Original Contribution

Pragmatic Randomized Trial Evaluating the Clinical and Economic Effectiveness of Acupuncture for Chronic Low Back Pain

Claudia M. Witt[1], Susanne Jena[1], Dagmar Selim[1], Benno Brinkhaus[1], Thomas Reinhold[1], Katja Wruck[1], Bodo Liecker[2], Klaus Linde[3], Karl Wegscheider[4], and Stefan N. Willich[1]

[1] Institute for Social Medicine, Epidemiology, and Health Economics, Charité University Medical Center, Berlin, Germany.
[2] Techniker Krankenkasse, Hamburg, Germany.
[3] Centre for Complementary Medicine Research, Department of Internal Medicine II, Technical University, Munich, Germany.
[4] Institute of Statistics and Econometrics, University of Hamburg, Hamburg, Germany.

In a randomized controlled trial plus a nonrandomized cohort, the authors investigated the effectiveness and costs of acupuncture in addition to routine care in the treatment of chronic low back pain and assessed whether the effects of acupuncture differed in randomized and nonrandomized patients. In 2001, German patients with chronic low back pain were allocated to an acupuncture group or a no-acupuncture control group. Persons who did not consent to randomization were included in a nonrandomized acupuncture group. All patients were allowed to receive routine medical care in addition to study treatment. Back function (Hannover Functional Ability Questionnaire), pain, and quality of life were assessed at baseline and after 3 and 6 months, and cost-effectiveness was analyzed. Of 11,630 patients (mean age = 52.9 years (standard deviation, 13.7); 59% female), 1,549 were randomized to the acupuncture group and 1,544 to the control group; 8,537 were included in the nonrandomized acupuncture group. At 3 months, back function improved by 12.1 (standard error (SE), 0.4) to 74.5 (SE, 0.4) points in the acupuncture group and by 2.7 (SE, 0.4) to 65.1 (SE, 0.4) points among controls (difference = 9.4 points (95% confidence interval 8.3, 10.5); $p < 0.001$). Nonrandomized patients had more severe symptoms at baseline and showed improvements in back function similar to those seen in randomized patients. The incremental cost-effectiveness ratio was €10,526 (euros) per quality-adjusted life year. Acupuncture plus routine care was associated with marked clinical improvements in these patients and was relatively cost-effective.

acupuncture; back pain; complementary therapies; cost-benefit analysis; health care economics and organizations; low back pain; randomized controlled trials

Low back pain has a lifetime prevalence of more than 70 percent of the population in Western industrialized countries and represents a major economic burden (1). Although 90 percent of acute episodes resolve within 6 weeks, up to 7 percent of patients develop chronic pain (1). The therapeutic management of chronic low back pain varies widely, and the effectiveness of a number of conventional standard treatments has yet to be clearly established (2). According to recent surveys, the number of patients with chronic pain who use complementary and alternative medicine is growing (3–5). Acupuncture is a widely used treatment, especially for chronic low back pain (3), and back pain is the most common reason for visits to acupuncturists (6). However, systematic reviews have shown that the evidence

Correspondence to Dr. Claudia M. Witt, Institute for Social Medicine, Epidemiology, and Health Economics, Charité University Medical Center, 10098 Berlin, Germany (e-mail: claudia.witt@charite.de).

produced by previous trials on the effectiveness of acupuncture is inconclusive because of inconsistent results, low methodological quality, and small sample sizes (7–9). In contrast, two recent meta-analyses demonstrated that acupuncture is more effective than sham treatment for short-term pain relief (10, 11). However, the majority of previous trials were designed as experimental studies; as a result, there is currently very little information about the effectiveness and cost-effectiveness of acupuncture in general medical practice.

In Germany, acupuncture is administered primarily by physicians. Before the year 2000, a number of German health insurance companies covered the costs of acupuncture treatment, at least in part. Under increasing budgetary pressure, however, the German Federal Committee of Physicians and Health Insurers proposed in 2000 that large research initiatives on acupuncture be conducted for several pain syndromes and recommended that patients be reimbursed for the costs of acupuncture only if they participated in such studies (12). We designed the present study as a pragmatic trial to investigate the effectiveness and cost of acupuncture in addition to routine care among patients with chronic low back pain, as compared with routine care alone. In addition, we examined whether the effects of acupuncture differ in randomized and nonrandomized patients, whether treatment effects last over a longer period of time in acupuncture patients, and whether specific patient and physician characteristics are associated with particular treatment outcomes.

MATERIALS AND METHODS

Design

The Acupuncture in Routine Care Study consisted of a multicenter, randomized controlled trial and a nonrandomized cohort. Patients who agreed to randomization were allocated to an acupuncture group that received immediate acupuncture treatment or a control group that received delayed acupuncture treatment 3 months later. Patients who declined to be randomized were included in a third arm of the study and also received immediate acupuncture treatment (nonrandomized acupuncture group). The study period per patient was 6 months and comprised a 3-month treatment phase followed by 3 months of follow-up.

The Acupuncture in Routine Care Study is part of a large acupuncture research initiative funded by a group of social health insurance funds that provide coverage to approximately 10 percent of the German population. The study protocol was approved by the local ethics review boards, and the study itself was conducted according to standard guidelines (i.e., the Declaration of Helsinki and European Epidemiology Federation Good Epidemiological Practice) (13). All study participants provided written, informed consent.

Patients

Patients insured by one of the participating social health insurance funds were recruited after they contacted a participating physician because of chronic low back pain. If a pa-

tient requested acupuncture or if the physician considered acupuncture to be a suitable treatment option, the patient was informed about the study. Subjects who met the inclusion criteria and provided informed consent were randomized using a central telephone randomization procedure. For randomization, we used blocks of 10 patients, and the random list was generated with SAS software (SAS Institute, Inc., Cary, North Carolina).

