5-5-2017

Validity of Robot-based Assessments of Upper Extremity Function

Alison McKenzie
Chapman University, amckenzi@chapman.edu

Lucy Dodakian
University of California - Irvine

Jill See
University of California - Irvine

Vu Le
University of California - Irvine

Erin Burke Quinlan
University of California - Irvine

See next page for additional authors

Recommended Citation

McKenzie A, Dodakian L, See J, Le V, Quinlan EB, Bridgford C, Head D, Han VL, Cramer SC, Validity of Robot-based Assessments of Upper Extremity Function, ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION (2017), doi: 10.1016/j.apmr.2017.02.033.

Validity of Robot-based Assessments of Upper Extremity Function

Comments
NOTICE: this is the author's version of a work that was accepted for publication in *Archives of Physical Medicine and Rehabilitation*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version will be subsequently published in *Archives of Physical Medicine and Rehabilitation* in 2017. DOI:

Authors
Alison McKenzie, Lucy Dodakian, Jill See, Vu Le, Erin Burke Quinlan, Claire Bridgford, Daniel Head, Vy L. Han, and Steven C. Cramer

Validity of Robot-based Assessments of Upper Extremity Function

Alison McKenzie, PT, DPT, PhD, Lucy Dodakian, MA, OTR/L, Jill See, PT, MPT, Vu Le, MS, Erin Burke Quinlan, PhD, Claire Bridgford, DPT, Daniel Head, DPT, Vy L. Han, MD, Steven C. Cramer, MD.

PII: S0003-9993(17)30299-X

DOI: 10.1016/j.apmr.2017.02.033

Reference: YAPMR 56881

To appear in: *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION*

Received Date: 4 October 2016

Revised Date: 10 February 2017

Accepted Date: 27 February 2017

Please cite this article as: McKenzie A, Dodakian L, See J, Le V, Quinlan EB, Bridgford C, Head D, Han VL, Cramer SC, Validity of Robot-based Assessments of Upper Extremity Function, *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION* (2017), doi: 10.1016/j.apmr.2017.02.033.

Validity of Robot-based Assessments of Upper Extremity Function

Alison McKenzie, PT, DPT, PhD, Lucy Dodakian, MA, OTR/L, Jill See, PT, MPT, Vu Le, MS, Erin Burke Quinlan, PhD, Claire Bridgford, DPT, Daniel Head, DPT, Vy L. Han, MD, Steven C Cramer, MD.

PLEASE DIRECT CORRESPONDENCE TO
Alison McKenzie, PT, DPT, PhD
Chapman University
Rinker Campus
9401 Jeronimo Road
Irvine, CA 92618
PHONE: 714-744-7827
FAX: 714-744-7621
EMAIL: amckenzi@chapman.edu

Acknowledgments: The authors thank Mr. Jeby Abraham, and Ms. Lisa Meng for their assistance with data entry and the nursing staff of the UCI Institute for Clinical and Translational Science for their assistance.

This work was supported by the National Institutes of Health (R01 NS059909, K24 HD074722, and NIH/NCRR UL1 TR000153).

Disclosure: Dr. Cramer reports financial relationships with GlaxoSmithKline, MicroTransponder, Dart Neuroscience, RAND Corporation, Roche, and personalRN outside the submitted work. The other authors have nothing to disclose.

clinicaltrials.gov # NCT01244243

Abstract

Objective. To examine the validity of 5 robot-based assessments of arm motor function post-stroke.

Design. Cross sectional.

Setting. Outpatient clinical research center.

Participants. Volunteer sample of 40 participants, age >18 years, 3-6 months post-stroke, with arm motor deficits that had plateaued.

Intervention. None.

Main Outcome Measures. Clinical standards included the Fugl-Meyer Arm Motor Scale (FMA), and 5 secondary motor outcomes: hand/wrist subsection of the FMA; Action Research Arm Test (ART); Box & Blocks test (B/B); hand subscale of Stroke Impact Scale-2 (SIS); and the Barthel Index (BI). Robot-based assessments included: wrist targeting; finger targeting; finger movement speed; reaction time; and a robotic version of the (B/B) test. Anatomical measures included percentage injury to the corticospinal tract (CST) and primary motor cortex (M1, hand region) obtained from MRI .

Results. Subjects had moderate-severe impairment (arm FMA scores = 35.6±14.4, range 13.5-60). Performance on the robot-based tests, including speed (r=0.82, p<0.0001), wrist targeting (r=0.72, p<0.0001), and finger targeting (r=0.67, p<0.0001) correlated significantly with the FMA scores. Wrist targeting (r=0.57 - 0.82) and finger targeting (r=0.49 - 0.68) correlated significantly with all 5 secondary motor outcomes and with percent CST injury. The robotic version of the B/B correlated significantly with the clinical B/B test but was less prone to floor effect. Robot-based assessments were comparable to FMA score in relation to percent CST injury and superior in relation to M1 hand injury.

Conclusions. The current findings support using a battery of robot-based methods for assessing the upper extremity motor function in subjects with chronic stroke.

Key Words: Stroke, Robot Therapy, Arm Outcome Measures

Stroke is a leading cause of disability, frequently resulting in the loss of wrist and hand function required for activities of daily living[1-3]. Emerging evidence supports the use of restorative therapies for improving patient outcomes, yet in typical clinical settings, therapists are often unable to deliver the type or amount of intensive intervention needed for optimal recovery[4 5 6 7] due to constraints in the healthcare delivery system [8-10]. To address this problem, researchers and clinicians are incorporating technology-based therapies (e.g., robotic therapy, computer-based games [11, 12] and home-based telerehabilitation systems[13, 14]) into stroke rehabilitation, but the results have been mixed[7, 15-19 20]. Interpreting and comparing the results of studies on stroke rehabilitation can be difficult due to the use of different outcome measures across investigations[21, 22 23 24]. The dearth of valid, technology-based outcome measures poses additional challenges to evaluating the effectiveness of these new approaches. Therefore, continuing progress in technology-based stroke rehabilitation depends upon the availability of valid instrumented assessments that are comparable to existing clinical outcome measures.

