# Supplementary Methods and Materials

*Statistical Analysis of Clinical Scores*

To test our primary hypothesis that greater clinical scores would be found in the anodal tDCS group, each score was subjected to a multiple regression using the General Linear Model (GLM). The model variables were group (sham treatment and anodal tDCS) and the baseline for the respective score. The residuals of this model were tested for normality using the Shapiro-Wilk test *(42)*. Whereas the normality of the residuals ensured that p-values and confidence intervals computed using normal (parametric) theory are valid and exact, the various tests performed introduced a multiple comparisons problem that was further aggravated by the fact that, since each subject was assessed multiple times, the measurements and respective tests could not be considered independent, and therefore a naïve Bonferroni correction, if applied, would be unduly conservative. Moreover, the aggregate of various tests is more powerful than tests conducted separately, but their non-independence also limits the use of many conventional combined tests. Modelling such non-independence with parametric tests is intractable in practice. Both situations (multiplicity of tests and their combination) can, however, be addressed with the use of permutation tests. If the permutations are performed synchronously across all tests, their correlation structure is implicitly taken into account, without the need for explicit modelling.

We implemented the Non-Parametric Combination *(39, 40)* to allow for combined tests that interrogate aggregate effects of anodal tDCS on scores and timepoints, irrespective of the covariance structure among these variables. The same permutation framework allows the use of the distribution of the maximum statistic to produce p-values adjusted to control the family-wise error rate *(43)*. We proceeded as follows: we first ran a single global test, pooling the clinical scores and the post-intervention timepoints using the Fisher's combined probability test. In the Non-Parametric Combination, this meant computing the one-tailed p-values for each score and each timepoint using synchronised permutations; these p-values were used to construct the Fisher's $\chi^2$ combined statistic, which had its significance assessed using the same set of permutations that yielded the initial not

combined p-values. We used 10,000 permutations. To assess which clinical measures were showing any effect, we next repeated the procedure for each measure, pooling all post-intervention timepoints, and computing family-wise error rate-corrected p-values for each Fisher's $\chi^2$ combined statistic using the distribution of the maximum, thus correcting for the 3 measures considered. Next, as we were primarily interested in long-lasting clinical effects, we ran a similar Non-Parametric Combination strategy using Fisher's test on data from the 3-month follow up, pooling the 3 clinical measures, and computing p-values family-wise error rate-corrected across the four post-intervention timepoints. Finally, for each clinical measure, we reported the mean absolute difference and parametric confidence intervals (95%) between anodal tDCS and sham groups at the 3-month follow up, and tested the significance of this difference using the distribution of the maximum statistic across all tests, from the same set of permutations used with the previous tests, thus producing family-wise error rate-corrected p-values. In all cases, the level to declare significance was $p<0.05$ (corrected). See Supplementary materials for subject-level clinical scores with baselines regressed out (all scores, all timepoints, all subjects) and the MATLAB code used to run this analysis.

### *MRI Analysis*

Analysis of MRI data was carried out using tools from the FMRIB Software Library (www.fmrib.ox.ac.uk/fsl) *(41)*. For all voxel-wise analyses, images from patients with right hemispheric strokes were flipped about the midline after registration to standard space so that all lesions appeared on the left side. Stroke lesion volume were defined by manual outlining of lesion tissue on each patient's T1-weighted images using fslview.

## Voxel-based morphometry (VBM)

A standard FSL-VBM pipeline was used to analyse longitudinal changes in grey matter *(44)*. For each subject, the following steps were taken: the three T1-weighted structural images (baseline, post-intervention, and one month follow-up) were aligned to one another and a "midpoint space" was created, equidistant from all three images. The scans were averaged together in mid-point space and non-brain tissue removed using BET. The resulting image was binarised and then transformed back to mask each original scan. Each scan was then segmented in its native space using FAST. The grey

matter maps for the three scans for each subject were then transformed to the subject's midpoint space, where they were averaged together. These average grey matter images were then nonlinearly aligned to standard MNI space using FNIRT.

The standard-space aligned, average grey matter images of all the subjects (i.e., one per subject) were then combined to build a study-specific template. The average grey matter images were then re-aligned to this study-specific template. The grey matter for each individual scan (i.e., three scans per subject) was then transformed to standard space using the combination of rigid transformation from native space to midpoint space and the warp from midpoint space to MNI space. The images were then modulated for local expansion and contraction caused by the non-linear registration and smoothed with a sigma of 4mm. These pre-processed images were then analysed using permutation-based testing, as implemented in Randomise within FSL, to test for statistical differences between timepoints and groups and for correlations with clinical scores. Threshold-free cluster enhancement was used to determine areas of significant differences at a level of $p<0.05$, corrected for family-wise error.

