Systematic Reviews

RESEARCH

Open Access

Effectiveness of scalp acupuncture and comparison with traditional acupuncture for stroke: an overview of systematic reviews and updated evidence

Sun-Young Park[1], In Heo[1,2], Man-Suk Hwang[1,2], Eui-Hyoung Hwang[1,2] and Byung-Cheul Shin[1,2*]

Abstract

Background Stroke recovery is a critical global-health priority; there is growing interest alternative therapies in scalp acupuncture (SA) to overcome the limitations of conventional treatments and improve outcomes. This study provides an overview of systematic reviews to evaluate the evidence on the effectiveness and safety of SA and to compare its therapeutic potential with traditional acupuncture (TA).

Methods A systematic search of 12 databases was conducted to identify systematic reviews and meta-analyses, completed on September 30, 2023, was performed without language restrictions. Selection criteria included adult stroke patients treated with SA, focusing on comparisons of effectiveness and safety in neurological deficits, motor function, disability, and total efficacy rate. Two reviewers independently screened studies and assessed methodological quality using AMSTAR-2, ROBIS, PRISMA-A, and GRADE frameworks. Data were synthesized to compare SA and TA for stroke outcomes, using total searched SA studies and TA data from the Cochrane review, followed by an analysis of high-quality studies to enhance evidence reliability.

Results After overviewing seven systematic reviews, the certainty of evidence supporting the standalone effectiveness and safety of SA remains low owing to methodological shortcomings. However, SA showed a greater effect size in the neurological deficits (-0.96 vs -0.53) in total studies and high-quality studies (-0.92 vs -0.48). Regarding motor function, SA had a higher effect size in total studies (0.94 vs 0.70), but TA outperformed it in high-quality studies (0.39 vs 0.82). Regarding disability outcomes, TA had a slightly larger effect size in total studies (1.27 vs 1.06), whereas SA surpassed it in high-quality studies (1.65 vs. 1.16).

Conclusions This overview highlights the potential of SA as an effective alternative therapy for stroke recovery, with high-quality studies demonstrating its effectiveness in improving neurological deficits and disability outcomes. This work guides clinicians on integrating SA for stroke recovery and offers insights for improving public health rehabilitation strategies. Despite limitations in the overall evidence owing to methodological shortcomings, the positive results from high-quality studies support SA as a possible approach for stroke recovery, underscoring the need for further rigorous research to strengthen its clinical application.

Keywords Systematic review, Evidence, Scalp acupuncture, Stroke rehabilitation, Traditional acupuncture

*Correspondence:
Byung-Cheul Shin
drshinbc@pusan.ac.kr
Full list of author information is available at the end of the article

Background

Stroke is a significant global health challenge and is ranked as the second and third leading cause of death and disability, respectively, according to the 2021 Global Burden of Disease Survey. From 1990 to 2019, the global stroke burden increased significantly, with incident strokes increasing by 70% and stroke-related fatalities increasing by 43% [1].

Traditional stroke recovery approaches include surgical intervention, pharmacotherapy, and rehabilitative exercises, each of which focuses on restoring independence and quality of life [2]. Owing to challenges in stroke treatment, there is growing interest in alternative treatments, such as acupuncture, which is considered safe and effective with fewer side effects than drugs [3]. In Traditional Chinese Medicine (TCM), symptom differentiation categorizes stroke symptoms into syndromes, guiding individualized acupuncture strategies. This approach aligns with TCM's holistic principles and supports its role in stroke rehabilitation.

Traditional acupuncture (TA) and scalp acupuncture (SA) use distinct systems; TA is based on the theory of meridians, whereas SA, a modern extension of TA developed in the 1970 s, targets specific scalp areas to stimulate brain regions. Although SA lacks a definitive theoretical basis, neuroimaging studies suggest it modulates brain activity in regions linked to cognition and motor function. Resting-state functional MRI (rs-fMRI) and electroacupuncture studies indicate that SA enhances executive control and sensorimotor network connectivity while reducing default mode network activity, [4, 5] supporting its role in stroke rehabilitation. This neuromodulatory effect provides a physiological basis for SA's therapeutic potential in treating neurological diseases such as migraine, stroke, and Parkinson's disease [6, 7].

Multiple randomized controlled trials (RCTs) have shown various benefits of SA for stroke recovery, with numerous meta-analyses (MAs) supporting these findings [6–12]. However, systematic reviews (SRs) provided inconsistent results, preventing clear recommendations [6–12]. The Cochrane SR indicates weak evidence for the effectiveness of TA in stroke rehabilitation and finds no evidence that SA is superior to TA [13]. Despite the widespread use of SA for neurological disorders, its comparative effectiveness with TA remains unclear.

This study aimed to update the evidence of SA for stroke recovery by reviewing current SRs and evaluating the comparative effectiveness of SA and TA in terms of their effectiveness on stroke. The results could provide essential guidance for healthcare practitioners and patients with stroke, aiding in treatment decisions and improving outcomes.

Methods

Registration

We conducted this overview of SRs in accordance with the Preferred Reporting Items for Overviews of Reviews guidelines (Supplement 1). This overview was registered in the PROSPERO database of York University (registration number CRD42022309463).

Search strategy

Two reviewers (SYP and IH) conducted comprehensive searches across multiple databases until 30 September 2023 without language restrictions. The databases used were MEDLINE (Medical Literature Analysis and Retrieval System Online, English), Embase (Excerpta Medica Database, English), Cochrane Library (The Cochrane Library, English), PubMed (Public/Publisher MEDLINE, English), Web of Science (Clarivate Web of Science, English), VIP (Chinese Scientific Journals Database, Chinese), CBM (Chinese Biomedical Literature Database, Chinese), CNKI (China National Knowledge Infrastructure, Chinese), Wan-Fang (Wan-Fang Data, Chinese), KISS (Korean Studies Information Service System, Korean), RISS (Research Information Sharing Service, Korean), and CiNii (Citation Information by National Institute of Informatics, Japanese). We searched for terms related to stroke, such as "Stroke," "Hemorrhage," and "Infarction," and terms related to SA, such as "scalp acupuncture" and "head acupuncture," focusing on "systematic review" and "meta-analyses." The search strategy was customized for each database, as exemplified by searches in the PubMed and CNKI databases.

