2014

Robotic Measurement of Arm Movements After Stroke Establishes Biomarkers of Motor Recovery

H. I. Krebs

M. Krams

D. K. Agrafiotis

A. DiBernardo

J. C. Chavez

See next page for additional authors

Published in final edited form as:

Stroke. 2014 January ; 45(1): 200–204. doi:10.1161/STROKEAHA.113.002296.

Robotic Measurement of Arm Movements After Stroke Establishes Biomarkers of Motor Recovery

Hermano I. Krebs, PhD, **Michael Krams, MD**, **Dimitris K. Agrafiotis, PhD**, **Allitia DiBernardo, MD**, **Juan C. Chavez, PhD**, **Gary S. Littman, PhD**, **Eric Yang, PhD**, **Geert Byttebier, MSc**, **Laura Dipietro, PhD**, **Avrielle Rykman, MA**, **Kate McArthur, MD**, **Karim Hajjar, MD**, **Kennedy R. Lees, MD**, and **Bruce T. Volpe, MD**

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge (H.I.K., L.D.); Janssen Research & Development, Titusville, NJ (M.K., A.D.B.); Covance, Princeton, NJ (D.K.A., E.Y.); Biogen-Idec, Experimental Medicine, Cambridge, MA (J.C.C.); GSL Statistical Consulting, Ardmore, PA (G.S.L.); BVBA Bioconstat, Gent, Oostakker, Belgium (G.B.); The Burke Medical Research Institute, White Plains, NY (A.R.); Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, United Kingdom (K.M.A., K.H., K.R.L.); Department of Neurology, University of Duisburg-Essen, Essen, Germany (K.H.); and The Feinstein Institute for Medical Research, Manhasset, NY (B.T.V.).

Abstract

Background and Purpose—Because robotic devices record the kinematics and kinetics of human movements with high resolution, we hypothesized that robotic measures collected longitudinally in patients after stroke would bear a significant relationship to standard clinical outcome measures and, therefore, might provide superior biomarkers.

Methods—In patients with moderate-to-severe acute ischemic stroke, we used clinical scales and robotic devices to measure arm movement 7, 14, 21, 30, and 90 days after the event at 2 clinical sites. The robots are interactive devices that measure speed, position, and force so that calculated kinematic and kinetic parameters could be compared with clinical assessments.

Results—Among 208 patients, robotic measures predicted well the clinical measures (cross-validated R^2 of modified Rankin scale=0.60; National Institutes of Health Stroke Scale=0.63; Fugl-Meyer=0.73; Motor Power=0.75). When suitably scaled and combined by an artificial neural network, the robotic measures demonstrated greater sensitivity in measuring the recovery of patients from day 7 to day 90 (increased standardized effect=1.47).

Conclusions—These results demonstrate that robotic measures of motor performance will more than adequately capture outcome, and the altered effect size will reduce the required sample size. Reducing sample size will likely improve study efficiency.

Correspondence to Hermano I. Krebs, PhD, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA. hikrebs@mit.edu.

The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.113.002296/-/DC1.

Disclosures

The other authors report no conflicts.

Keywords

biomarkers; motor skills; robotics; sensory motor performance; stroke

Stroke is the leading cause of permanent disability in the United States.[1] A combination of an aging, stroke-prone population, improved neurological and medical care, and limitation of treatments for acute stroke to intravenous thrombolysis or various endovascular approaches[2,3] has heightened the urgency for new and improved clinical trials.[4] Trials of neuroprotectants impose narrow treatment windows and make patient accession difficult and time sensitive.[5] Despite the establishment of networks of treatment centers, to be adequately powered given the variability of outcome measures, clinical studies in stroke require large sample sizes. The current successful treatment trials for stroke required 600 to 850 patients.[2,3] Effect size and group number are partially determined by measurement variability that, in turn, results from the disparate nature and extent of initial clinical deficits and from reliance on multiple raters using scalar assessment tools, despite rigorous quality control procedures.[6] Furthermore, extensive training of raters and centralization of outcome assessments reduce this variability; this partial solution carries its own costs.[7] Thus, clinical research in stroke would be aided by a reliable, repeatable, and speedy assessment of continuous measures of impairment and its change during recovery. Such measurements would also yield valuable, objective information for an emerging archive on motor behavior after a stroke and the effect of different treatment on recovery. Interactive robotic technology can perform these measurements. This powerful toolbox has also been used as a training device with demonstrated significant benefits for poststroke motor recovery of the upper but not the lower extremities.[1,7–13]

Here, we describe a robotic assessment of 208 patients who were also evaluated with 4 well-known clinical instruments (National Institutes of Health Stroke Scale [NIHSS], modified Rankin scale [mRS], Fugl-Meyer [FM], and Motor Power [MP]). We tested whether robot-assisted measurement of kinematic and kinetic (RMK2) predicted the clinical scales with sufficient accuracy to serve as their surrogates and whether it is possible to design a more sensitive RMK2-based end point to measure motor recovery and thus reduce the number of subjects required to power future clinical trials.

