



Published in final edited form as:

Ann Neurol. 2016 September ; 80(3): 342–354. doi:10.1002/ana.24734.

Dose-response of task-specific upper limb training in people at least 6 months post stroke: A Phase II, single-blind, randomized, controlled trial

Catherine E Lang, PT, PhD^{1,2,3}, Michael J Strube, PhD⁴, Marghuretta D Bland, PT, DPT, MSCI^{1,2,3}, Kimberly J Waddell, MSOT, OTR/L¹, Kendra M Cherry-Allen, PT, DPT¹, Randolph J Nudo, PhD⁵, Alexander W Dromerick, MD⁶, and Rebecca L Birkenmeier, OTD, OTR/L^{1,2,3}

¹Program in Physical Therapy, Washington University School of Medicine, Saint Louis MO, USA

²Program in Occupational Therapy, Neurology, Washington University School of Medicine, Saint Louis MO, USA

³Department of Neurology, Washington University School of Medicine, Saint Louis MO, USA

⁴Department of Psychological and Brain Sciences, Washington University, Saint Louis MO, USA

⁵Department of Rehabilitation Medicine, Kansas University Medical Center, Kansas City KS, USA

⁶Department of Rehabilitation Medicine, Georgetown University and MedStar National Rehabilitation Hospital, Washington DC, USA

Abstract

Objective—1) Determine if higher doses of motor therapy in chronic post-stroke hemiparesis result in better outcomes compared to lower doses, and 2) Evaluate potential modifiers of the dose-response relationship.

Methods—Eighty-five adults with UE paresis 6 months after stroke were randomized to one of four dose groups in this single-blind, parallel, RCT. The dosing parameter manipulated was amount of task-specific training, as indexed by the number of task repetitions. Groups received 3200, 6400, 9600, or Individualized Maximum (IM) repetitions, during 1 hr sessions, 4 days/week for 8 weeks. The intervention was an individualized, progressive task-specific upper limb training program designed to improve upper limb functional motor capacity. The primary outcome was the slope of the Action Research Arm Test (ARAT) during the intervention. Effects of dose and potential modifiers of the dose-response relationship were evaluated with hierarchical linear models.

Corresponding author: Catherine Lang, Program in Physical Therapy, 4444 Forest Park, Campus Box 8502, Washington University School of Medicine, St. Louis, MO 63108 USA, langc@wustl.edu, 314-286-1945.

Author Contributions

Conception and design: CEL, MJS, RJN, AWD, RLB

Acquisition and analyses of data: CEL, MJS, MDB, KJW, KMC, RLB

Drafting the manuscript and figures: CEL, MJS

Conflicts of Interest

There are no potential conflicts of interest between the authors and any commercial sponsors.

Results—ARAT scores for the 3200, 9600, and IM groups improved over time as indicated by slopes (ARAT/wk, mean \pm SEs) of 0.40 ± 0.15 , 0.31 ± 0.16 , and 0.66 ± 0.14 , respectively ($p < 0.05$). The slope of the 6400 group was smaller (-0.05 ± 0.15) and significantly different from the 3200 and IM groups ($p < 0.001$). Initial motor capacity, neglect, and other tested characteristics did not modify the dose-response relationship.

Interpretation—Overall, treatment effects were small. There was no evidence of a dose-response effect of task-specific training on functional capacity in people with long-standing upper limb paresis post stroke.

INTRODUCTION

The idea that the dose of neurorehabilitation may influence outcomes¹ emerges from three sources: 1) clinical trials showing efficacy of a higher dose over a lower dose of treatment;^{2–5} 2) small Phase I feasibility trials showing a consistent, moderate relationship between dose received and amount of change;^{6–8} and 3) a recent meta-analysis showing a small-to-moderate effect size attributable to dose across numerous stroke rehabilitation trials.⁹ The early period after stroke however may be different, with two studies suggesting negative effects from higher doses.^{10, 11} There is an urgent need to explicitly examine dose-response relationships between evidence-based therapies (e.g. task-specific training) and outcomes.

The purpose of this trial was to determine the range of doses of task-specific upper limb training that produce the greatest improvements in outcomes in people 6 months post stroke. Quantification of dose in rehabilitation is substantially more difficult than quantification of dose for pharmaceutical agents, as there is no easy assay for the amount of active ingredient delivered or its half-life.¹ Here, we manipulated one parameter of dose, amount, via the total number of repetitions of task-specific practice, and kept all other parameters constant. Our primary hypothesis was that people who received higher doses would have better outcomes compared to those who received lower doses. Our secondary hypothesis was that beneficial doses were likely to vary across individuals, based on the severity of motor deficits and the presence of non-motor deficits (e.g. hemispatial neglect, depression) in other domains. The hypotheses were tested in individuals 6 months or more after stroke to avoid the potential confound of varying recovery trajectories in the first few months after injury.

METHODS

This was a Phase II, single-blind, randomized, parallel, dose-response trial conducted at a single site (NCT 01146379). Participants were recruited from the Cognitive Rehabilitation Research Group and the Brain Recovery Core¹² databases at Washington University School of Medicine in St. Louis, which contain contact information for adults with stroke who consented to being contacted for participation in research studies. Participants were also recruited via physician referral, posted flyers, stroke support groups, and word-of-mouth. Inclusion criteria were: 1) ischemic or hemorrhagic stroke as determined by a stroke neurologist and consistent with neuroimaging; 2) time since stroke 6-months; 3) cognitive skills to actively participate, as indicated by scores of 0–1 on items 1b and 1c of the National

Institutes of Health Stroke Scale (NIHSS);¹³ 4) unilateral upper limb weakness, as indicated by a score of 1–3 on item 5 (arm item) on the NIHSS; and 5) mild-to-moderate functional motor capacity of the affected upper limb, as indicated by a score of 10–48 on the Action Research Arm Test (ARAT).^{14–16} The lower limit of 10 on the ARAT meant that participants had at least some ability to partially open the hand, grasp and lift off the table at least 2–3 test items. Exclusion criteria were: 1) participant unavailable for 2-month follow-up; 2) inability to follow 2-step commands; 3) psychiatric diagnoses; 4) current participation in other upper limb stroke treatments (e.g. Botox); 5) other neurological diagnoses; 6) participants living further than one hour away and unwilling to travel for assessment and treatment sessions; and 7) pregnancy. This study was approved by the Washington University Human Research Protection Office and all participants provided written informed consent prior to participation.

Prior to study initiation, a power estimate for the primary analysis was done with the following assumptions: 1) ability to detect a 6 point difference between groups on the ARAT, a whole-number estimate of the minimal clinically important change for people with chronic stroke;¹⁷ 2) a standard deviation of change scores from our feasibility study of 8 points;⁷ 3) power of 0.80; 4) 2-tailed alpha = 0.05; 5) a dropout rate of 15%; and 6) the assumption that we would find up to 3 additional predictor variables (e.g., covariates). This led to an estimated sample size of 22 participants per group to be included in analyses.