Eligibility criteria and exclusion criteria

Study types

SRs and MAs of RCTs that investigated the effectiveness of SA in treating stroke were included in the study.

Participants

The participants were diagnosed using magnetic resonance imaging (MRI), computed tomography (CT), or World Health Organization guidelines and underwent SA. The inclusion criteria were adults, with no limitations on race, sex, duration of illness, or stroke stage.

Interventions

The study included SRs and MAs, focusing solely on treatments involving scalp acupoints, excluding ear and body acupoints. SA had to be the primary intervention;

however, studies were included if additional treatments were applied equally in both groups.

Outcome measures

All outcome measures were based on stroke clinical trial guidelines.

1) Neurological deficit scales (NDS) were assessed using the National Institutes of Health Stroke Scale (NIHSS) and the Modified Edinburgh–Scandinavian Stroke Scale (MESSS) to measure stroke severity and predict stroke outcomes [14, 15].
2) Disability was evaluated using the Barthel Index (BI), and dependence among patients with stroke was assessed using the Functional Independence Measure (FIM) [16–18].
3) Quality of Life (QOL) was measured using the Short Form 36 (SF-36) [19] and EuroQol-5 Dimension (EQ-5D) [20] for straightforward, generic health assessment.
4) Motor function was evaluated using the Fugl–Meyer Assessment (FMA), which focuses on motor function, balance, sensory perception, and joint functionality [21].
5) The Total Effective Rate (TER) measured the proportion of patients achieving positive outcomes post-treatment, excluding non-responders. It is primarily used in Chinese RCTs [22].
6) Safety reports regarding SA treatment were also included.

Exclusion criteria

Exclusion criteria were: (1) repeated publications; (2) SRs/MAs with incomplete or irretrievable data; (3) use of SA in control group interventions; (4) absence of relevant stroke diagnostic criteria in studies; (5) editorial reviews or conference abstracts; and (6) unavailable full texts for review.

Study selection and data extraction

Studies that did not meet the eligibility criteria were excluded by their title and abstract. Two independent reviewers (SYP and IH) screened, selected, and extracted the data separately before cross-checking. A third researcher (BCS) resolved any disagreements. The information collected included the first author, year of publication, intervention, outcomes, and other relevant details.

Assessment method

Methodological quality

Two reviewers (SYP and MSH) independently evaluated the methodological quality using the Assessment of Multiple Systematic Reviews-2 (AMSTAR-2) tool, including cross-verification, and a third researcher (BCS) to resolve discrepancies. AMSTAR-2 was chosen as it assesses key methodological domains, ensuring a standardized evaluation of systematic review quality. AMSTAR-2, consisting of 16 items (seven critical and nine non-critical), was used to assess the review quality for essential tool critiquing SRs [23]. The overall quality of the reviews was classified into four levels based on deficiencies: high, moderate, low, or extremely low.

Risk of bias assessment

Two independent researchers (SYP and MSH) used the ROB in Systematic Reviews (ROBIS) tool with cross-validation to assess the risk of bias (ROB) for each included SR. ROBIS was included as it specifically evaluates bias at the systematic review level, complementing AMSTAR-2's methodological assessment. A third researcher (BCS) evaluated unresolved discrepancies. The ROBIS tool, which includes four domains across three phases, determined the risk of bias as"low,""high ,"or"unclear"[24].

Assessment of the reporting quality

Two independent evaluators (SYP and EHH) assessed the reporting quality of each included SR using the Preferred Reporting Items for Systematic Reviews and Meta-analyses for Acupuncture (PRISMA-A) checklist. PRISMA-A was used to ensure adherence to acupuncture-specific reporting standards, improving transparency and completeness. Scores were mutually verified, and discrepancies were resolved through discussion with a third evaluator (BCS). PRISMA-A comprises 27 items that assess the integrity of reviews, with responses categorized as "yes," "no," or "partial yes" [25].

Assessment of the evidence certainty

Two independent evaluators (SYP and IH) assessed evidence quality using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) system. GRADE was applied to systematically evaluate the certainty of evidence, considering factors that influence confidence in effect estimates. Initially rated as "high quality," RCTs'ratings were reduced based on ROB, imprecision, indirectness, inconsistency, or publication bias. Evidence strength was classified into four categories: "very low," "low," "moderate," and "high." Unresolved

discrepancies were resolved using a third evaluator (BCS) [26, 27].

Data synthesis

We conducted a thorough analysis of SA for stroke treatment using relevant data such as risk ratios, odds ratios (ORs), 95% confidence intervals (CIs), weighted mean differences (MDs), and standard mean differences (SMD) extracted from the SRs/MAs included. We included the heterogeneity of each study using I^2 and P values. We performed a descriptive analysis of the included SRs, and the findings were compiled and displayed as percentages and frequencies. The findings were reported in accordance with the PRIOR guidelines for overviews of healthcare intervention reviews.

Comparison with TA

To compare the effectiveness of SA with that of TA, we analyzed the effect sizes and quality of evidence across common stroke outcomes. Two comparisons were conducted. First, we compared the TA data from the Cochrane review with the total number of SA studies identified by our team. If this analysis indicated low evidence quality, we re-analyzed both the TA and SA datasets by focusing only on high-quality studies. High-quality studies were defined as those with a sample size > 40 and explicit random sequence generation to ensure statistical power and minimize selection bias. Hertzog (2008) suggests that a sample size of 40 improves effect size precision and reduces variability, supporting its use as a reliability criterion [28]. Random sequence generation was prioritized as it reduces selection bias, aligns with allocation concealment, and is more consistently reported than other risk-of-bias factors [29–32]. Both comparisons used a random-effects model to quantify effect sizes through SMD, with evidence quality assessed using the GRADE system to ensure consistent evaluation of evidence reliability.

Results

Study selection

After screening 418 studies, seven SRs, [6–12] our final overview included only those that met all the inclusion criteria and provided adequate data (Fig. 1).