Methods

Study Population

Patients with stroke (baseline NIHSS 7–20 recorded within 7 days of stroke onset) were enrolled and exposed to a battery of standard clinical assessments, including the NIHSS, the mRS, the FM, and the MP assessments. Each patient was also evaluated with the commercial version of the MIT-Manus robot[14] (Interactive Motion Technologies, Watertown, MA) but did not receive any robotic therapy. Both the affected (paretic) and unaffected limbs of patients were evaluated 7, 14, 21, 30, and 90 days after stroke onset. We enrolled 208 patients at 2 sites (146 patients at Burke Rehabilitation Center, White Plains, NY, and 62 at the Western Infirmary, Glasgow, Scotland, United Kingdom). To minimize inter-rater variability on the clinical scales, all the patients were assessed by the same set of

highly trained clinicians at each site. These patients were divided into 2 complementary populations: (1) those with complete data sets for days 7 and 90 after stroke onset (hereafter referred to as completers; n=87), and (2) those with some missing data on day 7 or 90 (referred to as noncompleters; n=121). Because there was no obvious advantage to patients with stroke to participate in the study (it did not include any incentives such as robotic therapy), many elected not to participate/return for all evaluations. In addition, to test the sensitivity of robot-assisted surrogate marker against historical data, we selected a subset of patients with stroke with a comparable degree of stroke severity as measured at day 7 NIHSS from the Virtual International Stroke Trials Archive (VISTA).[15,16] In brief, based on the frequency distribution of day 7 NIHSS from our study population, ranging from 0 to 24, ≈14 subjects were randomly chosen from the VISTA registry at the same clinical level measured at day 7 for each one of our study patients.

Robot-Assisted Measurement of Kinematics and Kinetics

The RMK2 assessment consists of a series of visually guided and visually evoked unconstrained reaching and circle-drawing movements and attempts to move against resistance that a patient with stroke completes in 30 minutes.[17] From these movements, the metrics include deviation from a straight line, aim, average speed, peak speed, and duration of movement while attempting to reach toward different targets. Three additional parameters characterizing the smoothness of movement are calculated: mean divided by the peak speed, mean magnitude of jerk normalized by the peak speed (best for discrete movements), and root mean square of jerk normalized by the duration of the movement (best for rhythmic movements). These unconstrained reaching movements are decomposed into submovements to produce 6 additional kinematic metrics (number, duration, overlap, peak, interval, and skewness of submovements).[18]

The circle-drawing task yields a metric that captures the coordination of arm movements.[19] The kinematic set is completed with 3 additional kinetic parameters that measure the strength of the shoulder and the ability of a patient to move against resistance or to hold against an externally applied force, leading to a total of 35 measurements for each patient (the parameter that measures shoulder strength has a ceiling effect and is measured only on the impaired side).[17] All parameters were linearly normalized between 0 and 1.

Model Generation

Given the large number of robot measures and possible combinations of measures, we built a model that identified the optimal set of robot metrics to correlate with clinical outcome or predicted clinical outcome. To minimize the effect of magnitude on the selection of parameters, we normalized the parameters via *z* scoring. We used a standard approach to build nonlinear models of the clinical scales with 3-layer artificial neural networks using logistic functions for the hidden and output layers.[20,21] The models were derived independently for each clinical scale and were trained to predict the clinical scores of a given patient on a given day from the respective RMK2 metrics using the completer population as a training set. To minimize overfitting, subsets of relevant RMK2 metrics were first identified using a feature selection algorithm.[22,23] Once the relevant features were identified, ensemble models comprising 10 neural network predictors were constructed

using the same network topology and training parameters but initialized with a different random number seed. The predictions of these 10 models were averaged to produce an ensemble prediction.[24] All models were cross-validated using the standard jackknife approach that divided the training data into 10 disjoint subsets containing 10% of the patterns each, systematically removing each subset from the training set, building a model with the remaining patterns, and predicting the clinical scores of the removed patterns using the optimized network parameters. The resulting predictions were compared with the original clinical scores to determine their degree of agreement (R^2_{CV}). This process was repeated 10× to obtain more robust cross-validation statistics. Finally, the best models identified by cross-validation were validated on the noncompleters, who formed an independent validation set.

