Journal of
Neurology, Neurosurgery
& Psychiatry

Intensive Upper Limb Neurorehabilitation in Chronic Stroke - Outcomes from the Queen Square Programme

Journal:	*Journal of Neurology, Neurosurgery, and Psychiatry*
Manuscript ID	jnnp-2018-319954.R2
Article Type:	Research paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Ward, Nick; UCL Institute of Neurology, Department of Clinical and Motor Neurosciences; The National Hospital for Neurology and Neurosurgery Brander, Fran; The National Hospital for Neurology and Neurosurgery Kelly, Kate; The National Hospital for Neurology and Neurosurgery
Keywords:	
Specialty:	

Intensive Upper Limb Neurorehabilitation in Chronic Stroke – Outcomes from the Queen Square Programme

Nick S Ward[1,2,3], Fran Brander[2,3], Kate Kelly[2,3]

[1]*Department of Clinical and Motor Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG.*

[2]*The National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG.*

[3]*UCLP Centre for Neurorehabilitation, Queen Square, London WC1N 3BG.*

Number of characters in the title (with spaces): 67

Number of words in the abstract: 249

Word count (excluding title page, abstract, references, figures, tables): 3845

Number of figures: 6

Number of tables: 4

Correspondence should be addressed to Professor N S Ward

Department of Clinical and Motor Neuroscience

UCL Institute of Neurology,

33 Queen Square, London WC1N 3BG,

Tel: +44 (0) 20 3448 8762

e-mail n.ward@ucl.ac.uk

Abstract

Objective: Persistent difficulty in using the upper limb remains a major contributor to physical disability post-stroke. There is a nihilistic view about what clinically relevant changes are possible after the early post-stroke phase. The Queen Square Upper Limb Neurorehabilitation programme delivers high quality, high dose, high intensity upper limb neurorehabilitation during a 3-week (90 hour) programme. Here, we report clinical changes made by the chronic stroke patients treated on the programme, factors that might predict responsiveness to therapy, and the relationship between changes in impairment and activity.

Methods: Upper limb impairment and activity were assessed on admission, discharge, 6 weeks and 6 months after treatment, with modified upper limb Fugl-Meyer, (FM-UL, max-54), Action Research Arm Test (ARAT, max-57) and Chedoke Arm and Hand Activity Inventory (CAHAI, max-91). Patient reported outcome measures were recorded with the Arm Activity Measure (ArmA) parts A (0-32) and B (0-52), where lower scores are better.

Results: 224 patients (median time post-stroke 18 months) completed the 6 month programme. Median scores on admission were: FM-UL=26 (IQR 16-37), ARAT=18 (IQR 7-33), CAHAI=40 (28-55), ArmA-A=8 (IQR 4.5-12), ArmA-B=38 (IQR 24-46). The median scores 6 months after the programme were: FM-UL=37 (IQR 24-48), ARAT=27 (IQR 12-45), CAHAI=52 (IQR 35-77), ArmA-A=3 (IQR 1-6.5), ArmA-B=19 (IQR 8.5-32). We found no predictors of treatment response beyond admission scores.

Conclusion: With intensive upper limb rehabilitation, chronic stroke patients can change by clinically important differences in measures of impairment and activity. Crucially, clinical gains continued during the 6 month follow up period.

Keywords: Stroke, rehabilitation, upper limb

INTRODUCTION

Stroke remains common[1] and persistent difficulty in using the upper limb is a major contributor to ongoing physical disability[2]. The general consensus remains that most spontaneous recovery of the upper limb occurs over the first three months after stroke and current levels of rehabilitation result in little improvement after that, particularly at the level of impairment[3]. Improving outcomes through higher dose (time in rehabilitation or number of repetitions) and intensity (dose per session) of rehabilitation is an attractive option[4]. However, clinical trials of higher dose upper limb rehabilitation have generally not produced the magnitude of improvement that will change clinical practice[5], whether delivered in the early[6] or chronic stages post-stroke[7–9]. A common factor in these trials is that the dose (in hours) of additional therapy remained relatively low (18-36 hours). Despite scepticism that stroke patients could tolerate much higher doses[8], one study managed to deliverer 300 hours of upper limb therapy to chronic stroke patients over 12 weeks and reported changes in measures of both impairment and activity that were far greater than those in lower dose studies[10]. Three hundred hours represents an order of magnitude higher than any dose of rehabilitation offered in previous upper limb rehabilitation trials and deserves further consideration. However, this idea is challenging because of the logistics of setting up such a trial in health care settings where the ethic of high dose, high intensity rehabilitation is not supported. In this context, it is important to report the findings of clinical services that are able to deliver higher doses than conventionally seen. The Queen Square Upper Limb (QSUL) Neurorehabilitation programme is a single centre clinical service that provides 90 hours of timetabled treatment focusing on the post-stroke upper limb in chronic (> 6 months post-stroke) stroke patients. Here we report (i) outcomes for patients admitted to this programme at the National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Trust (UCLH), including 6 month follow up data to look at whether any clinical benefits were maintained, (ii) the characteristics of the patients admitted and any predictors of response and (iii) the relationship between changes in impairment and activity.

METHODS

Participants

Patients were referred by primary care physicians for advice about ongoing management of the paretic upper limb following stroke. There were no criteria around time since stroke. Clinical criteria warranting admission to the programme were broad but were focused on whether we felt able to help patients achieve their goals for the affected upper limb. Barriers to admission included factors precluding the ability to work in l upper limb tasks; (i) absent movements, throughout the limb; (ii) a painful shoulder limiting an active forward reach (mostly due to adhesive capsulitis); (iii) severe spasticity or non-neural loss of range; (iv) unstable medical conditions. These patients were not offered admission to the treatment programme, but were provided with appropriate advice or referred for other treatment approaches. Re-referral to the programme was then considered (for example after successful management of adhesive capsulitis of the shoulder). Features that were more likely to result in admission to the programme included at least some ability to flex the shoulder and at least visible flickers of movement in finger and/or wrist extensors, although this was not always the case as reflected in baseline upper limb scores. Patients were admitted to the programme as day attenders, either from home or from the UCLH dedicated patient accommodation and as such were either self-caring, or self-caring with the support of one other person.