## Task fMRI analysis

The following pre-processing steps were applied: high-pass temporal filtering at 100 seconds; motion correction using MCFLIRT *(45)*; removal of non-brain tissue using BET; spatial smoothing using a Gaussian kernel of 5mm; $B_0$ unwarping with fieldmaps using FUGUE. To remove artefactual components, the pre-processed EPI data were then denoised using Independent Component Analysis from within MELODIC *(46)*.

Statistical analyses were performed on this denoised data using FILM with local autocorrelation correction *(47)*. Registration of each patient's individual EPI and structural images to standard space was performed using FNIRT and Boundary Based Registration (BBR) *(48)*. Within FEAT, a boxcar regressor modelling the 30-second task and rest blocks was used to create first-level statistical activation maps for each patient at each timepoint: baseline (base), post-intervention on Day 10 (post), one month follow-up (month).

Higher-level, mixed effects analyses were then run using FLAME to compare activation maps across groups and timepoints and to test for correlations with clinical scores. Z statistic images with an initial cluster-forming threshold of $Z = 2.0$, and a corrected cluster extent threshold of $p < 0.01$.

## Diffusion MRI analysis

Diffusion data was analysed using FMRIB's Diffusion Toolbox (FDT). For each subject, the two sets of diffusion images were merged into a single 4D file. Eddy current correction was then run on the merged dataset to reduce the distortions induced by the gradient coils, and to correct for head motion by affine registering all images within the file to one another. DTIfit then fit a diffusion tensor at each brain voxel, which was subsequently used to generate voxel-wise estimates of fractional anisotropy (FA). FA maps were registered into standard space using FLIRT. For each patient and timepoint, mean FA values were extracted from within standard-space corticospinal tract regions of interest (see below). Asymmetry of the corticospinal tracts was calculated as (Contralesional CST FA – Ipsilesional CST FA) / (Contralesional CST FA + Ipsilesional CST FA), with larger asymmetry reflecting relatively higher FA values, i.e., structural integrity, in the contralesional CST. CST asymmetry scores were compared between groups and timepoints using repeated measures ANOVA, and were tested for correlations with clinical scores using Pearson's correlation.

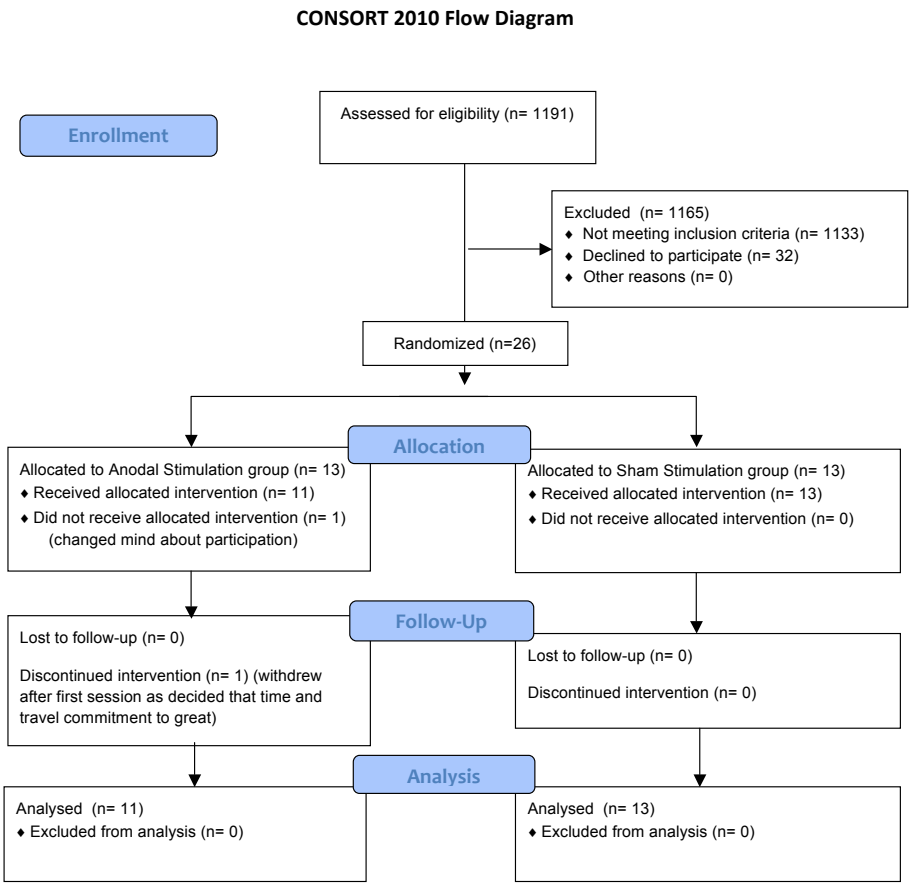
To generate corticospinal tract regions of interest, diffusion data from 10 healthy volunteers with a similar age range as the patients (7F, age range: 44-69 years, mean age: 55.9 years) was processed using multifiber probabilistic tractography *(49, 50)*. Masks of the left M1, right M1, and a single slice of the pons (at the level of z=-32) were drawn on a standard space brain. For each subject, probtrackx was then used to generate tracts for the left and right corticospinal tract (CST) separately. The pons was entered as the seed from which streamlines were initiated; M1 as a waypoint through which the streamlines had to pass. A midline exclusion mask was also used to maintain hemispheric independence. Each subject's probtrackx-generated left and right CST were thresholded at 1% of the maximum value to remove implausible tracts. Individual subject masks of the CST were binarised,