Study characteristics

Seven SRs, all based on RCTs published between 2012 and 2021, included six in English and one in Chinese. Among the included 7 SRs, five were conducted in China, and two were conducted in South Korea. Given that SA is most widely practiced in China, the majority of included RCTs were published in China. Among the seven SRs included in this review, the number of RCTs and participants varied significantly, ranging from a minimum of 7 RCTs with 230 participants [6] to a maximum of 27 RCTs with 2,741 participants [12]. These SRs searched between 3 and 15 databases. Among the SRs, one focused on ischemic stroke [11], one on both stroke types [9], one on intracerebral hemorrhage [6], and the rest did not specify. All patients were diagnosed using CT/MRI, with conditions ranging from acute [6, 11] to various stages of stroke [7–10, 12]. The treatment duration ranged from 10 days to 180 days. Western conventional medicine (WCM) served as the comparison group in all studies. The NDS criteria were assessed using MESSS [6, 11] and NIHSS [10], while some SRs did not specify the scale [8, 9, 12]. Motor function was mainly measured by FMA [7, 8, 10], but one SR did not specify the tool [9]. Disability was evaluated using BI [8, 10], FIM [8, 10], and Rankin Scale [8], with some SRs not specifying the measure [9]. Five studies used TER, [8–12] and QOL was measured using SF-36 and EQ-5D [10]. Methodological quality was predominantly assessed using the Cochrane ROB tool, [7, 10–12] whereas others used different methods, such as blinding [6, 9] or the Modified Jadad score [8]. Most studies reported adverse events, [6–11] and several performed subgroup [10, 12] and sensitivity analyses [7, 8, 10, 12] (Table 1).

Quality of the included SRs

Quality of the methodology

The analysis of SRs using AMSTAR-2 criteria showed significant variability in methodological quality, with deficiencies in critical domains such as protocol registration, justification for exclusions, and publication bias assessment. Issues such as inadequate literature searches, poor ROB assessments, and inconsistent meta-analytical methods further compromised the reliability of the studies. Only one study was rated"Extremely Low Quality,"[8] two as"Moderate Quality,"[10, 11] and four as"Low Quality"[6, 7, 9, 12] (Supplement 2).

ROB

The ROBIS assessments revealed that most SRs had low ROB in certain domains: 85.7% for study eligibility criteria and data collection/appraisal but only 42.9% for study identification/selection and synthesis of results. Overall, 57.1% of the SRs were judged to have low ROB across all phases (Supplement 3).

Reporting quality

The PRISMA-A assessment showed that 18 out of 27 reporting items had a reporting eligibility rate of at least 70%, with an overall reporting completeness exceeding 80%. However, deficiencies were noted in the protocol registration (introduction), search methods (methods),

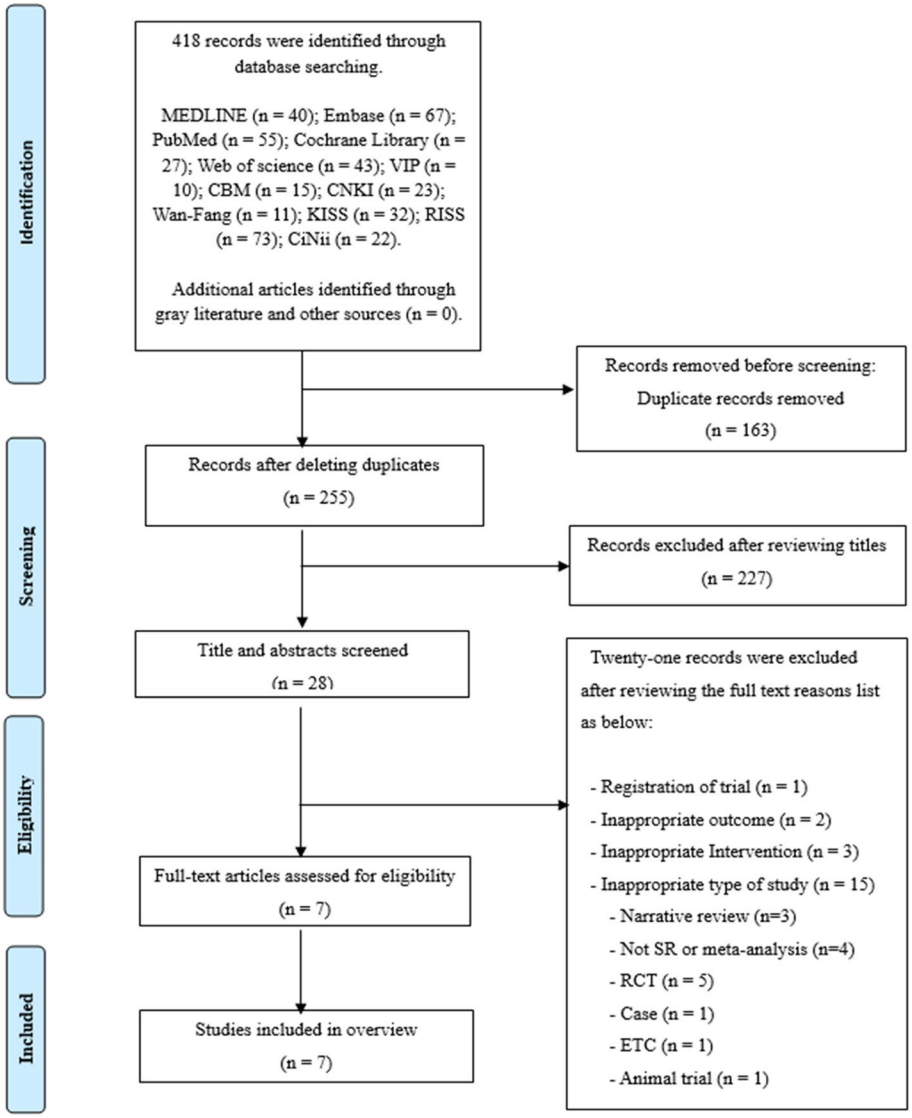

Fig. 1 Flowchart of the literature screening and selection process

ROB across studies (results and methods), additional analyses (results and methods), and study selection (results). None of the SRs reported protocol registration; three used a comprehensive search strategy, [6, 10, 11] four detailed their ROB assessment processes, [7, 8, 10, 11] and five assessed publication bias. [7–10, 12] Only two additional analyses have been previously reported [7, 10] (Supplement 4).

