Acupuncture for Subacute Stroke Rehabilitation

A Sham-Controlled, Subject- and Assessor-Blind, Randomized Trial

Jongbae Park, KMD, PhD; Adrian R. White, MA, BM, BCH, MD; Martin A. James, MD, FRCP; Anthony G. Hemsley, MRCP; Paul Johnson, MRCP; John Chambers, MRCP; Edzard Ernst, MD, PhD, FRCP, FRCPEd

Background: Any adjunctive therapy that may reduce persistent disability after stroke should be considered. Acupuncture is used for this purpose, but there is conflicting evidence on its effectiveness.

Methods: Patients with a recent (<4 weeks) episode of stroke were randomized to receive 12 sessions of either real or sham acupuncture during 2 weeks. The primary outcome was the change in Barthel activities of daily living score at the end of treatment. Secondary outcome measures included National Institutes of Health Stroke Scale score, motoricity index, and quality of life (EQ-5D [EuroQol–5 Dimensional form] and EQ-VAS [EuroQol–Visual Analog Scale]). Assessments were carried out by blinded physicians.

Results: A total of 116 patients (56 in the real acupuncture group and 60 in the sham group) were randomized, and 98 (real, 48; sham, 50) completed treatment and the 2-week assessment. Patient blinding by means of the sham acupuncture device was successful. Acupuncture was well tolerated except for 1 seizure during a real acupuncture session. The improvements in the Barthel scores were 4 points (interquartile range [IQR], 0-8) vs 3 points (IQR, 0-7) in the real and sham acupuncture groups, respectively \( (P = .38) \). The secondary outcome measures also essentially showed no significant effect of acupuncture. Post hoc analysis by baseline severity showed a greater improvement in leg function in the subgroup with baseline Barthel score less than the median (median score, 6): 22 points (IQR, 0-37) vs 4 points (IQR, 0-4) in the acupuncture and sham control groups, respectively \( (P = .02) \).

Conclusions: Acupuncture is not superior to sham treatment for recovery in activities of daily living and health-related quality of life after stroke, although there may be a limited effect on leg function in more severely affected patients.

Arch Intern Med. 2005;165:2026-2031
a recent (<4 weeks) clinically or radiologically confirmed stroke (ischemic or hemorrhagic) who were able to give informed consent were eligible for inclusion. Patients were excluded if they had preexisting disability leading to modified Rankin score of 3 or more, recent history of other serious diseases such as cancer or diseases transmissible by blood, fear of needling, stroke that had occurred under general anesthesia, history of previous acupuncture, or the likelihood of full recovery within 2 weeks. All patients were given full details of the study in both oral and written form. The study received approval from the North and East Devon Research Ethics Committee.

SAMPLE SIZE

We estimated that 63 patients were required in each group to detect a clinically relevant 2-point difference in Barthel activities of daily living (ADL) score (scale, 0 [entirely dependent] to 20 [fully independent]) with a power of 80% and an α value of .05. This magnitude of effect was based on an unblinded study of acupuncture in stroke rehabilitation. Assuming a 30% drop-out rate, we planned to randomize 82 patients to each group. However, recruitment was much slower than planned because many patients were excluded for not meeting the inclusion criteria (stroke too severe or too mild, other abnormality requiring surgery) or not screened because they were admitted to wards other than the main study site for logistical reasons. Recruitment was closed for operational reasons before the required sample size was achieved.

RANDOMIZATION

Randomization was stratified by baseline National Institutes of Health Stroke Scale score of less than 20 or greater than or equal to 20, age less than 70 years or greater than or equal to 70 years, and whether the affected side was dominant or non-dominant. Block randomization was by sequential, sealed, opaque envelopes, and it occurred after the acupuncturist’s evaluation (concealed allocation).

