INPLASY

INPLASY202690064

doi: 10.37766/inplasy2026.9.0064

Received: 14 September 2026

Published: 14 September 2026

Dry Needling for Upper-Limb Spasticity and Hypertonia After Stroke: A Systematic Review and Meta-Analysis

Wang, XP; Yang, R.

Corresponding author:
Rui Yang

yangrui_171@163.com

Author Affiliation:
Shanxi Province Academy of Traditional Chinese Medicine, Shanxi Province Hospital of Traditional Chinese Medicine.

ADMINISTRATIVE INFORMATION

Support - No support.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202690064

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 14 September 2026 and was last updated on 14 September 2026.

INTRODUCTION

Review question / Objective To synthesize within-treatment change, controlled continuous effects, and controlled responder effects after upper-limb dry needling.

Condition being studied The condition being studied is upper-limb spasticity and muscle hypertonia after stroke in adults. Stroke remains a leading cause of mortality and long-term disability worldwide. Spasticity is a clinically important manifestation of the upper motor neuron syndrome after stroke. A recent systematic review estimated that approximately one quarter of stroke survivors develop spasticity, with higher prevalence among individuals with motor paresis. Spasticity refers to a velocity- and muscle-length-dependent increase in resistance to externally imposed stretch related to altered descending control and stretch-reflex hyperexcitability. Muscle hypertonia is a broader clinical finding of increased resistance to passive movement and may reflect neural overactivity, non-neural stiffness, or both.

In the upper limb, flexor overactivity may interfere with hand opening, dressing, hygiene, splint tolerance, reaching, grasping, and task-specific practice. The functional consequences are nevertheless patient specific: reducing tone may facilitate care or movement in some individuals but may have little effect on activity when weakness, impaired selective motor control, pain, contracture, or sensory loss is the predominant limitation.

Current recommendations emphasize early recognition, goal-oriented assessment, and repeated reassessment throughout post-stroke spasticity management. Upper-limb rehabilitation guidelines favor intensive task-specific practice and individualized combinations of physical, pharmacological, and technological interventions rather than a single universal treatment. Best-practice guidance supports multimodal care, including positioning, stretching, splinting,

strengthening, task training, oral medication, focal chemodenervation, and selected physical modalities. International consensus recommends that botulinum toxin A be delivered within an integrated rehabilitation program and linked to explicit patient-centered goals. Nevertheless, real-world access to focal treatment varies in timing, dosage, and continuity.

The review evaluates adults aged 18 years or older with ischemic or hemorrhagic stroke and documented upper-limb spasticity or muscle hypertonia using a defined clinical assessment. Eligible outcome assessments include the Modified Ashworth Scale, Modified Modified Ashworth Scale, Resistance to Passive Movement Scale, or a prespecified binary responder definition for reduction in upper-limb spasticity or hypertonia. Samples spanning acute, subacute, and chronic stroke are eligible. Studies limited to healthy participants, children, cerebral palsy, traumatic brain injury, spinal cord injury, multiple sclerosis, or non-neurological pain conditions are excluded.

METHODS

Participant or population The review addressed adults aged 18 years or older with ischemic or hemorrhagic stroke and documented upper-limb spasticity or muscle hypertonia using a defined clinical assessment. Data for the stroke population had to be reported separately from other neurological groups. Samples spanning acute, subacute, and chronic stroke were eligible. Studies limited to healthy participants, children, cerebral palsy, traumatic brain injury, spinal cord injury, multiple sclerosis, or non-neurological pain conditions were excluded.

Intervention The intervention evaluated in this review was invasive dry needling with a solid filiform needle applied to an upper-limb muscle, motor point, taut band, or myofascial trigger point. Eligible techniques included manual or deep dry needling, ultrasound-guided dry needling, and dry needling combined with intramuscular electrical stimulation. Dry needling could be delivered as a stand-alone intervention or as an adjunct to rehabilitation, exercise, or botulinum toxin A, provided that outcomes for a dry-needling-treated arm were numerically identifiable. Uncontrolled studies were eligible only for the within-treatment synthesis and were not interpreted as evidence of comparative efficacy.

Comparator Eligible comparators included sham dry needling, usual rehabilitation, exercise, botulinum toxin A without dry needling, extracorporeal shock-wave therapy, or another needling condition. Uncontrolled studies were eligible only for the within-treatment synthesis and were not interpreted as evidence of comparative efficacy. In the controlled continuous synthesis, comparator conditions included standard rehabilitation, botulinum toxin A plus exercise without dry needling, extracorporeal shock-wave therapy, and conventional physical therapy. In the controlled responder synthesis, dry needling was compared with sham needling. The within-treatment synthesis did not include a common concurrent counterfactual and described change observed after a dry-needling-containing intervention rather than a causal comparative effect.

Study designs to be included Randomized trials, crossover trials, prospective controlled studies, quasi-experimental studies, and single-group before-after studies were eligible. Duplicate reports, protocols, reviews, editorials, conference abstracts without usable numerical results, lower-limb-only studies, case reports without an estimable variance, and secondary economic analyses were excluded.