then combined across all subjects. These group images were then thresholded such that the remaining voxels were common to at least 7 of the 10 subjects.

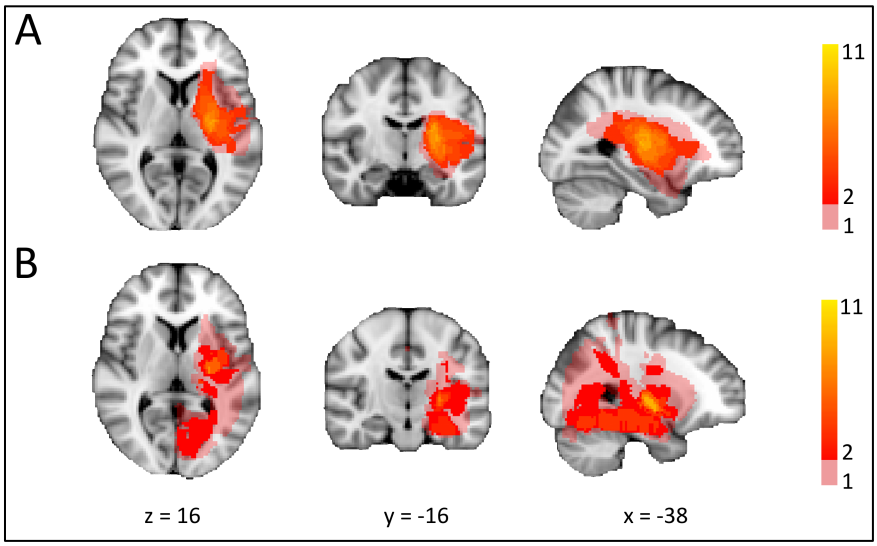
*Transcranial Magnetic Stimulation (TMS)*

During the first baseline assessment, TMS (80mm wing diameter figure-of-eight coil, oriented 45˚ to the mid-sagittal axis, Magstim 200 Monophasic Stimulator (Magstim, Carmarthenshire, UK)) was applied to the ipsilesional hemisphere to test whether muscle responses could be elicited in the affected hand. From a starting location 5cm lateral to the vertex, single pulses of TMS were applied and the affected hand was monitored for visible muscle twitches. If no responses were elicited more locations were tested and TMS intensity was increased up to 100% of maximum stimulator output. The presence or absence of muscle response was recorded.