Patients were included in the study only if we received both the physician's baseline questionnaire and the patient's consent form following randomization. Upon successful inclusion in the study, patients were sent the baseline questionnaire by standard mail. To be included in the study, a patient had to meet the following criteria: clinical diagnosis of chronic low back pain with disease duration of more than 6 months; age ≥ 18 years; and provision of written informed consent. The exclusion criteria were: protusio or prolapse of one or more intervertebral discs with concurrent neurologic symptoms; prior vertebral column surgery; infectious spondylopathy; low back pain caused by inflammatory, malignant, or autoimmune disease; congenital deformation of the spine, except for slight lordosis or scoliosis; compression fracture caused by osteoporosis; spinal stenosis; and spondylolysis or spondylolisthesis.

Interventions

Physicians interested in participating in the study were required to have at least an A-diploma, a German diploma representing 140 hours of certified acupuncture education. This education and other training include wide variations in style and acupuncture technique.

Each patient received a maximum of 15 acupuncture sessions. To assess the effectiveness of acupuncture in general medical practice, we left the acupuncture points and the number of needles used to the discretion of each physician. Only needle acupuncture (with disposable one-time needles and manual stimulation) was allowed; other forms of acupuncture treatment, such as laser acupuncture, were not permitted. In all three treatment groups, the patients were allowed to use additional conventional treatments as needed.

In accordance with German federal regulations, the participating social health insurance funds covered 100 percent of the acupuncture costs for patients who agreed to randomization and 90 percent of the costs for patients who participated in the study but did not agree to randomization.

Outcome measurements

Patients completed standardized questionnaires, which included questions on sociodemographic characteristics, at baseline and after 3 or 6 months. Because of the large number of patients participating in the nonrandomized acupuncture group, a random sample of 50 percent received the questionnaire after 6 months. The primary outcome measure was back function at 3 months, as assessed by the validated Hannover Functional Ability Questionnaire (HFAQ; in German, Funktionsfragebogen Hannover Rücken) (14). The HFAQ rates back function on a scale from 0 to 100, with 100 representing perfect back function.

In order to give this variable the same orientation as the back pain scale (higher values indicating worse outcomes), we used the difference between 100 and the HFAQ value for analysis and called it "back function loss." As an outcome measure, we used the percent reduction of back function loss, since in all three groups reductions were roughly proportional to the baseline back function loss. If back function loss increased for an individual patient during follow-up, the percentage was calculated with respect to the maximum possible loss and given a negative sign. Patients who showed an improvement of at least 20 percent for the variable "back function loss" were considered to be treatment responders.

As further secondary outcome parameters after 3 months, we calculated percent reductions for the Low Back Pain Rating Scale (15) and absolute changes from baseline for Medical Outcomes Study 36-Item Short Form (SF-36) component scales (16) to assess health-related quality of life. Data on the adjunctive use of analgesics were provided by the participating health insurance companies. Side effects were evaluated using patient and physician questionnaires after 3 months. In order to study the durability of any therapeutic benefit in the acupuncture groups and the effect of delayed acupuncture treatment in the control group, changes from baseline to 6 months were calculated analogously.

Statistics

Confirmatory testing of the primary and secondary outcome measures (carried out with SPSS, version 11.5; SPSS, Inc., Chicago, Illinois) was based on the entire study population using the full data set. Sensitivity analyses were performed for the primary outcome measure, either by replacing missing data according to the "last value carried forward" principle or by using various hot deck methods or regression-based multiple imputation. The test procedure was performed in order to maintain a global significance level of $\alpha = 5$ percent. Using covariance analysis, we tested the two-sided null hypothesis, H_0: mean HFAQ score after 3 months for the acupuncture group = mean HFAQ score after 3 months for controls. With 366 patients per group and $\alpha = 5$ percent, the study would have had 90 percent power to detect a difference of 12 percent (9.36 points) in the back function score (HFAQ), assuming a mean score of 78 points and a standard deviation of 39 in the control/acupuncture group. However, we decided to increase the number of patients per group to 1,500, thus allowing a larger number of physicians to participate and increasing the applicability of our findings, as well as the reliability of our analysis of possible predictors.

To identify factors affecting improvements in back function or back pain and to better understand the patient selection due to the acceptance of randomization, we fitted linear mixed models for back function loss and back pain to the data of all study patients. We chose mixed models to comply with the potential cluster structure of the data, because several patients were included by the same physician. As potential regressors, we prespecified several characteristics of the physicians (age, years of professional experience, type of acupuncture diploma, hours of acupuncture training, years of acupuncture experience, diagnosis in the context

of traditional Chinese medicine, and percentage of practice time with acupuncture treatment) and several characteristics of the patients (sex, age, education, baseline physical and mental quality-of-life scores, back function and back pain, the duration of complaints prior to the study, and the study group to which each patient was assigned) before the study started. For the final model, we selected significant variables in a stepwise backwards procedure based on likelihood ratio tests. In addition, we considered the selected regressors as potential modifiers of the acupuncture effect and added the corresponding interaction terms to the model, backwards-selecting them if they were significant. All reported p values are two-sided.

Economic analyses

The cost perspective was societal. Data analysis included 1) overall costs during the 3 months after randomization (including costs not related to chronic low back pain) and 2) only diagnosis-specific costs, using *International Classification of Diseases*, Tenth Revision, codes to identify costs due to chronic low back pain and related conditions. Direct health-related costs for physician visits, hospital stays, medication, acupuncture treatment, and number of sick-leave days were provided by the participating social health insurance funds. Because the observation period was 6 months in length, there was no need to discount any costs or effects.