INTERVENTION

Real acupuncture was performed by means of standard stainless-steel needles (0.30-mm diameter, 40-mm length; AcuPrime; Dong Bang Acupuncture Inc, Chungnam, Korea), manually stimulated to elicit needle sensation (de qi). Sham acupuncture was performed with a newly described nonpenetrating sham needle that is blunt and telescopic, giving the impression of insertion but without penetrating the skin (Park Sham Device; 0.30-mm diameter; fully extended length, 40 mm; Dong Bang Acupuncture Inc). This device has been shown not to elicit the sensation of de qi, which is believed to be necessary for the therapeutic effect. Recognized points were used for the real acupuncture group, and sham points at least 1.5 cm away from real points were used for the sham acupuncture group. In addition to conventional multidisciplinary stroke rehabilitation, all randomized patients received between 9 and 12 sessions of real or sham acupuncture during 2 weeks from a physician of Korean medicine (J.P.) with 5 years’ experience in acupuncture. Each treatment session lasted for at least 20 minutes in both groups. Prescriptions of acupuncture points were individually tailored according to 1 of 4 subgroups of Korean acupuncture (defined according to the excess or deficiency of the yin and yang aspect of the patients). Each of the 4 prescriptions comprised 10 points: 6 points determined by the patient’s classification (BL66, LI1, HT3, HT4, GB43, GB44; LU8, SP3, HT8, HT9, Ki17, ST36, LI5, LI11, HT7, HT8, SI5; and SP1, SP2, HT8, HT9, LR1, CV4) and 4 points common to all (ST40 [bilateral], CV12, GV20, and GV26). Thus, each treatment involved a total of 11 needles. At the end of 2 weeks of treatment, the success of blinding was tested by asking all patients the following question: “When you volunteered for the trial, you were informed that you had an equal chance of receiving acupuncture or sham (pretend) acupuncture. Which acupuncture do you think you received?”

OUTCOME MEASURES

The primary outcome was the change in Barthel ADL score. Secondary outcome measures included National Institutes of Health Stroke Scale score, motoricity index, quality of life (EQ-5D [EuroQoL–5 Dimensional form] and EQ-VAS [EuroQoL–Visual Analog Scale]), Nottingham Extended ADL score, Ashworth scale for muscle spasticity, timed 10-m walk, 0-hole peg test, swallowing status (“safe” or “unsafe” swallow based on a bedside swallow screening test), and the patient’s blinding regarding treatment. Blinded physician assessors (M.A.J., A.G.H., P.J., or J.C.) screened and recruited patients, and carried out baseline and posttreatment assessments. The first acupuncture treatment was given within 48 hours of screening assessment, followed by the baseline assessment within 72 hours of screening; the posttreatment assessment was done within 2 days of the completion of treatment.

STATISTICAL ANALYSIS

Primary and secondary outcome measures were compared between groups on the 2-week data. The analyses were performed on an intention-to-treat basis, with missing data replaced by baseline values. In addition, on the basis of the data from a previous unblinded study, a post hoc analysis of the primary outcome measure per pair of the real and sham subgroups was performed to look for a differential effect according to baseline severity. Treatment responses were compared between those with baseline Barthel score at or above the median value vs those below the median value (6/20). For within-group comparisons, the Wilcoxon signed-rank test was used, and for between-group comparisons, either the Mann-Whitney test or χ² test was used. Blinding indices in both groups were calculated using the method by Bang et al, and their statistical significance was tested using 2 independent nonparametric tests. Most data are presented as median (interquartile range [IQR]) unless otherwise specified, as they were nonparametrically distributed, and P<.05 (2 sided) was regarded as statistically significant. Statistical analyses were conducted with the SAS software version 9.0 (SAS Institute Inc, Cary, NC).

RESULTS

Recruitment commenced on November 20, 1999, and was terminated on August 30, 2003, after 116 patients had been enrolled (56 to receive real acupuncture treatment and 60, sham treatment). Figure 1 shows the flow of participants through the trial. We recruited 9.4% of all patients admitted to the hospital with stroke during the entire study period. End-of-treatment data were complete for 98 (84.5%) of subjects randomized (48 in the real treatment group and 50 in the sham group). Three patients died during the study because their condition deteriorated after randomization, which had been based on the clinical expectation that they would survive. The underlying disease in these deaths was so apparent that the stroke physician in charge ruled out association with acupuncture treatment.
Adverse events were assessed by the blinded assessors at the posttreatment assessment, asking whether the patients had experienced any adverse events. Most patients who received acupuncture tolerated it well, except for 1 adverse event, a seizureslike reaction during the real acupuncture session. The baseline characteristics and data are shown in Table 1.