Participants meeting criteria were randomized to one of four doses of individualized, progressive, upper limb task-specific training. Randomization was done by entering participant information into a custom-built secure web application hosted by the Washington University Division of Biostatistics. The website used an adaptive scheme to balance selected baseline characteristics, modified from procedures described by others.¹⁸ The baseline characteristics balanced were: age, time since stroke, baseline ARAT, and the presence/absence of depression and hemispatial neglect (see Study Measures below).

Dose Groups and the Intervention

The four dose groups were quantified by the total number of repetitions of upper limb tasks. The four doses tested were 3200 (100 repetitions/session), 6400 (200 repetitions/session), 9600 (300 repetitions/session), and Individualized Maximum repetitions (300 repetitions/session and sessions continuing until meeting stopping criteria, see below). Total doses were chosen based on pilot data⁷ and on doses used in animal studies in which neural plasticity was documented.¹⁹ The 3200 repetition dose served as the reference dose, as this is the approximate number of repetitions per session (100) observed in earlier studies, when repetitions of all types of activities (active exercise, passive exercise, functional training etc.) were summed.^{20, 21} The 6400 group and the 9600 group doubled and tripled the reference dose, respectively. The fourth group, the Individualized Maximum dose, was chosen because of the desire to probe how much and how long upper limb task-specific practice is tolerated and beneficial in humans with stroke.

The dose-controlled interventions were scheduled for 1 hour/day, 4 days/week, for 8 weeks. For participants randomized to the Individualized Maximum group, the intervention could extend beyond 8 weeks until they reached one or more stop criteria. These stop criteria were:

1) a change of 1 point or not exceeding a previously higher value on the ARAT for 2 consecutive weekly assessments; 2) a desire by the participant to stop the intervention; or 3) achievement of 100,000 repetitions.

The intervention consisted of supervised, massed practice of functional daily tasks which were appropriately graded and progressed for each participant.^{7, 8, 22} Most functional upper limb tasks require four essential movement components: reaching for, grasping, moving/manipulating, and then releasing an object.²³ What varies across the repertoire of daily upper limb tasks is how the combinations of the components are strung together and the specifics of the component (e.g. direction of reach, type of grasp, manipulative forces required). This intervention provided progressive training of combinations of these essential components through repeated practice of various tasks, with the desired goal of building the participant's capacity to perform a multitude of upper limb functions. Participants were given the Canadian Occupational Performance Measure²⁴ to assist in determining activities of interest and specific tasks in which to practice reaching, grasping, moving/manipulating, and releasing components during treatment sessions. The treatment approach in each group was identical except for the number of repetitions. The intervention was delivered by study therapists who were trained and monitored to ensure adherence to the protocol.

Using information from the baseline assessments, selected tasks were graded in difficulty to match the motor capabilities of the participant. The job of the study therapist was to grade tasks such that they engage and challenge, but not over- or underwhelm, the motor capabilities of each participant. Guiding principles for delivering the intervention, rules for progression, and example tasks are provided elsewhere.²² Repetitions per session were split across 3 tasks to allow for variability in task practice and to avoid the boredom that might come from practicing a single task. During each session, the treating therapist documented the tasks, specific grading levels, the number of repetitions, patient active time (i.e. time in active task practice),²⁵ and participant ratings of exertion (0–10 numeric rating scale).²⁶ Because exertion scales are typically used to assess effort during whole-body cardiorespiratory training, it was anticipated that perceived exertion during training of arm and hand movements would not be high.

Study Measures

Study measures assessed outcomes, quantified factors that might be modifiers of the dose-response relationship, monitored adherence to the protocol (see previous section), and provided further description of the sample. Measures were assessed at baseline prior to randomization, post-intervention, and then two months later by trained raters who were blinded to group assignment. A brief battery, including the primary outcome was assessed each week to obtain individual dose-response data. The primary outcome measure was the ARAT,^{14–16} a test that quantifies upper limb functional capacity with 19 items and a total score of 57. The ARAT was chosen because: 1) it has demonstrated strong psychometric properties across multiple studies, including sensitivity to change in people with stroke; 2) the time to administer is short compared to other, similar measures; and 3) it is widely used in upper limb rehabilitation trials around the world.

Secondary outcome measures included: 1) the Stroke Impact Scale – hand & ADL subscales,²⁷ as a measure of self-perceived daily performance; 2) the Canadian Occupational Performance Measure,²⁴ as a measure of success in achieving self-identified goals; and 3) a 7-point Likert scale^{28, 29} evaluating whether or not the individual thought he/she had changed and if the change was meaningful.

Potential modifiers of the dose-response relationship examined were: 1) initial severity of the functional deficit, quantified by the baseline ARAT score; 2) hemispatial neglect, quantified by the unstructured version of the Mesulam Cancellation test;³⁰ 3) depressive symptomatology, quantified by the Center for Epidemiologic Studies Depression Scale;^{31, 32} 3) aphasia, quantified by the Boston Naming Test – 15 item screen;³³ 4) cognitive loss, quantified by the Short Blessed test;³⁴ 5) impaired somatosensation, quantified by monofilament testing to the index finger;³⁵ and 6) muscle tone, quantified by the Modified Ashworth Scale at the elbow.³⁶ Additional demographic and clinical characteristics were age, gender, race, dominant side, side affected, time post stroke, type of stroke, stroke location, and number of strokes.

Statistical Analyses

The planned primary analytic approach used hierarchical linear modeling, also referred to as linear mixed effects regression.^{37–41} This approach was chosen instead of the traditional ANOVA approach because: 1) it allows flexible modeling of individual trajectories over time, including the modeling of moderators of intercepts and slopes; 2) data can be modeled even if some assessment points are missing, allowing all data to be used and not requiring data imputation; and 3) it does not require the same number of assessments or that the assessments are taken at the same time point, i.e. it can handle the longer intervention and more assessments in the Individualized Maximum group. Statistical analyses were done in R, with statistical significance predetermined at $p < 0.05$. A series of hierarchical models was constructed using the ARAT data. We first tested a model in which the individual intercepts and slopes (linear effect of time, out to the post-intervention assessment, P1) were included as random effects, with time centered on the treatment week 2 assessment, so that the intercepts represented individual ARAT performance at the start of treatment week 2. Group assignment (dummy coded) was included as a moderator of the intercepts and slopes. Subsequent models tested the inclusion of curvilinear time components (quadratic and cubic), alternative ways of modeling the repeated measures error structure (default diagonal error variance-covariance matrix, compound symmetry model, autocorrelation model, unstructured covariance matrix), and different centering points. A linear change model was found to be the best fit to the data. Alternative error structures did not alter the conclusions so the default diagonal model was retained. Other centering points did not produce inferences that differed from the original centering. Model comparisons were made either using likelihood ratio tests (if the models were nested) or using Akaike Information Criterion (if the models were not nested). We also examined the residuals for normality and the presence of outliers. Departures from normality were mild; nonetheless we verified the inferences reported here using bootstrapping procedures and permutation tests; no differences in inferences were found across analytic method. After establishing the basic model, other potential moderators of slopes and intercepts (e.g., baseline ARAT, depression)

were added, each in a separate model. Results presented in the tables include the intercepts and slopes from the best fit model, in units of the ARAT. The 3200 group was used as the reference group, such that presented group effects for the other three groups are their effect with respect to the 3200 group.