For technology-based therapies to gain widespread acceptance, they must render outcome data that are consistent with valid outcome measures such as the Fugl-Meyer arm motor test (FMA), which is considered a gold standard assessment[25-27]. Outcomes also should be validated against other anatomical measures of stroke severity, such as corticospinal tract (CST) integrity via neuroimaging. Administering standardized clinical behavioral outcome measures to assess arm and hand recovery adds to the cost and inconvenience of technology-based therapies. Therefore is it advantageous to incorporate the use of technology into home-based models of care to assess patients remotely. Consequently, developing reliable, valid outcome measures that are comparable to valid clinical behavioral outcome measures is a key step toward integrating technology into clinical practice, particularly when access to care is limited. To that end, researchers are working toward identifying instrumented assessments that can serve in lieu of standardized behavioral outcome measures administered by trained professionals[28 29]. Krebs et al. (2014) [24] demonstrated that kinetic measures of upper extremity movements performed during robotic therapy correlated well with clinical measures, however, such measures may involve a level of complexity not feasible for wide-spread use in patients' homes. Using scores of performance on technology-based therapies as indicators of function could be a viable alternative to standardized assessments, providing that those scores accurately reflect arm motor function. Ultimately, having a more comprehensive understanding of the relationships among clinical behavioral indicators, technology-based-assessments, and anatomical measures (e.g., corticospinal tract integrity)[30] of stroke-related motor deficits may lead to the development of new and better patient-centered therapies that target specific motor deficits.

As the use of technology-based therapies increases, another factor to consider is incorporating simple, accurate tests of arm motor function post-stroke that address the spectrum of the World Health Organization's (WHO) International Classification of Functioning Disability and Health (ICF). To capture the full extent of the effects of stroke-related disability, the ICF model includes limitations of body structure/function, activities, and participation in society, in addition to personal and environmental factors [31]. Using the ICF model may enhance clinicians' abilities to relate the effects of impaired movement due to dysfunction of a limb (e.g., arm and hand weakness) to the specific activities that are affected by those impairments (e.g., dressing and eating) and how limitations in those activities influence one's ability to carry out one's usual roles in life (e.g., working) [32]. Having accurate measures of movement function across ICF domains may enhance clinicians' abilities to determine the full impact of individuals' stroke-related motor deficits and develop more effective treatment strategies. Using robot-based scores across ICF domains may provide a safe, simple alternative to time-intensive behavioral examinations by therapists.

As an initial step, the current study examined the validity of 5 robot-based assessments of arm motor status by exploring the relationships between these instrumented assessment scores and established clinical and anatomical measures pertaining to stroke-induced upper extremity deficits across the ICF. Specifically, we hypothesized that the robot-based assessment scores would demonstrate construct validity across the ICF domains when compared to standard clinical behavioral outcome measures and would also correlate with CST integrity, thereby demonstrating validity with respect to anatomy following stroke. Further, we aimed to demonstrate that robot-based assessments could be administered more rapidly than clinical behavioral assessments, thereby saving clinicians' time. Ultimately, if technology-based assessments can be administered in patients' homes, clinicians may be able to track patient performance remotely.

Methods

Study Design. The current study was a cross-sectional objective analysis of baseline data collected as part of a larger clinical trial (clinicaltrials.gov # NCT01244243).

Subjects. Subjects were recruited from the surrounding area through flyers sent to rehabilitation facilities, healthcare providers, and individuals who had contacted the laboratory directly to participate in a study of robotic therapy for arm weakness after stroke. All subjects provided informed consent, in accordance with the University of California Irvine Institutional Review Board, and were contacted by telephone and screened by the study coordinator (LD) to determine eligibility. Entry criteria included age >18 years, stroke with onset 11-26 weeks prior to initial study assessments, arm motor deficits that had reached a stable plateau, and absence of any condition that would confound study participation. All data in the current report were obtained at baseline, prior to any therapy.

Procedures. Subjects (or their proxy, for those who were unable to complete the forms due to motor deficits) completed questionnaires about demographic information (age, sex, ethnicity, level of education), medical and rehabilitation history, and prior level of function. Subjects were examined by licensed therapists with established inter-rater reliability (JS, LD, and AM) via clinical measures as well as robot-based assessments[19]. The primary clinical measure for current analyses was the total FMA scale[25, 33, 41], a measure of upper extremity impairment. Five secondary clinical measures also were examined: (1) the hand/wrist subsection of the FMA; (2) Action Research Arm Test (ARAT)[34, 35]; (3) Box & Blocks test (B/B)[36], a second measure of upper extremity function with different psychometric qualities that lends itself to implementation in a robotic setting; (4) hand motor subscale of Stroke Impact Scale-2 (SIS)[37], a patient-reported measure of hand usage; and (5) the Barthel Index (BI)[38]. The primary behavioral measure (FMA) and four of the five secondary behavioral measures (hand/wrist subsection of FMA, ARAT, B/B, SIS-hand) are modality-specific for arm motor status; the BI is a global measure of function [39]. In terms of the ICF categories, restrictions in: 1) ***body/structure function*** were assessed by FMA and the hand/wrist subsection of the FMA; ***activity*** were assessed by B/B, ARAT, and BI; and ***participation*** in society were assessed by SIS-hand (Supplement A).

Data from five robotic assessments also were collected (Figure 1 and Supplement B). The Hand Wrist Assistive Rehabilitation Device (HWARD) robot focuses on distal upper extremity motor function and is described in greater detail in Takahashi et al. [19]. For the current study, a second (mirror-image) robot was built to allow inclusion of subjects with left-sided upper extremity involvement. Briefly, the forearm was supported and stabilized in a cradle to prevent extraneous movements; subjects moved their wrists and fingers while the robot sensors measured movement across the 3 degrees of freedom. Scores on the robot assessments were obtained without robot actuation (i.e., the pneumatically actuated assistance provided by the robot during therapy was disabled during testing). Participants were required to move on their own as the robot sensors recorded the five robot-based metrics (below) while participants moved in response to the cues provided on a computer monitor. After a brief practice period during which subjects demonstrated their understanding of each of the games, subjects were asked to complete the tasks described in Figure 1 and Supplement B. The robot-based assessments focus on wrist and finger movement (flexion and extension), accuracy, and speed.

The software dictated the time required for administering the robot-based tests. Robot-based wrist movement test data were collected from 38 of the 40 subjects, as that test was introduced beginning with the third subject; otherwise, clinical and robotic data were collected from all subjects.