Maximizing Effect Size With Novel Composites

We further tested whether the sensitivity of the clinical end points could be improved by using a novel RMK2-based composite optimized for effect size. Effect size is a reliable indicator of the ability of a scale to detect changes. It was assessed using Cohen's *d* for paired observations, defined as the mean divided by the SD of the day 7 to day 90 changes over all the completers. Because optimizing nonlinear composites does not lead to a unique solution (ill-posed mathematical problem), we restricted ourselves to using an algorithm constructed by identifying among the preselected RMK2 metrics the feature that yielded the maximum effect size and then adding 1 feature at a time until all preselected RMK2 metrics were included.

Results

Descriptive Statistics

Descriptive statistics by clinical site and completion status are provided in Tables 1 and 2. In this noninterventional observational study, no statistically significant difference in baseline severity was seen between the completer and noncompleter populations.

Correlation Analysis

The correlation structure of the RMK2 data set is illustrated in Figure 1. The nonlinear map was constructed using stochastic proximity embedding so as to preserve as much as possible the correlation distances among features, defined as $d_{ij}=1-\text{abs}(R_{ij})$, where R_{ij} is the Pearson correlation coefficient between the *i*th and *j*th features.[25,26] The 4 clinical scales (red) show a substantial degree of correlation to each other versus the majority of the RMK2 parameters, with FM and MP exhibiting a high level of correlation (R=0.933), as previously demonstrated.[6,17] Furthermore, for our patients, the RMK2 parameters on the affected side (blue) are more correlated to the clinical scales than those on the nonaffected side (green).

Prediction of Clinical Scales

Figure 2 shows the asymptotic behavior with respect to the number of robot-derived features. It is apparent that the robot-derived models for the FM and MP display comparable predictive power; furthermore, they are better predictors than those for NIHSS and mRS. The models retain much of their predictive power on the noncompleters as illustrated by the

dotted lines in Figure 2. These are patients who were not used by the model during training and represent an independent validation set. The weights and biases of the best model for each clinical scale are summarized in Table I of the online-only Data Supplement.

Maximizing Effect Size

Based on the frequency distribution of day 7 NIHSS scores from our study population, ranging from 0 to 24, we matched historical subjects from the VISTA database. This was done using simple random sampling to provide the same proportion of patients within each NIHSS point. This randomly selected group of 2937 patients had a day 7 mean NIHSS score of 5.7±4.1 that improved by 2.1±3.1 points to day 90, giving a standardized effect size of 0.67 for the changes from day 7 to day 90 in completers. Figure 3 shows the average effect size for the training and validation sets as a function of the number of robot-derived features included in the composite (using a similar cross-validation procedure to the one described earlier) and the historical value for NIHSS effect size.

Discussion

Limitations on Measurement Tools

The choice of primary outcome measure must be responsive to the targeted intervention to reflect a clinically meaningful change. Therefore, clinical evaluations must allow measurement of performance across multiple levels of function that correspond to components of the International Classification of Functioning and Disability, namely, body functions (specifically, movement abilities), activities, and participation. We selected the FM assessment[27,28] and the Medical Research Council MP[29,30] to assess motor impairment (International Classification of Functioning and Disability body function) and the NIHSS[31,32] and the mRS[33,34] to assess clinician-based measures of ability and activity, respectively (used in standard pharmacological interventions). Because of the practical limitation of the 1-hour span to measure both clinical and RMK2 scales, we did not include any additional assessment of functional use for the upper limb nor any patient-based assessment of activity (International Classification of Functioning and Disability activity level). Certainly, every scale has weaknesses as outlined by the ESO (European Stroke Organization) Outcomes Working Party[35] and potential for different forms of biases.

The RMK2 also has limitations: it has a floor effect in that subjects must be able to comprehend and follow basic instructions and partially move their impaired limb through a visually evoked and visually guided task in a gravity-compensated environment. In an effort to evaluate hemiplegic patients, we also included the evaluation of the so-called unaffected arm. We will evaluate the effectiveness of the approach of including the unaffected arm in a follow-up subgroup analysis including only patients with severe deficits.

Predicting Clinical Scales and Nonlinear Model

Here, we developed nonlinear models as predictors of the clinical scales. These nonlinear models are markedly better than our past efforts using multilinear regression models for chronic stroke,[17] in which we used principal component analysis for feature selection before model building. Here, we used the Ant algorithm for feature selection and artificial neural

networks.[22] The observed differences in predictive power between these different models could be because of several factors, including the type and number of patients used in the 2 studies, the use of neural networks as opposed to multilinear regression, and the use of a more elaborate feature selection algorithm. A direct comparison between these different models revealed that most of the present gains stem from the use of nonlinear modeling.