We compared costs between the two randomized groups and performed a cost-effectiveness analysis. The SF-36 (16) data at baseline and after 3 months were transformed to Short Form-6D data using the algorithm developed by Brazier et al. (17). Only patients with complete SF-36 data were included in the cost-effectiveness analysis. We calculated the number of quality-adjusted life years (QALYs) gained by adopting the area-under-the-curve method (18, 19), using the following formula:

$$
\begin{aligned}
\text{QALY utility}_{\text{gained}} \\
= \left(\alpha_{\text{Acupuncture}} + \left(\frac{\beta_{\text{Acupuncture}} - \alpha_{\text{Acupuncture}}}{2} \right) \right) \\
- \left(\alpha_{\text{Control}} + \left(\frac{\beta_{\text{Control}} - \alpha_{\text{Control}}}{2} \right) \right) \\
= \left(\frac{\alpha_{\text{Acupuncture}} + \beta_{\text{Acupuncture}}}{2} \right) \\
- \left(\frac{\alpha_{\text{Control}} + \beta_{\text{Control}}}{2} \right).
\end{aligned}
$$

This type of analysis is based on the utility values at each time point (α = baseline utility, β = utility after 3 months) and uses the common assumption of a linear change over time (18). Because the health economic section of our study was designed to focus on estimation rather than on hypothesis testing, we calculated the incremental cost-effectiveness ratio (20). Nonparametric bootstrapping was used to create a measure of uncertainty around the estimated incremental cost-effectiveness ratio. We bootstrapped the original sample 1,000 times in order to obtain a graphical overview
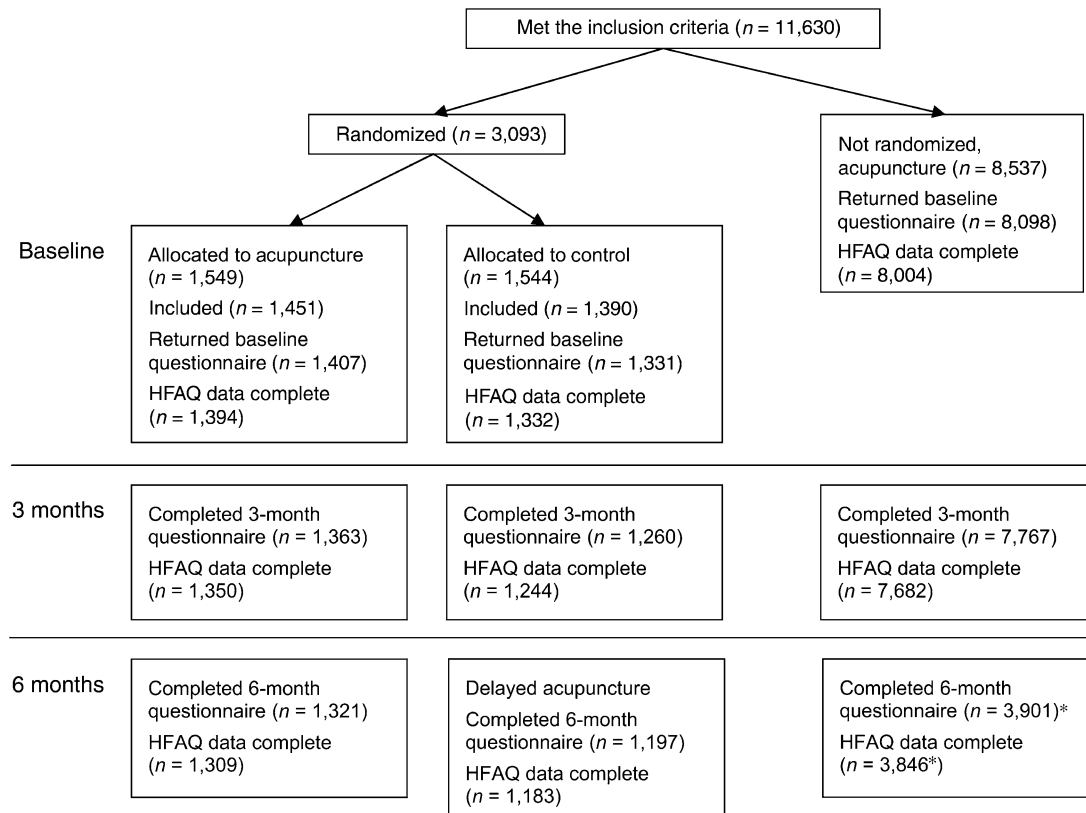
FIGURE 1. Trial flow chart of a study on the effectiveness and cost of acupuncture in the treatment of chronic low back pain, Germany, January 2001–October 2004. (*In the nonrandomized acupuncture group, a random sample of 50% received the questionnaire after 6 months.) HFAQ, Hannover Functional Ability Questionnaire.

of the position of such bootstrapped incremental cost-effectiveness ratios on the cost-effectiveness plane.

The net benefit approach (21) was used to measure the acceptability curve of incremental cost-effectiveness against a societal threshold value λ, which is often described as society's willingness to pay for one extra QALY gained. In the United Kingdom, a threshold of £30,000 (pounds) per QALY is found to be consistent with decisions for adopting new technologies (22). Because such a threshold does not yet exist in Germany, we used an arbitrary and hypothetical threshold of €50,000 (euros) per QALY. For a given value of λ, an intervention would be considered cost-effective if its net benefit were greater than zero—in other words, if the incremental cost-effectiveness ratio were below λ. Thus, a new treatment should be adopted if the net benefit under λ is greater than zero (23).