Interventions

Initial assessment consisted of analysis of both movement and performance in activities of daily living. Subsequent treatment was aimed at reducing impairment and promoting re-education of motor control within activities of daily living. Individualised meaningful tasks were practiced repeatedly in order to facilitate task mastery with a focus on quality of movement. This was achieved through (i) adaptation of the task, e.g. decomposing tasks into individual components to be practiced; (ii) adaptation of the environment, e.g. fabrication of functional splints and adaptation of tools such as cutlery or screwdrivers, to enable integration of the affected hand in meaningful activities; (iii) assistance, e.g. de-weighting the arm to allow strengthening and training of movement quality and control through increased range; (iv) independent task practice. Coaching was considered a key component of the programme, used throughout to embed new skills and knowledge into individual daily routines. Consequently, individuals increase participation and confidence in their desired goals, enhancing self-efficacy and motivation to sustain behavioural change beyond the end of the active treatment period.

This overall approach was achieved through two daily sessions each of physiotherapy and occupational therapy, supplemented with tailored, individualized interventions, including repetitive practice with a rehabilitation assistant or robotic device, sensory retraining, use of dynamic and functional orthoses, neuromuscular electrical stimulation, and group work. Furthermore, patients were encouraged to work on cardiovascular fitness during the programme. A 6 hour timetable was implemented 5 days a week for 3 weeks (total therapy time, 90 hours). Motor tasks could be described as passive or active, assisted or unassisted, functional or non-functional. Over the three weeks, the aim was to increase the time spent on active, unsupported functional tasks, depending on a patient's level of impairment and progress. The programme was staffed with a 1:1 staff to patient ratio (2 physiotherapists, 2 occupational therapists, 2 rehabilitation assistants for 6 patients).

Quantitative assessment

All patients had the following baseline scores measured on admission (table 1): modified Rankin scale (mRS), Barthel Index (BI), Neurological Fatigue Index (NFI), Hospital Anxiety and Depression Scale (HADS), and sensation (as indexed by light touch on palm as described in Fugl-Meyer assessment of the upper extremity).

The affected upper limb was assessed on admission (T1), discharge (T2), 6 weeks (T3) and 6 months (T4) post-discharge, using the following measures: Fugl-Meyer (upper limb) (FM-UL), Action Research Arm Test (ARAT), Chedoke Arm and Hand Activity Inventory (CAHAI-13), Arm Activity Measure

The FM-UL is a stroke-specific, performance-based impairment index (where impairment refers to loss of body structure and function) with good validity and intra- and inter-rater reliability[11,12]. Here we have used a modified version of the FM-UL, excluding measures of co-ordination and reflexes based on hierarchical properties of the scale[13], and as these do not relate to the upper limb synergies of interest[14]. The minimum clinically important difference (MCID) has been reported as 5.25 points [15].

The ARAT assesses a patient's ability to handle objects differing in size, weight and shape and therefore can be considered to be an arm-specific measure of activity limitation[16]. The MCID has been suggested as 5.7 points[17].

The CAHAI-13 is a validated, upper-limb measure that uses a 13-point quantitative scale in order to assess recovery of the arm and hand in performing activities of daily living after a stroke[18]. It is a performance test using day to day items. It is not designed to measure the patient's ability to complete the task using only their unaffected hand, but rather to encourage bilateral upper limb use. No MCID has been reported for CAHAI-13 (although the minimum detectable change is reported as 6.2[18]).

The Arm Activity Measure (ArmA) is a patient reported outcome score[19] with two components. The ArmA-A asks patients about their ability to care for their affected arm either themselves with their unaffected arm or by a carer or a combination of both of these. ArmA-B asks patients about how easy or hard it is to use the affected arm in activities of daily living. For ArmA-A and ArmA-B, note that lower scores are better. No MCID has been reported for ArmA-A or ArmA-B.

Analysis plan

Our primary aim was to measure changes in upper limb deficit over time were examined using the Friedman test for all outcome scores. Post-hoc analysis to test differences between individual time points was carried out with the Wilcoxon signed-ranks test. As a secondary aim, we were then interested in whether baseline characteristics correlated with final (T4) outcomes. We initially examined for correlations between individual baseline characteristics and outcomes using Spearman's rank correlation. We also tested whether median T4 outcomes were different depending on gender or side of lesion using Wilcoxon rank sum test. Time since stroke was not considered as a covariate as once in the chronic phase, there is no indication that time since stroke has a linear effect on outcome. Lastly we performed multiple linear regression to look for predictors of either absolute scores at T4 or the change in scores T1 to T4. Lastly, we also examined the relationship between the changes in outcome scores from admission to 6 month follow up using Spearman's rank correlation.

RESULTS

Baseline characteristics

Between January 2015 and December 2017, 268 patients were admitted to the QSUL programme (representing 46% of patients assessed in out-patient clinic). 40.4% of patients were admitted from home, whilst 59.6% stayed in the UCLH dedicated patient accommodation. Note, only 16 patients were admitted to the programme prior to 6 months post-stroke, and only 1 prior to 3 months, reflecting overall referral patterns to the programme.

Of those, 30 were non-stroke (12 traumatic brain injury, 8 brain tumour, 3 peripheral neurological conditions, 3 spinal cord injury, 4 inflammatory central nervous system conditions) and are excluded from this analysis. A total of 238 stroke patients were admitted, of whom 224 completed follow up assessments at 6 weeks and 6 months after discharge from the programme. Of the 14 incomplete follow ups, 5 patients considered it too far to travel for follow up, 4 patients suffered intercurrent illness (recurrent stroke, fractured hip, complications of surgery, seizures), and 5 could not be contacted. Differences in median values between complete and missing data groups was tested with Wilcoxon Signed Rank test. Where appropriate, differences in proportions between complete and missing data groups were tested with Chi-squared test. There were no differences in baseline characteristics between patients who completed follow up and those who were lost to follow up (table 1). Because (i) there was no systematic difference between subjects with complete data and those with missing data, and (ii) only outcome data (not explanatory variables, which were collected at T1) were missing, then missing data were dealt with by performing a complete case analysis (i.e. only using subjects where complete data were available).