RMK2 and Smaller Clinical Trials

Similar to the individual items of the clinical scales, the individual RMK2 features are not as sensitive individually (data not shown). However, optimized RMK2 composites with only 4 features increased the effect size compared with the clinical scales by as much as 107% for the training and 83% for the validation set. This result cannot be attributed to motor learning because we demonstrated the stability of the RMK2 measurements on both people with chronic stroke and age-matched controls. An increase of 83% in effect size compared with the validated set would result in a 70% reduction in the number of patients required to achieve the typical 80% statistical power in a clinical trial. This potential reduction of the sample size is even larger compared with the training set.

One must take these results with the appropriate caveats. The population enrolled in this study was highly selected, with a day 7 mean NIHSS score of 5.7±4.1. Clearly, any gains that increased statistical power will need to be balanced against lower enrollment rates imposed by the selection criteria and against potential failure to complete follow-up or to comply with the RMK2 measurements. Furthermore, some of the apparent advantages of RMK2 may arise from the opportunity to repeat measurements and to use the mean of several. The same might be true if we use ordinal analysis, central adjudication by multiple raters, and global testing procedures that combine complementary scales across clinical domains and across time. Finally, in the current pool of completers, the NIHSS achieved a much larger effect size than historical data (VISTA). This larger effect size was because of a larger improvement among our completers with comparable variability. The pool of 2937 patients selected in the VISTA registry to match the distribution of our pool of 208 patients at day 7 had an improvement of 2.1±3.1 by day 90, whereas our completers had an improvement of 3.7±3.3. One might consider this difference with the appropriate caveats, particularly considering that the historical patient data were treated at a different time and in different institutions, and so there is inevitable bias.

Despite robotics limited penetration to date in the poststroke neurorehabilitation arena (eg, only 200 InMotion Arm robots were produced to date), taken together, our data suggest that robotic measurements may enable early decision making in clinical testing, reduce required sample sizes, and offer a more reliable method to track longitudinal change in patients affected by stroke than using current clinical scales. Furthermore, this study marks a novel beginning for the technology-based measurement of outcomes. It represents a proof of principle, with other robotic and wearable devices potentially affording even further efficiencies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding

This work was supported, in part, by grants from Wyeth-Ayerst Research, Johnson & Johnson Pharmaceutical Research & Development, and NIH R01 HD069776.

Dr. Krebs is a coinventor in several Massachusetts Institute of Technology–held patents for the robotic technology used in this work. He holds equity positions in Interactive Motion Technologies, Watertown, MA, the company that manufactures this type of technology under license to MIT.

References

1. Miller EL, Murray L, Richards L, Zorowitz RD, Bakas T, Clark P, et al. American Heart Association Council on Cardiovascular Nursing and the Stroke Council. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the American Heart Association. Stroke. 2010; 41:2402–2448. [PubMed: 20813995]
2. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. N Engl Med. 1995; 333:1581–1587.
3. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et al. ECASS Investigators. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. N Engl J Med. 2008; 359:1317–1329. [PubMed: 18815396]
4. Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, et al. Early stroke treatment associated with better outcome: the NINDS rt-PA stroke study. Neurology. 2000; 55:1649–1655. [PubMed: 11113218]
5. Cheng YD, Al-Khoury L, Zivin JA. Neuroprotection for ischemic stroke: two decades of success and failure. NeuroRx. 2004; 1:36–45. [PubMed: 15717006]
6. Krebs HI, Volpe BT, Ferraro M, Fasoli S, Palazzolo J, Rohrer B, et al. Robot-aided neurorehabilitation: from evidence-based to science-based rehabilitation. Top Stroke Rehabil. 2002; 8:54–70. [PubMed: 14523730]
7. 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. [PubMed: 20400552]
8. Aisen ML, Krebs HI, Hogan N, McDowell F, Volpe BT. The effect of robot-assisted therapy and rehabilitative training on motor recovery following stroke. Arch Neurol. 1997; 54:443–446. [PubMed: 9109746]
9. Volpe BT, Krebs HI, Hogan N, Edelstein OTRL, Diels C, Aisen M. A novel approach to stroke rehabilitation: robot-aided sensorimotor stimulation. Neurology. 2000; 54:1938–1944. [PubMed: 10822433]
10. Management of Stroke Rehabilitation Working Group. VA/DOD clinical practice guideline for the management of stroke rehabilitation. J Rehabil Res Dev. 2010; 47:1–43.
11. Kwakkel G, Kollen BJ, Krebs HI. Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review. Neurorehabil Neural Repair. 2008; 22:111–121. [PubMed: 17876068]
12. Norouzi-Gheidari N, Archambault PS, Fung J. Effects of robot-assisted therapy on stroke rehabilitation in upper limbs: systematic review and meta-analysis of the literature. J Rehabil Res Dev. 2012; 49:479–496. [PubMed: 22773253]
13. Mehrholz J, Hädrich A, Platz T, Kugler J, Pohl M. Electromechanical and robot-assisted arm training for improving generic activities of daily living, arm function, and arm muscle strength after stroke. Cochrane Database Syst Rev. 2012; 6:CD006876. [PubMed: 22696362]
14. Krebs HI, Hogan N, Aisen ML, Volpe BT. Robot-aided neurorehabilitation. IEEE Trans Rehabil Eng. 1998; 6:75–87. [PubMed: 9535526]
15. Ali M, Bath PM, Curram J, Davis SM, Diener HC, Donnan GA, et al. The Virtual International Stroke Trials Archive. Stroke. 2007; 38:1905–1910. [PubMed: 17446430]