RESULTS

Patient inclusion, baseline characteristics, and treatment

Between January and September of 2001, a total of 11,630 patients with chronic low back pain were recruited

for the study by 3,486 study physicians (see figure 1 for patient selection). A total of 3,093 patients accepted randomization and were allocated to the acupuncture group or the control group. Two hundred and fifty-two patients (98 acupuncture, 154 control) could not be included in the analysis because the study office did not receive their consent forms. The remaining 11,378 patients (1,451 acupuncture, 1,390 control, and 8,537 nonrandomized acupuncture) were included in the analysis. After 3 months, data were available for 91 percent of the patients (1,363 acupuncture, 1,260 control, and 7,767 nonrandomized acupuncture).

The randomized groups were comparable with regard to all baseline characteristics (see table 1). Between the randomized and nonrandomized acupuncture groups, however, there were significant differences. Patients who did not consent to randomization were more likely to have had more than 10 years of schooling than randomized patients. Nonrandomized patients were also more likely to have experienced more severe complaints, and they had, on average, suffered from chronic low back pain for 1 year less than randomized patients. During the first 3 months, patients in the acupuncture groups received 10.4 (standard deviation (SD), 3.0) acupuncture sessions (randomized acupuncture, 10.4 (SD, 2.6); nonrandomized acupuncture, 10.3 (SD, 3.1); $p = 0.40$). Most patients (74 percent) received 5–10

TABLE 1.   Baseline characteristics of subjects in a study of acupuncture for treatment of chronic low back pain, Germany, January 2001–October 2004*

Parameter	Randomized groups			Acupuncture groups		p value†	Total (N = 11,378)
	Acupuncture (n = 1,451)	Control (n = 1,390)	p value	Randomized (n = 2,841)	Nonrandomized (n = 8,537)		
Female sex (no. and %)	837 (57.7)	791 (56.9)	0.675	1,627 (57.3)	5,061 (59.3)	0.167	6,689 (58.8)
Age (years)	53.1 (13.5)	52.6 (13.2)	0.370	52.8 (13.3)	52.9 (13.8)	0.731	52.9 (13.7)
>10 years of schooling (%)	25.8	29.2	0.083	27.4	30.8	0.008	26.0
Duration of disease (years)	7.2 (8.0)	7.2 (7.8)	0.820	7.2 (7.9)	6.1 (7.6)	<0.001	6.3 (7.8)
Back function score (HFAQ‡)	61.8 (21.0)	63.3 (20.8)	0.067	62.5 (20.9)	60.6 (22.0)	<0.001	61.1 (21.7)
Back pain score (Low Back Pain Rating Scale§)	25.5 (12.3)	25.0 (12.1)	0.327	25.3 (12.2)	26.0 (12.4)	0.005	25.8 (12.4)
Quality of life score (SF-36‡)							
Physical component	34.3 (9.0)	34.6 (9.6)	0.463	34.5 (9.3)	33.8 (9.1)	0.002	34.0 (9.2)
Mental component	43.3 (10.3)	43.5 (10.1)	0.544	43.4 (10.2)	43.5 (10.2)	0.574	43.5 (10.2)

* All data shown are mean values with standard deviations in parentheses, except for sex (number and percentage) and schooling (percentage).

† Two-sided t test or χ^2 test.

‡ HFAQ, Hannover Functional Ability Questionnaire; SF-36, Medical Outcomes Study 36-Item Short Form.

§ Lower values indicate less pain.

sessions, whereas 21 percent received more than 10 sessions and 5 percent received fewer than five sessions.

Randomized comparisons

The course of back function in the two randomized groups, as well as in the nonrandomized group, is depicted in figure 2. In the primary analysis after 3 months, back function improvement was more pronounced in the acupuncture group than in the control group (mean HFAQ scores increased by 12.1 (standard error (SE), 0.4) to 74.5 (SE, 0.4) points in the acupuncture group and by 2.7 (SE, 0.4) to 65.1 (SE, 0.4) points in the control group; difference = 9.4 points (95 percent confidence interval (CI): 8.3, 10.5); $p < 0.001$), after adjustment for baseline differences.

This improvement was robust in the sensitivity analyses for missing data; indeed, the smallest difference between the acupuncture and control groups was 7.0 ($p < 0.001$). The proportion of responders was 52.6 percent in the acupuncture group as compared with 26.8 percent in the control group ($p < 0.001$), the absolute risk reduction was 25.8 percent, and the number needed to treat was 4.

For back function loss, back pain, and quality of life (on both SF-36 component scores), 3-month improvement was significantly more pronounced in the acupuncture group than in the control group (see table 2). There were no significant differences between the acupuncture and control groups with regard to the number of patients prescribed analgesics during the 3 months following randomization (acupuncture group, 21.1 percent of patients; control group, 22.7 percent of patients; $p = 0.29$).
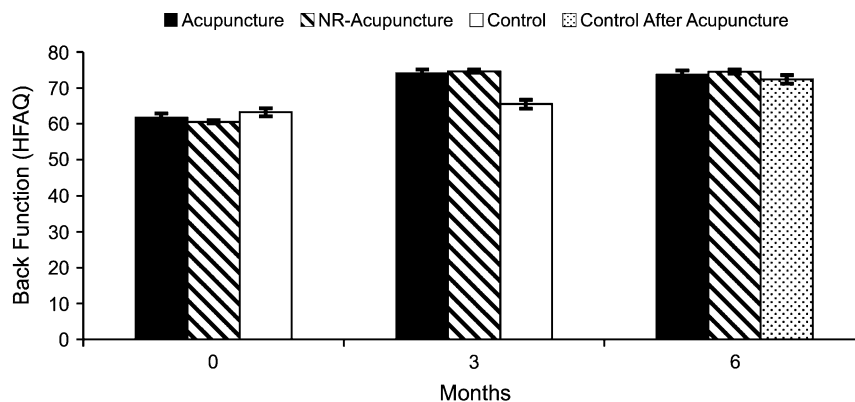
FIGURE 2.   Course of back function (mean Hannover Functional Ability Questionnaire (HFAQ) score) in three groups of participants at baseline, 3 months, and 6 months (between 3 months and 6 months, the controls received acupuncture), Germany, January 2001–October 2004. Bars, 95% confidence interval. NR, nonrandomized.