Changes in upper limb deficit

The median scores for the affected upper limb at admission (T1), discharge (T2), 6 weeks (T3) and 6 months post-discharge (T4) are shown in table 2. The Friedman test was used to demonstrate a significant effect of time since admission for all outcome scores; FM-UL (χ^2 (3) = 431.8, p<0.0001); ARAT (χ^2 (3) = 383.2, p<0.0001); CAHAI (χ^2 (3) = 371.6, p<0.0001); ArmA-A (χ^2 (3) = 238.4, p<0.0001); ArmA-B (χ^2 (3) = 305.6, p<0.0001). Post hoc analysis with Wilcoxon signed-ranks test was conducted to test the hypothesis that the difference between the paired values (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4) comes from a distribution whose median is zero (i.e. testing whether the change in score is significantly greater than zero). A Bonferroni correction was applied to the significance level for the 6 comparisons made for each of the 5 different outcome scores ($p<0.05/30 = 0.0017$). For each outcome score, there was a significant difference between each pair of time points compared, except T2-T3 for the ArmA-A (table S1). Individual scores at each time point are shown in figures 1 and 2. Individual changes in scores in comparison to admission are shown in table 3 and figures 1 and 2. These results demonstrate improvement during the 3 week programme, but also continued improvement after discharge.

By 6 months after the programme, 68.3% of patients had achieved greater than the MCID of 5.25 points on the FM-UL. For ARAT, this figure was 61.6%. No published MCID is published for the other scores. MCID is often quoted as 10% of the maximum score. If this were the case, then by 6 months, MCID would have been reached by 59.4% of patients for CAHAI, 53.8% for ArmA-A and 72.3% for ArmA-B.

Predictors of outcome

Firstly, we were interested to see whether any baseline characteristics (T1) correlated with final outcome at T4. Spearman's rank correlation was used to show that admission values at T1 for FM-UL, ARAT, CAHAI, ArmA-A, ArmA-B all correlated with T4 values for FM-UL, ARAT, CAHAI, ArmA-A, ArmA-B. In addition, admission BI correlated with T4 CAHAI; admission mRS correlated with T4 FM-UL, ARAT and CAHAI; admission sensation correlated with T4 CAHAI. Age, NFI and HADS at T1 did not correlate with any scores at T4. All p-values were corrected for multiple comparisons by Bonferroni correction (Table 4). Figure 3 helps to visualise the influence of baseline admission scores on final score at 6 month follow up. Furthermore, a Wilcoxon ranked sum test showed that T4 outcomes did not differ by gender or affected hemisphere. There was a trend for patients with non-dominant affected limbs to achieve better (lower) ArmA-B scores in comparison to those with dominant affected limbs (median 17 for non-dominant vs -27 for dominant, $p = 0.027$, uncorrected). There were no differences in T4 scores between patients with dominant or non-dominant affected limbs for any of the other outcome scores.

Secondly, we were interested to see whether any baseline characteristics (T1) correlated with magnitude of change from admission to 6 months follow up (T1 to T4). Spearman's rank correlation was used to show that FM-UL, ARAT, CAHAI and ArmA-B at T1 correlated with change in ARAT and CAHAI (but not FM-UL) from T1 to T4. FM-UL at T1 also correlated with change in ArmA-B from T1 to T4. ArmA-A at T1 correlated with change in both CAHAI and ArmA-A from T1 to T4. ArmA-B at T1 correlated with change in both ArmA-A and ArmA-B from T1 to T4. Age, NFI, HADS, BI and mRS at T1 did not correlate with any scores at T4. All p-values were corrected for multiple comparisons by Bonferroni correction (table 5). We also performed a Wilcoxon ranked sum test to show that change in scores from T1 to T4 were not different depending on gender or affected hemisphere. There was a trend for patients with non-dominant affected limbs to improve on the ArmA-B between T1 and T4 to a greater degree than those with dominant affected limbs (median changes of -17.5 for non-dominant vs -12 for dominant, p = 0.017, uncorrected). There were no differences in change scores between patients with dominant or non-dominant affected limbs for any of the other outcome scores.

We then performed multiple linear regression to look for predictors of either absolute scores at 6 months after treatment (T4) or the change in scores from admission to 6 months after treatment (T1 to T4) using age, BI, mRS, NFI, HADS and sensation on admission together with the matching initial score on admission (e.g. FM1 if attempting to predict FM4 or change from FM1 to FM4). Models predicting T4 scores for FM-UL, ARAT, CAHAI, ArmA-A and ArmA-B were all highly significant (p<0.0001). Models predicting change from T1 to T4 for ARAT, CAHAI, ArmA-A and ArmA-B (but not FM-UL) were also highly significant (p<0.0001). However, the only factor that significantly contributed to these models was the matched initial score on admission (e.g. ARAT at T1 predicting ARAT at T4), but not age, BI, mRS, NFI, HADS or sensation on admission. As such, this approach did not contribute further to the correlation results shown in table 4 and 5, with regression models not able to account for a higher proportion of variance in scores at T4 or change in scores from T1 to T4 than the admission score (see also Figure 3).

Relationship between outcome measures

Each outcome score assesses a different aspect of recovery. Using Spearman's rank correlation, we examined the relationship between the changes in outcome scores from admission to 6 month follow up. A Bonferroni correction was applied to the significance level for the 10 comparisons made between the 5 different outcome scores (p<0.05/10 = 0.005) (Table 6 and Figure 4). Changes in FM-UL, ARAT and CAHAI correlated weakly with one another. The ArmA-B change correlated with changes in both ArmA-A and ARAT.

DISCUSSION

We report the results from a single centre clinical service delivering high dose and intensity of upper limb focused neurorehabilitation for people with stroke. Almost all of these patients were in the chronic (>6 months after stroke) phase. The key messages from our results are (i) patients were able to complete the 90 hours of the programme, despite exhibiting a range of impairments and fatigue levels, (ii) large clinical improvements in upper limb impairment (FM-UL) and activity (ARAT, CAHAI, ArmA) were observed, and (iii) these changes were maintained, or even improved upon, 6 months after treatment.

The QSUL programme is based on the hypothesis that a high dose and intensity of upper limb neurorehabilitation can lead to large clinically meaningful improvements in chronic stroke patients. It is worth comparing the magnitude of the changes with previous reported clinical trials of upper limb rehabilitation. Most obviously, the study by McCabe et al[10] investigated the effects of 300 hours of upper limb rehabilitation in 48 chronic stroke patients with similar characteristics to those reported here (baseline FM-UL 24, compared to 26 in our patients) and reported an increase of 8-11 points on the FM-UL at the end of 300 hours of upper limb treatment (although later follow up results have not been reported). In the QSUL programme, the median FM-UL (modified) improvement was 6 points after the 90-hour programme, increasing to 9 points by 6 months follow up, achieving a similar overall magnitude of change in comparison to McCabe et al. In addition to reductions in impairment (FM-UL), our patients also improved at the level of activity (ARAT and CAHAI) and participation (using the patient reported outcome measure, ArMA). Importantly, we report that these group level effects continued to improve at 6 weeks and then 6 months after treatment had stopped, (figure 1 and 2).