Confidence in study outcomes

Studies were most frequently downgraded mainly owing to ROB, followed by imprecision and inconsistency, with no downgrades for indirectness. NDS outcome studies were downgraded for ROB, imprecision, and inconsistency, resulting in six low-quality and one very low-quality study. Motor function studies were affected by ROB, imprecision, and inconsistency, and all were rated as very low quality. Disability studies faced downgrades owing to ROB and imprecision, leading to two low-quality and one very low-quality study. TER outcomes were significantly impacted by ROB, imprecision, and inconsistency, resulting in a majority of low- and very-low-quality studies and one moderate-quality study. TER and NDS faced the most significant methodological challenges, whereas motor function and disability were less affected (Table 2).

Table 1 Characteristics of systematic reviews and meta-analyses

First author (year)	Initial language	Number of RCTs (total participants)	Duration of illness	Period of treatment	Intervention	Comparison	Outcomes	Assessment of methodological quality	Adverse event	Subgroup analysis	Sensitivity analysis
Zheng (2011) [6]	English	7 (230)	Acute	10–28 (d)	SA + WCM	WCM	①	Random/blind	Yes	No	No
Huang (2021) [7]	English	12 (1,043)	Acute to chronic	14–180 (d)	SA + WCM + RH	WCM + RH	②	Cochrane	Yes	No	Yes
Young (2018) [8]	English	21 (2,231)	Acute to chronic	10–60 (t)	SA SA + WCM SA+BA	BA WCM BA	①②③④	Modified Jadad/ Cochrane	Yes	No	Yes
Zhou (2013) [9]	English	13 (1,395)	Acute to chronic	10–60 (t)	SA + WCM SA	WCM BA	①②③④	Random/blind	Yes	No	No
Lee (2013) [10]	English	21 (2,172)	Acute to chronic	10–60 (t)	SA SA SA + WCM SA + RH	WCM RH WCM RH	①②③④⑤	Cochrane	Yes	Yes	Yes
Wang (2012) [11]	English	8 (538)	Acute	10–60 (d)	SA+ WCM	WCM	①④	Cochrane	Yes	No	No
Wang (2015) [12]	Chinese	27 (2,741)	Acute to chronic	10–60 (d)	SA SA + HM SA + WCM SA	HM HM WCM BA	①④	Cochrane	NR	Yes	Yes

Outcome measures: ① neurological deficit scales; ② motor function score; ③ disability; ④ total effective rate; ⑤ quality of life

RCTs Randomized controlled trials, *d* Day, *t* Times, *SA* Scalp acupuncture, *WCM* Western conventional medication, *RH* Rehabilitation, *BA* Body acupuncture, *HM* Herbal medicine, *NR* Not reported

Table 2 Assessment of the study quality using GRADE

Outcomes	First Author (year)	Number of studies (sample size)	Interventions	Comparison	Effect (95% CI)	*P* value	Certainty assessment					Certainty
							Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	
Neurological deficit scales	Young (2018) [8]	4 (302)	SA + WCM	WCM	SMD 0.52 (0.29, 0.75)	<0.0001	−1②	0	0	−1①	none	⊕⊕OO Low
	Zhou (2013)[9]	3 (196)	SA + WCM	WCM	WMD −2.96 (−4.00, −1.92)	<0.0001	−1②	0	0	−1①	none	⊕⊕OO Low
	Wang (2012) [11]	7 (436)	SA + WCM	WCM	MD −3.89 (−5.36, −2.43)	<0.00001	−1②	−1③	0	−1①	strongly suspected	⊕OOO Very low
	Zheng (2011) [6]	7 (458)	SA + WCM	WCM	WMD −3.57 (−4.92, −2.22)	<0.00001	0	−1③	0	−1①	strongly suspected	⊕⊕OO Low
	Lee (2013) [10]	6 (422)	SA + WCM	WCM	SMD −0.61 (−0.81, −0.40)	<0.00001	−1②	0	0	−1①	none	⊕⊕OO Low
	Wang (2015) [12]	2 (144)	SA + WCM	WCM	MD −2.11 (−3.31, 0.91)	0.0006	−1②	0	0	−1①	none	⊕⊕OO Low
	Wang (2015) [12]	8 (691)	SA + HM	SA + HM	MD −5.33 (−6.71, −3.96)	<0.00001	−1②	−1③	0	0	none	⊕⊕OO Low
Motor function	Huang (2021) [7]	8 (659)	SA + WCM + RH	WCM + RH	MD 11.16 (8.09, 14.23)	< 0.01	−1②	−2④	0	0	strongly suspected	⊕OOO Very low
	Young (2018) [8]	2 (129)	SA + BA	BA	SMD 0.33 (−0.04, 0.71)	0.077	−1②	0	0	−2⑤	none	⊕OOO Very low
	Lee (2013) [10]	2 (150)	SA	RH	MD 1.22 (−1.38, 3.82)	0.36	−1②	0	0	−2⑤	none	⊕OOO Very low
	Lee (2013) [10]	2 (150)	SA + RH	RH	MD −4.11 (−11.82, 3.60)	0.30	−1②	0	0	−2⑤	none	⊕OOO Very low
Disability	Lee (2013) [10]	2 (150)	SA	RH	SMD −0.16 (−0.50, 0.18)	0.34	−1②	0	0	−2⑤	none	⊕OOO Very low
	Lee (2013) [10]	2 (150)	SA + RH	RH	MD 13.41 (11.05, 15.76)	<0.00001	−1②	0	0	−1①	none	⊕⊕OO Low
	Lee (2013) [10]	2 (123)	SA + WCM	WCM	SMD 0.78 (0.40, 1.17)	<0.0001	−1②	0	0	−1①	none	⊕⊕OO Low
Total Effective Rate	Young (2018) [8]	3 (272)	SA	BA	RR 1.21 (1.09, 1.31)	0.0002	−1②	0	0	−1①	none	⊕⊕OO Low
	Zhou (2013) [9]	5 (549)	SA	BA	RR 1.10 (0.93, 1.28)	0.26	−1②	−2④	0	−2⑤	none	⊕OOO Very low
	Wang (2015) [12]	4 (541)	SA	BA	OR 0.27 (0.14, 0.51)	<0.0001	−1②	0	0	−1①	none	⊕⊕OO Low