# Supplementary Figures

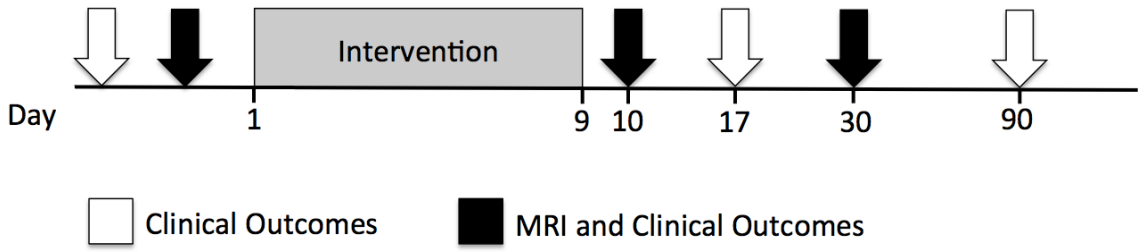



*Supplementary Figure 1: CONSORT flow diagram for study*



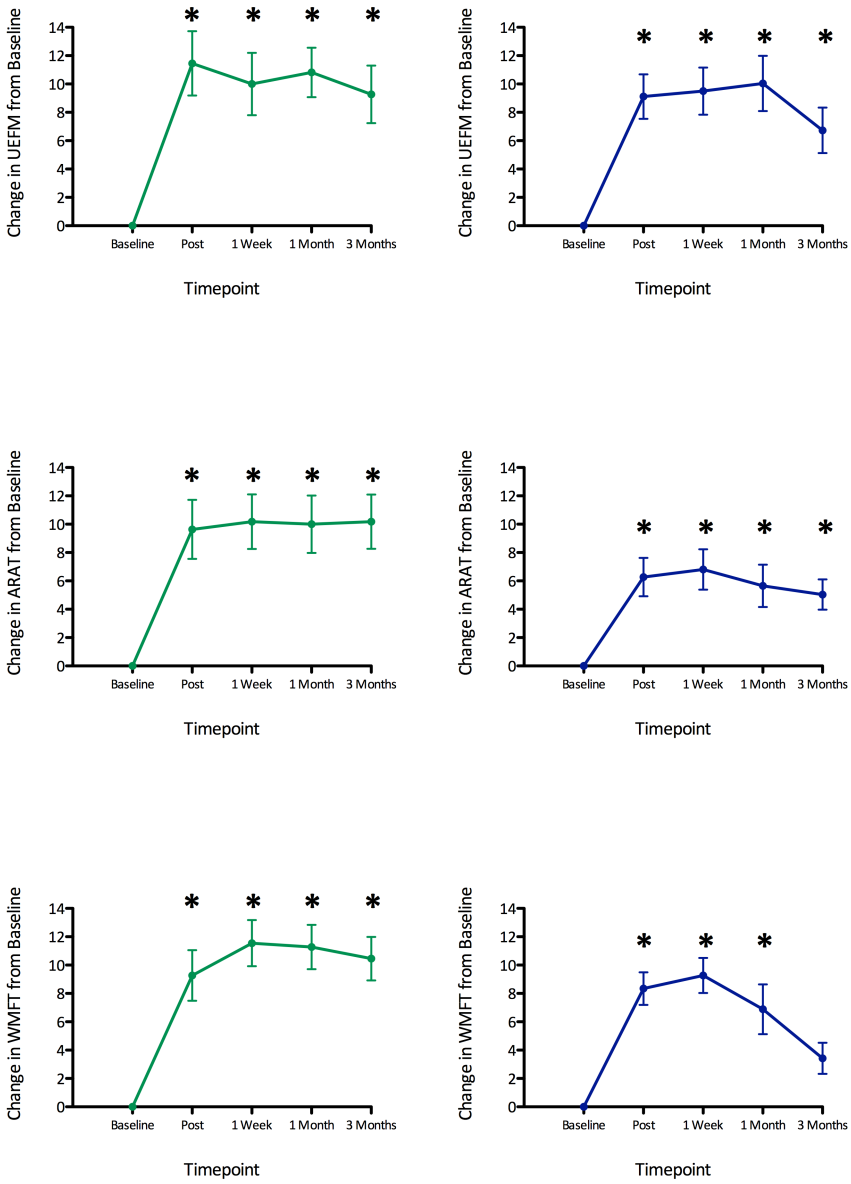

*Supplementary Figure 2: Overlap in lesion volumes across patients for each group*

Map showing lesion locations for each subject for the (A) sham and (B) anodal groups. Images are in radiological convention and, where necessary, data have been flipped about the midline after registration to standard space so that all lesions appeared on the left side.