The primary focus was on three of these tests: (1) precision of *wrist targeting* movements (speed and accuracy of flexing or extending the wrist while moving toward a circular target); (2) precision of 4-*finger targeting* movements (ability to flex or extend fingers quickly and accurately while reaching and maintaining position over a target); and (3) maximum *speed* of finger movements in response to a 'go' signal. In addition, (4) a robot-based version of the *B/B* test was also scored, during which subjects manipulated virtual blocks on the computer screen using the same instructions as with the clinically tested B/B test; and (5) a simple test of *reaction time*. To ensure that the motor behavioral outcome measures were stable (indicating that subjects had plateaued), two assessments of the FMA, ARAT, and B/B were performed between 1 and 3 weeks of one another at baseline, and the scores were averaged; subjects whose total FMA scores varied by more than 2 points were excluded. All clinical assessments were performed by the same licensed physical therapist (JS); intra-rater and inter-rater reliability for the ARAT and the FMA were established previously for the laboratory[35, 40] and the average duration of the testing procedures was determined.

In addition to the behavioral and robotic assessments, anatomical data were collected from an MRI scan (3T, Philips Achieva system) obtained at baseline, prior to any treatment, and included high resolution T1-weighted images (repetition time = 8.5 ms, echo time = 3.9 ms, slices =150, voxel size = 1 x 1 x 1 mm^3). Infarct volume was outlined, binarized, then transformed into Montreal Neurologic Institute (MNI) stereotaxic space. The extent of injury to the hand region of the primary motor cortex (M1) injury was determined by measuring the degree of overlap that each infarct mask had with an MNI-space map of the hand region of M1[41]. The percent injury to the corticospinal tract (CST) was determined as described previously[30 41].

Data Analysis.

Descriptive statistics (means, standard deviations, and ranges) and non-parametric (Spearman's rho) correlations were calculated between the clinical behavioral outcome measures (FMA, hand/wrist FMA, ARAT, B/B, BI) and the robot-based scores on finger targeting, wrist targeting, reaction time, speed, and robot-based B/B using JMP, version 8; Bonferroni correction was made for multiple comparisons between the measures of interest ($p<0.007$). All r values are reported as absolute value because better motor status is the higher score for some scales and lower for others; moderate correlations were considered to be those in the range of 0.5 to 0.7, with strong correlations being >0.7 [42].

Results

Study subjects: A total of 40 subjects (29 male/11 female; average age=58 years (±14)) were studied. Demographic information and clinical and robotic assessments are presented in Table 1. All subjects successfully generated scores on the instrumented assessments, which were rapidly and successfully obtained in all subjects (11-20.5 minutes per session for robotic assessments vs. 29-49 minutes for behavioral assessments). Restrictions in movement ranged from mild to severe motor impairment (Table 1). The five robotic assessment scores also reflected mild to severe deficits (Table 1). Anatomical measures of injury were concordant, showing that M1 and CST injury ranged from mild to severe (Table 1).

Validity of Robot-based Assessments across the ICF: All of the scores on the clinical outcome measures correlated with the robot-based scores, however, different patterns emerged with regard to the ICF domains of Body Structure/Function, Activity, and Participation (Table 2). Across ICF domains, motor behavioral assessments focused on the upper extremity showed the strongest correlation with the robotic assessment of *speed* and the poorest with *reaction time* (Table 2).

ICF domain of Body Structure/Function Limitation: The FMA total score measures body structure/function and correlated most closely with the robot-based *speed* test (r= 0.82, p<0.0001), followed by *wrist targeting* (r = 0.72, p<0.0001); and *finger targeting* (r = 0.67, p<0.0001). Likewise, scores on the hand/wrist subset of the FMA correlated with the *speed* test (r = 0.79, p<0.001), but in this case, *finger targeting* (r = 0.68, p<0.001) was slightly more correlated than wrist targeting (r= 0.66, p<0.001).

ICF domain of Activity Limitation: The ARAT is a modality-specific measures of upper extremity activity limitation, and was significantly correlated with the *speed* test (r= 0.84, p<0.0001), *wrist targeting* (r= 0.76, p<0.0001), and *finger targeting* (r= 0.65, p<0.0001); the B/B, another modality-specific measure of upper extremity activity limitation, correlated most strongly with the *wrist targeting* (r= 0.85, p<0.0001), *speed* (r= 0.84, p<0.0001), and *finger targeting* (r= 0.65, p<0.0001) tests.

The Barthel Index is a global measure of activity limitation and had a unique profile of correlations with robotic assessments, being strongest for *finger targeting* (0.58, p< 0.0001) and weakest for *speed* (0.37, p< 0.05-0.007).

ICF domain of Participation Limitation: The SIS-hand correlated with robotic *wrist targeting* (r=0.68, p< 0.0001), followed by *speed* (r=0.65, p< 0.0001) tests.

Ceiling/Floor effects. The robotic tests performed well with regard to ceiling and floor effects. There was at least one robotic test without a ceiling effect (*finger targeting*) and at least one without a floor effect (*B/B*). The robust performance of robotic assessments with regard to this issue was particularly apparent when comparing the two versions of the B/B: while 12 subjects had the lowest score (zero blocks) on the clinically tested B/B test (30%), only 3 (7.5%) subjects had the lowest score (zero blocks) with the *robotic B/B test* (Figure 2).

Relationship between robotic assessments and anatomy. Each of the robot-based assessment scores significantly correlated with the percent CST injury (Table 3), indicating that that greater the injury to the CST, the worse the performance on those robot-based assessments. The robotic assessment scores of *finger targeting* (r=-0.56, p <0.007-0.0001) and *reaction time* (r=0.55, p <0.007-0.0001) were moderately correlated with percent CST injury. These correlations were stronger than the relationship between the primary clinical assessment (total FMA) and percent CST injury, which was r=-0.46, p < 0.006. A similar picture emerged when examining the amount of injury to the hand region of the primary motor cortex (M1), with which *finger targeting* and *reaction time* significantly correlated with amount of injury to the hand region of M1, while the relationship between the primary clinical assessment (total

FMA) and amount of injury to the hand region of M1 did not show a significant relationship (r=-0.16, p = 0.37).

Discussion

In this study, we **e**xplored the validity of five robot-based assessments of arm motor status by comparing them to established clinical and anatomical measures of stroke-induced upper extremity deficits. All of the robot-based assessment scores were rapidly obtained and demonstrated good construct validity with respect to several established clinical outcome measures across the ICF domains of Body Structure/Function, Activity, and Participation, but the results were less robust with respect to anatomical measures of motor system injury. The robot-based assessments strongly correlated with the total FMA score and the secondary clinical outcome measures (FMA hand/wrist, ARAT, B/B, BI, SIS-hand). The utility of robot-based testing is most apparent when using a panel of tests, including speed, wrist and finger targeting, and B/B, however, as no single test by itself was sufficient.