Analyses of SIS and COPM data used traditional ANOVAs with a between-subjects effect of group and a within-subject effect of time (baseline, P1), because these measures were assessed at baseline and post-intervention. Additional descriptive analyses included: 1) calculation of largest change (maximum ARAT value attained at any time point minus baseline ARAT) with the 95% confidence interval of that change for each group; 2) frequency of individuals with each score on the Likert scale quantifying overall perception of change; 3) calculation of ARAT change scores with 95% confidence intervals of those people who perceived a meaningful change versus those who did not;²⁸ and 4) correlation of change scores with total repetitions received. Finally, to determine if changes were retained after the intervention had ended, post-intervention (P1) and 2 month follow-up scores on the ARAT, SIS, and COPM subscales were compared using ANOVAs.

RESULTS

Eighty-five participants were enrolled between July 2011 and June 2015, with the last participant completing study assessments in October 2015. Figure 1 shows the CONSORT diagram. Demographic and clinical characteristics of the participants are provided in Table 1. In general, the sample consisted of people with long-standing mild and moderate paresis post stroke, and included people with non-motor deficits, such as hemispatial neglect, aphasia, and somatosensory loss. The four groups were well-matched, with the exception of concordance (dominant side = affected side) in the 6400 group, which had fewer people affected on their dominant side. Two additional characteristics, concordance and time post stroke, were explored post hoc as potential modifiers. A summary of the intervention delivered to each group is provided in Table 2. As intended, the groups attained the targeted number of repetitions. Patient active time, i.e. time spent physically practicing, increased over the four groups and was strongly correlated to the number of repetitions ($r = 0.92$, $p < 0.0001$). For the Individualized Maximum group, one participant stopped treatment after three sessions per his wishes; the remaining participants completed 32 sessions or more, stopping after meeting the criteria for lack of change on the ARAT. Participants reported a low-to-moderate level of exertion, as might be expected with an UE training program where many of the activities are seated, non-aerobic exercise.

Primary outcome

Overall, there was modest change in motor function but no clear difference in response based on treatment dosage. Figure 2 provides the individual response trajectories on the ARAT for each participant for each time point, displayed by group. The best fit mixed effects model included the linear effect of time, the effect of group, and the group by time interaction. The addition of time to the unconditional model accounted for 37% of the residual variance and produced a substantially better fitting model, $\chi^2(df=3) = 137$, $p < .001$ (AIC= 3006 vs. 3137). The addition of the treatment group dummy codes accounted for 0%

of the intercept variance (unsurprising given the centering occurred at the second treatment week) and 18% of the slope variance, producing a better fit compared to the model that included only the intercepts and slopes, $\chi^2(df=6) = 13.5$, $p < .05$ (AIC= 3005 vs. 3006). The correlation between intercepts and slopes was 0.03 ($p > 0.05$), indicating at the individual level, that rate of change was not related to ARAT score. The top portion of Table 3 shows the estimates (in units of the ARAT, using centering at the second treatment week) of model intercepts and slopes (ARAT/week), degrees of freedom, and p values, with the 3200 group used as the reference group. Modeled initial ARAT scores (intercepts) across the four groups were different from zero, but not different from each other, as expected based on the inclusion criteria, subsequent randomization, and centering. The effect of time (slope) in the 3200 group was significantly different from zero, but the magnitude of change was < 1 point per week. Slopes in the 9600 (absolute estimate = 0.31 ± 0.22) and Individualized Maximum (absolute estimate = 0.66 ± 0.20) groups were not different from the 3200 group. The slope of the 6400 group however was smaller (absolute estimate = -0.05 ± 0.21) than the 3200 and Individualized Maximum groups and not different from zero. Change scores (mean with 95% CI) over the intervention period for each group were 5.8 (3.9 – 7.7), 5.1 (3.1 – 7.1), 5.5 (3.4 – 7.6), and 8.4 (5.7 – 11.1) for the 3200, 6400, 9600, and Individualized Maximum groups respectively. No relationship between total repetitions and ARAT change scores was found (Spearman rho = 0.19, $p > 0.05$).

Potential modifiers to the basic model were then examined, one at a time. The bottom half of Table 3 show the estimates of change in intercepts (at the second treatment week) and change in slopes on the ARAT outcome for each potential modifier. Values provided represent overall effects on the intercepts and slopes for each moderator; additional models tested differential effects for treatment group (i.e., inclusion of two- and three-way interactions involving groups and moderators). Tests for the contributions that these interaction effects made to explanation of ARAT outcomes were not significant, with one exception, an interaction indicating that the slope in the 6400 group increased 0.07 ARAT pts/wk as depression increased. Three of nine potential modifiers had an influence on the intercept. First, for every 1 point increase in the baseline ARAT, participants had a 1.07 increase in the intercept, as one would expect. Second, for every 1 point increase on the Modified Ashworth Scale, participants had a decrease of 4.89 points on the intercept, indicating that people with greater muscle tone had lower ARAT scores. And third, if there was concordance between then dominant and affected limbs, then intercepts were increased by 7.04 points, indicating that people with an affected dominant limb had higher functional capacity at the start of treatment. None of the modifiers interacted with time to influence the slopes, i.e. the dose-response relationship. We note that baseline ARAT did not modify the dose-response relationship, indicating that the rates of change in those with initial low capacity (e.g. ARAT 10–20 points with minimal finger extension) were not different from the rates of change in those with higher initial capacity (e.g. ARAT 35–45 points and nearly full finger extension). Likewise, concordance did not modify the dose-response relationship, indicating that rates of change were not different between those affected on the dominant vs. non-dominant sides.

Secondary outcomes and retention over time

Overall, subjective measures indicated that patients felt they had improved despite the small objective motor improvement described above. Secondary outcome results are provided in Table 4. For the Stroke Impact Scales subscales where participants report the ability to perform standard items, there were small, but not statistically-different changes over time, with no differences between groups and no group \times time interactions. For the Canadian Occupational Performance Measure subscales where participants report the ability to perform self-selected goals, there were small, significant improvements over time, but no differences between groups or group \times time interactions. A large proportion of the sample (90%) perceived that they had experienced a meaningful change due to therapy. ARAT change scores however, were not different for those who thought they had made a meaningful change compared to those who did not. There were no changes on the ARAT or the secondary measures from the post-intervention (P1) assessment to the follow-up assessment (P2), indicating that changes were retained over time (all p values > 0.05 , data not shown)

DISCUSSION

We found no convincing data to indicate that more movement training resulted in better functional outcomes than less movement training in people with long-standing paresis post stroke. In this explicit test of the dose-response relationship for task-specific upper limb training, the intervention was delivered as intended, with manipulation of number of repetitions and other aspects carefully controlled. Three of the four dose groups made small improvements over time (slopes < 1 point per week) but the higher dose groups (9600 and Individualized Maximum groups) were not different from the 3200 repetition group. The 6400 group did not improve over time. None of the hypothesized potential modifiers had effects on the dose-response relationship. The analytic approach used here helped to minimize the inherent noise from clinical measures by modeling the trajectories of change, but even with this analytic approach, only small improvements were found. This Phase II RCT was powered to detect a 6-point difference between groups, as that has been suggested as a clinically-important difference on the primary outcome. Although a larger sample could modify statistical significance, enrollment of additional participants would not impact the clinical interpretation of the small rates of change and even smaller differences between higher versus lower doses.