16. Kerr DM, Fulton RL, Lees KR. VISTA Collaborators. Seven-day NIHSS is a sensitive outcome measure for exploratory clinical trials in acute stroke: evidence from the Virtual International Stroke Trials Archive. Stroke. 2012; 43:1401–1403. [PubMed: 22308254]
17. Bosecker C, Dipietro L, Volpe B, Krebs HI. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. Neurorehabil Neural Repair. 2010; 24:62–69. [PubMed: 19684304]
18. Krebs HI, Aisen ML, Volpe BT, Hogan N. Quantization of continuous arm movements in humans with brain injury. Proc Natl Acad Sci U S A. 1999; 96:4645–4649. [PubMed: 10200316]
19. Dipietro L, Krebs HI, Fasoli SE, Volpe BT, Stein J, Bever C, et al. Changing motor synergies in chronic stroke. J Neurophysiol. 2007; 98:757–768. [PubMed: 17553941]
20. Haykin, SS.; Haykin, SS. Neural Networks and Learning Machines. New York, NY: Prentice Hall; 2009.
21. Haykin, SS. Neural Networks: A Comprehensive Foundation. Upper Saddle River, NJ: Prentice Hall; 1999.
22. Izrailev S, Agrafiotis DK. Variable selection for QSAR by artificial ant colony systems. SAR QSAR Environ Res. 2002; 13:417–423. [PubMed: 12184383]
23. Izrailev S, Agrafiotis DK. A method for quantifying and visualizing the diversity of QSAR models. J Mol Graph Model. 2004; 22:275–284. [PubMed: 15177079]
24. Agrafiotis DK, Cedeño W, Lobanov VS. On the use of neural network ensembles in QSAR and QSPR. J Chem Inf Comput Sci. 2002; 42:903–911. [PubMed: 12132892]
25. Agrafiotis DK. Stochastic proximity embedding. J Comput Chem. 2003; 24:1215–1221. [PubMed: 12820129]
26. Yang E, Liu P, Rassokhin DN, Agrafiotis DK. Stochastic proximity embedding on graphics processing units: taking multidimensional scaling to a new scale. J Chem Inf Model. 2011; 51:2852–2859. [PubMed: 21961974]
27. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The poststroke hemiplegic patient. 1. a method for evaluation of physical performance. Scand J Rehabil Med. 1975; 7:13–31. [PubMed: 1135616]
28. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. Phys Ther. 1983; 63:1606–1610. [PubMed: 6622535]
29. Gregson JM, Leathley MJ, Moore AP, Smith TL, Sharma AK, Watkins CL. Clinical measures: reliable or not? Age Ageing. 2001; 30:86–87. [PubMed: 11322682]
30. Gregson JM, Leathley MJ, Moore AP, Smith TL, Sharma AK, Watkins CL. Reliability of measurements of muscle tone and muscle power in stroke patients. Age Ageing. 2000; 29:223–228. [PubMed: 10855904]
31. Pallicino P, Snyder W, Granger C. The NIH stroke scale and the FIM in stroke rehabilitation. Stroke. 1992; 23:919. [PubMed: 1595117]
32. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. Arch Neurol. 1989; 46:660–662. [PubMed: 2730378]
33. Burn JP. Reliability of the modified Rankin Scale. Stroke. 1992; 23:438. [PubMed: 1610453]
34. Quinn TJ, McArthur K, Dawson J, Walters MR, Lees KR. Reliability of structured modified Rankin scale assessment. Stroke. 2010; 41:e602. author reply e603. [PubMed: 21030706]
35. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. European Stroke Organization Outcomes Working Group. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. Stroke. 2012; 43:1163–1170. [PubMed: 22426313]