TABLE 2. Back function and secondary outcomes for three treatment groups in a study of acupuncture and chronic low back pain (3- and 6-month changes from baseline), Germany, January 2001–October 2004*

Parameter	Randomized groups				Nonrandomized group	Nonrandomized vs. randomized	
	Acupuncture	Control	Acupuncture vs. control				
			Difference	p value†		Difference	p value†
Back function and back pain	Reduction (%)	Reduction (%)			Reduction (%)		
3-month change from baseline							
Back function loss (HFAQ‡)	33.3 (31.4, 35.3)	11.3 (9.5, 13.1)	22.0 (19.3, 24.7)	<0.001	35.5 (34.7, 36.4)	2.2 (0.1, 4.4)	0.044
Back pain loss (Low Back Pain Rating Scale§)	37.0 (35.2, 38.9)	9.8 (7.9, 11.7)	27.2 (24.5, 20.9)	<0.001	39.4 (38.6, 40.1)	2.3 (0.3, 4.3)	0.023
6-month change from baseline¶							
Back function loss (HFAQ)	32.4 (30.3, 34.4)	28.6 (26.5, 30.8)	3.7 (0.7, 6.7)	0.015	35.5 (34.3, 36.8)	3.2 (0.7, 5.7)	0.012
Back pain loss (Low Back Pain Rating Scale§)	33.5 (31.4, 35.7)	30.8 (28.7, 33.0)	2.7 (−0.3, 5.7)	0.082	37.0 (35.7, 38.2)	3.4 (0.9, 5.9)	0.007
Quality of life (SF-36‡)	Mean increase (points)	Mean increase (points)			Mean increase (points)		
3-month change from baseline							
Physical component score	7.0 (6.5, 7.5)	2.3 (1.8, 2.8)	4.7 (4.0, 5.4)	<0.001	7.9 (7.7, 8.1)	0.8 (0.3, 1.4)	0.004
Mental component score	2.4 (1.9, 2.9)	0.3 (−0.2, 0.8)	2.1 (1.4, 2.8)	<0.001	2.1 (1.9, 2.4)	−0.2 (−0.8, 0.3)	0.356
6-month change from baseline¶							
Physical component score	6.9 (6.3, 7.4)	6.3 (5.8, 6.9)	0.6 (−0.2, 1.3)	0.154	8.0 (7.6, 8.3)	1.1 (0.4, 1.7)	0.002
Mental component score	1.5 (1.0, 2.0)	1.3 (0.8, 1.9)	0.2 (−0.6, 1.0)	0.607	1.7 (1.4, 2.0)	0.2 (−0.4, 0.8)	0.535

* All data are percentages or mean point increases, with 95% confidence intervals shown in parentheses.
† Two-sided *t* test.
‡ HFAQ, Hannover Functional Ability Questionnaire; SF-36, Medical Outcomes Study 36-Item Short Form.
§ Lower values indicate less pain.
¶ The control group also received acupuncture.

Nonrandomized comparisons

At 3 months, back function improvement was more pronounced in the nonrandomized acupuncture group than in the randomized acupuncture group (mean HFAQ scores increased by 12.1 (SE, 0.4) to 74.5 (SE, 0.4) points in the randomized acupuncture group and by 14.6 (SE, 0.3) to 75.9 (SE, 0.2) points in the nonrandomized acupuncture group; difference = 1.5 points (95 percent CI: −2.4, −0.5); $p < 0.003$).

Comparing the randomized and nonrandomized acupuncture groups (table 2) for the other outcome parameters after 3 months revealed that the effect of acupuncture was more pronounced in nonrandomized patients with regard to back function loss, back pain, and physical quality of life, whereas mental quality of life was similar in both groups. The proportion of responders was 53.0 percent in the nonrandomized acupuncture group and 52.6 percent in the randomized acupuncture group ($p = 0.75$).

Factors affecting 3-month back function or pain improvement scores

After adjustment for all other variables, back function loss was significantly more reduced among men, younger patients, patients with more education (>10 years), patients with reduced baseline back function or less back pain, patients with higher baseline physical or mental quality of life, and (regardless of whether the patient received acupuncture) patients whose treating physician had greater experience in acupuncture treatment, as measured by the percentage of acupuncture treatments performed as part of daily clinical practice. Back pain reduction was significantly more pronounced among women, patients with more education, younger patients, patients with higher baseline back pain, patients with a higher baseline physical or mental quality of life, and patients whose treating physician had greater experience in acupuncture treatment.

After inclusion of these baseline variables, the back function loss or back pain differences between the randomized and nonrandomized acupuncture patients disappeared. In other words, the selection due to the randomization requirement could be completely explained by the model. However, three acupuncture effect modifiers could be identified: The acupuncture effects on back function as well as on back pain were more pronounced in patients with worse initial back function ($p < 0.001$) and younger patients ($p < 0.001$). Furthermore, the acupuncture effect on back function was higher in patients with more than 10 years of schooling ($p = 0.01$). The physician's acupuncture qualifications (hours of training, years of experience) had no significant influence on the effect of the treatment.