Other than the study by McCabe et al, the general view has been that large changes, especially in impairment, have not been seen for behavioural interventions of the post-stroke upper limb[3]. Direct comparison with other studies is difficult because some may have been conducted in less severely affected patients, the dose (measured in time) of the intervention may have been lower, and the numbers of patients enrolled has been much lower than we report here. Whilst this report of our single-centre clinical outcomes is not a replication of the McCabe study, it suggests that the next wave of upper limb rehabilitation trials must investigate much higher doses of the treatment than is currently attempted. Whilst we will hear the refrain that it is not possible to deliver such high doses in current health care settings (the pragmatic view), it is in fact precisely the role of clinical research to challenge what we currently do in order to reshape and improve our clinical services to make them better (the aspirational view). Those involved in stroke medicine will recognise that this is exactly the effect that thrombolysis had on acute management of stroke, leading to the high quality hyperacute stroke services we see today. Post-stroke neurorehabilitation must adopt the same aspirational approach[20].

Whilst the appropriate dose of upper limb rehabilitation is debated, the content of the intervention receives less attention. Our approach was based on analysis of movement and performance in activities of daily living, reduction of impairment and re-education of quality and control of movement within activities of daily living, all performed in an iterative and non-linear fashion. Education, self-efficacy and goal setting were integral components of the programme, and we can speculate that the education and focus on self-efficacy contributed to the continued increase in scores after treatment stopped. The therapeutic approach taken towards specific impairments was the same across patients, but approaches at the activity and participation level will necessarily vary as they are tailored to an individual's specific meaningful goals. We acknowledge that an immediate challenge for us and for the field in general is to be able to define what the 'active ingredients' of neurorehabilitation are, so that these elements can be further tested. Our approach appears to have similarities with that taken by McCabe et al, with an emphasis on the quality of the movement as well as the final goal of the movement. The latter is often the sole focus of task-specific repetition, which has been used in many upper limb rehabilitation studies to date. The evidence for repetitive task-specific training is of low to moderate quality[21] and the effects have been disappointingly small even when the number of repetitions has been greatly increased[8]. It is difficult to tell whether this lack of effect is specific to task-specific training, because in general the dose (in terms of hours of active training) has been quite low, whether delivered by a therapist[21] or robotic device[22]. There have been suggestions that chronic stroke patients might not tolerate more training than has been offered in these trials[8], but the study of McCabe et al[10] and now our own clinical experience undermine this view.

Although this result is exciting, we must acknowledge that it is a single centre clinical service with no randomisation and no control group. Although there are some reports of late improvement up to 12 months after lacunar stroke[23], we expect that chronic stroke patients who are often several years after their stroke, will not change across this range of outcome scores without treatment. We suggest that it is unlikely that this magnitude of change would be seen without therapeutic intervention, but this assertion requires empirical confirmation. Our work however, provides strong justification for undertaking such high dose clinical trials. Clinic data on which to base these future studies is sorely lacking and this has led to the proliferation of low dose pragmatic rehabilitation studies. However, there is still much to be learned from published data arising from clinical services in neurorehabilitation[24], making up for lack of clinical trial rigour with large sample sizes that are rarely achieved in clinical trials. Data from clinical services therefore have the potential to refine what experimental questions are addressed in future randomised clinical trials. Another limitation is the unblinded nature of the assessments, introducing possible bias into the results. All assessors are also therapists on the programme, although we ensure that assessments are carried out by therapists from another team who are not treating the patient. Fidelity of outcome scores is always an important issue in clinical trials. In our programme, we have instituted regular in-service training for FM-UL, ARAT and CAHAI to ensure high levels of reproducibility between assessors.

A common complaint regarding single centre clinical services is that they select only patients most likely to succeed. It is important to acknowledge that the QSUL programme does not have strict inclusion/exclusion criteria. Although our patients are generally younger than the average stroke patient and have high levels of independence they have a wide range of upper limb impairments. It can be seen from figures 1 and 2 that the range of severity of the upper limb on the QSUL programme was very broad, quite unlike a clinical trial, although the median scores suggest that patients were weighted towards the moderate-severe end of upper limb impairment and activity limitation. We would argue that the QSUL programme is aspirational in its goals (high quality, high dose, high intensity treatment) and pragmatic in its application (wide range of patients admitted onto the programme). It's interesting to consider the potential impact of initial upper limb severity on our outcomes, but in fact, although initial score correlates well with the final score at 6 months (table 4), the change in score correlates only weakly with initial score (table 5, Figure 3). There are two ways this 'mixed-bag' approach may have potentially reduced the overall magnitude of the clinical effects we observed. Firstly, there may be a ceiling effect from patients with relatively high scores on admission. Secondly, many of our patients were quite severely impaired. Indeed, a quarter of our patients had an admission ARAT below 7, which suggests they have great difficulty in using their hand in any kind of activity of daily living. The treatment goals in these type of patients might initially be very different to the moderate to mild type patients, with benefits less likely to be reflected in changes in FM-UL and ARAT compared to the CAHAI or ArMA for example. However, the wide range of patients treated in the programme means that no clear patterns, in relation to differential effects on various scores, could be seen (Table 6, Figure 4).

Some might argue that the wide range of responses, including those patients who changed very little, points to the need for stratification and selection of those patients most likely to respond, and this is certainly a reasonable view. However, the first point to make is that our results show that with large enough numbers and a high enough dose of the intervention, it is possible to demonstrate large group-level effects in relatively unselected populations of chronic stroke patients. Indeed, if the former (stratification) approach is to be adopted, then these kind of large data sets will be required to determine what the characteristics of the 'responders' are likely to be. The second point to make is that we found no strong clinical indicators, amongst our measured explanatory variables, of which patients were most likely to change, including sensory loss in the hand, fatigue[25] and depression (tables 4 and 5). It might be that neurophysiological (presence or absence of motor evoked potentials)[26], neuroimaging (assessment of corticospinal tract, whole brain damage or brain function)[27–30] or cognitive (sustained attention, memory, motivation)[31,32] measures are required for accurate stratification based on likely response, and this can certainly be tested in future.