Table 2 (continued)

Outcomes	First Author (year)	Number of studies (sample size)	Interventions	Comparison	Effect (95% CI)	*P* value	Certainty assessment					Certainty
							Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	
	Young (2018) [8]	5 (430)	SA + BA	BA	RR 1.12 (1.05, 1.20)	<0.0001	$-1^{②}$	$-2^{④}$	0	$-1^{①}$	none	⊕OOO Very low
	Lee (2013) [10]	3 (401)	SA	WCM	RR 1.30 (1.19, 1.42)	0.04	$-1^{②}$	0	0	$-1^{①}$	none	⊕⊕OO Low
	Young (2018) [8]	4 (363)	SA + WCM	WCM	RR 1.62 (1.40, 1.87)	<0.00001	$-1^{②}$	$-1^{③}$	0	$-1^{①}$	none	⊕OOO Very low
	Zhou (2013) [9]	5 (650)	SA + WCM	WCM	RR 1.27 (1.06, 1.51)	0.009	$-1^{②}$	$-2^{④}$	0	$-1^{①}$	none	⊕OOO Very low
	Wang (2012) [11]	4 (308)	SA + WCM	WCM	RR 1.23 (1.11, 1.37)	0.0001	$-1^{②}$	0	0	$-1^{①}$	none	⊕⊕OO Low
	Lee (2013) [10]	3 (243)	SA + WCM	WCM	RR 1.19 (1.05, 1.36)	0.007	$-1^{②}$	0	0	$-1^{①}$	none	⊕⊕OO Low
	Wang (2015) [12]	2 (150)	SA + WCM	WCM	OR 0.12 (0.04, 0.32)	<0.0001	$-1^{②}$	$-1^{③}$	0	$-1^{①}$	none	⊕OOO Very low
	Lee (2013) [10]	2 (130)	SA	RH	RR 1.00 (0.88, 1.13)	0.96	$-1^{②}$	0	0	$-2^{⑤}$	none	⊕OOO Very low
	Lee (2013) [10]	2 (150)	SA + RH	RH	RR 1.12 (1.01, 1.23)	0.02	$-1^{②}$	0	0	$-1^{①}$	none	⊕⊕OO Low
	Wang (2015) [12]	4 (340)	SA	HB	OR 0.16 (0.03, 0.97)	0.05	$-1^{②}$	$-1^{③}$	0	$-1^{①}$	none	⊕OOO Very low
	Wang (2015) [12]	12 (1,260)	SA + HB	HB	OR 0.20 (0.14, 0.29)	<0.00001	$-1^{②}$	0	0	0	none	⊕⊕⊕O Moderate

①: The sample size was too small, $n<300$; ②: The design of the included studies had a bias due to the randomization, blinding, allocation concealment, completeness of data, or selective reporting; ③: I^2 value of the combined results was large, confidence intervals overlapped difference, $50\% \leq I^2 \leq 75\%$, and the *P* value of heterogeneity test is very small; ④: I^2 value of the combined results was large, confidence intervals overlapped difference, $75\% \leq I^2$, and the *P* value of heterogeneity test is very small ⑤: The confidence interval was not narrow enough, and the sample size was too small, $n<300$

GRADE: Grading of Recommendations, Assessment, Development, and Evaluation; SA: scalp acupuncture; WCM: western conventional medication; RH: rehabilitation; BA: body acupuncture; HM: herbal; WMD: weighted mean difference; SMD: standard mean difference; MD: mean difference; CI: confidence interval; RR: relative risk; OR: odds ratio

Evidence quality of included SRs

The stroke effectiveness data are summarized in Table 2 based on the results of the included SRs. Reported effectiveness primarily encompasses the following outcomes: NDS, motor function, disability, and TER. To date, none of the SRs have reported the pooled safety outcomes.

Effectiveness of SA on NDS

SA plus WCM comparison with WCM Six studies reported on NDS. Six SRs [6, 8–12] reported that there was a significant difference in improving neurological deficits for SA + WCM compared to WCM ($P < 0.0001$; [8, 9] $P = 0.0006$; [12] $P < 0.00001$, [6, 10, 11] with evidence quality ranging from low [6, 8–10, 12] to very low [11] because of bias and imprecision.

SA plus HM comparison with HM One study [12] reported improved neurological deficits with low-quality evidence owing to bias and inconsistency ($P < 0.00001$).

Effectiveness of SA on motor function

All studies provided low-quality evidence of improvements in motor function. One study compared SA + WCM and RH with WCM and RH alone ($P < 0.01$). [7] Another study compared SA + BA with BA alone ($P = 0.077$) [8], and another study [10] compared SA alone with RH ($P = 0.36$) and SA + RH against RH alone ($P = 0.3$).

Effectiveness of SA on disability

One SR [10] reported disability in three different intervention groups. SA showed no significant effect compared to RH ($P = 0.34$), with low-quality evidence. However, SA combined with WCM ($P < 0.0001$) with a very low quality of evidence and SA combined with RH ($P < 0.00001$) with a low quality of evidence showed significant improvements.

Effectiveness of SA on TER

SA plus BA comparison with BA For TER, two SRs showed significant improvement for SA versus BA ($P = 0.0002$; [8] $P < 0.0001$[12]) with low quality, whereas another study reported no significant difference ($P = 0.26$ [9]) with very low quality. SA combined with BA was more effective than BA alone ($P < 0.0001$[8]) with very low quality.

SA plus WCM comparison with WCM Studies comparing SA with WCM showed significant effects ($P = 0.04$ [10]) with low-quality evidence. For SA + WCM versus WCM, significant effects were reported ($P < 0.0001$; [11] $P = 0.007$; [10] $P = 0.009$; [9] $P = 0.0001$; [11] $P < 0.00001$ [8]) with low-to very-low-quality evidence.

SA plus RH comparison with RH alone SA combined with RH showed a significant effect over RH alone ($P = 0.02$) [10] but with low-quality evidence. SA alone versus RH showed no significant difference ($P = 0.96$), [10] with very low-quality evidence.