Durability of acupuncture effects over 6 months

The 6-month follow-up results in the treatment groups are shown in table 2. The 6-month effects in the randomized and nonrandomized acupuncture groups were only slightly lower than they had been at 3 months. In the randomized acupuncture group, the 6-month response rate was 50.0 percent. In the nonrandomized acupuncture group, the response rate was 52.8 percent.

Delayed acupuncture

Of the 1,390 control patients included in the study, 1,205 (87 percent) received delayed acupuncture. Following delayed acupuncture, the improvement in back function loss seen in control patients almost equaled that observed in patients who had been randomized to receive immediate acupuncture therapy (table 2). However, immediate acupuncture was significantly superior to delayed acupuncture with regard to back function loss and the 6-month response rate (50.0 percent vs. 45.5 percent; $p = 0.02$). In contrast, back pain reduction was not significantly different between the groups.

Side effects

In total, 6 percent of patients ($n = 646$) reported experiencing side effects after acupuncture; the total number of side effects was 767 (54 percent of patients had minor local bleeding or hematoma, 17 percent had pain (e.g., needling pain), 8 percent had vegetative symptoms, and 21 percent had other side effects). No life-threatening side effects were reported.

Economic analyses

Data on QALYs were available for 2,388 of the 2,841 randomized patients (84 percent; 1,231 acupuncture, 1,157 control). As a result, 2,388 patients were included in the economic analysis.

From baseline to 3 months, we observed significant differences in overall and diagnosis-specific costs between the acupuncture and control groups (€1,062.46 (SD, 1,539.74) vs. €782.36 (SD, 1,728.80) ($p < 0.001$) and €557.15 (SD, 872.94) vs. €251.91 (SD, 1,065.41) ($p < 0.001$), respectively). The mean difference between the two treatment groups (€280.10 (95 percent CI: 148.42, 411.78) vs. €305.24 (95 percent CI: 226.79, 383.68)) was essentially due to the costs of acupuncture (€366.95 (SD, 84.90)) in the acupuncture group, whereas no significant differences were observed for other cost components.

Table 3 shows QALY utility values at baseline and after 3 months. There were no significant differences between the two randomized groups at baseline. After 3 months, QALY utility values were higher in the acupuncture group than in the control group (0.65 QALYs (SD, 0.10) vs. 0.62 QALYs (SD, 0.10); $p < 0.001$).

The incremental cost-effectiveness ratios were estimated to be €10,526 per QALY gained (overall cost perspective) and €11,470 per QALY gained (diagnosis-specific cost per-

TABLE 3. QALY* utility values at different time points in a study of acupuncture and chronic low back pain, Germany, January 2001–October 2004

| Time point	Acupuncture group		Control group		*p* value†
	Mean	SD*	Mean	SD	
At baseline (α)	0.60	0.11	0.61	0.11	0.455
At 3 months (β)	0.69	0.12	0.63	0.12	<0.001
Over duration of study: $(\alpha - \beta)/2$	0.65	0.10	0.62	0.10	<0.001

* QALY, quality-adjusted life years; SD, standard deviation.
† Two-sided *t* test.

spective). Bootstrapping of the original sample showed that acupuncture in addition to routine care was more effective and more costly than routine care alone (figure 3). The net benefit for a benchmark of €50,000 was €1,050, and the probability that acupuncture was cost-effective was close to 100 percent (figure 4).

DISCUSSION

Patients with chronic low back pain treated with acupuncture in addition to routine care showed significant improvements in symptoms and quality of life compared with patients who received routine care alone. Acupuncture plus routine care was associated with higher costs but was
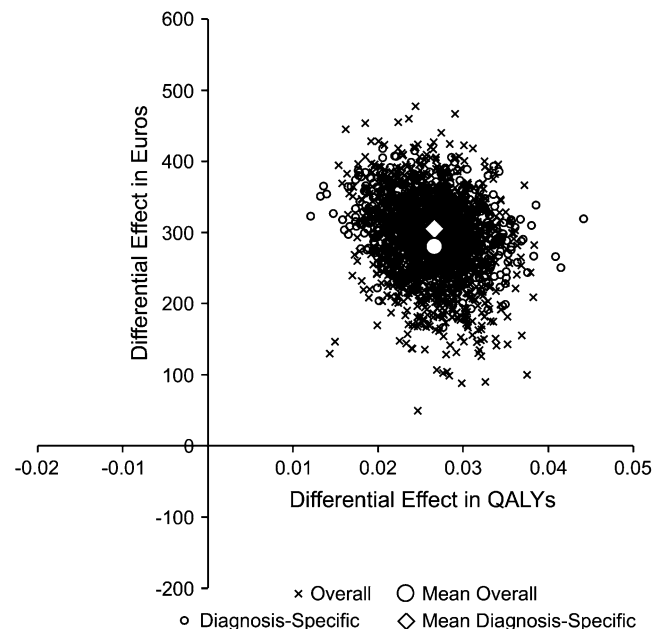
FIGURE 3. Cost-effectiveness of acupuncture in the treatment of chronic low back pain, Germany, January 2001–October 2004. The graph shows diagnosis-specific and overall cost-effectiveness ratios (1,000 bootstrapped samples each). QALYs, quality-adjusted life years.
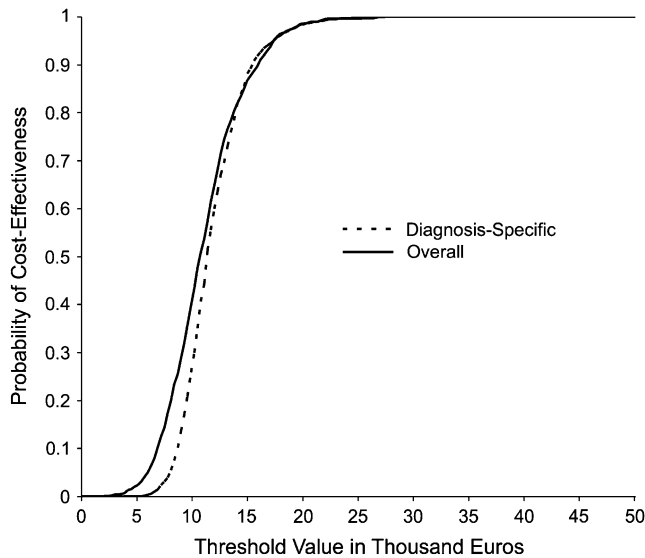
FIGURE 4. Acceptability curve showing the cost-effectiveness of acupuncture in the treatment of chronic low back pain for threshold values of up to €50,000, Germany, January 2001–October 2004.