In summary, here we present the results from a single centre clinical service dedicated to post-stroke upper limb neurorehabilitation in chronic stroke patients. Despite treating patients with a wide range of impairment, we were able to show a large group-level magnitude of change at both the impairment and activity level. Our experience suggests that much higher doses and intensity of upper limb neurorehabilitation can be delivered and our results should inform future clinical trial design.

ACKNOWLEDGEMENTS

Thanks to all the physiotherapists and occupational therapists at The National Hospital for Neurology and Neurosurgery, Queen Square, who have treated patients on this programme. Thanks to UCLH Charities, Friends of UCLH and The National Brain Appeal for funding to purchase equipment used in this programme.

REFERENCES

1. Licher S, Darweesh SKL, Wolters FJ, et al. Lifetime risk of common neurological diseases in the elderly population. *J Neurol Neurosurg Psychiatry.* 2018 Oct 2; doi: 10.1136/jnnp-2018-318650. [Epub ahead of print].
2. Broeks JG, Lankhorst GJ, Rumping K, et al. The long-term outcome of arm function after stroke: results of a follow-up study. Disabil Rehabil. 1999;21(8):357–64.
3. Krakauer JW, Carmichael ST, Corbett D, et al. Getting neurorehabilitation right: what can be learned from animal models? *Neurorehabil Neural Repair.* 2012;26(8):923–31.
4. Lohse KR, Lang CE, Boyd LA. Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke.* 2014;45(7):2053–8.
5. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. *Lancet Neurol.* 2009;8(8):741–54.
6. Winstein CJ, Wolf SL, Dromerick AW, et al. Effect of a Task-Oriented Rehabilitation Program on Upper Extremity Recovery Following Motor Stroke: The ICARE Randomized Clinical Trial. *JAMA*. 2016;315(6):571–81.
7. Lo AC, Guarino PD, Richards LG, et al. Robot-assisted therapy for long-term upper-limb impairment after stroke. *N Engl J Med*. 2010;362(19):1772–83.
8. Lang CE, Strube MJ, Bland MD, et al. Dose-response of task-specific upper limb training in people at least 6 months post stroke: A Phase II, single-blind, randomized, controlled trial. *Ann Neurol*. 2016;80(3):342-54.
9. Klamroth-Marganska V, Blanco J, Campen K, et al. Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomised trial. *Lancet Neurol.* 2014;13(2):159–66.
10. McCabe J, Monkiewicz M, Holcomb J, et al. Comparison of robotics, functional electrical stimulation, and motor learning methods for treatment of persistent upper extremity dysfunction after stroke: a randomized controlled trial. *Arch Phys Med Rehabil.* 2015;96(6):981–90.
11. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Phys Ther.* 1983;63(10):1606–10.
12. Fugl-Meyer AR, Jääskö L, Leyman I, et al. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7(1):13–31.
13. Crow JL, Harmeling-van der Wel BC. Hierarchical properties of the motor function sections of the Fugl-Meyer assessment scale for people after stroke: a retrospective study. *Phys Ther*. 2008;88(12):1554–67.

14. Hsieh Y-W, Hsueh I-P, Chou Y-T, et al. Development and validation of a short form of the Fugl-Meyer motor scale in patients with stroke. *Stroke*. 2007;38(11):3052–4.
15. Page, S. J., Fulk, G. D. & Boyne, P. Clinically important differences for the upper-extremity Fugl-Meyer Scale in people with minimal to moderate impairment due to chronic stroke. *Phys Ther* **92**, 791–798 (2012).
16. Platz T, Pinkowski C, van Wijck F, et al. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clin Rehabil*. 2005;19(4):404–11.
17. Van der Lee, J. H. *et al.* The intra- and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke. *Arch Phys Med Rehabil* **82**, 14–19 (2001).
18. Barreca SR, Stratford PW, Lambert CL, et al. Test-retest reliability, validity, and sensitivity of the Chedoke arm and hand activity inventory: a new measure of upper-limb function for survivors of stroke. *Arch Phys Med Rehabil.* 2005;86(8):1616–22.
19. Ashford S, Turner-Stokes L, Siegert R, et al. Initial psychometric evaluation of the Arm Activity Measure (ArmA): a measure of activity in the hemiparetic arm. *Clin Rehabil.* 2013;27(8):728–40.
20. Bernhardt J, Borschmann K, Boyd L, et al. Moving rehabilitation research forward: Developing consensus statements for rehabilitation and recovery research. *Int J Stroke.* 2016;11(4):454–8.
21. French B, Thomas LH, Coupe J, et al. Repetitive task training for improving functional ability after stroke. *Cochrane Database Syst Rev.* 2016 14;11:CD006073.
22. Veerbeek JM, Langbroek-Amersfoort AC, van Wegen EEH, et al. Effects of Robot-Assisted Therapy for the Upper Limb After Stroke. *Neurorehabil Neural Repair.* 2017;31(2):107–21.
23. Ganesh A, Gutnikov SA, Rothwell PM, Oxford Vascular Study. Late functional improvement after lacunar stroke: a population-based study. *J Neurol Neurosurg Psychiatry*. 2018 Jul 21; doi: 10.1136/jnnp-2018-318434. [Epub ahead of print]
24. Freeman JA, Hobart JC, Playford ED, et al. Evaluating neurorehabilitation: lessons from routine data collection. *J Neurol Neurosurg Psychiatry.* 2005;76(5):723–8.
25. De Doncker W, Dantzer R, Ormstad H, et al. Mechanisms of poststroke fatigue. *J Neurol Neurosurg Psychiatry.* 2018;89(3):287-293.
26. Agarwal S, Koch G, Hillis AE, et al. Interrogating cortical function with transcranial magnetic stimulation: insights from neurodegenerative disease and stroke. *J Neurol Neurosurg Psychiatry.* 2018 Jun 4; doi: 10.1136/jnnp-2017-317371. [Epub ahead of print].
27. Aguilar OM, Kerry SJ, Ong Y-H, et al. Lesion-site-dependent responses to therapy after aphasic stroke. *J Neurol Neurosurg Psychiatry.* 2018 Apr 17; doi: 10.1136/jnnp-2017-317446. [Epub ahead of print].