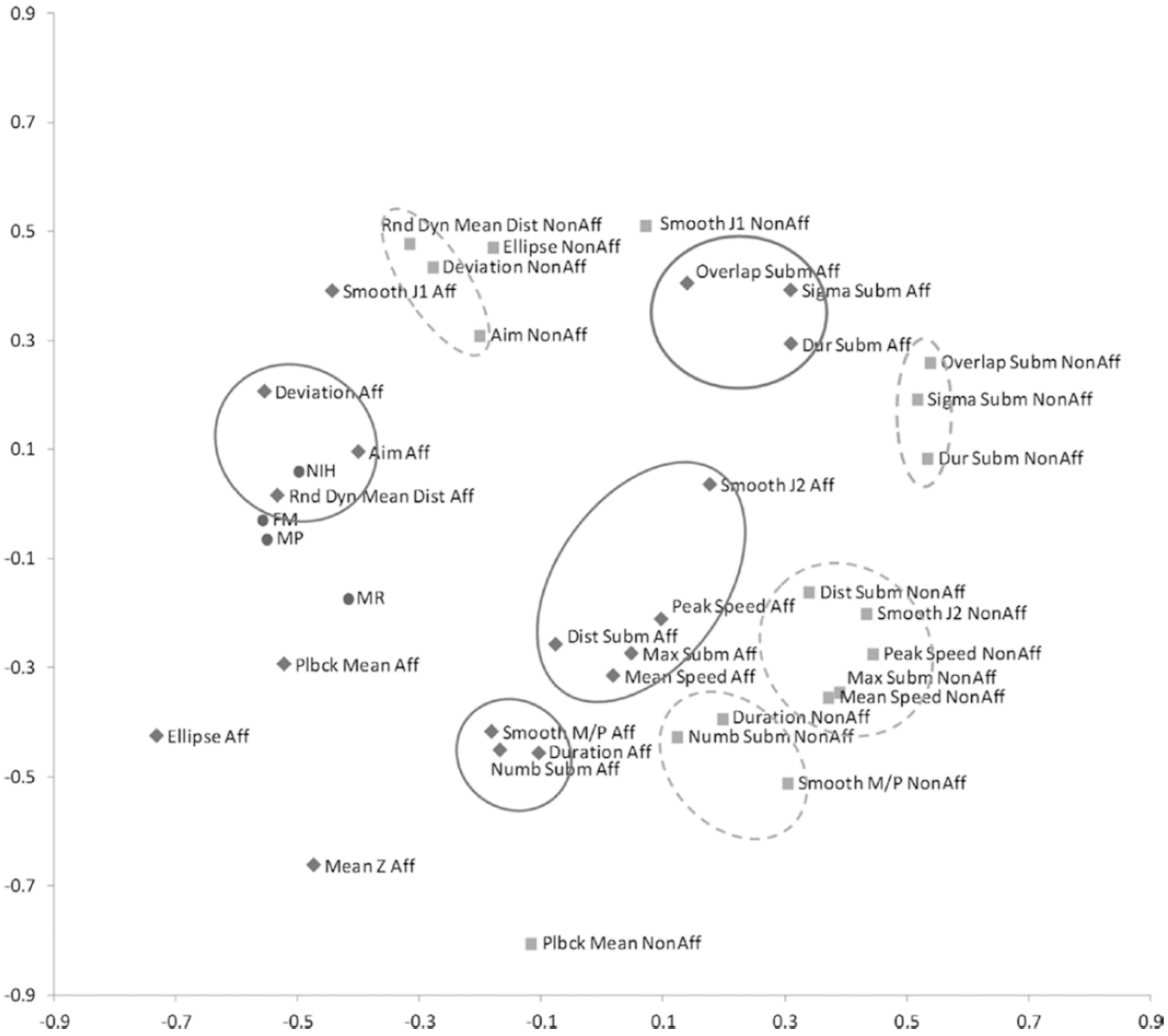

Figure 1.

Stochastic proximity embedding map of the correlation distances of the clinical and RMK2 parameters for the completers cohort. The map was derived by computing the pairwise Pearson correlation coefficients (*R*) for all pairs of features, converting them to correlation distances (1−abs(*R*)), and embedding the resulting matrix into 2 dimensions in such a way that the distances of the points on the map approximate as closely as possible the correlation distances of the respective features. The clinical parameters are displayed as circles, the RMK2 parameters on the affected side as diamonds, and the RMK2 parameters on the unaffected side as squares. The map also shows distinct clusters of correlated variables that are preserved on both the affected and unaffected sides (outlined by dark continuous and light dashed ellipses, respectively). Aff indicates affected or paretic limb; FM, Fugl-Meyer;

MP, Motor Power; MR, modified Rankin scale; M/P, the smoothness metric calculated from the mean and peak movement speed; and NIH, National Institutes of Health.