estimated to be cost-effective. In patients who consented to randomization, treatment outcomes after acupuncture were similar to those seen in patients who declined randomization. All of the differences observed can be explained by differences at baseline.

The present study is by far the largest randomized trial of acupuncture in patients with chronic low back pain to date, including 12 percent of physicians specializing in acupuncture and a full 3 percent of all primary care physicians in Germany. We took a pragmatic approach, aiming to evaluate acupuncture in a manner that would reflect as closely as possible the conditions of daily medical practice and maximize external validity. The additional inclusion of patients who declined randomization allowed us to investigate any potential selection effects. Although the study had high follow-up rates, we nevertheless used conservative methods to deal with missing data.

Obviously, such an approach also has its methodological limitations. In this study, neither providers nor patients were blinded to treatment. Therefore, a bias due to unblinding cannot be ruled out. To minimize social acceptability bias, all questionnaires were sent directly to and from the coordinating research institute. The numbers of patients who used analgesic agents during the study were similar in the two randomized groups. Because both the specifics of acupuncture treatment and the specifics of any cointerventions were left to the discretion of the physicians, the treatment regimens in our study were highly variable. Moreover, inclusion criteria were broad, which resulted in a heterogeneous patient sample and possibly some diagnostic misclassification. While these issues might be considered limitations from an experimental perspective, the study design was chosen to reflect general medical practice. Another limitation of our study was that it contained comparative data for only 3 months.

Patient self-selection in randomized studies of complementary and alternative medical treatments could be a relevant problem (24). Although a variety of study designs have been recommended for including both randomized and nonrandomized patients, few studies have actually employed them to date (25). In our investigation, approximately three out of four eligible patients refused randomization, in spite of a minor financial incentive and the seemingly slight disadvantage of having a 50 percent chance of a 3-month delay before starting acupuncture treatment (following an average disease duration of 6.3 years). Although differences in baseline characteristics and treatment outcomes between randomized and nonrandomized patients were small in absolute numbers, our findings indicate that randomization was associated with some selection. This pragmatic study included a large sample of "normal" patients with low back pain in Germany. The patients received acupuncture from physicians who also offered other treatments, including conventional medicine. As a result, our study may have been less subject to selection bias than smaller experimental studies of acupuncture. Therefore, the use of study designs that also include nonrandomized patients appears to be desirable.

It is notable, however, that the benefits of treatment were similar in the randomized and nonrandomized acupuncture groups after adjustment for baseline differences. This suggests that the results of randomized trials can be representative for routine medical care situations, at least in large pragmatic studies. This finding is supported by a review which found that randomized controlled trials and observational studies of acupuncture had comparable results (26).

In both randomized and nonrandomized patients, the improvements seen in back function were clinically relevant (14). An important finding of our study is that the improvements seen immediately after 3 months of treatment continued for at least another 3 months.

Our finding that the formal qualifications of the physician and his/her number of years of acupuncture experience had no significant influence on treatment outcome could be interpreted as a further indication that formal acupuncture training plays only a limited role with regard to treatment effect. However, these results should be interpreted with caution, because the indicators in the present study may not adequately reflect the quality of treatment offered by individual physicians. In general, our regression analyses identified only three variables that predicted treatment outcome, including younger age, more severe complaints, and more than 10 years of schooling.

Our study provides further evidence that acupuncture is a safe intervention. This is in agreement with large, previously published surveys (27, 28). When interpreting these findings, however, one must keep in mind that all acupuncture in this study was administered by physicians.

Acupuncture is a relatively resource-intensive intervention because of the time involved for physicians and patients alike. One randomized trial, a United Kingdom study of acupuncture in the treatment of chronic headache, has included a methodologically sound analysis of cost-effectiveness (23). Those authors concluded that acupuncture

is relatively cost-effective in comparison with a number of other interventions provided in the United Kingdom. Our study showed that acupuncture was associated with additional costs but was cost-effective according to international threshold values (22, 29). These results are comparable to the findings of Thomas et al. (30), who calculated a cost-effectiveness ratio of approximately €6,500 for acupuncture treatment of low back pain in the United Kingdom. Although a number of studies have claimed to investigate the cost-effectiveness of other treatment options for patients with chronic low back pain (31–34), no incremental cost-effectiveness ratios have been published to date.