28. Park C-H, Kou N, Ward NS. The contribution of lesion location to upper limb deficit after stroke. *J Neurol Neurosurg Psychiatry.* 2016;87(12):1283-1286.

29. Rondina JM, Park C-H, Ward NS. Brain regions important for recovery after severe post-stroke upper limb paresis. *J Neurol Neurosurg Psychiatry.* 2017;88(9):737-743.

30. Burke Quinlan E, Dodakian L, See J, et al. Neural function, injury, and stroke subtype predict treatment gains after stroke. *Ann Neurol.* 2015;77(1):132–45.

31. Robertson IH, Ridgeway V, Greenfield E, et al. Motor recovery after stroke depends on intact sustained attention: a 2-year follow-up study. *Neuropsychology.* 1997;11(2):290–5.

32. Quattrocchi G, Greenwood R, Rothwell JC, et al. Reward and punishment enhance motor adaptation in stroke. *J Neurol Neurosurg Psychiatry.* 2017;88(9):730-736.

FIGURE LEGENDS

Figure 1. Outcome scores for all patients. Each data point represents a single patient. Top row shows individual scores at admission, discharge, 6 weeks and 6 months after discharge. Bottom row shows the individual difference scores for admission to discharge, admission to 6 weeks post-discharge, and admission to 6 months post-discharge. Scores are shown for modified Fugl-Meyer (upper limb), Action Research Arm Test and Chedoke Arm and Hand Activity Inventory (CAHAI). Median (solid line) and upper and lower quartiles (dotted lines) are shown.

Figure 2. Self-reported outcome scores for all patients. Lower scores are better. Each data point represents a single patient. Top row shows individual scores at admission, discharge, 6 weeks and 6 months after discharge. Bottom row shows the individual difference scores for admission to discharge, admission to 6 weeks post-discharge, and admission to 6 months post-discharge. Scores are shown for the Arm Activity Measure part A and B (ArmA-A and ArmA-B). Median (solid line) and upper and lower quartiles (dotted lines) are shown.

Figure 3. Admission score plotted against final scores at 6 months post-discharge for Fugl-Meyer (FM-UL), Action Research Arm Test (ARAT), Chedoke Arm and Hand Activity Inventory (CAHAI), Arm Activity Assessment A & B (ArmA-A and ArmA-B). The black line represents the line of 'no change'. Patients above this line for FM-UL, ARAT and CAHAI have improved. Patients below this line for ArmA-A and ArmA-B have improved.

Figure 4. Change in scores from admission to 6 months post-discharge for Fugl-Meyer (FM-UL), Action Research Arm Test (ARAT), Chedoke Arm and Hand Activity Inventory (CAHAI), plotted against each other.

Table 1. Demographic and clinical characteristics on admission

Variable	**Patients with complete data (n = 224)**	**Patients with missing data (n=14)**	**Difference between median values**
Age, Median (IQR), yrs	52 (38-60)	54.5 (43-61)	P = 0.29
Gender,			
Male, n (%)	138 (61.6%)	7 (50%)	P = 0.30
Female, n (%)	86 (38.4%)	7 (50%)	
Affected Limb*			
Left, n (%)	117 (51.3%)	8(57.1%)	P = 0.39
Right, n (%)	107 (48.7%)	6 (42.9%)	
Non-dominant	110 (49.1%)	7 (50.0%)	P = 0.87
Dominant	114 (50.9%)	7 (50.0%)	
Median time since Stroke Months (IQR)	18 (12-51)	16 (12-22)	P = 0.29
Modified Rankin Scale, n (%)			
Median (IQR)	2 (2 - 3)	2 (2 – 2.75)	P = 0.32
No Significant Disability	20 (9.9%)	2 (14.3%)	
Slight Disability	106 (47.4%)	8 (57.1%)	
Moderate Disability	95 (42.4%)	4 (28.6%)	
Moderately Severe	3 (1.3)	0 (0%)	
The Barthel Index, Median (IQR), max = 20	19 (17-19)	19 (18-20)	P = 0.12
The Neurological Fatigue Index, Median (IQR), max = 62	35 (25 - 42)	33 (23-38)	P = 0.34
Hospital Anxiety and Depression Scale, Median (IQR), max = 34	12 (8-16)	13 (8.5-18)	P = 0.64
Sensory Loss, n, %			
0 = Severe	13 (5.8%)	2 (14.3%)	
1 = Mild	111 (49.6%)	7 (50%)	P = 0.36
2 = Normal	100 (44.6%)	5 (35.7%)	
Modified Fugl-Meyer (Upper Limb), Median (IQR) max = 54	26 (16-37)	27.5 (19-34)	P = 0.92
Action Research Arm Test, Median (IQR), Max = 57	18 (7-33)	15 (9.5-23)	P = 0.92
Chedoke Arm and Hand Activity Inventory, Median (IQR), Max = 91	40 (28-55)	33.5 (27-40)	P = 0.37
Arm Activity Measure A, Median (IQR), Max = 28	8 (5-12)‡	8 (6-12)‡‡	P = 0.55
Arm Activity Measure B, Median (IQR), Max = 52	38 (24-46)‡	36 (29-41)‡‡	P = 0.65

Difference in medians tested with Wilcoxon Signed Rank test. Difference in proportions was tested with Chi-squared test. Performed with ‡ 195 patients out of 224 had ArMA-A and ArMA-B measured on admission. ‡‡11 patients out of 14 had ArMA-A and ArMA-B measured on admission. *202 (90.2%) patients were right handed, 22 (9.8%) patients were left handed.