The lack of a dose-response effect was unexpected given evidence suggesting that more is better,^{2, 4, 6, 9} including implications from animal models (for example see⁴²) and work from our own group.^{1, 7, 8} Despite the accumulation of indirect evidence, very few studies have explicitly examined responses to more than two different doses. Four different doses of an upper limb intervention were examined in another dose-response trial of people with more severe paresis post stroke enrolled 8 – 84 days post stroke.⁴³ They too failed to find a dose-response effect, which was partially attributed to difficulty delivering the two higher doses early after stroke. We were successful in delivering the higher doses, with a large percentage of time in those groups spent in active practice (Table 2). Indeed, if one considers the typical portion of active time during therapy sessions (8 – 50%),^{25, 44} our patient active time in the

highest dose group is likely equivalent to 65 hours of scheduled therapy. This duration is equivalent to scheduled therapy time in the original constraint-induced movement therapy protocols.⁴⁵ Compared to robotic studies, the number of repetitions per session in our higher dose groups may be about one third of the number of reaches per session with the robot.⁴⁶ Our repetitions were compound movements including reaching, grasping, manipulating and releasing, so it is hard to compare the numbers of repetitions directly. Interestingly, results from robotic trials in this same later phase post stroke have shown improvements in upper limb capacity which are similar or smaller than those found here. The task-specific training used here was developed as a human-adapted version of the training experienced by animal models in studies of neuroplasticity, and practiced tasks were just as carefully titrated to develop skill but focused on actions that are of individual interest and are functionally-relevant for daily life. The intervention was delivered beyond the time when spontaneous biological recovery generally occurs. Thus, these results suggest that larger doses of task-specific training may be of little benefit for humans at this later time point post stroke.

The small improvements in functional capacity and self-reported performance seen here were likely driven by the social and psychological effects of being in therapy, not by the amount of training provided. There must be a powerful effect of attending therapy, perhaps stemming from the close human interaction, given that 90% of our participants perceived a meaningful change and their perceptions were not linked to measured behavioral changes. It is also possible that the changes seen here were driven by repeated testing on the ARAT. While we cannot rule this out completely, the non-significant slope in the 6400 group suggests that there is not a repeated testing effect on the ARAT. Further, multiple baselines in a previous sample failed to show a change in scores due to repeated testing.⁷ Nonetheless, the magnitude of change seen here is similar to the magnitudes of change seen in many other motor intervention studies at this later time point post stroke (for examples from recent studies see⁴⁷⁻⁴⁹) and to changes in routine outpatient care,⁵⁰ but is substantially smaller than magnitudes of change seen in interventions delivered earlier after stroke.^{10, 51} The lack of a dose effect in this study highlights the importance of explicitly testing dose-response hypotheses with more than two dose groups in carefully controlled Phase II trials. Indeed, if we had chosen only the two dose groups of 6400 and 9600, then we would have reached the incorrect conclusion that higher doses were better than lower doses.

The lack of significant change in the 6400 group is difficult to explain. It cannot be explained by difference in group exposure to the intervention, since the number of participants completing the expected 32 sessions was similar to the other groups (Table 2). It also cannot be explained by concordance, i.e. when the dominant side is the affected side. While concordance was lower in the 6400 group (i.e. fewer people affected on their dominant side, Table 1), this did not result in a difference in baseline scores (Table 1) or model intercepts (Table 3). The presence of concordance did not affect response to the intervention, as indicated by a non-significant interaction of concordance and time (Table 3). There was also no unique effect of concordance in the 6400 group, because no group by concordance by time interactions were found. Conceptually, it seems unlikely that dose would have an inverted 'U' shape such that total repetitions around 6400 would yield poorer outcomes than 3200 and 9600 total repetitions. Given rates of change of < 1 point per week, perhaps the 6400 group response to intervention could simply be viewed as the low end of

possible responses to task-specific training at this later time period post stroke (i.e. no change).

There are several limitations to consider when interpreting these results. First, one wonders if the doses provided were simply too low. Some have speculated, based on rodent models of recovery, that much higher amounts of training in patients may be more effective. While we cannot eliminate this possibility, the presence of the Individualized Maximum group makes it less likely. We are skeptical that patients could achieve another order of magnitude of dose in the context of a traditional therapy session. Also, we found that participants in the Individual Maximum group reached a performance plateau prior to stopping, thus more training delivered over several weeks did not yield any additional benefit. Second, we did not use neurophysiological measures, such as the presence or absence of a motor evoked potential, as a randomization variable or as a potential modifier of the response.⁵² MRI data was obtained on a portion of the sample; that information may yield future insights into the dose-response question. And third, we manipulated only the amount of movement training and not the frequency (days/week) or intensity (e.g. cardiorespiratory effort). We cannot rule out the possibility that altering another dosing parameter would influence outcomes, as has been suggested for intensity of mobility interventions.⁵³ For upper limb training however, the influence of frequency may be minimal based on the relatively equivalent efficacy in original and modified constraint-induced movement therapy studies.⁴⁵ These last questions will require other trial designs to efficiently answer.

In conclusion, we found no evidence of a dose-response effect of task-specific training on functional capacity in people with long-standing upper limb paresis post stroke. The sample studied included people with motor and non-motor stroke-induced deficits, which is relatively uncommon in motor intervention research, but has more potential to generalize to routine clinical practice. The presence of non-motor deficits had no or negligible effects on responsiveness to the intervention. Overall improvements were small, with changes of < 1 point per week of therapy provided.

Acknowledgments

We thank Brittany Hill, Jill DeGeeter, Ryan Bailey, Michael Urbin, and Sydney Schafer for their invaluable assistance in running the trial. Funding was provided by NIH R01 HD068290.