Overall, the robotic *speed* and *wrist targeting* tests were the most consistent modality-specific (i.e., arm motor function) performers, regardless of ICF level, followed by *finger targeting scores*, but this relationship did not hold true for the anatomical measures. With regard to injury to the CST and M1 hand area, both anatomical measures were most correlated with *reaction time* and *finger targeting* scores, whereas *speed* and *wrist targeting* were least correlated. As a result, these differences in scoring patterns may reveal some of the complex and differential effects of lesion size and location on behavior.

The relationships between scores on the robot-based assessments of arm motor behavior across the spectrum of WHO ICF domains were particularly interesting. For the ICF domain of Body structure/Function, the robot-based *speed* test was most highly correlated with scores on both the total FMA and the hand/wrist subsection of the FMA. For the Activity domain, the robot-based *speed* test was again correlated with the modality-specific tests of B/B, and ARAT; the robot-based *wrist targeting* test also highly correlated with B/B. Likewise, the robotic and clinical versions of the B/B, although slightly different, also correlated. The more global BI scores were most closely correlated with robot-based *finger targeting* and *wrist targeting* scores, but least correlated with *speed* and *reaction time* scores. Thus, the relationships between behavioral and robotic assessments clustered relative to modality-specificity vs. global function, not just according to ICF level. The arm motor modality-specific FMA, B/B, and ARAT are all timed tests, so speed likely plays a prominent role in performance. Since the items on the BI are not speed dependent, the motor control and coordination required for the targeting tests may be more relevant than speed for overall function. For the ICF domain of Participation, the SIS-hand scores were most correlated with *wrist targeting,* again suggesting that motor control may be more important than speed for overall function. These findings illustrate the relevance of robot-based assessments with respect to the ICF domains and modality-specific vs. global function deficits, providing a comprehensive picture of the full impact of stroke on individuals’ ability to function.

The correlations between robot-based assessments and anatomical measures of injury were generally weaker than those for the clinical outcome measures and the pattern of correlations differed somewhat. Robot-based assessments may offer some advantages over standardized clinical or neuroimaging measures of injury for capturing the effects of stroke. Overall, the anatomical results suggest that the robot-based assessments are of approximately similar value compared to the FMA total score in relation to percent CST injury, and indeed may be of greater validity than the FMA total score with respect to amount of M1 hand region injury. Since the robotic assessments did not require individuated fine finger movements, which would likely be more significantly impaired with damage to the hand region of M1 than other motor cortical areas contributing to the CST, [43, 44] the robotic assessment scores may better reflect the integrity of the CST than M1. These findings suggest that perhaps a more specific, patient-centered treatment approach may be developed by considering both the anatomy involved and the types of motor deficits measured by robot-based tests.

If valid outcome measures of upper extremity function that address ICF domains can be administered quickly, the time and cost of performing assessments may be reduced. Although previous investigators have demonstrated that kinematic measures derived from technology-based systems correlate well with standardized clinical measures [24], using simple, easy-to administer instrumented performance measures to assess the full spectrum of function across the ICF may prove to be more utilitarian in the long-run, particularly for individuals with stroke. Eventually, using robot-based assessments in lieu of standardized behavioral tests administered by a skilled clinician may provide opportunities for remote testing, such as in the context of telerehabilitation settings.

The results of this study were consistent across a variety of motor assessments, including instrumented, robot-based assessments of distal motor function; clinical outcome measures of impairment and activity, including modality-specific (arm motor) and global measures; and patient-reported measures of participation related to hand function. Valid and technology-based assessments that address the full spectrum of the ICF, and that are also related to anatomical measures of injury, may prove to be useful in driving the next generation of therapeutic interventions. For example, being able to track patient performance and progress quickly, easily, and remotely may make it easier for therapists to develop more patient-centered treatment plans that identify and address task-specific deficits.

In our sample population, language and cognitive deficits were mild and did not interfere with subjects' ability to use the instrumented assessments, thereby reinforcing the robot's utility as a device for measuring motor function in many individuals post-stroke. The specific threshold for cognitive and language deficits that might limit patients' abilities to participate in this type of testing is as yet undetermined, however.

Future work will explore an analysis of the potential cost benefit of using robot- or related technology-based assessments. Robot-based assessments have the potential to provide valid and highly consistent outcome assessments that can be used in emerging models of care, but further studies are needed to explore the full capabilities of this type of assessment strategy. Investigations into the use of instrumented assessments that are incorporated into Telerehabilitation systems and other game-based therapies are currently ongoing. While technology is unlikely to replace clinicians or clinical assessments, it is already playing a role in augmenting and expanding more typical rehabilitation provided one-on-one by therapists on-site, thereby off-setting current limitations in access to optimal care. As clinicians and researchers seek to clarify the relationships between and among lesion location and size, patients' scores on outcome measures, the selection of appropriate interventions, and the prognosis for recovery, so too must the appropriate use of technology be factored in to future models of healthcare delivery.

Limitations of the study. Some of the clinical outcome measures used in this study have floor (e.g., B/B) or ceiling (e.g., FMA, BI) effects. Nonetheless, they represent the current standards and are widely used in research in the field. The robot-based assessments used in this study may be prone to similar limitations, which is why using this battery of tests is preferable to using a single outcome measure. Also, the two versions of the B/B tests, while correlated, are different; the robotic version does not require proximal arm and shoulder movement and it allows more time overall, limiting the user's rate of grasp and release. As a result, the robot version may be slightly easier and less fatiguing than the clinical version. Future technology-based therapies also could benefit from incorporating measures of sensory function[45] to provide a more comprehensive assessment of upper extremity function. Finally, language and cognitive deficits were mild in the current population, so the extent to which current results generalize to a more globally impaired population remains to be determined. The use of technology-based assessment and treatment interventions may be restricted to those with minimal cognitive impairment until specific guidelines are established.