Table 2. Upper limb scores before and after admission

	Admission	Discharge	6 weeks	6 months
FM-UL	26 (16 - 37)	34 (23 - 44)	35 (23 - 46)	37 (24 - 48)
ARAT	18 (7 - 33)	25 (10 - 42)	26 (11 - 44)	27 (11 - 45)
CAHAI	40 (28 - 55)	49 (36 - 70)	50 (35 - 73)	52 (35 - 77)
ArmA-A	8 (4.5 - 12)	5 (2 - 8)	4 (2 -7)	3 (1 - 6.5)
ArmA-B	38 (24 - 46)	26 (13 - 37)	21 (11 - 35)	19 (8.5 - 32)

All scores given as median (IQR)

Table 3. Changes in upper limb scores

	Adm - Dis	Adm - 6 weeks	Adm - 6 months
FM-UL	6 (3 to 9)	8 (4 to 11)	9 (4 to 12)
ARAT	6 (2 to 8)	7 (2 to 10)	8 (3 to 11)
CAHAI	9 (5 to 13.5)	10 (5 to 16)	12 (6 to 17)
ArmA-A	-2 (-5 to -1)	-3 (-6 to -1)	-4 (-8 to -1)
ArmA-B	-7 (-14 to -3)	-10 (-17 to -4)	-11 (-19 to -5)

All scores given as median (IQR)

Table 4. Correlation between baseline scores and 6 month outcomes

	FM-UL (T4)	ARAT (T4)	CAHAI (T4)	ArmA-A (T4)	ArmA-B (T4)
Age	n.s.	n.s.	n.s.	n.s.	n.s.
Barthel Index	n.s.	n.s.	0.28	n.s.	n.s.
modified Rankin Scale	-0.29	-0.25	-0.32	n.s.	n.s.
Neurological Fatigue Index	n.s.	n.s.	n.s.	n.s.	n.s.
Hospital Anxiety and Depression Scale	n.s.	n.s.	n.s.	n.s.	n.s.
Sensation	n.s.	n.s.	0.22	n.s.	n.s.
FM-UL (T1)	0.92	0.84	0.83	-0.42	-0.63
ARAT (T1)	0.87	0.95	0.91	-0.48	-0.71
CAHAI (T1)	0.84	0.88	0.92	-0.46	-0.71
ArmA-A (T1)	-0.39	-0.44	-0.44	0.56	0.41
ArmA-B (T1)	-0.53	-0.58	-0.61	0.40	0.71

Correlation coefficient (rho) given only when significant at p<0.05 corrected for 55 multiple comparisons (p< 0.0009). n.s. = not significant

Table 5. Correlation between baseline scores and change scores from admission to 6 months

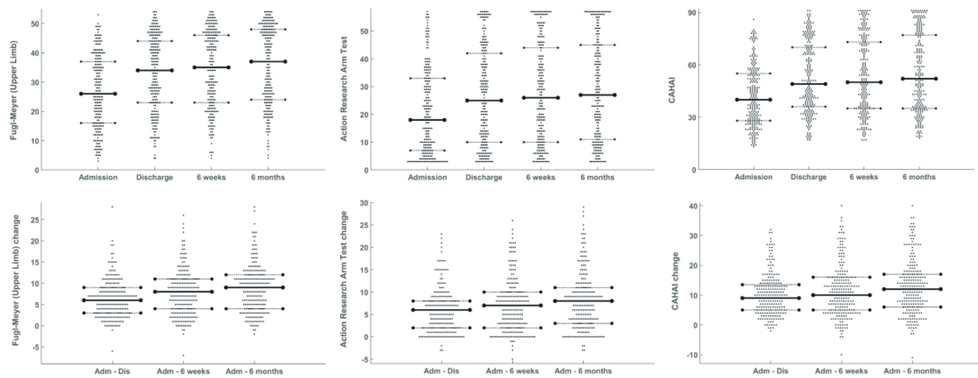
	FM-UL (T1-T4)	ARAT (T1-T4)	CAHAI (T1-T4)	ArmA-A (T1-T4)	ArmA-B (T1-T4)
Age	n.s.	n.s.	n.s.	n.s.	n.s.
Barthel Index	n.s.	n.s.	n.s.	n.s.	n.s.
modified Rankin Scale	n.s.	n.s.	n.s.	n.s.	n.s.
Neurological Fatigue Index	n.s.	n.s.	n.s.	n.s.	n.s.
Hospital Anxiety and Depression Scale	n.s.	n.s.	n.s.	n.s.	n.s.
Sensation	n.s.	n.s.	n.s.	n.s.	n.s.
FM-UL (T1)	n.s.	0.40	0.45	n.s.	-0.24
ARAT (T1)	n.s.	0.36	0.52	n.s.	n.s.
CAHAI (T1)	n.s.	0.42	0.35	n.s.	n.s.
ArmA-A (T1)	n.s.	n.s.	-0.31	-0.72	n.s.
ArmA-B (T1)	n.s.	-0.26	-0.38	-0.28	-0.26

Correlation coefficient (rho) given only when significant at p<0.05 corrected for 55 multiple comparisons (p< 0.0009). n.s. = not significant

Table 6: Correlations between changes in different outcome scores

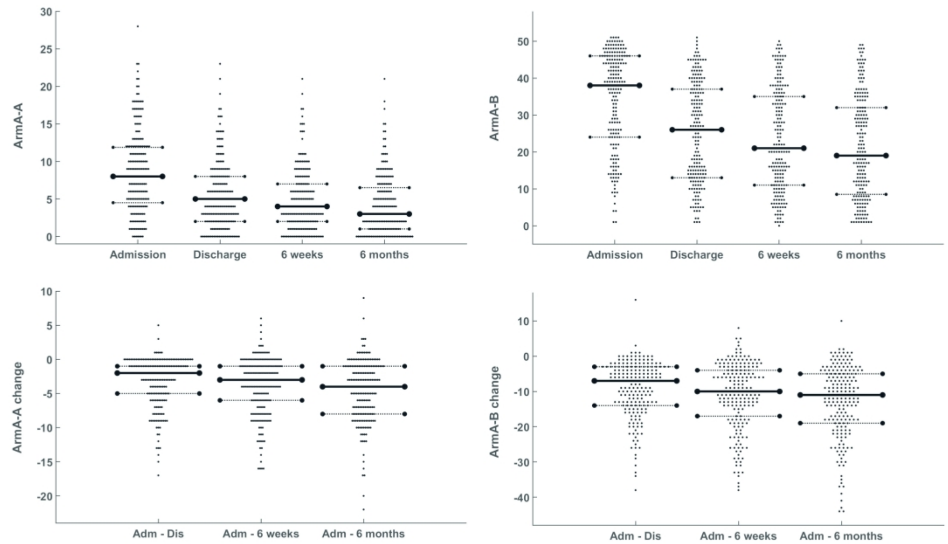
	FM-UL (T1-T4)	ARAT (T1-T4)	CAHAI (T1-T4)	ArmA-A (T1-T4)	ArmA-B (T1-T4)
FM-UL (T1-T4)	-	0.32	0.34	n.s.	n.s.
ARAT (T1-T4)	0.32	-	0.37	n.s.	-0.25
CAHAI (T1-T4)	0.34	0.37	-	n.s.	-0.20
ArmA-A(T1-T4)	n.s.	n.s.	n.s.	-	0.29
ArmA-B (T1-T4)	n.s.	-0.25	-0.20	0.29	-