REFERENCES

1. Lang CE, Lohse KR, Birkenmeier RL. Dose and timing in neurorehabilitation: Prescribing motor therapy after stroke. *Current opinion in neurology*. 2015
2. Duncan PW, Sullivan KJ, Behrman AL, Azen SP, Wu SS, Nadeau SE, et al. Body-weight-supported treadmill rehabilitation after stroke. *N Engl J Med*. 2011; 364:2026–2036. [PubMed: 21612471]
3. Ross LF, Harvey LA, Lannin NA. Do people with acquired brain impairment benefit from additional therapy specifically directed at the hand? A randomized controlled trial. *Clinical rehabilitation*. 2009; 23:492–503. [PubMed: 19321521]
4. Hsu SS, Hu MH, Wang YH, Yip PK, Chiu JW, Hsieh CL. Dose-response relation between neuromuscular electrical stimulation and upper-extremity function in patients with stroke. *Stroke; a journal of cerebral circulation*. 2010; 41:821–824.
5. Harris JE, Eng JJ, Miller WC, Dawson AS. A self-administered graded repetitive arm supplementary program (grasp) improves arm function during inpatient stroke rehabilitation: A

- multi-site randomized controlled trial. *Stroke; a journal of cerebral circulation*. 2009; 40:2123–2128.
6. Moore JL, Roth EJ, Killian C, Hornby TG. Locomotor training improves daily stepping activity and gait efficiency in individuals poststroke who have reached a "plateau" in recovery. *Stroke; a journal of cerebral circulation*. 2010; 41:129–135.
 7. Birkenmeier RL, Prager EM, Lang CE. Translating animal doses of task-specific training to people with chronic stroke in 1-hour therapy sessions: A proof-of-concept study. *Neurorehabilitation and neural repair*. 2010; 24:620–635. [PubMed: 20424192]
 8. Waddell KJ, Birkenmeier RL, Moore JL, Hornby TG, Lang CE. Feasibility of high-repetition, task-specific training for individuals with upper-extremity paresis. *Am J Occup Ther*. 2014; 68:444–453. [PubMed: 25005508]
 9. Lohse KR, Lang CE, Boyd LA. Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke; a journal of cerebral circulation*. 2014; 45:2053–2058.
 10. Dromerick AW, Lang CE, Birkenmeier RL, Wagner JM, Miller JP, Videen TO, et al. Very early constraint-induced movement during stroke rehabilitation (vectors): A single-center rct. *Neurology*. 2009; 73:195–201. [PubMed: 19458319]
 11. Efficacy and safety of very early mobilisation within 24 h of stroke onset (avert): A randomised controlled trial. *Lancet*. 2015
 12. Lang CE, Bland MD, Connor LT, Fucetola R, Whitson M, Edmiaston J, et al. The brain recovery core: Building a system of organized stroke rehabilitation and outcomes assessment across the continuum of care. *J Neurol Phys Ther*. 2011; 35:194–201. [PubMed: 22027474]
 13. Brott T, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: A clinical examination scale. *Stroke; a journal of cerebral circulation*. 1989; 20:864–870.
 14. Lang CE, Bland MD, Bailey RR, Schaefer SY, Birkenmeier RL. Assessment of upper extremity impairment, function, and activity after stroke: Foundations for clinical decision making. *J Hand Ther*. 2013; 26:104–115. [PubMed: 22975740]
 15. 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. [PubMed: 7333761]
 16. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabilitation and neural repair*. 2008; 22:78–90. [PubMed: 17704352]
 17. van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Deville WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients: Results from a single-blind randomized clinical trial. *Stroke; a journal of cerebral circulation*. 1999; 30:2369–2375.
 18. Signorini DF, Leung O, Simes RJ, Beller E, Gebiski VJ, Callaghan T. Dynamic balanced randomization for clinical trials. *Stat Med*. 1993; 12:2343–2350. [PubMed: 8134737]
 19. Nudo RJ, Milliken GW, Jenkins WM, Merzenich MM. Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 1996; 16:785–807. [PubMed: 8551360]
 20. Lang CE, MacDonald JR, Gnip C. Counting repetitions: An observational study of outpatient day treatment for people with hemiparesis. *Journal of Neurologic Physical Therapy*. 2007; 31:1–8.
 21. Lang CE, Macdonald JR, Reisman DS, Boyd L, Jacobson Kimberley T, Schindler-Ivens SM, et al. Observation of amounts of movement practice provided during stroke rehabilitation. *Archives of physical medicine and rehabilitation*. 2009; 90:1692–1698. [PubMed: 19801058]
 22. Lang, CE.; Birkenmeier, RL. Upper-extremity task-specific training after stroke or disability: A manual for occupational therapy and physical therapy. Alexandria, VA: AOTA Press, Inc.; 2013.
 23. Lang, CE. Impaired motor control. In: Guccione, A.; Wong, R.; Avers, D., editors. *Geriatric physical therapy*. Elsevier; 2012.
 24. Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The canadian occupational performance measure: An outcome measure for occupational therapy. *Can J Occup Ther*. 1990; 57:82–87. [PubMed: 10104738]
 25. Host HH, Lang CE, Hildebrand MW, Zou D, Binder EF, Baum CM, et al. Patient active time during therapy sessions in postacute rehabilitation: Development and validation of a new measure. *Phys Occup Ther Geriatr*. 2014; 32:169–178. [PubMed: 25419032]

26. Lee IM, Sesso HD, Oguma Y, Paffenbarger RS Jr. Relative intensity of physical activity and risk of coronary heart disease. *Circulation*. 2003; 107:1110–1116. [PubMed: 12615787]
27. 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; a journal of cerebral circulation*. 1999; 30:2131–2140.
28. Lang CE, Edwards DF, Birkenmeier RL, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Archives of physical medicine and rehabilitation*. 2008; 89:1693–1700. [PubMed: 18760153]
29. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol*. 1998; 16:139–144. [PubMed: 9440735]
30. Rengachary J, d'Avossa G, Sapir A, Shulman GL, Corbetta M. Is the posner reaction time test more accurate than clinical tests in detecting left neglect in acute and chronic stroke? *Archives of physical medicine and rehabilitation*. 2009; 90:2081–2088. [PubMed: 19969172]
31. Radloff LS. The ces-d scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401.
32. Shinar D, Gross CR, Price TR, Banko M, Bolduc PL, Robinson RG. Screening for depression in stroke patients: The reliability and validity of the center for epidemiologic studies depression scale. *Stroke; a journal of cerebral circulation*. 1986; 17:241–245.
33. Goodglass, H.; Kaplan, E.; Barresi, B. Boston diagnostic aphasia examination. Philadelphia, PA: Lippincott, Williams & Wilkins; 2001.
34. Katzman R, Brown T, Fuld P, Peck A, Schechter R, Schimmel H. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am J Psychiatry*. 1983; 140:734–739. [PubMed: 6846631]
35. Bell-Krotoski, JA. Light touch - deep pressure testing using semmes-weinstein monofilaments. In: Hunter, JM.; Schneider, LH.; Mackin, EJ.; Callahan, AD., editors. *Rehabilitation of the hand: Surgery and therapy*. St. Louis, MO: Mosby Co.; 1990. p. 585-593.
36. Bohannon RW, Smith MB. Interrater reliability of a modified ashworth scale of muscle spasticity. *Physical therapy*. 1987; 67:206–207. [PubMed: 3809245]
37. Long, JD. *Longitudinal data analyses for the behavioral sciences using r*. Thousand Oaks, CA: Sage Publications; 2012.
38. Snijders, T.; Bosker, R. *Multilevel analyses: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications; 2012.
39. Raudenbush, SW.; Bryk, AS. *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage; 2002.
40. Maxwell, SE.; Delaney, HD. *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth; 1990.
41. Heck, RH.; Thomas, SL. *An introduction to multilevel modeling techniques*. New York, NY: Routledge; 2009.
42. Bell JA, Wolke ML, Ortez RC, Jones TA, Kerr AL. Training intensity affects motor rehabilitation efficacy following unilateral ischemic insult of the sensorimotor cortex in c57bl/6 mice. *Neurorehabilitation and neural repair*. 2015; 29:590–598. [PubMed: 25323461]
43. Hunter SM, Hammett L, Ball S, Smith N, Anderson C, Clark A, et al. Dose-response study of mobilisation and tactile stimulation therapy for the upper extremity early after stroke: A phase I trial. *Neurorehabilitation and neural repair*. 2011; 25:314–322. [PubMed: 21282528]
44. Hayward KS, Brauer SG. Dose of arm activity training during acute and subacute rehabilitation post stroke: A systematic review of the literature. *Clinical rehabilitation*. 2015
45. Kwakkel G, Veerbeek JM, van Wegen EE, Wolf SL. Constraint-induced movement therapy after stroke. *Lancet Neurol*. 2015; 14:224–234. [PubMed: 25772900]
46. 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]
47. Ackerley SJ, Byblow WD, Barber PA, MacDonald H, McIntyre-Robinson A, Stinear CM. Primed physical therapy enhances recovery of upper limb function in chronic stroke patients. *Neurorehabilitation and neural repair*. 2015