In both the real and sham acupuncture groups, all outcome measures improved significantly at 2 weeks (P < .001). There was, however, no significant effect of acupuncture treatment. The improvements in the Barthel scores were 4 points (IQR, 0-8) vs 3 points (IQR, 0-7) in the real and sham acupuncture groups, respectively (P = .38, Table 2). The secondary outcome measures followed the same pattern of significant improvements during the 2-week period but without a significant effect of acupuncture treatment. The exception was the number of patients with an unsafe swallow, which was significantly higher in the experimental group (6/48, 13%) than in the control group (2/49, 4%) (P = .04).

Post hoc comparison of treatment effects by baseline severity showed a greater improvement in leg function (leg subscale of the motoricity index) in the subgroup with baseline Barthel score less than the median: 22 points (IQR, 0-37) vs 3 points (IQR, 0-4) in the acupuncture and sham control groups, respectively (P = .02) (Figure 2). No similar changes were seen with the arm subscale of the motoricity index, or with other secondary end points.

Table 3 summarizes the results of patient blinding with the sham acupuncture device. Of the 94 patients who gave an answer, 4 (4%) in the sham group believed that they had received sham acupuncture. The blinding index was 0.47 (95% CI, 0.33 to 0.61) in the acupuncture group and −0.31 (95% CI, −0.49 to −0.13) in the sham group. The two proportions of unblinding were not significantly different (P = .16).

Recent systematic reviews of acupuncture in stroke rehabilitation have not drawn firm conclusions regarding the efficacy of treatment, mainly owing to problems with the quality of previous studies, and have recommended the conduct of adequately controlled and methodologically sound studies. Previous studies have largely not used blinding, which is a particular difficulty with a physical treatment such as acupuncture, or have been confounded by comparison with another treatment with potential effects. In general, studies of complementary therapies have not been subjected to tests of efficacy and safety equivalent to those for conventional or pharmacologic treatments. In this context, our randomized, subject- and assessor-blinded clinical trial of 2 weeks of acupuncture as an adjunct to conventional multidisciplinary stroke rehabilitation has not shown a significant effect of acupuncture on recovery of ADL when compared with control treatment with a novel validated sham acupuncture needle. Our results indicate that almost all of the participants either believed they were receiving genuine acupuncture or were unable to tell, and as such we believe our study in stroke rehabilitation to be the first to adequately control for the nonspecific effects of treatment and thus to come as close as possible to testing exclusively the therapeutic effect of needling.

Our principal finding of no difference in the primary outcome measure of Barthel score after 2 weeks of acupuncture treatment is open to a number of possible explanations, one of which is that acupuncture had a smaller effect size than that suggested by the previous unblinded study.
We present the statistical power of 80% at the .05 significance level to test the clinically significant yet plausible treatment effect of a difference of 2 points on the 20-point Barthel score. For comparison, a difference of 1 point on the Barthel score has been regarded as a significant effect of occupational therapy for subacute stroke.\textsuperscript{29} The effect size (Cohen’s $d$) of acupuncture in the unblinded Chinese study was 0.72 and 1.29 at 2 weeks and 4 weeks, respectively.\textsuperscript{28} We were unable, for various external reasons, to achieve our a priori recruitment target, although this was mitigated by a lower dropout rate than we anticipated, and our study still has the largest group size to date among published randomized studies of manual acupuncture in stroke. Alternatively, it might be argued that our primary outcome measure was insensitive to some of the other potential benefits that acupuncture may have on mood or communication, although we found no indication of this across a range of secondary outcome measures intended to detect other, more subtle effects on impairments or abilities. Our choice of Barthel score was based on the strong support of its frequent use in clinical trials. However, in hindsight, we regret that we did not use another mobility-related outcome measure, such as the Fugl-Meyer index, which is asserted to be more sensitive, as the primary outcome measure.