Conclusions

Robot-based assessment scores were valid across all domains of the ICF, correlating with both established clinical outcome measures and anatomical measures of motor system injury. Using a battery of robot-based, instrumented assessments (i.e., speed, finger targeting, wrist targeting, and B/B) of post-stroke upper extremity motor function may be a viable option for both patients and therapist

References

1. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, et al. Heart disease and stroke statistics--2012 update: A report from the american heart association. *Circulation*. 2012;125:e2-e220
2. Mayo NE, Wood-Dauphinee S, Ahmed S, Gordon C, Higgins J, McEwen S, et al. Disablement following stroke. *Disabil Rehabil*. 1999;21:258-268
3. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics--2015 update: A report from the american heart association. *Circulation*. 2015;131:e29-322
4. Nudo RJ. Adaptive plasticity in motor cortex: Implications for rehabilitation after brain injury. *J Rehabil Med*. 2003:7-10
5. Carey JR, Kimberley TJ, Lewis SM, Auerbach EJ, Dorsey L, Rundquist P, et al. Analysis of fmri and finger tracking training in subjects with chronic stroke. *Brain*. 2002;125:773-788
6. Kimberley TJ, Samargia S, Moore LG, Shakya JK, Lang CE. Comparison of amounts and types of practice during rehabilitation for traumatic brain injury and stroke. *J Rehabil Res Dev*. 2010;47:851-862
7. 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
8. Teasell RW FN, Salter KL, Jutai JW. A blueprint for transforming stroke rehabilitation care in canada: The case for change. *Arch Phys Med Rehabil.* 2008;89:575-578
9. Teasell R MS PS, McIntyre A, Janzen S, Allen L, Lobo L, Viana R. . Time to rethink long-term rehabilitation management of stroke patients. *Top Stroke Rehabil.* 2012;19:457-462
10. S. ROB. Trends in inpatient rehabilitation stroke outcomes before and after advent of the prospective payment system: A systematic review. . *JNPT.* 2010;34:17-23
11. Henderson A, Korner-Bitensky N, Levin M. Virtual reality in stroke rehabilitation: A systematic review of its effectiveness for upper limb motor recovery. *Top Stroke Rehabil*. 2007;14:52-61
12. Saposnik G, Levin M, Outcome Research Canada Working G. Virtual reality in stroke rehabilitation: A meta-analysis and implications for clinicians. *Stroke*. 2011;42:1380-1386
13. Reinkensmeyer DJ, Pang CT, Nessler JA, Painter CC. Web-based telerehabilitation for the upper extremity after stroke. *IEEE Trans Neural Syst Rehabil Eng*. 2002;10:102-108
14. Johansson T, Wild C. Telerehabilitation in stroke care--a systematic review. *Journal of telemedicine and telecare*. 2011;17:1-6
15. Volpe BT, Huerta PT, Zipse JL, Rykman A, Edwards D, Dipietro L, et al. Robotic devices as therapeutic and diagnostic tools for stroke recovery. *Arch Neurol*. 2009;66:1086-1090
16. Krebs HI, Volpe BT, Williams D, Celestino J, Charles SK, Lynch D, et al. Robot-aided neurorehabilitation: A robot for wrist rehabilitation. *IEEE Trans Neural Syst Rehabil Eng*. 2007;15:327-335
17. Lo AC, Guarino PD, Richards LG, Haselkorn JK, Wittenberg GF, Federman DG, et al. Robot-assisted therapy for long-term upper-limb impairment after stroke. *N Engl J Med*. 2010;362:1772-1783
18. Prange GB, Jannink MJ, Groothuis-Oudshoorn CG, 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
19. Takahashi CD, Der-Yeghiaian L, Le V, Motiwala RR, Cramer SC. Robot-based hand motor therapy after stroke. *Brain*. 2008;131:425-437
20. Mehrholz J, Platz T, Kugler J, Pohl M. Electromechanical and robot-assisted arm training for improving arm function and activities of daily living after stroke. *Stroke*. 2009

21. Chen SY, Winstein CJ. A systematic review of voluntary arm recovery in hemiparetic stroke: Critical predictors for meaningful outcomes using the international classification of functioning, disability, and health. *J Neurol Phys Ther*. 2009;33:2-13
22. Johnson MJ. Recent trends in robot-assisted therapy environments to improve real-life functional performance after stroke. *J Neuroeng Rehabil*. 2006;3:29
23. Sullivan JE, Andrews AW, Lanzino D, Perron AE, Potter KA. Outcome measures in neurological physical therapy practice: Part ii. A patient-centered process. *J Neurol Phys Ther*. 2011;35:65-74
24. Krebs HI, Krams M, Agrafiotis DK, DiBernardo A, Chavez JC, Littman GS, et al. Robotic measurement of arm movements after stroke establishes biomarkers of motor recovery. *Stroke*. 2014;45:200-204
25. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: A critical review of its measurement properties. *Neurorehabil Neural Repair*. 2002;16:232-240
26. Woodbury ML, Velozo CA, Richards LG, Duncan PW, Studenski S, Lai SM. Dimensionality and construct validity of the fugl-meyer assessment of the upper extremity. *Arch Phys Med Rehabil*. 2007;88:715-723
27. 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
28. Einav O, Geva D, Yoeli D, Kerzhner M, Mauritz KH. Development and validation of the first robotic scale for the clinical assessment of upper extremity motor impairments in stroke patients. *Topics in stroke rehabilitation*. 2011;18:587-598
29. Balasubramanian S, Colombo R, Sterpi I, Sanguineti V, Burdet E. Robotic assessment of upper limb motor function after stroke. *Am J Phys Med Rehabil*. 2012;91:S255-269
30. Riley JD, Le V, Der-Yeghiaian L, See J, Newton JM, Ward NS, et al. Anatomy of stroke injury predicts gains from therapy. *Stroke; a journal of cerebral circulation*. 2011;42:421-426
31. WHO. World health organisation (who) international classification of functioning, disability and health: Icf. . 2001;Geneva:
32. Mulroy SJ WC, Kulig K, Beneck GJ, Fowler EG, DeMuth SK, Sullivan KJ, Brown DA, Lane CJ; Physical Therapy Clinical Research Network. Secondary mediation and regression analyses of the ptclinresnet database: Determining causal relationships among the international classification of functioning, disability and health levels for four physical therapy intervention trials. *Phys Ther. 2011 Dec;91(12):1766-79.* **2011**;91:1766-1779
33. Fugl-Meyer AR, Jaasko 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;7:13-31
34. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res*. 1981;4:483-492
35. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair*. 2008;22:78-90
36. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the box and block test of manual dexterity. *Am J Occup Ther*. 1985;39:386-391
37. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The stroke impact scale version 2.0. Evaluation of reliability, validity, and sensitivity to change. *Stroke*. 1999;30:2131-2140
38. Collin C, Wade DT, Davies S, Horne V. The barthel adl index: A reliability study. *Int Disabil Stud*. 1988;10:61-63
39. Cramer SC, Koroshetz WJ, Finklestein SP. The case for modality-specific outcome measures in clinical trials of stroke recovery-promoting agents. *Stroke*. 2007;38:1393-1395.