48. Young BM, Nigogosyan Z, Walton LM, Remsik A, Song J, Nair VA, et al. Dose-response relationships using brain-computer interface technology impact stroke rehabilitation. *Frontiers in human neuroscience*. 2015; 9:361. [PubMed: 26157378]
49. Wolf SL, Sahu K, Bay RC, Buchanan S, Reiss A, Linder S, et al. The haapi (home arm assistance progression initiative) trial: A novel robotics delivery approach in stroke rehabilitation. *Neurorehabilitation and neural repair*. 2015; 29:958–968. [PubMed: 25782693]
50. Lohse KR, Bland MD, Lang CE. Quantifying change during outpatient stroke rehabilitation: A retrospective regression analysis. in review.
51. Wolf SL, Dromerick AW, Lane CJ, Nelsen MA, Lewthwaite R, Cen SY, et al. Icare primary results: A phase iii stroke rehabilitation trial. *International Stroke Conference*. 2015:LB15.
52. Stinear CM, Barber PA, Smale PR, Coxon JP, Fleming MK, Byblow WD. Functional potential in chronic stroke patients depends on corticospinal tract integrity. *Brain : a journal of neurology*. 2007; 130:170–180. [PubMed: 17148468]
53. Hornby TG, Straube DS, Kinnaird CR, Holleran CL, Echaz AJ, Rodriguez KS, et al. Importance of specificity, amount, and intensity of locomotor training to improve ambulatory function in patients poststroke. *Topics in stroke rehabilitation*. 2011; 18:293–307. [PubMed: 21914594]

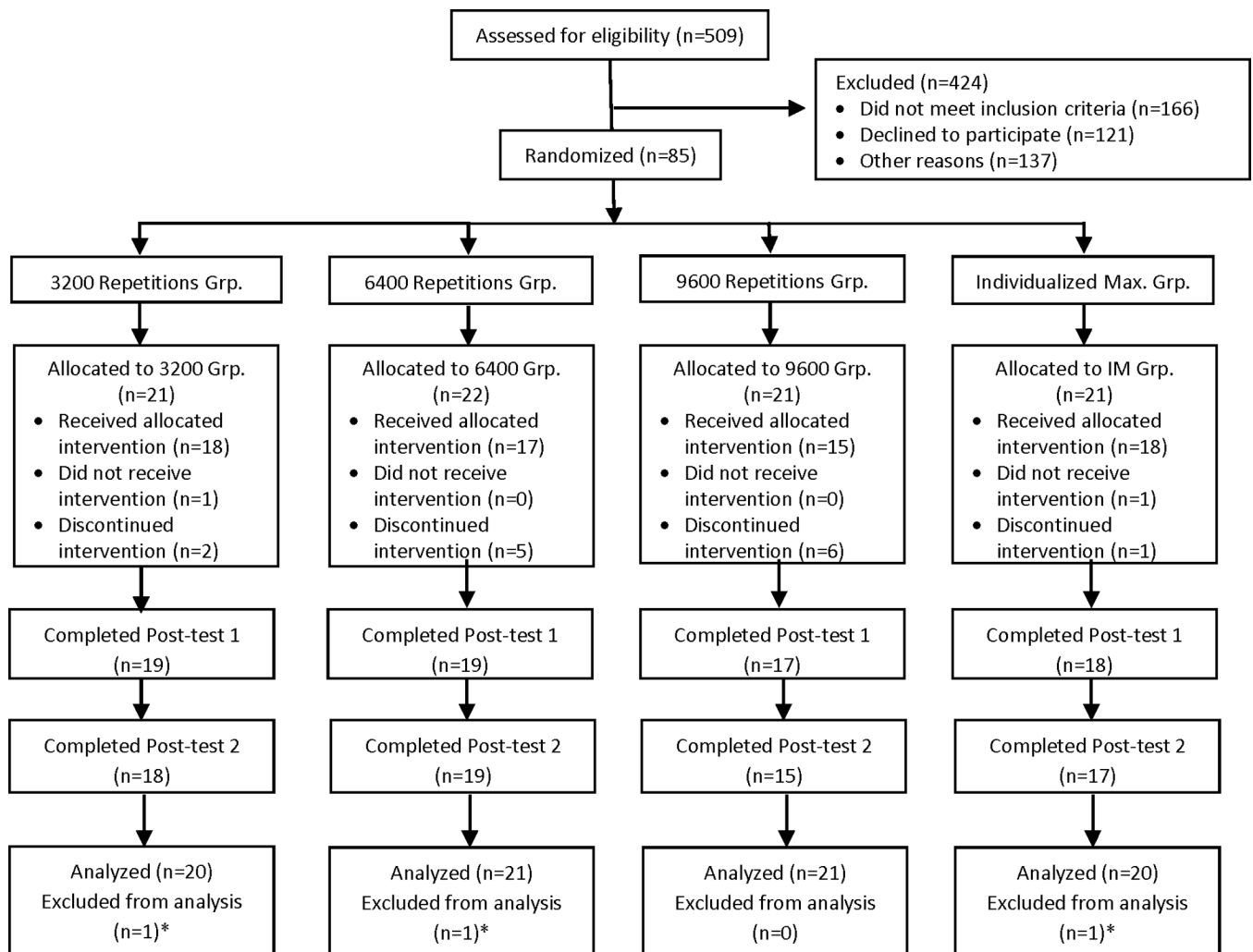


Figure 1. CONSORT diagram. *The analytic approach allowed inclusion of all participants with the first (baseline) and second (beginning of treatment week 2) assessments.