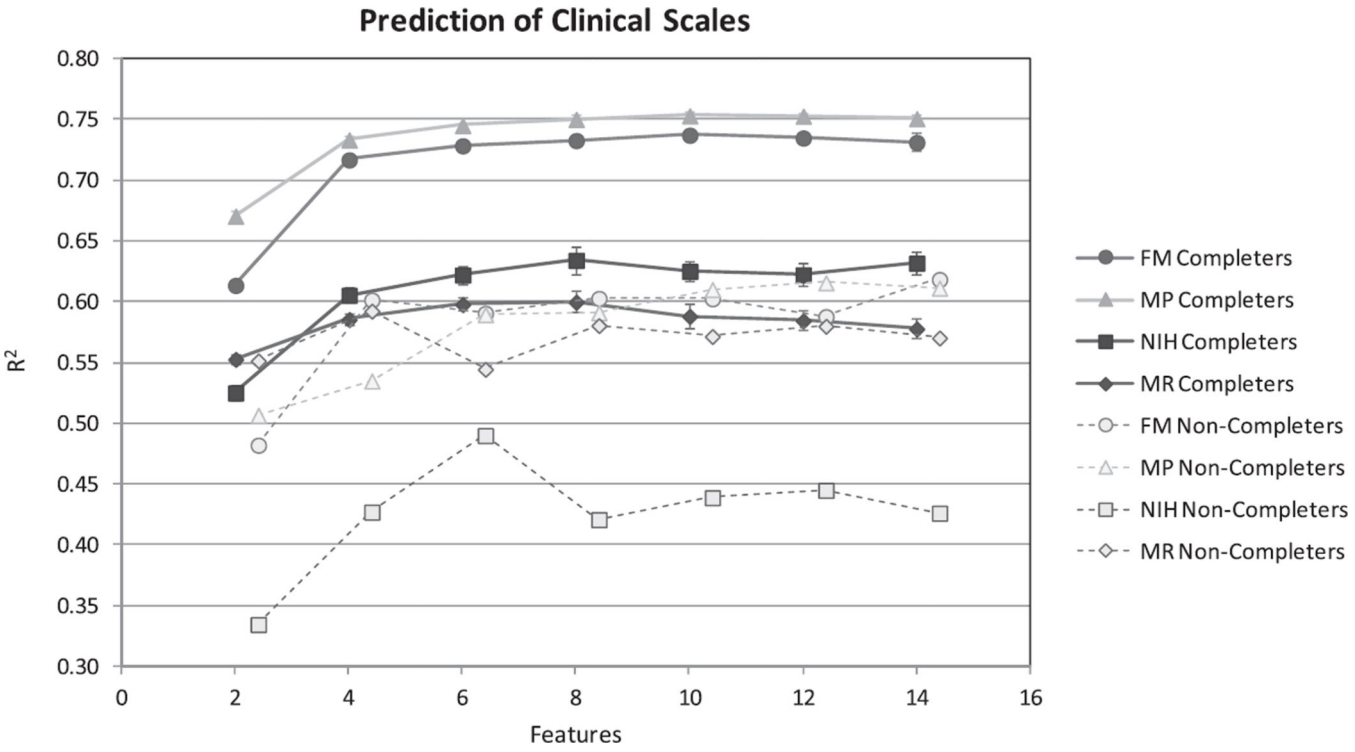

Figure 2.

Cross-validated R^2 of the best models derived from the completers (solid lines) and validated with the noncompleters (dashed lines) for each of the 4 clinical scales using 2, 4, 6, 8, 10, 12, and 14 robot-derived RMK2 features. The figure shows the ability of the robot-derived RMK2 models with increasing features to predict the clinical scales. The model performance exhibits an asymptotic behavior with respect to the number of RMK2 features, reaching the point of diminishing returns at ≈8 features for all 4 clinical scales. Note that the small variance in the prediction of the trained data as shown by the small whiskers, which for the most part, is not visible in the figure. FM indicates Fugl-Meyer; MP, Motor Power; and NIH, National Institutes of Health.