SA plus HM comparison with HM Comparing SA with HM showed no significant difference ($P = 0.05$) [12] and was rated as very low quality. However, SA combined with HM showed a significant effect ($P < 0.00001$) [12] and was rated as moderate quality.

Comparison with TA

We initially compared the total TA data from the Cochrane review with the SA studies identified by our team, including those where Western conventional treatment was part of the intervention group, and synthesized the results for analysis. The common outcomes assessed were the NDS, FMA, and BI. Due to the very low quality of evidence found in this analysis, we focused our re-analysis exclusively on high-quality studies. In our analysis of the NDS, SA demonstrated a greater negative effect size in total studies (−0.96) than TA (−0.53), although the evidence quality for SA was very low, and for TA, it was low. In high-quality studies, SA maintained a stronger effect size (−0.92) than TA (−0.48), with SA showing low- and moderate-quality evidence. For the FMA, SA had a higher effect size in total studies (0.94) than TA (0.70), with both showing very low-quality evidence. However, in high-quality studies, TA outperformed SA (0.82 vs 0.39), with moderate- and low-quality evidence. Regarding BI, TA exhibited a slightly higher effect size in total studies (1.27) than SA (1.06), both with very low-quality evidence. In high-quality studies, SA showed an increased effect size (1.65) compared with TA (1.16), although the evidence quality remained very low for SA and low for TA (Table 3).

Heterogeneity analysis

The included SRs exhibited substantial heterogeneity across various outcomes. The I^2 values ranged from 0% to 96.7%, indicating low to high heterogeneity across studies. While the I^2 statistic reflects statistical variability, it does not explain the underlying causes of heterogeneity. Therefore, we further examined potential clinical and methodological heterogeneity to better understand the sources of variation.

Table 3 Comparison of the effectiveness of SA versus TA and quality of the evidence

	Outcomes	Type of intervention (A) + WCT vs WCT	No. of participants (No. of studies)	Effect size, SMD	95% CI	Quality of the evidence
Total studies	NDS	SA	1517 (22 RCTs)	−0.96	−1.27, −0.64	⊕⊖⊖⊖Very Low
		TA	363 (6 RCTs)	−0.53	−0.83, −0.23	⊕⊕⊖⊖Low
	FMA	SA	809 (10 RCTs)	0.94	0.47, 1.40	⊕⊖⊖⊖Very Low
		TA	245 (4 RCTs)	0.70	0.31, 1.08	⊕⊖⊖⊖Very Low
	BI	SA	273 (4 RCTs)	1.06	0.22, 1.91	⊕⊖⊖⊖Very Low
		TA	616 (9 RCTs)	1.27	0.54, 1.99	⊕⊖⊖⊖Very Low
High-quality studies[a]	NDS	SA	747 (12 RCTs)	−0.92	−1.31, −0.52	⊕⊕⊖⊖Low
		TA	240 (4 RCTs)	−0.48	−0.82, −0.13	⊕⊕⊕O Moderate
	FMA	SA	318 (3 RCTs)	0.39	−0.18, 0.96	⊕⊕⊖⊖Low
		TA	205 (3 RCTs)	0.82	0.42, 1.21	⊕⊕⊕O Moderate
	BI	SA	153 (2 RCTs)	1.65	0.34, 2.96	⊕⊖⊖⊖Very Low
		TA	536 (8 RCTs)	1.16	0.37, 1.94	⊕⊕⊖⊖Low

A Acupuncture, *SA* Scalp acupuncture, *TA* Traditional acupuncture, *No.* Number, *SMD* Standard mean difference, *CI* Confidence interval, *WCT* Western conventional treatment, *NDS* Neurological deficit scale, *FMA* Fugl–Meyer Assessment, *BI* Barthel Index

[a] sample size ≥40, clear mention of "Random Sequence Generation"

Clinical Heterogeneity: Variability in stroke type (ischemic, hemorrhagic, or unspecified) [6, 9, 11], disease stage (acute or. acute to chronic) [6–12], and intervention protocols (SA alone vs. SA + WCM, SA + RH, SA + HM, SA + BA) [6–12] contributed to differences in treatment effects. Additionally, the use of different comparators (WCM, RH, BA, HM) [6–12] further increased heterogeneity.

Methodological Heterogeneity: Differences in risk-of-bias assessments (Cochrane ROB, Modified Jadad) [7, 8, 10, 11], outcome measurement tools (e.g., NIHSS, MESSS, FMA, BI) [6, 7, 9, 10, 12] also contributed to variability. While some studies conducted subgroup [10, 12] and sensitivity analyses [8, 12], others did not explore heterogeneity in detail, further complicating the interpretation of results.

Adverse events

Six SRs/Mas [6–11] assessed adverse reactions in the included studies. Of the seven RCTs reviewed, only four conducted safety assessments, and none reported significant adverse events [6]. One study among 30 RCTs found a slightly lower, but not statistically significant, rate of dizziness and skin redness in the SA group [7]. Three of the 21 RCTs documented adverse reactions, with limited details provided [8]. Only one study reported adverse events without elaborating on their causes [9]. Of the 21 RCTs, only three reported adverse events potentially linked to SA, two found no adverse events, and one documented two participant deaths without establishing a causal link [10]. Another study reported no adverse effects, [11] and one SR did not analyze adverse events. [12] Due to the inconsistent and incomplete reporting of adverse reactions across studies, neither quantitative nor qualitative analyses were performed.

Discussion

Main findings

SA operates based on the neurostimulation theory, effectively targeting specific scalp areas to stimulate brain regions and the central nervous system, making it particularly promising for neurological conditions [33–35]. This mechanism positions SA as a potentially valuable therapeutic option for stroke recovery, particularly when addressing neurological deficits and disabilities. After overviewing 7SRs, we found that although SA demonstrated significant positive effects on neurological deficits, disability, and TER, particularly when combined with therapies such as WCM, the evidence supporting its standalone effectiveness and safety remains generally low. Methodological shortcomings, such as missing protocol registrations and inconsistent bias assessments, have led to low confidence despite the use of robust evaluation tools.