Correlation coefficient (rho) given only when significant at p<0.05 corrected for 10 multiple comparisons (p< 0.005). n.s. = not significant



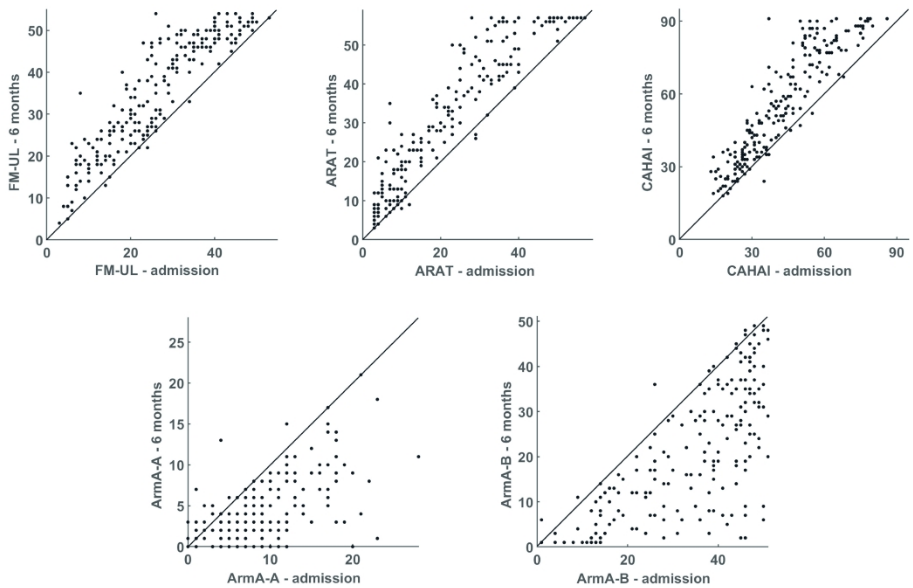
Outcome scores for all patients. Each data point represents a single patient. Top row shows individual scores at admission, discharge, 6 weeks and 6 months after discharge. Bottom row shows the individual difference scores for admission to discharge, admission to 6 weeks post-discharge, and admission to 6 months post-discharge. Scores are shown for modified Fugl-Meyer (upper limb), Action Research Arm Test and Chedoke Arm and Hand Activity Inventory (CAHAI). Median (solid line) and upper and lower quartiles (dotted lines) are shown.

600x250mm (300 x 300 DPI)



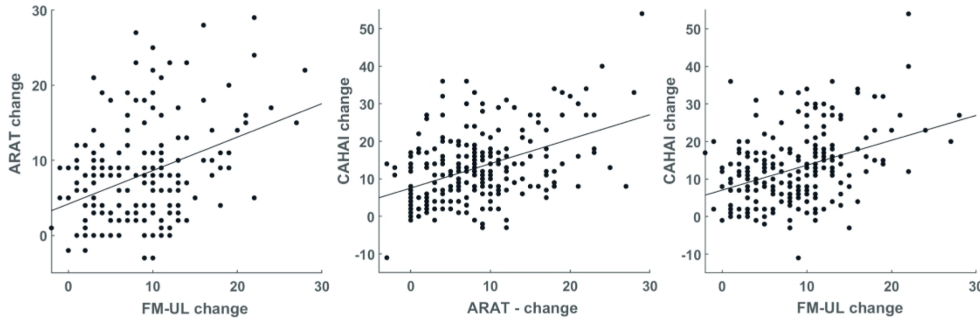
Self-reported outcome scores for all patients. Lower scores are better. Each data point represents a single patient. Top row shows individual scores at admission, discharge, 6 weeks and 6 months after discharge. Bottom row shows the individual difference scores for admission to discharge, admission to 6 weeks post-discharge, and admission to 6 months post-discharge. Scores are shown for the Arm Activity Measure part A and B (ArmA-A and ArmA-B). Median (solid line) and upper and lower quartiles (dotted lines) are shown.

399x250mm (300 x 300 DPI)



Admission score plotted against final scores at 6 months post-discharge for Fugl-Meyer (FM-UL), Action Research Arm Test (ARAT), Chedoke Arm and Hand Activity Inventory (CAHAI), Arm Activity Assessment A & B (ArmA-A and ArmA-B). The black line represents the line of 'no change'. Patients above this line for FM-UL, ARAT and CAHAI have improved. Patients below this line for ArmA-A and ArmA-B have improved.

299x199mm (300 x 300 DPI)



Change in scores from admission to 6 months post-discharge for Fugl-Meyer (FM-UL), Action Research Arm Test (ARAT), Chedoke Arm and Hand Activity Inventory (CAHAI), plotted against each other.

299x119mm (300 x 300 DPI)

Table S1. Significance of changes in outcome scores between each time point

	T1-T2	T1-T3	T1-T4	T2-T3	T3-T4	T2-T4
FM-UL	Z=11.9, p<0.0001	Z=11.8, p<0.0001	Z=12.1, p<0.0001	Z=5.8, p<0.0001	Z=8.1, p<0.0001	Z=4.5, p<0.0001
ARAT	Z=11.4, p<0.0001	Z=11.4, p<0.0001	Z=11.4, p<0.0001	Z =5.4, p<0.0001	Z=7.0, p<0.0001	Z=4.2, p<0.0001
CAHAI	Z=12.1, p<0.0001	Z=11.8, p<0.0001	Z=11.9, p<0.0001	Z=3.6, p=0.0003	Z=4.9, p<0.0001	Z=3.2, p=0.0014
ArmA-A	Z=10.1, p<0.0001	Z=10.0, p<0.0001	Z=10.4, p<0.0001	Z=2.6, p=N.S.	Z=5.6, p<0.0001	Z=3.4, p<0.0001
ArmA-B	Z=10.4, p<0.0001	Z=10.9, p<0.0001	Z=11, p<0.0001	Z=4.1, p<0.0001	Z=6.9, p<0.0001	Z=3.8, p=0.0002

Significance determined with Wilcoxon's signed ranks test