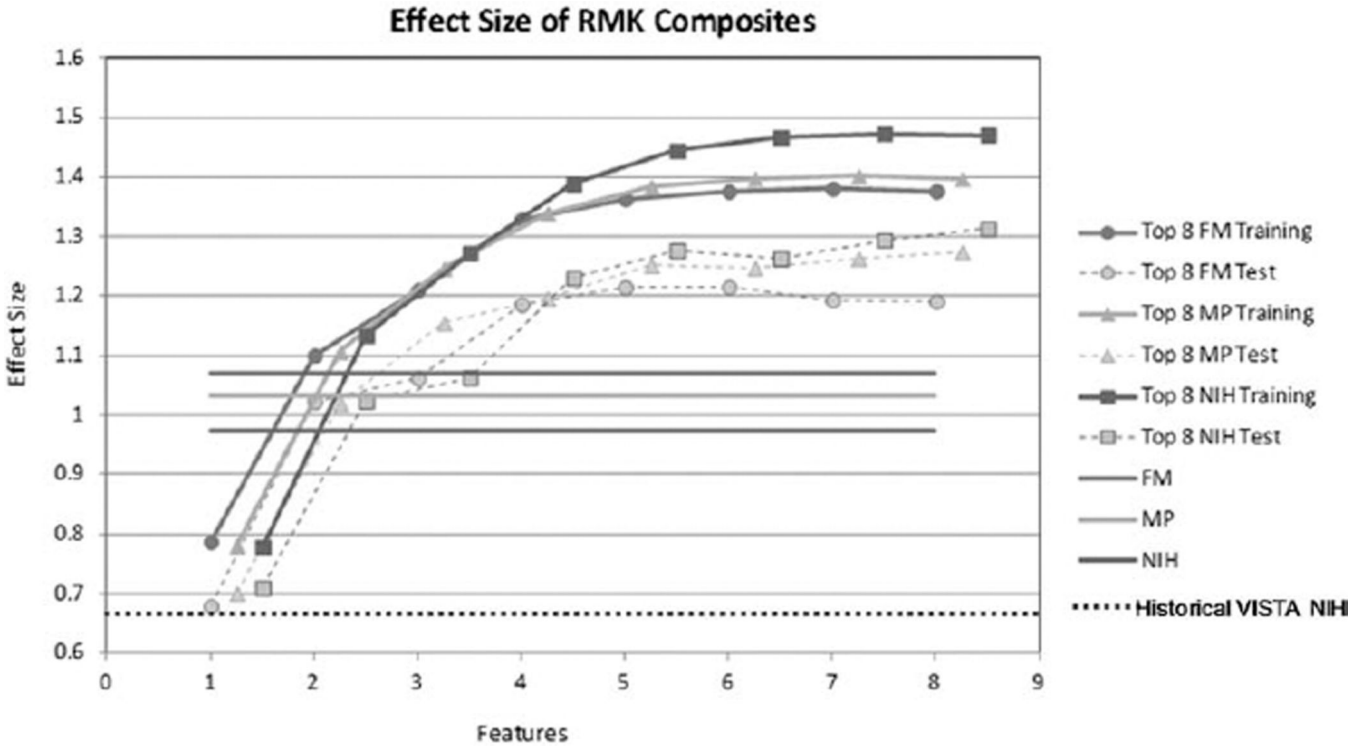

Figure 3.

Optimization of effect size for robot-derived RMK2 metrics. The horizontal lines show the day 7 to day 90 effect size for comparable patients of the historical Virtual International Stroke Trials Archive (VISTA) data for the National Institutes of Health Stroke Scale (NIHSS), as well the effect sizes for the NIHSS, Fugl-Meyer (FM), and Motor Power (MP) assessment scales for our completers cohort. The figure also shows the performance of the robot-derived RMK2 composites optimized for effect size for the trained (solid lines) and cross-validated sets (dashed lines). Note the increase of >20% in cross-validated effect size for the RMK2 composites compared with the clinical scales with 4 features for this study (and >70% over the historical data).

Table 1

Demographics and Descriptive Statistics

Characteristic	Burke (n=146)	Western Infirmary (n=62)
Sex, male	49%	55%
Age, y	74.4 (12.84)	67.27 (14.35)
Lesion side, right hemisphere	54%	58%
Handedness, right	88%	84%
Ethnicity (white, Hispanic, Asian, black)	70%, 9.5%, 2%, 18.5%	100%
NIHSS day 7 (/42)	6.9 (4.8)	3.9 (3.2)
FM day 7 (/66)	35.9 (21.0)	51.2 (20.2)
MP day 7 (/70)	38.4 (19.8)	62.5 (15.2)
NIHSS day 90 (/42)	2.4 (2.7)	2.3 (2.0)
FM day 90 (/66)	51.5 (19.0)	58.6 (17.5)
MP day 90 (/70)	53.9 (16.8)	65.4 (12.4)
mRS day 90 (/5)	3.2 (1.1)	1.8 (1.1)

Values represent the percentage or mean (SD). FM indicates Fugl-Meyer; MP, Motor Power; mRS, modified Rankin scale; and NIHSS, National Institutes of Health Stroke Scale.

/ indicates what is the maximum score in that particular scale.

Table 2

Training and Validation Data Sets

Characteristic	Completers (n=87)	Noncompleters (n=121)
Sex, male	52%	51%
Age, y	69.7 (13.5)	75.7 (13.0)
Lesion side, right hemisphere	51%	56%
NIHSS day 7 (/42)	5.6 (4.3)	6.5 (4.9)
FM day 7 (/66)	40.3 (21.7)	38.5 (21.9)
MP day 7 (/70)	45.5 (20.1)	39.4 (21.7)

Values represent the percentage or mean (SD). FM indicates Fugl-Meyer; MP, Motor Power; and NIHSS, National Institutes of Health Stroke Scale.