Eligibility criteria Additional inclusion and exclusion criteria, beyond those defined in the PICOS sections, were as follows. Quantitative eligibility required sufficient numerical information to calculate an effect estimate and its sampling variance. For each synthesis, one effect estimate was selected per non-overlapping participant cohort within an original study. When multiple eligible muscles, outcomes, time points, or intervention arms were reported, selection followed the prespecified hierarchy of the investigator-designated primary upper-limb outcome, the muscle targeted by the needling protocol, the earliest post-treatment assessment, and the dry-needling arm with the least additional co-intervention. A study could contribute to more than one synthesis when it reported distinct eligible estimands, but statistically dependent effects from the same participants were not entered simultaneously within the same model. Duplicate reports, protocols, reviews, editorials, conference abstracts without usable numerical results, lower-limb-only studies, case reports without an estimable variance, and secondary economic analyses were excluded; when multiple publications described the same cohort, the report with the most complete eligible outcome data was retained.

Information sources PubMed/MEDLINE, Embase, Web of Science Core Collection, Scopus, and the

Cochrane Central Register of Controlled Trials were searched for records published from January 1, 2010, through July 31, 2026. The start date was selected to capture the emergence of dry needling as an intervention for post-stroke spasticity while avoiding an unnecessarily broad pre-intervention period. No language restriction was applied at the search stage. The complete database-specific strategies are provided in Appendix 1.

Reference lists of eligible articles and relevant reviews were hand-searched, and journal websites, DOI records, PubMed related-article links, and available full-text repositories were used to verify publication details and retrieve numerical information. Reference lists of eligible articles and relevant reviews were also screened for additional records.

Search results were imported into the review library and deduplicated using DOI, title, author, and publication year. Potential multiple reports from the same participant cohort were linked using author names, recruitment setting, participant characteristics, intervention, and sample size; only one report per cohort was included in each synthesis, prioritizing the report with the most complete eligible outcome data.

Main outcome(s) The review evaluated upper-limb spasticity or hypertonia outcomes after dry needling in adults after stroke. Eligible outcomes included the Modified Ashworth Scale (MAS), Modified Modified Ashworth Scale (MMAS), Resistance to Passive Movement Scale (REPAS), or a prespecified binary responder definition for reduction in upper-limb spasticity or hypertonia. Quantitative eligibility required sufficient numerical information to calculate an effect estimate and its sampling variance. For each synthesis, one effect estimate was selected per non-overlapping participant cohort within an original study; when multiple eligible muscles, outcomes, time points, or intervention arms were reported, selection followed the prespecified hierarchy of the investigator-designated primary upper-limb outcome, the muscle targeted by the needling protocol, the earliest post-treatment assessment, and the dry-needling arm with the least additional co-intervention.

Separate random-effects meta-analyses were conducted by estimand. Model A estimated the standardized change in upper-limb tone from baseline to the selected post-treatment assessment among participants who received dry needling, expressed as Hedges g with positive values indicating improvement. Model B estimated the standardized post-treatment difference between a dry-needling-containing arm and a concurrent comparator, expressed as Hedges g with positive values favoring the dry-needling-containing intervention. Model C synthesized study-defined binary responder outcomes as risk ratios. Within-person pre-post correlations were rarely reported; the primary sampling variance for Model A assumed r = 0.50, with sensitivity analyses using r = 0.30 and r = 0.70. Functional outcomes and adverse events were inconsistently reported.

Quality assessment / Risk of bias analysis A domain-based risk-of-bias assessment was undertaken because the evidence included several study designs. Randomized studies were assessed with RoB 2; nonrandomized comparative studies with ROBINS-I; and single-group before-after studies with the JBI Critical Appraisal Checklist for Quasi-Experimental Studies. Judgments followed the domains and categories specified by each tool. Data-reliability flags for converted summary statistics, inferred counts, or reported standardized effects without raw data were recorded separately.

The included evidence was methodologically heterogeneous. Although seven articles reported randomized allocation, allocation concealment, participant blinding, and assessor blinding were incompletely or inconsistently described. Sham needling was used in only a subset of controlled studies, and successful blinding may be difficult when local twitch responses or post-needling soreness occur. Four studies contributed noncomparative within-group data and were therefore vulnerable to temporal effects, regression to the mean, and concurrent therapy. Several trials included fewer than 20 participants per arm. Cointerventions were particularly relevant in the Cuenca Zaldívar and Kösem studies, in which dry needling was embedded in rehabilitation or botulinum toxin-based care.

The reliability of quantitative data also varied: some effects were calculated from directly reported means and standard deviations, whereas others required reconstruction from ordinal frequencies, conversion of medians, inference of responder counts from rounded percentages, or use of a reported standardized effect without raw data. These limitations were addressed in prespecified sensitivity analyses. Study-level judgments from RoB 2, ROBINS-I, and the JBI checklist are shown in Supplementary Figure S1.