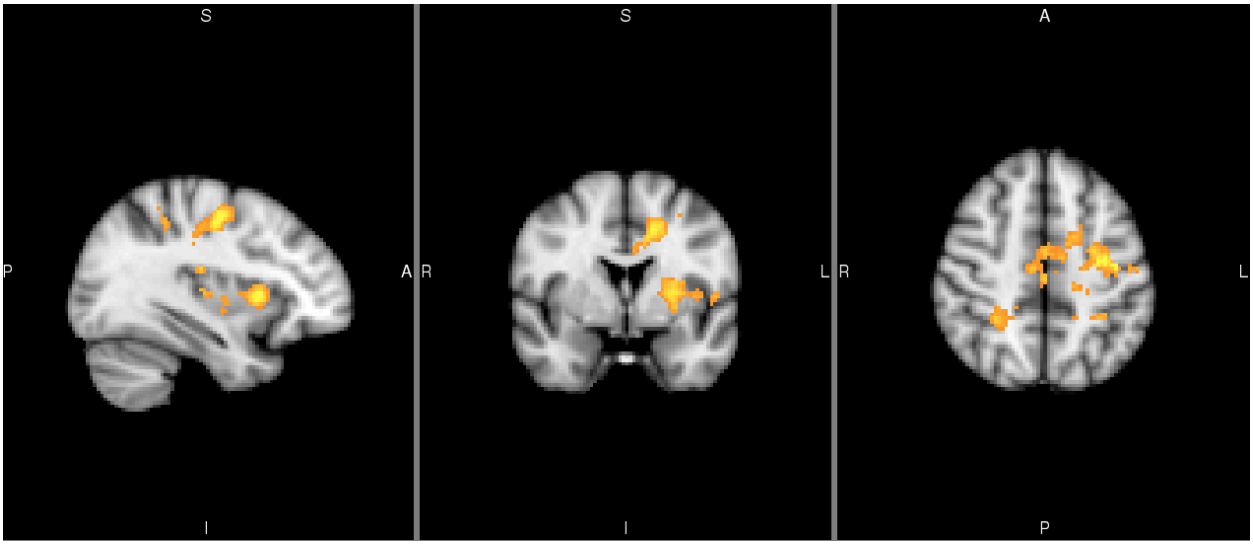

*Supplementary Figure 3: Study timeline.*

Participants received clinical assessments twice before the intervention, and at four timepoints after the intervention (immediate, one week, one month, three month). MRI scans were performed at baseline, immediately after the intervention, and at the one month follow-up.




*Supplementary Figure 4: Change in clinical scores compared to baseline for anodal and sham groups separately*

Change in clinical scores for anodal tDCS (green) and sham (blue) groups separately. * - Timepoints when the clinical score was significantly greater than at baseline ($p < 0.05$, corrected)




*Supplementary Figure 5: No effect of correction for grey matter on functional MRI results*

Higher level FMRI analysis was re-run including voxel-wise regressors of gray matter partial volume estimates for each subject. This controls for any variation due to changes in grey matter. Figure shows brain regions showing significantly greater increases in activity during affected hand movement from baseline to immediately post-intervention for the anodal group versus the sham group (i.e., post>pre, anodal>sham). These results are very similar to those reported in the main paper (Figure 3).

# Supplementary Table

| Anatomical Region | Cluster size (voxels) | Maximum Z score | Coordinates of maximum Z statistic | | |
|---|---|---|---|---|---|
| | | | x | y | z |
| *Baseline to Post* | | | | | |
| Cluster 1 | 5567 | | | | |
| L insula | | 4.19 | -40 | -10 | -2 |
| L putamen | | 4.08 | -28 | 4 | 4 |
| L ventral premotor cortex | | 4.06 | -40 | -4 | 38 |
| L caudate | | 3.83 | -18 | -6 | 24 |
| L sensory cortex | | 3.83 | -52 | -14 | 44 |
| L frontal pole | | 3.70 | -16 | 56 | -8 |
| Cluster 2 | 1529 | | | | |
| R putamen | | 3.83 | 24 | 10 | 8 |
| R cerebral white matter | | 3.71 | 36 | -32 | 0 |
| R thalamus | | 3.49 | 18 | -8 | 18 |
| R supramarginal gyrus | | 3.38 | 50 | -38 | 8 |
| R caudate | | 3.38 | 14 | 16 | 20 |
| R thalamus | | 3.37 | 14 | -14 | 18 |
| *Baseline to Month* | | | | | |
| Cluster 1 | 13586 | | | | |
| L lingual gyrus | | 4.64 | -14 | 56 | -8 |
| L putamen | | 4.51 | -22 | 6 | 10 |
| R frontal pole | | 4.41 | 12 | 56 | -10 |
| L paracingulate gyrus | | 4.39 | -10 | 28 | 38 |
| L angular gyrus | | 4.28 | -30 | -60 | 26 |
| L superior parietal lobule | | 4.28 | -34 | -42 | 46 |
| Cluster 2 | 845 | | | | |
| R lateral occipital cortex | | 3.96 | 36 | -68 | 2 |
| R lateral occipital cortex | | 3.74 | 50 | -70 | 6 |
| R intracalcerine cortex | | 3.57 | 22 | -82 | 6 |
| R lateral occipital cortex | | 3.55 | 32 | -80 | -2 |
| R lateral occipital cortex | | 3.36 | 28 | -68 | 24 |
| R occipital pole | | 3.31 | 28 | -90 | -4 |

*Supplementary Table 1: Location and Z statistic of peak voxels from functional MRI results*

Regions showing greater increases in task-based activation following the intervention compared to baseline, for anodal compared to sham tDCS groups.