It could be argued that an effect might have emerged if treatment had been continued longer, as treatment effects of acupuncture may be incremental during a longer period.\textsuperscript{30} The timing of assessments affected the result of the treatment effect of both sham and real acupuncture, with the primary outcome measure of the Fugl-Meyer index being more sensitive, as the primary outcome measure. However, our secondary outcomes included the Barthel Index, which was insensitive to some of the other potential benefits of acupuncture. This may simply represent a welcome secular trend in current UK practice, given that all patients in our study also received specialist stroke rehabilitation, but in older studies the lack of an effective intervention for the control group may have cast any treatment in a more favorable light.

Yet, nonetheless, we did, in a post hoc analysis, detect a potential beneficial effect of treatment on leg function in a subgroup of more severely affected patients with baseline Barthel score below the median. This differential effect on leg function was also reported in the previous study by Hu et al\textsuperscript{13} and the recent study by Alexander et al.\textsuperscript{31} There was

Table 2. Outcome Measures 2 Weeks After Treatment

<table>
<thead>
<tr>
<th>Outcome Measures</th>
<th>No. of Patients</th>
<th>Posttreatment (2 wk)</th>
<th>Change Between Baseline and Posttreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real Acupuncture</td>
<td>Sham Acupuncture</td>
<td>Real Acupuncture</td>
</tr>
<tr>
<td>Barthel ADL score, median (IQR)</td>
<td>56</td>
<td>60</td>
<td>11 (5-17)</td>
</tr>
<tr>
<td>NIHSS score, median (IQR)</td>
<td>56</td>
<td>60</td>
<td>5.5 (2-10.3)</td>
</tr>
<tr>
<td>Motoricity index, median (IQR)</td>
<td>56</td>
<td>60</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Arm score</td>
<td>51.5 (0-85.3)</td>
<td>71 (21.3-87.3)</td>
<td>3 (0-25)</td>
</tr>
<tr>
<td>Leg score</td>
<td>75 (39-99)</td>
<td>75 (54-89)</td>
<td>8 (0-38)</td>
</tr>
<tr>
<td>Trunk score</td>
<td>121 (100)</td>
<td>74 (48-100)</td>
<td>12 (0-37)</td>
</tr>
<tr>
<td>Cannot swallow, No. (%)</td>
<td>48</td>
<td>50</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Spastic (ASS &gt;1), No. (%)</td>
<td>48</td>
<td>50</td>
<td>16 (33)</td>
</tr>
<tr>
<td>NEADL score, median (IQR)</td>
<td>48</td>
<td>50</td>
<td>4 (2-12.1)</td>
</tr>
<tr>
<td>Time to walk 10 m, s</td>
<td>22</td>
<td>19</td>
<td>17 (12-23.5)</td>
</tr>
<tr>
<td>9-Hole peg test, s</td>
<td>20</td>
<td>23</td>
<td>37.5 (21.5-92.3)</td>
</tr>
<tr>
<td>EQ-5D, median (IQR)</td>
<td>48</td>
<td>50</td>
<td>0.64 (0.03-0.8)</td>
</tr>
<tr>
<td>EQ-VAS, median (IQR)†</td>
<td>48</td>
<td>50</td>
<td>60 (48.6-72.5)</td>
</tr>
</tbody>
</table>

Abbreviations: ADL, activities of daily living; ASS, Ashworth spasticity score; EQ-5D, EuroQol (quality of life)–5 Dimensional form; EQ-VAS, EuroQoL–Visual Analog Scale; IQR, interquartile range; NA, not applicable; NEADL, Nottingham Extended Activities of Daily Living; NIHSS, National Institutes of Health Stroke Scale; VAS, visual analog scale.

†EQ-VAS offers a simple method for obtaining a self-rating of current health-related quality of life by generating a score.

Figure 2. Change in leg score of motoricity index of subgroups categorized by the baseline severity (greater than or equal to or less than the mean baseline Barthel score [BBS] of 6).