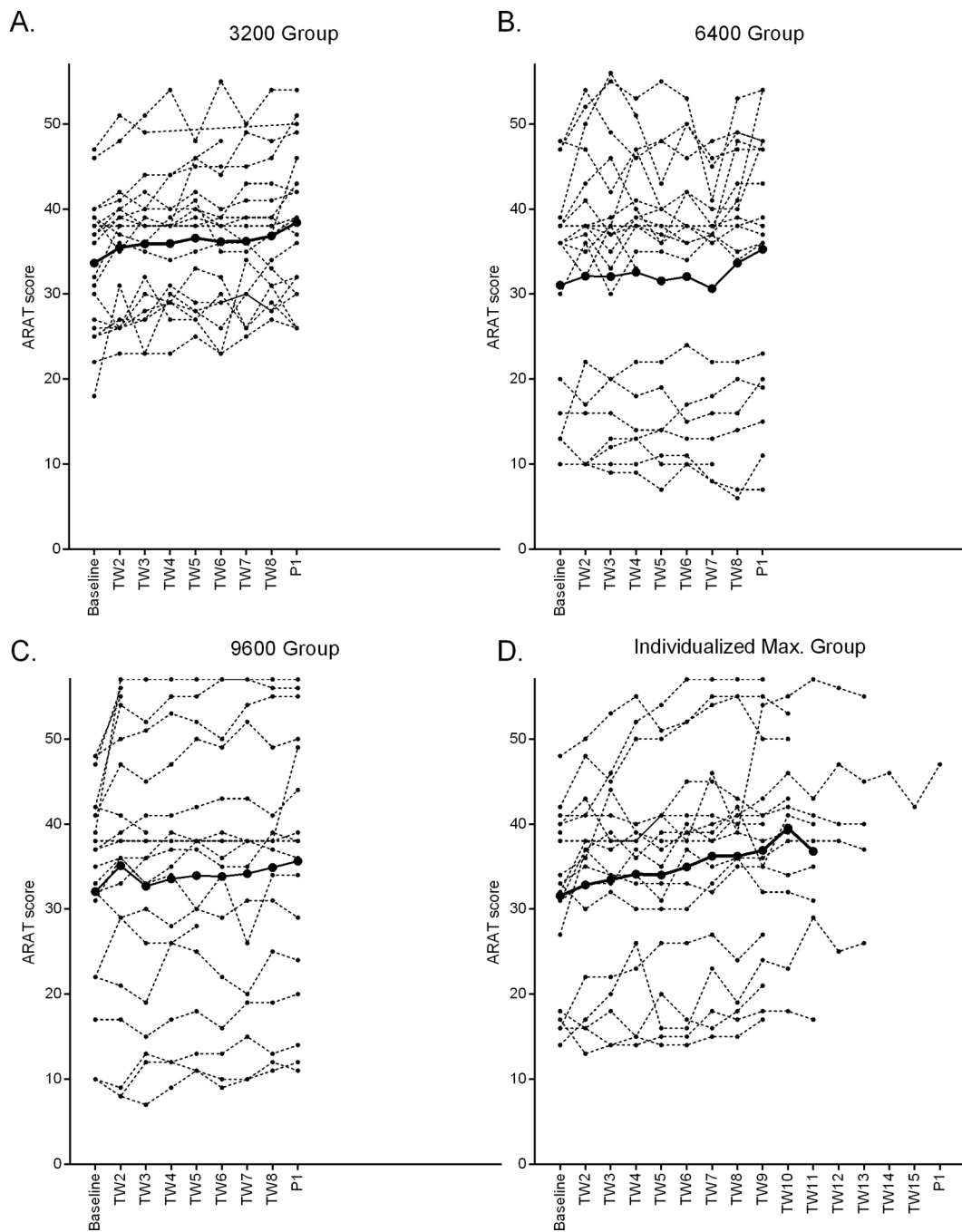


Figure 2.

Individual (thin, dashed lines) and average (thick, solid lines) trajectories of upper limb functional capacity (Action Research Arm Test, ARAT). TW: treatment week, such that TW2 is the assessment taken at the beginning of the second treatment week. P1: post-intervention assessment. For the Individualized Maximum grp, the post-intervention assessment for most participants occurs between TW9 and TW15.

Table 1

Demographic and clinical characteristics of the sample. Values are means \pm SDs, median (minimum, maximum), or N

	3200 grp. (n=20)	6400 grp. (n=21)	9600 grp. (n=21)	Individualized Maximum grp. (n=20)	P- value*
Age (yrs)	59.9 \pm 12.8	62.1 \pm 8.6	60.0 \pm 8.3	60.9 \pm 13.4	0.9
Gender	7 F, 13 M	5 F, 16 M	10 F, 11 M	8 F, 12 M	0.4
Race	10 Caucasian 10 Afr. American	11 Caucasian 10 Afr. American	10 Caucasian 9 Afr. American 1 Asian 1 Multi-race	11 Caucasian 9 Afr. American	0.9
Dominant side	2 L, 18 R	5 L, 16 R	2 L, 19 R	2 L, 18 R	0.5
Affected side	9 L, 11 R	11 L, 10 R	11 L, 10 R	7 L, 13 R	0.6
Concordance, [†] dominant=affected	11	7	10	15	0.04
Time since stroke [†]	12.0 (6, 180)	13.0 (6, 221)	13.0 (6, 54)	11.5 (6, 144)	0.6
Type of stroke	14 ischemic	17 ischemic	15 ischemic	14 ischemic	0.2
Stroke location	5 cortical 5 subcortical 2 cort. & subcort. 4 post. circ. 4 unknown	7 cortical 1 subcortical 7 cort. & subcort. 0 post. circ. 6 unknown	6 cortical 5 subcortical 4 cort. & subcort. 1 post. circ. 5 unknown	6 cortical 3 subcortical 4 cort. & subcort. 2 post. circ. 5 unknown	0.5
1 st stroke	17	17	14	17	0.3
Baseline scores of potential modifiers					
ARAT	33.7 \pm 7.9	31.0 \pm 13.4	32.1 \pm 12.3	31.6 \pm 10.3	0.9
Mesulam, total errors	1 (0, 8)	2 (0, 30)	1 (0, 26)	2 (0, 60)	0.5
CES-D	10 (1, 33)	12 (0, 35)	9 (0, 52)	8 (2, 23)	0.8
Boston naming test – 15	14 (10, 15)	14 (9, 15)	14 (0, 15)	14 (0, 15)	0.1
Short Blessed	2 (0, 28)	2 (0, 15)	3 (0, 10)	2 (0, 10)	0.9
LT sensation, index finger, filament sensed	3.61 (2.83, none)	2.83 (2.83, none)	2.83 (2.83, 6.65)	3.61 (2.83, none)	0.5
Modified Ashworth, elbow	1 (0, 4)	1 (0, 3)	1 (0, 4)	1 (0, 3)	0.9

* From comparisons between groups, using Chi square tests for categorical data and parametric and non-parametric ANOVAs for continuous data.