Comparative analysis of SA versus TA for stroke treatment revealed that, although SA showed a larger effect size in reducing neurological deficits and motor function recovery than TA in total studies, the overall quality of evidence was low or very low. Nevertheless, when focusing on high-quality studies, SA demonstrated notable advantages, particularly in improving neurological deficits and disability outcomes. While the overall evidence base has limitations owing to methodological inconsistencies, positive results from high-quality studies support

the potential of SA as complementary or alternative therapy. These findings emphasize the importance of further rigorous and high-quality research to strengthen the evidence and validate SA's broader application in stroke rehabilitation.

Strengths and weaknesses of the review

A key limitation of this study is the lack of standardization in SA localization and needling techniques. Variations in acupoint selection (WHO lines, Zhu's, YNSA), stimulation methods (manual vs. electroacupuncture), needle depth, and retention time create heterogeneity, making comparisons challenging and affecting treatment efficacy. Future research should establish standardized protocols to improve reproducibility and clinical applicability.

Despite these limitations, this study has notable strengths. The systematic and transparent approach used ensures a comprehensive evaluation of SA for stroke rehabilitation. The study utilized robust methodologies, including AMSTAR-2, ROBIS, PRISMA-A, and GRADE, to assess the quality of systematic reviews, ensuring a high level of reliability. Additionally, pre-registration in PROSPERO enhanced transparency and minimized potential biases. The inclusion of a multilingual literature search broadened the scope of analysis, allowing for a more comprehensive understanding of SA's global application in stroke recovery. Furthermore, this study compared SA with TA using high-quality studies, reinforcing its potential as an effective complementary or alternative therapy. By conducting key comparisons between Cochrane review TA data and total SA studies, we ensured a robust comparative analysis that adds value to existing evidence.

Another limitation is the heterogeneity among included studies, with variations in intervention protocols, stroke types (ischemic vs. hemorrhagic), and outcome measures. Subgroup analyses and meta-regression should be incorporated in future research to refine clinical applicability. Furthermore, inconsistencies in adverse event reporting limit a comprehensive safety assessment, highlighting the need for standardized reporting guidelines.

SA is widely utilized in stroke rehabilitation due to its simplicity and reported benefits, yet inconsistencies in acupuncture literature make its effectiveness difficult to assess. Some studies lacked rigorous controls, impacting data reliability, and only a few SRs distinguished between stroke types. Given the different rehabilitation prognoses, failure to differentiate ischemic from hemorrhagic stroke limits precise conclusions. Future studies should stratify stroke types to refine clinical recommendations.

Implications for future studies

In future studies, it is crucial to enhance methodological rigor by implementing clearly defined and registered protocols, adhering to the AMSTAR-2 criteria, and using standardized randomization and blinding methods. Expanding sample sizes, including diverse participants, and improving search strategies, especially for gray literature, will enhance the generalizability and accuracy of SRs. Longitudinal studies with extended follow-up are necessary to assess the long-term effects and potential side effects. Comparative and sham-controlled studies should be conducted to distinguish between the actual effectiveness of SA and placebo effects. Additionally, detailed SR reporting, including RCT descriptions, funding sources, and ROB, is essential. To ensure reliable evidence for SA, future research should emphasize transparency and strict methodological controls. Addressing potential biases through rigorous trial designs and independent validation will be essential for accurately assessing SA's effectiveness and facilitating its integration into clinical practice. Finally, economic evaluations comparing the cost-effectiveness of SA with conventional treatments will offer valuable insights for healthcare decision-making. This comprehensive approach could pave the way for integrating SA into evidence-based clinical guidelines for stroke recovery.

Conclusions

This overview highlights SA as a therapeutic option for stroke recovery, particularly in improving neurological deficits and disability outcomes, as evidenced in high-quality studies. While the overall certainty of evidence remains low owing to methodological limitations, SA demonstrates notable potential in specific domains, suggesting its viability as a complementary or alternative therapy. Comparisons with TA did not yield conclusive evidence of superiority but emphasized the need for a tailored approach based on patient-specific recovery goals. To solidify the clinical application of SA, further high-quality research, addressing methodological shortcomings and incorporating robust designs, is essential to establish definitive evidence of its efficacy and safety.

Abbreviations

AMSTAR-2	Assessment of Multiple Systematic Reviews-2
BI	Barthel Index
CIs	Confidence intervals
CT	Computed tomography
EQ-5D	EuroQol-5 Dimension
FIM	Functional Independence Measure
FMA	Fugl–Meyer Assessment
GRADE	Grading of Recommendations, Assessment, Development, and Evaluation
MAs	Meta-analyses
MDs	Mean differences
MESSS	Modified Edinburgh–Scandinavian Stroke Scale
MRI	Magnetic resonance imaging

NDS	Neurological deficit scales
NIHSS	National Institutes of Health Stroke Scale
ORs	Odds ratios
PRISMA-A	Preferred Reporting Items for Systematic Reviews and Meta-analyses for Acupuncture
QOL	Quality of life
RCTs	Randomized controlled trials
ROB	Risk of bias
ROBIS	ROB in Systematic Reviews
SA	Scalp acupuncture
SF-36	Short Form 36
SMD	Standard mean differences
SRs	Systematic reviews
TA	Traditional acupuncture
TER	Total Effective Rate
WCM	Western conventional medicine

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13643-025-02819-x.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
SYP and BCS designed the study. IH and SYP developed the search strategy for this study. SYP, MSH, and BCS conducted AMSTAR; SYP, MSH, and BCS conducted ROBIS; SYP, EHH and BCS conducted PRISMA-A; and SYP, IH, and BCS conducted the GRADE analysis. All the authors critically revised the protocol and approved the final manuscript.

Funding
This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2021-NR056450).

Data availability
All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
None.

Author details
[1]School of Korean Medicine, Pusan National University, Yangsan, Republic of Korea. [2]Department of Korean Medicine Rehabilitation, Pusan National University Korean Medicine Hospital, Yangsan, Republic of Korea.