©2005 American Medical Association. All rights reserved.
also a trend with treatment toward a faster timed 10-m walk in the few patients who were ambulant at the 2-week assessment (% = .09) (Table 2), and these 2 potential effects of treatment may be related. A previous study reported a higher probability of early discharge in patients with a faster timed 10-m walk,32 but our study found no difference in length of hospital stay between the groups (data not shown). Such post hoc analysis, however, should always be interpreted with caution, particularly with regard to the risk of a type I statistical error, but at the very least it does suggest that any further studies of acupuncture in stroke rehabilitation should concentrate on the patients most severely affected. The significant difference in the number of patients with an unsafe swallow at 2 weeks (Table 2) may similarly have been caused by chance, in the absence of any other plausible mechanism.

Our study took place during a period of significant change in the provision of stroke care in the United Kingdom,33 and it suffered from practical difficulties with slow recruitment, and with patients with acute stroke being well enough and available for treatment when not receiving other conventional treatments or diagnostic tests. All acupuncture treatments were performed by a single acupuncturist, a feature that generates high internal validity but limits the external validity in that other practitioners use different approaches. We hypothesized that any potential benefit of acupuncture would be greatest if the treatment were given at the earliest practicable stage in stroke recovery, and this reflects customary practice where acupuncture is routinely used.34 As such, our study represents a realistic test of what it might be like to deliver such complementary therapies in the setting of a busy acute stroke unit, rather than in selected patients in a subacute rehabilitation setting. We originally planned to conduct a 6-month follow-up including detailed assessment of balance but were prevented from doing so by a poor response rate, due to death or further stroke; patients unable to attend because of persistent handicap; and patients moving from the area. We obtained 6-month data in a skewed sample of only 13 patients. It could perhaps be reasonably argued that in the absence of a significant effect at the end of 2 weeks of acute treatment, there was even less likelihood of a significant effect of treatment being detected at 6 months.

Blinding indices, 0.47 (95% CI, 0.33 to 0.61) and −0.31 (95% CI, −0.49 to −0.13) in the acupuncture and sham groups, respectively, imply that 47% of participants correctly guessed the treatment identity beyond chance, while 31% of participants in the sham group made incorrect guesses, ie, they thought that they had received acupunct-

ture although they received the sham treatment. Following the logic of Bang et al.,23 instead of declaring it a failure in blinding, these results indicate high “response bias.” This typifies the situation in which a majority of study participants tends to (or wants to) believe that they were assigned an active or a more effective intervention.35

In conclusion, we contend that our study of the efficacy of acupuncture in subacute stroke rehabilitation, the first of its kind to use a credible sham acupuncture control, has adequately addressed the issue of a specific effect of acupuncture needling alone on recovery of ADL. Our findings are consistent with those from other randomized trials of acupuncture in stroke rehabilitation in identifying no beneficial effect on recovery in ADL and health-related quality of life.24-26 In addition, our data suggest that, should further studies be considered (which would itself be of debatable value), it may be appropriate to confine them to more severely affected patients.

Accepted for Publication: April 25, 2005.
Correspondence: Jongbae Park, KMD, PhD, Division for Research and Education in Complementary and Integrative Medical Therapies, Harvard Medical School Osher Institute, Landmark Center, 401 Park Dr, Suite 22A-West, Boston, MA 02215 (jongbae_park@hms.harvard.edu).
Financial Disclosure: Dr Park developed the Park Sham Device and supplied the samples for use in the study.
Funding/Support: This study was supported by an equipment grant from the Stroke Association, London, England; a grant from the Teresa Rosenbaum Golden Charitable Trust, Edgware, England, for travel expenses; a Chevening scholarship from the British Foreign Ministry–Korean Government, Seoul (Dr Park); a Korean Medicine Development Scholarship, provided by Hye Jung Lee, KMD, PhD, of Kyung Hee University, Seoul, Korea; a Baekrok Scholarhip, provided by Youngkwan Kim, KMD, PhD; the Dr Susil Kumar and Jamila Mitra Charitable Trust (UK), London; and ILMAEK Korean Research Fellowship from the ILMAEK Medical Foundation, Seoul.

Acknowledgment: We acknowledge the contribution of Vanessa Stedman, MRCP, in the blinded assessment of the patients and Heejung Bang, PhD, for statistical analysis of blinding assessment.