40. See J. D-YL, Chou C., Chan, V., McKenzie A., Reinkensmeyer D., Cramer S. A standardized approach to stroke outcome assessments. Ii. The fugl-meyer motor assessment. *Neurorehabil Neural Repair*. 2012

41. Burke E, Dodakian L, See J, McKenzie A, Riley JD, Le V, et al. A multimodal approach to understanding motor impairment and disability after stroke. *J Neurol*. 2014;261:1178-1186

42. Swinscow T. Chapter 11. Correlation and regression. *Statistcis at Square One*. 1997;2017

43. Lawrence DG KH. The functional organization of the motor system in the monkey. I. The effects of bilateral pyramidal lesions. *Brain*. 1968;91:1-14

44. Schieber MH SM. Hand function: Peripheral and central constraints on performance. *J APpl Physiol.* 2004;96:2293–2300

45. Semrau JA, Herter TM, Scott SH, Dukelow SP. Robotic identification of kinesthetic deficits after stroke. *Stroke*. 2013;44:3414-3421

Figure Legends

Figure 1. Description of Robot Assessments:

A. Hand Wrist Assistive Rehabilitation Device (HWARD) Robot. The subject's forearm and hand are stablized in the cradle to allow flexion and extension of the wrist and hand in the plane of gravity. (Image from: Takahashi et al.,Instrumented hand motor therapy after stroke, *Brain* (2008); 131 (2): 425-437**,** used with permission from Oxford University Press.)

B. Wrist targeting task: Subject flexes and extends the affected wrist in the plane of gravity to align the cursor (white circle), over the colored balls, achieving 90% overlap of the target (blue ball) and holding the position for 1 sec. The balls flash at a set rate, alternating between red and blue, beginning at 3 sec intervals; in subsequent trials, the rate is increased or decreased, depending upon the subject's performance.

C. Finger targeting task: Subject flexes and extends the affected fingers in the plane of gravity to move the red bar inside blue box and keep it inside the blue box until the yellow bar fills for 3 sec, as represented by the yellow bar timer. The easiest level (Level 1) is shown above; with increasing levels of difficulty (up to level 25), the size of the target blue box is reduced.

D. Robotic Box and Blocks task: Subject must open their hand for a block to appear inside the image of the virtual hand on the computer screen. The subject then closes the hand for the virtual hand on the computer screen to grasp the virtual block until it clears the barrier, after which the subject's hand must open to release the virtual block.

(Reaction Time and Speed Tests not shown.)

Figure 2. Correlations Between Standard Box and Blocks and Robotic Box and Blocks Assessment: Scores on the instrumented version of the Box/Blocks test were significantly correlated with scores obtained by a therapist using the standard approach to this test (r=0.53, p<0.001). Note that the lowest score (zero blocks, floor effect) was found in 12 subjects (31.6%) using the standard B/B test but only 3 (7.5%) subjects with the instrumented B/B test.

Table 1. Characteristics of Subjects with Stroke.

N	40
Affected side	21 R / 19 L
Handedness	38 R / 2 L
Gender	29M / 11F
Age (Years)	58 ± 14 [21-86]
Time post-stroke (weeks)	19.2 ± 4.6 [10.9-26.0]
Total NIH Stroke Scale score (normal =0)	4.3 ± 2.2 [0-11]
Mini Mental Status Examination (normal = 30)	27.2 ± 2.8 [19-30]
Modified Rankin Score	2.3 ± 0.7 [range: 1-4]
Motor Behavioral Assessments (Affected Side):	
Total arm motor Fugl-Meyer Score (FMA) (normal=66)	35.6 ± 14.4 [13.5-60]
FMA-Hand/wrist Subsection (normal = 24)	10.5 ± 7.8 [1-24]
Action Research Arm Test (normal = 57)	25.1 ± 18.7 [0-57]
Box/Blocks (# blocks in 60 seconds) (normal = 75.2)	13.2 ± 15.5 [0-59]
Stroke Impact Scale II-hand motor (normal = 5)	2.1 ± 1.0 [1-4.2]
Barthel Index (normal =100)	88.5 ± 9.1 [60-100]
Robotic Assessments for Affected Side:	
Wrist Targeting (Worst Score = 6; Best Score = 1)	4.4 ± 1.3 [2.4-6]
Finger Targeting (Worst Score = 1; Best Score =25)	9.7 ± 10.0 [0-25]
Box and Blocks (Number of Blocks)	19.8 ± 7.6 [0-27]
Speed (Number of times across threshold)	4.2 ± 4.9 [0-19]
Reaction Time in seconds (Lower score is better)	0.6 ± 0.2 [0.1-1.3]
Anatomic Measures of Injury	
Infarct area, hand region primary motor (M1) cortex	1.8cm^3± 3.5 [0-13.5]
% CST injury	35.7% ± 25.8 [10-100]

Table 2. Correlations Between Motor Behavior and Robotic Assessments

Motor Behavior	Robotic Assessment				
	Finger Targeting	Wrist Targeting	Box and Blocks	Speed	Reaction Time
WHO ICF Level = Body/Structure function:					
FMA Total	0.67***	0.72***	0.53**	0.82***	0.37*
FMA Hand/wrist	0.68***	0.66***	0.55**	0.79***	0.34*
WHO ICF Level =Activity:					
ARAT	0.65***	0.76***	0.54**	0.84***	0.42**
B/B	0.65***	0.85***	0.52**	0.84***	0.41*
Barthel Index	0.58***	0.57**	0.51**	0.37*	0.44**
WHO ICF Level = Participation:					
SIS-hand motor	0.49*	0.68***	0.40*	0.65***	0.34*

*p< 0.05-0.007; **p< 0.007-0.0001; ***p< 0.0001. Absolute values are given for r

Table 3. Correlations Between Robotic Assessment and Injury Measures

Anatomic Measure	**Robotic Assessment**				
	Finger Targeting	**Wrist Targeting**	**Box and Blocks**	**Speed**	**Reaction Time**
Injury to Hand Region Primary Motor Cortex (M1)	0.37*	0.11	0.31	0.17	0.44*
Percent Corticospinal Tract Injury	0.56**	0.34*	0.52**	0.39*	0.55**

*p< 0.05-0.007; **p <0.007-0.0001. Absolute values are given for r

Figure 1. Description of Robot Assessments.

