(2):178-187. doi:10.1177/15459683 09345267

42. Lowrey CR, Jackson CPT, Bagg SD, Dukelow SP, Scott SH. A novel robotic task for assessing impairments in bimanual coordination post-stroke. *Int J Phys Med Rehabil*. 2014;s3(01). doi:10.4172/2329-9096.S3-002

43. Bourke TC, Lowrey CR, Dukelow SP, Bagg SD, Norman KE, Scott SH. A robot-based behavioural task to quantify impairments in rapid motor decisions and actions after stroke. *J Neuroengineering Rehabil*. 2016;13(1):91. doi:10.1186/ s12984-016-0201-2

44. Heilman KM, Bowers D, Coslett HB, Whelan H, Watson RT. Directional hypokinesia: prolonged reaction times for leftward movements in patients with right hemisphere lesions and neglect. *Neurology*. 1985;35(6):855-855. doi:10.1212/ WNL.35.6.855

45. Mutha PK, Haaland KY, Sainburg RL. The effects of brain lateralization on motor control and adaptation. *J Mot Behav*. 2012;44(6):455-469. doi:10.1080/00222895.2012.747482

46. Konczak J, Pierscianek D, Hirsiger S, et al. Recovery of upper limb function after cerebellar stroke: lesion symptom mapping and arm kinematics. *Stroke*. 2010;41(10):2191-2200. doi:10.1161/STROKEAHA.110.583641

47. Box, Cox. An analysis of transformations. *J R Stat Soc B Methodol*. 1964;26(2):211-252.

48. Hartigan JA, Hartigan PM. The dip test of unimodality. *Ann Stat*. 1985;13(1). doi:10.1214/aos/1176346577

49. Bowman H, Bonkhoff A, Hope T, Grefkes C, Price C. Inflated estimates of proportional recovery from stroke: the dangers of mathematical coupling and compression to ceiling. *Stroke*. 2021;52(5):1915-1920. doi:10.1161/STROKEAHA .120.033031

50. Lee HH, Kim DY, Sohn MK, et al. Revisiting the proportional recovery model in view of the ceiling effect of Fugl-Meyer assessment. *Stroke*. 2021;52:3167-3175. doi:10.1161/ STROKEAHA.120.032409

51. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. Proportional motor recovery after stroke: implications for trial design. *Stroke*. 2017;48(3):795-798. doi:10.1161/STROKEAHA.116.016020

52. Veerbeek JM, Winters C, van Wegen EEH, Kwakkel G. Is the proportional recovery rule applicable to the lower limb after a first-ever ischemic stroke? *PLOS ONE*. 2018;13(1):e0189279. doi:10.1371/journal.pone.0189279

53. Simmatis LE, Jin AY, Taylor SW, Bisson EJ, Scott SH, Baharnoori M. The feasibility of assessing cognitive and motor function in multiple sclerosis patients using robotics. *Mult Scler J Exp Transl Clin*. 2020;6(4):205521732096494. doi:10.1177/2055217320964940

54. Simmatis LER, Early S, Moore KD, Appaqaq S, Scott SH. Statistical measures of motor, sensory and cognitive performance across repeated robot-based testing. *J Neuro Engineering Rehabil*. 2020;17(1):86. doi:10.1186/s12984-020-00713-2

55. Beer RF, Dewald JPA, Rymer WZ. Deficits in the coordination of multijoint arm movements in patients with hemiparesis: evidence for disturbed control of limb dynamics. *Exp Brain Res*. 2000;131(3):305-319. doi:10.1007/s002219900275

56. Laczko J, Scheidt RA, Simo LS, Piovesan D. Inter-joint coordination deficits revealed in the decomposition of endpoint jerk during goal-directed arm movement after stroke. *IEEE Trans Neural Syst Rehabil Eng Publ IEEE Eng Med Biol Soc*. 2017;25(7):798-810. doi:10.1109/TNSRE.2017.2652393

57. Wood MD, Simmatis LER, Gordon Boyd J, Scott SH, Jacobson JA. Using principal component analysis to reduce complex datasets produced by robotic technology in healthy participants. *J NeuroEngineering Rehabil*. 2018;15(1):71. doi:10.1186/s12984-018-0416-5

58. Backenroth D, Goldsmith J, Harran MD, Cortes JC, Krakauer JW, Kitago T. Modeling motor learning using heteroscedastic functional principal components analysis. *J Am Stat Assoc*. 2018;113(523):1003-1015. doi:10.1080/01621459.2017.1379403

59. Bosecker C, Dipietro L, Volpe B, Igo Krebs H. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. *Neurorehabil Neural Repair*. 2010;24(1):62-69. doi:10.1177/1545968309343214

60. Agrafiotis DK, Yang E, Littman GS, et al. Accurate prediction of clinical stroke scales and improved biomarkers of motor impairment from robotic measurements. *PLOS ONE*. 2021;16(1):e0245874. doi:10.1371/journal.pone.0245874

61. Kenzie JM, Semrau JA, Hill MD, Scott SH, Dukelow SP. A composite robotic-based measure of upper limb proprioception. *J NeuroEngineering Rehabil*. 2017;14(1):114. doi:10.1186/s12984-017-0329-8

62. Smith MC, Byblow WD, Barber PA, Stinear CM. Proportional recovery from lower limb motor impairment after stroke. *Stroke*. 2017;48(5):1400-1403. doi:10.1161/STROKEAHA.116.016478

63. Lazar RM, Minzer B, Antoniello D, Festa JR, Krakauer JW, Marshall RS. Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke*. 2010;41(7):1485-1488. doi:10.1161/STROKEAHA.109.577338

64. Marchi NA, Ptak R, Di Pietro M, Schnider A, Guggisberg AG. Principles of proportional recovery after stroke generalize to neglect and aphasia. *Eur J Neurol*. 2017;24(8):1084-1087. doi:10.1111/ene.13296