[†] Characteristics explored post hoc as potential modifiers.

ARAT: Action Research Arm Test; LT: light touch; CES-D: Center for Epidemiologic Studies – Depression Scale;

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Summary of the task-specific intervention delivered to each group. Statistics were computed from participants who completed any portion of the intervention. Values are medians (minimum, maximum) because of skewed distributions.

	3200 grp.	6400 grp.	9600 grp.	Individualized Maximum grp.
# sessions	32 (12, 32)	32 (1, 32)	32 (7, 32)	36 (3, 58)
# of participants completing 32 sessions (or more for IM grp.)	18	17	15	18
Repetitions	3200 (1164, 3376)	6398 (203, 6506)	9582 (2095, 9609)	10808 (899, 16298)
Hours of active training	13.6 (6.3, 21.3)	20 (0.53, 28.3)	26.3 (4.5, 31.9)	32.8 (2.3, 49.2)
% scheduled time active per session	43% (53, 67)	63% (53, 88)	82% (64, 99)	91% (77, 85)
# tasks switched	7 (1, 14)	7 (0, 11)	8 (0, 11)	10 (0, 17)
# tasks upgraded	10 (2, 20)	9 (0, 15)	10 (0, 16)	10 (1, 19)
# tasks down-graded	2 (0, 7)	2 (0, 6)	2 (0, 7)	2 (0, 13)
Perceived exertion (0–10 scale)	3 (1, 7)	3 (1, 7)	4 (1, 7)	3 (1, 9)

Table 3

Primary outcome results. Model results use the 3200 group as the reference group.

Model estimates without modifiers				
	<i>Estimate ± SE</i>	<i>Estimate with respect to 3200 grp.</i>	<i>Df</i>	<i>P value</i>
<i>Intercepts (in ARAT units)</i>				
3200 grp	35.62 ± 2.83	--	75.9	< 0.0001 *
6400 grp		-3.47±3.95	75.9	0.382
9600 grp		-0.61 ± 3.95	76.0	0.878
IM grp		-3.00 ± 4.10	75.7	0.466
<i>Slopes (ARAT/week)</i>				
Time	0.40 ± 0.15	--	77.6	0.010 *
Time × 6400 grp		-0.45 ± 0.21	77.5	0.036 †
Time × 9600 grp		-0.09 ± 0.22	77.6	0.679
Time × IM grp		0.26 ± 0.20	61.4	0.209
Contributions of modifiers				
<i>Change in intercept</i>				
Functional severity: Baseline ARAT	1.07 ± 0.04		76.1	< 0.0001 *
Neglect: Mesulam	0.08 ± 0.18		73.9	0.66
Depression: CES-D	0.13 ± 0.14		73.9	0.38
Aphasia: BNT	0.58 ± 0.45		73.8	0.21
Cognition: Short Blessed	0.29 ± 0.29		72.9	0.33
Sensation: Light touch	0.52 ± 0.94		74.7	0.58
Muscle tone: Modified Ashworth	-4.89 ± 1.23		74.9	0.00015 *
Chronicity: Months post stroke	-0.04 ± 0.04		74.8	0.34
Concordance: dom. = aff.	7.04 ± 2.89		74.9	0.017 *
<i>Change in slope</i>				
Baseline ARAT × Time	-0.0005 ± 0.007		69.5	0.945
Mesulam × Time	-0.01 ± 0.01		73.6	0.26
CES-D × Time	0.001 ± 0.008		71.3	0.86
BNT × Time	-0.02 ± 0.02		62.0	0.46
Short Blessed × Time	0.02 ± 0.02		70.0	0.22
Light touch × Time	0.02 ± 0.05		61.2	0.68
Modified Ashworth × Time	-0.04 ± 0.07		65.8	0.55
Months post stroke × Time	-0.002 ± 0.002		69.1	0.30
Concordance × Time	0.10 ± 0.16		66.9	0.54

* Significantly different from zero.

† Significantly different from the 3200 group.

Table 4Secondary outcome results. Values are means \pm SEs

Rating Scales	Baseline	Post-intervention (P1)	P values
<i>SIS – ADL subscale</i>			
3200 grp	58.0 \pm 4.4	66.2 \pm 5.0	Grp = 0.42
6400 grp	61.5 \pm 4.3	69.0 \pm 5.1	Time = 0.11
9600 grp	67.8 \pm 4.4	75.3 \pm 5.4	Grp \times Time = 0.77
Individualized Max. grp	65.4 \pm 4.5	70.1 \pm 5.2	
<i>SIS – Hand function subscale</i>			
3200 grp	46.4 \pm 5.3	55.6 \pm 6.01	Grp = 0.24
6400 grp	41.7 \pm 5.2	51.7 \pm 6.2	Time = 0.06
9600 grp	56.1 \pm 5.3	62.7 \pm 6.5	Grp \times Time = 0.74
Individualized Max. grp	44.0 \pm 5.5	55.5 \pm 6.3	
<i>COPM – Performance</i>			
3200 grp	2.7 \pm 0.3	5.4 \pm 0.4	Grp = 0.43
6400 grp	3.0 \pm 0.3	5.3 \pm 0.4	Time < 0.0001
9600 grp	3.4 \pm 0.3	6.0 \pm 0.5	Grp \times Time = 0.89
Individualized Max. grp	3.0 \pm 0.3	5.6 \pm 0.5	
<i>COPM – Satisfaction</i>			
3200 grp	2.2 \pm 0.3	5.5 \pm 0.5	Grp = 0.49
6400 grp	2.5 \pm 0.3	5.4 \pm 0.5	Time < 0.0001
9600 grp	1.8 \pm 0.3	5.5 \pm 0.5	Grp \times Time = 0.76
Individualized Max. grp	2.1 \pm 0.3	5.1 \pm 0.5	
Overall perception of change (score on Likert scale)	Number* (%)	Mean ARAT change (min, max)	95% CI of mean change
A little worse, meaningful (-2)	1 (1%)	5	--
No change (0)	4 (5%)	5.8 (3, 10)	0.8 – 10.7
A little better, not meaningful (1)	2 (3%)	4 (0, 8)	--
A little better, meaningful (2)	36 (49%)	4.1 (-3, 15)	2.6 – 5.7
A lot better, meaningful (3)	30 (41%)	7.0 (-2, 18)	4.8 – 9.2

* 73 people (89% of those included in primary analysis) completed the overall rating of change at the post-intervention visit.