In conclusion, our study showed that acupuncture, in addition to routine care, resulted in a clinically relevant benefit and was cost-effective among patients with chronic low back pain from primary care practices in Germany. Thus, acupuncture should be considered a viable option in the management of patients with chronic low back pain.

REFERENCES

1. Speed C. Low back pain. BMJ 2004;328:1119–21.
2. Deyo RA, Weinstein JN. Low back pain. N Engl J Med 2001;344:363–70.
3. Eisenberg DM, Davis RB, Ettner SL, et al. Trends in alternative medicine use in the United States, 1990–1997: results of a follow-up national survey. JAMA 1998;280:1569–75.
4. Härtel U, Volger E. Use and acceptance of classical natural and alternative medicine in Germany—findings of a representative population-based survey. (In German). Forsch Komplementärmed Klass Naturheilkd 2004;11:327–34.
5. Thomas KJ, Nicholl JP, Coleman P. Use and expenditure on complementary medicine in England: a population based survey. Complement Ther Med 2001;9:2–11.
6. Cherkin DC, Deyo RA, Sherman KJ, et al. Characteristics of visits to licensed acupuncturists, chiropractors, massage therapists, and naturopathic physicians. J Am Board Fam Pract 2002;15:463–72.
7. Ernst E, White AR. Acupuncture for back pain: a meta-analysis of randomized controlled trials. Arch Intern Med 1998;158:2235–41.
8. van Tulder MW, Cherkin DC, Berman B, et al. The effectiveness of acupuncture in the management of acute and chronic low back pain. A systematic review within the framework of the Cochrane Collaboration Back Review Group. Spine 1999;24:1113–23.
9. Smith LA, Oldman AD, McQuay HJ, et al. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. Pain 2000;86:119–32.
10. Furlan AD, van Tulder MW, Cherkin DC, et al. Acupuncture and dry-needling for low back pain. Cochrane Database Syst Rev 2005;(1):CD001351.
11. Manheimer E, White A, Berman B, et al. Meta-analysis: acupuncture for low back pain. Ann Intern Med 2005;142:651–63.
12. German Ministry of Health. Bekanntmachung einer Änderung der Richtlinien über die Bewertung ärztlicher Untersuchungs- und Behandlungsmethoden gemäß. §135 Abs. 1. Fünftes Buch Sozialgesetzbuch (SGB V) (BUB-Richtlinien). Bundesanzeiger 2001;18:685.
13. IEA European Epidemiology Federation. Good Epidemiological Practice (GEP): proper conduct in epidemiological research. Dundee, United Kingdom: IEA European Epidemiology Federation, 2004. (http://www.dundee.ac.uk/iea/GoodPract.htm).
14. Kohlmann T, Raspe H. Hannover Functional Questionnaire in ambulatory diagnosis of functional disability caused by backache. (In German). Rehabilitation (Stuttg) 1996;35:I–VIII.
15. Manniche C, Asmussen K, Lauritsen B, et al. Low Back Pain Rating Scale: validation of a tool for assessment of low back pain. Pain 1994;57:317–26.
16. Bullinger M, Kirchberger I. SF-36 Fragebogen zum Gesundheitszustand. Göttingen, Germany: Hogrefe, 1998.
17. Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol 1998;51:1115–28.
18. Richardson G, Manca A. Calculation of quality-adjusted life years in the published literature: a review of methodology and transparency. Health Econ 2004;13:1203–10.
19. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ 2000;320:1197–200.
20. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ 1999;18:341–64.
21. Zethraeus N, Johannesson M, Jonsson B, et al. Advantages of using the net-benefit approach for analysing uncertainty in economic evaluation studies. Pharmacoeconomics 2003;21:39–48.
22. Lothgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. Health Econ 2000;9:623–30.
23. Wonderling D, Vickers AJ, Grieve R, et al. Cost-effectiveness analysis of a randomised trial of acupuncture for chronic headache in primary care. BMJ 2004;328:747–9.
24. Gerhard I, Abel U, Loewe-Mesch A, et al. Problems of randomized studies in complementary medicine demonstrated in a study on mistletoe treatment of patients with breast cancer. (In German). Forsch Komplementärmed Klass Naturheilkd 2004;11:150–7.
25. Britton A, McKee M, Black N, et al. Choosing between randomised and non-randomised studies: a systematic review.

(Health Technology Assessment 1998, vol 2, no. 13). Southampton, United Kingdom: Health Technology Assessment Programme, United Kingdom Department of Health, 1998: 1–124. (http://www.hta.nhsweb.nhs.uk/execsumm/summ213. htm).

26. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000;342:1887–92.
27. White A, Hayhoe S, Hart A, et al. Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists. BMJ 2001;323:485–6.
28. MacPherson H, Thomas K, Walters S, et al. The York acupuncture safety study: prospective survey of 34 000 treatments by traditional acupuncturists. BMJ 2001;323:486–7.
29. Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. BMJ 2001;323:1300–3.
30. Thomas KJ, MacPherson H, Ratcliffe J, et al. Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain. Health Technol Assess 2005;9:iii–iv, ix–x, 1–109.
31. Cherkin DC, Eisenberg D, Sherman KJ, et al. Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage, and self-care education for chronic low back pain. Arch Intern Med 2001;161:1081–8.
32. Cherkin DC, Deyo RA, Battie M, et al. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. N Engl J Med 1998;339:1021–9.
33. Kovacs FM, Llobera J, Abraira V, et al. Effectiveness and cost-effectiveness analysis of neuroreflexotherapy for subacute and chronic low back pain in routine general practice: a cluster randomized, controlled trial. Spine 2002;27: 1149–59.
34. Moffett JK, Torgerson D, Bell-Syer S, et al. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. BMJ 1999;319:279–83.