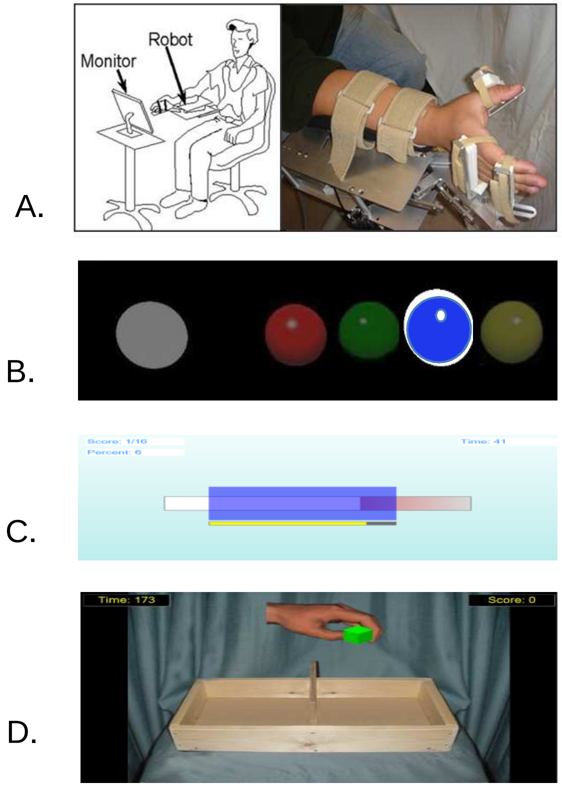
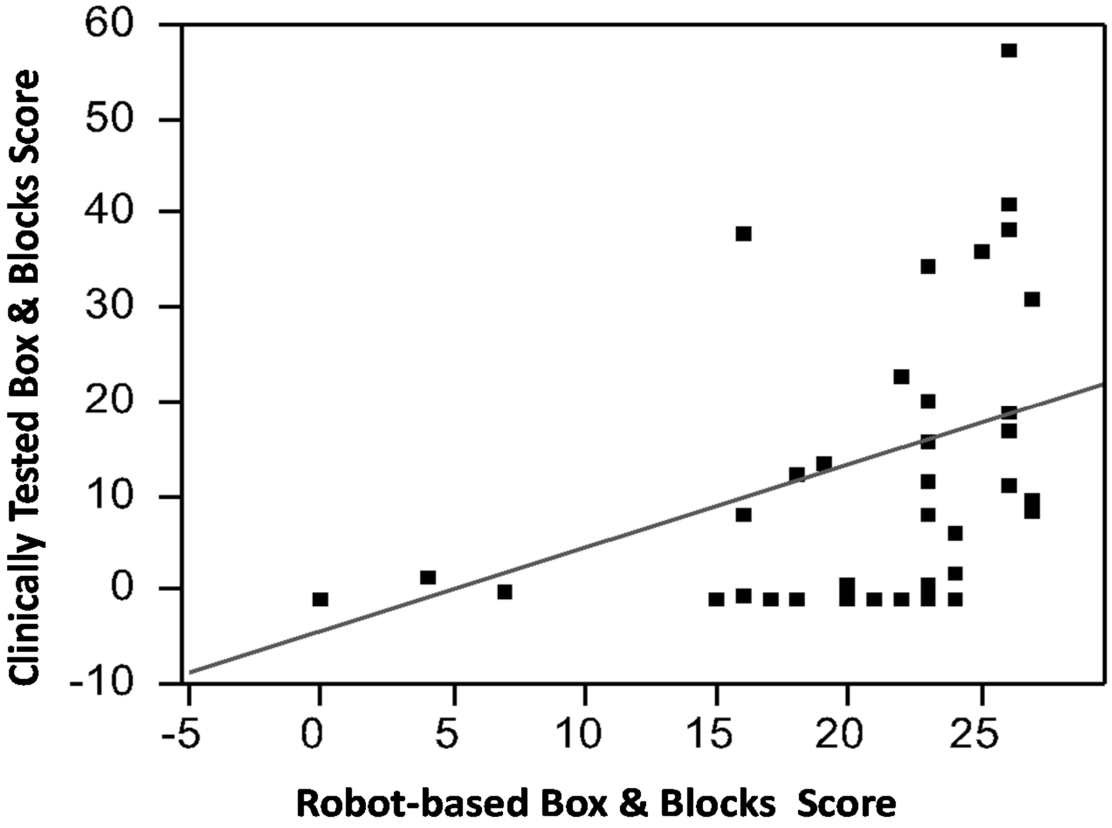


Figure 2. Correlations Between Standard Box and Blocks and Robotic Box and Blocks Assessment



Supplement B.

Description of Robot:

The HWARD device uses a lever design and air cylinders to achieve movement. Each air cylinder and limb interface is mounted on opposite ends of a lever, with a revolute joint in between. Midori CP-2FB low friction rotary potentiometers were used to translate the 360º endless mechanical rotation angles into a 0-5V range that was read by the computer using a the National Instruments PCI-6229 data acquisition card. This voltage value was used in the games to sense the degree of rotation.

The HWARD device allows 3-degrees-of-freedom (3-DOF) of rotational movement of the fingers, thumb, and wrist. The four fingers move as a single unit about the metacarpophalangeal (MCP) joint , allowing a range of movement (ROM) of approximately 25 to 90 degrees of flexion. Thumb movement out of the plane of the palm and fingers ranges from approximately 90% full extension to 75% of full flexion ROM. Wrist ßmovement ranges from approximately 20 degrees of extension to 15 degrees of flexion.

Description of Robot Assessments:

For all games, maximum finger extension/flexion and maximum wrist extension/flexion were recorded ahead of time. This enabled each game and assessment to be normalized to each subject's active range of movement.

1. Wrist targeting game: Images of four colored (red, green, blue, yellow) circles were aligned in a row on the computer screen. Subjects extended and flexed

the wrist, moving a round, white circular cursor on the screen that was normalized to their active range of motion. They then attempted to superimpose the moving cursor over the red and blue targets (positions 1 and 3), alternating between the two in response to a visual cue of the targets flashing (go signal). The starting rate was 3 seconds between go signals. If subjects scored greater than 60% accuracy at that level, they were advanced to a more difficult level (2 second intervals, then 1 second interval). If subjects were unable to meet the initial 3 second interval target, the level of difficulty was reduced to 4 second intervals (i.e., slower rate, up to a maximum of 6 seconds).