Strategy of data synthesis Separate random-effects meta-analyses were prespecified for

Models A, B, and C because substantial clinical and methodological heterogeneity was anticipated. Between-study variance (τ^2) was estimated using restricted maximum likelihood. For continuous outcomes, confidence intervals around the pooled effect were calculated using a Hartung-Knapp adjustment to reduce overconfidence when few studies were available. Model C included only two studies; therefore, an exploratory REML-Wald confidence interval was used for the primary display and the Hartung-Knapp-adjusted interval was reported as a sensitivity analysis. Statistical heterogeneity was summarized using Cochran Q, τ^2, and I^2. A prespecified subgroup analysis compared single-session with multiple-session studies in Model A. Influence was examined using leave-one-out analyses. Funnel-plot inspection and an Egger-type regression were performed only for Model A and were considered exploratory. All confidence intervals were two-sided and reported at the 95% level; P values were interpreted descriptively.

Model A estimated the standardized change in upper-limb tone from baseline to the selected post-treatment assessment among participants who received dry needling, with positive values indicating improvement. The pre-post mean difference was standardized using the baseline standard deviation and corrected for small-sample bias to obtain Hedges g. The primary sampling variance assumed a correlation of r = 0.50; sensitivity analyses used r = 0.30 and r = 0.70. Model B estimated the standardized post-treatment difference between a dry-needling-containing arm and a concurrent comparator, with positive values favoring the dry-needling-containing intervention. Means and standard deviations were converted to Hedges g. When only a standardized between-group effect was reported, the estimate was converted to Hedges g using the reported group sizes; unreported raw means or standard deviations were not reverse-engineered. Sensitivity analyses excluded the study based on median/range conversion, the study reporting a standardized effect without raw group data, and both studies simultaneously. Model C synthesized study-defined binary responder outcomes as risk ratios; event counts and group denominators were used to calculate study-specific risk ratios.

When exact frequencies across MAS or MMAS grades were reported, means and sample standard deviations were reconstructed by treating the ordered categories as numerical scores. Medians with interquartile ranges or ranges were converted to approximate means and standard deviations using sample-size-appropriate methods and were included in the primary synthesis. Model A was repeated in a sensitivity analysis restricted to studies reporting direct means and standard deviations or exact ordinal-frequency distributions. For each synthesis, one effect estimate was selected per non-overlapping participant cohort within an original study. A study could contribute to more than one synthesis when it reported distinct eligible estimands, but statistically dependent effects from the same participants were not entered simultaneously within the same model.

Subgroup analysis A prespecified subgroup analysis compared single-session with multiple-session studies in Model A. Within Model A, five single-session studies yielded a pooled Hedges g of 0.84 (95% CI, 0.11 to 1.57; I^2 = 59.8%). Five multiple-session studies produced a larger point estimate (g = 1.36) but substantially greater uncertainty (95% CI, -0.24 to 2.95; I^2 = 87.7%). This numerical difference should not be interpreted as evidence of a dose-response relationship because session number was confounded by stroke phase, target muscle, comparator, co-intervention, study design, outcome scale, and data-conversion method (Figure 5).

The subgroup analysis did not establish an optimal dose. Multiple-session studies had a larger average effect than single-session studies, but the confidence interval included no effect and heterogeneity approached 88%. Session number was inseparable from differences in target muscle, stroke chronicity, comparator, co-intervention, outcome scale, and data quality. It remains unclear whether repeated needling produces cumulative physiological adaptation, creates repeated opportunities for a transient response, or facilitates subsequent exercise. Dose-finding trials should standardize needle diameter, insertion depth, pistoning or retention technique, target for local twitch responses, electrical-stimulation parameters, treatment interval, total number of sessions, and rehabilitation delivered before and after needling.

Sensitivity analysis Sensitivity analyses were performed to assess the robustness of the pooled estimates. For Model A, varying the assumed pre-post correlation from r = 0.30 to r = 0.70 had negligible influence on the pooled estimate, which remained approximately 1.13 with lower confidence limits near 0.42. Restriction to seven studies reporting direct means and standard deviations or exact ordinal-frequency distributions reduced the pooled effect to g = 0.71 (95% CI, 0.30 to 1.12; I^2 = 53.0%). Exclusion of studies with

major co-interventions yielded g = 1.05 (95% CI, 0.25 to 1.84; I^2 = 77.2%), and restriction to studies published from 2020 through 2026 yielded g = 1.25 (95% CI, 0.31 to 2.20; I^2 = 82.9%) (Figure 7).

Model B was less robust. Excluding the study that required median/range conversion reduced the pooled estimate to g = 1.37 (95% CI, -0.50 to 3.25). Excluding the study that reported a standardized effect without raw group data yielded g = 1.88 (95% CI, -1.74 to 5.49). Excluding both studies produced g = 1.21 with a very wide 95% CI from -8.17 to 10.59. These analyses indicate substantial sensitivity to data-source assumptions in the controlled continuous synthesis (Supplementary Figure S3).

Model C included only two studies; therefore, an exploratory REML-Wald confidence interval was used for the primary display and the Hartung-Knapp-adjusted interval was reported as a sensitivity analysis.

Country(ies) involved China.

Keywords Stroke; Dry Needling; Muscle Spasticity; Muscle Hypertonia; Upper Extremity.

Contributions of each author
Author 1 - Xupeng Wang.
Author 2 - Rui Yang.