Received: 5 December 2024 Accepted: 12 March 2025
Published online: 10 May 2025

References

1. Institute for Health Metrics and Evaluation (IHME). Global burden of disease 2021.2024. Available from: https://www.healthdata.org.
2. Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, et al. Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2016;47:e98–169.
3. Wu HM, Tang JL, Lin XP, Lau J, Leung PC, Woo J, et al. Acupuncture for stroke rehabilitation. Cochrane Database Syst Rev. 2006;3:CD004131.
4. Chen H, Jann K, Li Y, Huang J, Chen Y, Kang Y, et al. A true response of the brain network during electroacupuncture stimulation at scalp acupoints: An fMRI with simultaneous EAS study. Brain Behav. 2023;13(1):e2829. https://doi.org/10.1002/brb3.2829.
5. Chung WY, Liu SY, Gao JC, Jiang YJ, Zhang J, Qu SS, et al. Modulatory effect of international standard scalp acupuncture on brain activation in the elderly as revealed by resting-state fMRI. Neural Regen Res. 2019;14(12):2126–31. https://doi.org/10.4103/1673-5374.262590.
6. Zheng GQ, Zhao ZM, Wang Y, Gu Y, Li Y, Chen XM, et al. Meta-analysis of scalp acupuncture for acute hypertensive intracerebral hemorrhage. J Altern Complement Med. 2011;17:293–9.
7. Huang YJ, Huang CS, Leng KF, Sung JY, Cheng SW. Efficacy of scalp acupuncture in patients with post-stroke hemiparesis: meta-analysis of randomized controlled trials. Front Neurol. 2021;12:746567.
8. Young-Nim Y, Gwang-Cheon P, Myung-Rae C, Min-Yeong S, Chang-Su N, Jae-Young H, et al. Meta-analysis on randomized controlled trials for scalp acupuncture treatment of stroke: a systematic review. J Tradit Chin Med. 2018;38:465–79.
9. Zhou JW, Li J, Zhao JJ, Xie H, Wang M. Meta analysis on ischemic stroke treated with scalp acupuncture. World J Acupunct Moxibustion. 2013;23:41–7.
10. Lee SJ, Shin BC, Lee MS, Han CH, Kim JI. Scalp acupuncture for stroke recovery: a systematic review and meta-analysis of randomized controlled trials. Eur J Integr Med. 2013;5:87–99.
11. Wang Y, Shen J, Wang XM, Fu DL, Chen CY, Lu LY, et al. Scalp acupuncture for acute ischemic stroke: a meta-analysis of randomized controlled trials. Evid Based Complement Alternat Med. 2012;2012:480950.
12. Wang JH, Zhang XH, Du XZ, Zhao M, Bao YC, Shang JF, et al. Effectiveness of scalp acupuncture for ischemic stroke: a meta-analysis. Chin J Evid Based Med. 2015;15:1161–7. ZHU bo-wen.
13. Yang A, Wu HM, Tang JL, Xu L, Yang M, Liu GJ. Acupuncture for stroke rehabilitation. Cochrane Database Syst Rev. 2016;2016:CD004131.
14. Zhao XJ, Li QX, Liu TJ, Wang DL, An YC, Zhang J, et al. Predictive values of CSS and NIHSS in the prognosis of patients with acute cerebral infarction: a comparative analysis. Medicine (Baltimore). 2018;97:e12419.
15. Runde D. Calculated decisions: NIH stroke scale/score (NIHSS). Emerg Med Pract. 2020;22:CD6-7.
16. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. Md State Med J. 1965;14:61–5.
17. Kidd D, Stewart G, Baldry J, Johnson J, Rossiter D, Petruckevitch A, et al. The functional independence measure: a comparative validity and reliability study. Disabil Rehabil. 1995;17:10–4.
18. Sulter G, Steen C, De Keyser J. Use of the Barthel index and modified Rankin scale in acute stroke trials. Stroke. 1999;30:1538–41.
19. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30:473–83.
20. Dorman P, Slattery J, Farrell B, Dennis M, Sandercock P. Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 questionnaires after stroke. United Kingdom collaborators in the international stroke trial. Stroke. 1998;29:63–8.
21. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. Scand J Rehabil Med. 1975;7:13–31.
22. Lu ZH, Yang CL, Yang GG, Pan WX, Tian LG, Zheng JX, et al. Efficacy of the combination of modern medicine and traditional Chinese medicine in pulmonary fibrosis arising as a sequelae in convalescent COVID-19 patients: a randomized multicenter trial. Infect Dis Poverty. 2021;10:31.
23. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.
24. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016;69:225–34.

25. Wang X, Chen Y, Liu Y, Yao L, Estill J, Bian Z, et al. Reporting items for systematic reviews and meta-analyses of acupuncture: the PRISMA for acupuncture checklist. BMC Complement Altern Med. 2019;19:208.
26. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–6.
27. Granholm A, Alhazzani W, Møller MH. Use of the GRADE approach in systematic reviews and guidelines. Br J Anaesth. 2019;123:554–9.
28. Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health. 2008;31(2):180–91.
29. Savović J, Turner RM, Mawdsley D, Jones HE, Beynon R, Higgins JPT, et al. Association between risk-of-bias assessments and results of randomized trials in Cochrane reviews: the ROBES meta-epidemiologic study. Am J Epidemiol. 2012;176(7):592–607.
30. Armijo-Olivo S, Saltaji H, da Costa BR, Fuentes J, Ha C, Cummings GG. What is the influence of randomisation sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open. 2015;5(9):e008562.
31. Barcot O, Boric M, Poklepovic Pericic T, Puljak L. Risk of bias judgments for random sequence generation in Cochrane systematic reviews were frequently not in line with Cochrane Handbook. BMC Med Res Methodol. 2019;19(1):170.
32. Bandolier. quality and validity. Bandolier Professional: Independent Evidence-Based Health Care. 2003. Available from: www.ebandolier.com.
33. Lee HS, Park HL, Lee SJ, Shin BC, Choi JY, Lee MS. Scalp acupuncture for Parkinson's disease: a systematic review of randomized controlled trials. Chin J Integr Med. 2013;19:297–306.
34. Hao JJ, Cheng W, Liu M, Li H, Lü X, Sun Z. Treatment of multiple sclerosis with Chinese scalp acupuncture. Glob Adv Health Med. 2013;2:8–13.
35. Junhua Z, Menniti-Ippolito F, Xiumei G, Firenzuoli F, Boli Z, Massari M, et al. Complex traditional Chinese medicine for poststroke motor dysfunction: a systematic review. Stroke. 2009;40:2797–804.

Publisher's Note