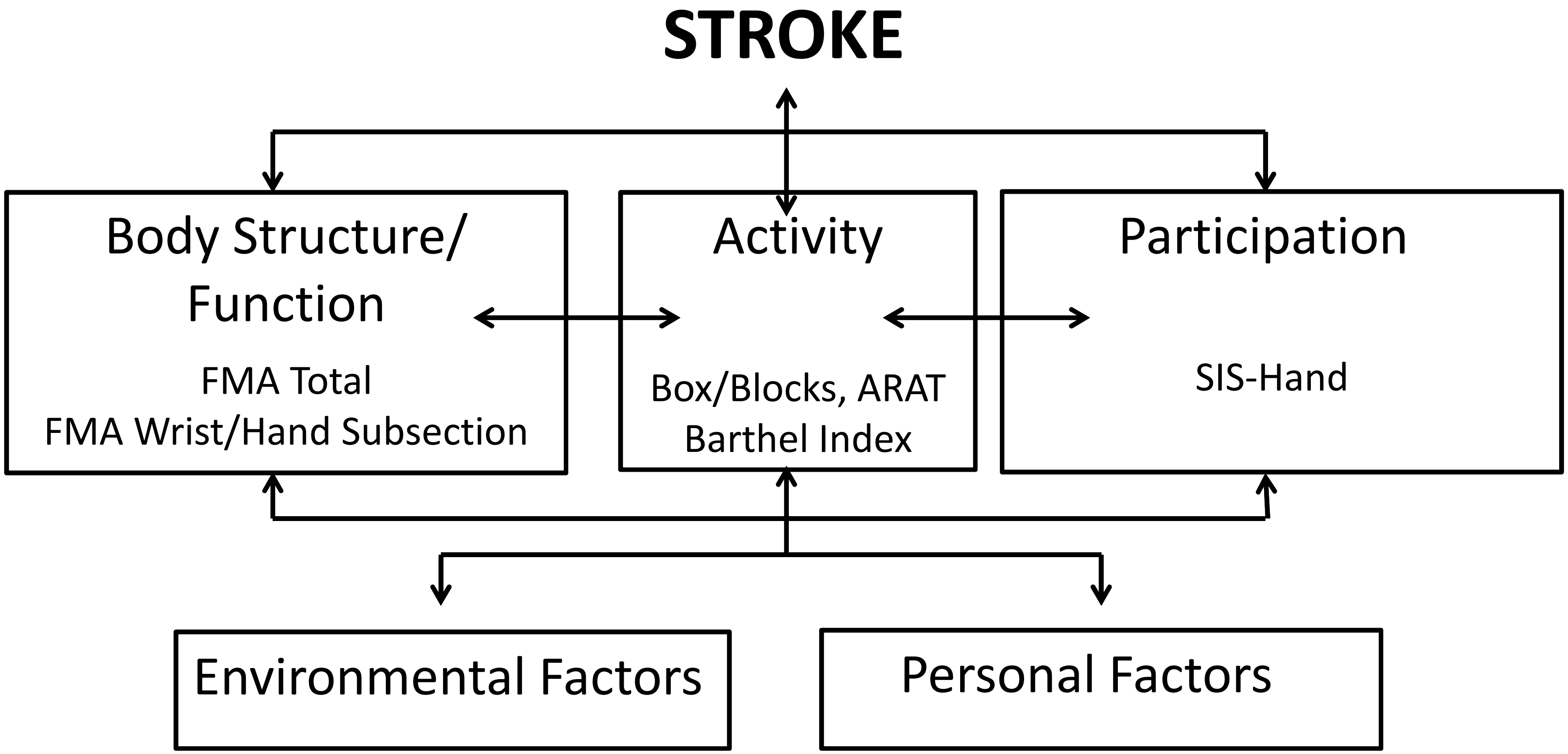
2. Finger Targeting Task. For this task, subjects moved the fingers (i.e., MCP flexion or extension) to move a cursor along a target (status bar) that was normalized to their active range of motion. Random targets would appear at various locations on the bar and the subject would be asked to flex or extend until their cursor moved into that location. They would then have to hold the position for a set amount of time. Each successful completion would be awarded 1 point. Total play time was 48 seconds and subjects were told to score as high as possible. The level of difficulty ranged from 1 (least difficult, large target box) to 25 (most difficult, small target box), depending upon the size of the target box. Testing began at level 10 and moved up or down, based upon the subject’s ability to achieve a score of > 60%.
3. Speed. For the speed game, the each specific appendage was placed in a set starting range: 50º - 90º for fingers and 25º - 50º for wrist. A line

corresponding to that degree was rendered on the computer screen, and a secondary line was rendered 10º below. The subjects were then asked to oscillate back and forth between these two set points over a duration of 20 seconds. Each time a successful alternation occurred between high and low, 1 point was awarded. Thus, higher scores were indicative of higher oscillation speed.

4. Reaction Time. Subjects self-selected their preferred motion, based on which was easiest, from among the motions of finger extension, finger flexion, wrist extension or wrist flexion. The goal of this assessment was to perform the selected motion as quickly as possible in response to a visual cue (rest, get ready, go signals). The specific appendage was again placed into a starting range: 33º -102º for fingers, 6º - 55º for wrist. The subject was then told to wait for a cue. When the cue was displayed, depending on the motion, the program would monitor for 2º of movement in the proper direction. For each of the 20 trials, the subject was allowed 21 seconds to try to cross the threshold. This assessment was then repeated, the number of trials set by the therapist, and the final score was the averaged response time. Lower scores indicated faster reaction times.
5. Box and Blocks. Box & block was a virtual representation of the real world assessment. A combination of either or both of the wrist and finger sensors were used to determine open and close hand positions. This threshold was determined by the therapist and corresponded to each subject's active range

of motion. When virtual blocks appeared on screen, the subject would have to move the proper appendages into the closed position. The block would then be moved virtually on screen over a vertical divider. Once past the divider, the subject would have to move to the open position, thereby releasing the block and scoring 1 point for each successful drop. If at any time the hand moved into the open position before crossing the divider, the virtual block would drop and return to the starting position with no score being awarded. The subjects were given 3 minutes to score as high as possible; the score is based upon the number of virtual blocks that the subject is able to get to the other side and release. The robotic version of the Box and Blocks test varies from the clinical version in that it: A) is based on finger grasp and release and does not require shoulder movements to move the block over the barrier, as the clinically tested version does; B) limits the maximum speed of block availability; and C) occurs over 3 minutes, rather than 1 minute for the clinical version.

Supplement B. Adapted ICF Framework



FMA, Box/Blocks, ARAT= Arm Motor Modality-Specific Outcome Measures; Barthel Index=Global Measure

Modified from: WHO (2001). "World Health Organisation (WHO) International Classification of Functioning, Disability and Health: ICF.Geneva.