Spine

SPINE Volume 37, Number 8, pp 701–710

HEALTH SERVICES RESEARCH

Implementation of a Guideline for Low Back Pain Management in Primary Care

A Cost-Effectiveness Analysis

Annette Becker, MD, MPH,* Heiko Held, MD,* Marcus Redaelli, MD,† Jean F. Chenot, MD, MPH, PhD,‡ Corinna Leonhardt, PhD,§ Stefan Keller, PhD,‖ Erika Baum, MD, PhD,* Michael Pfingsten, PhD,¶ Jan Hildebrandt, MD, PhD,¶ Heinz-Dieter Basler, PhD,§ Michael M. Kochen, MD, MPH, PhD, FRCGP,# Norbert Donner-Banzhoff, MD, MHSc, PhD,* and Konstantin Strauch, PhD,**††‡‡

Study Design. Cost-effectiveness analysis alongside a cluster randomized controlled trial.

Objective. To study the cost-effectiveness of 2 low back pain guideline implementation (GI) strategies.

Summary of Background Data. Several evidence-based guidelines on management of low back pain have been published. However, there is still no consensus on the effective implementation strategy. Especially studies on the economic impact of different implementation strategies are lacking.

Methods. This analysis was performed alongside a cluster randomized controlled trial on the effectiveness of 2 GI strategies (physician education alone [GI] or physician education in combination with motivational counseling [MC] by practice nurses)—both compared with the postal dissemination of the guideline (control group, C). Sociodemographic data, pain characteristics, and cost data were collected by interview at baseline and after 6 and 12 months. low back pain–related health care costs were valued for 2004 from the societal perspective.

Results. For the cost analysis, 1322 patients from 126 general practices were included. Both interventions showed lower direct and indirect costs as well as better patient outcomes during follow-up compared with controls. In addition, both intervention arms showed superiority of cost-effectiveness to C. The effects attenuated when adjusting for differences of health care utilization prior to patient recruitment and for clustering of data.

Conclusion. Trends in cost-effectiveness are visible but need to be confirmed in future studies. Researchers performing cost-evaluation studies should test for baseline imbalances of health care utilization data instead of judging on the randomization success by reviewing non-cost parameters like clinical data alone.

Key words: low back pain, cost-effectiveness, primary care, guideline implementation. **Spine 2012;37:701–710**

From the *Department of General Practice, Preventive and Rehabilitation Medicine, University of Marburg, Marburg, Germany; †Institute of General Practice and Family Medicine, Private University Witten/Herdecke gGmbH, Witten, Germany; ‡Department of General Practice, Institute for Community University, University Medicine of Greifswald, Greifswald, Germany; §Institute for Medical Psychology, University of Marburg, Marburg, Germany; ‖Department of Public Health Sciences, University of Hawaii at Manoa, HI; ¶Department of Anaesthesiology, Pain Clinic, University of Göttingen, Göttingen, Germany; #Department of General Practice, University of Göttingen, Göttingen, Germany; ††Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; and ‡‡Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research center for Environmental Health, Neuherberg, Germany.

Acknowledgment date: March 16, 2011. First revision date: June 8, 2011. Acceptance date: June 14, 2011.

The manuscript submitted does not contain information about medical device(s)/drug(s).

German Ministry for Education and Research (BMBF, FKZ 01 EM 0113) funds were received to support this work.

No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Annette Becker, MD, MPH, Department of General Practice, Preventive and Rehabilitation Medicine, University of Marburg, Karl-von-Frisch-St. 4, Marburg 35032, Germany; E-mail: Annette.Becker@staff.uni-marburg.de

DOI: 10.1097/BRS.0b013e31822b01bd

Low Back Pain (LBP) is one of the most prominent problems in the industrialized countries. Epidemiological studies report lifetime prevalences of 70% to 90%.[1,2] Although most patients recover spontaneously, there is a high rate of recurrences (70%) and 7% to 10% of patients develop chronic LBP, which leads to high health care cost due to sick leave and early retirement.[3]

Facing the problem of the characteristic diversity of providers and possible treatment options, current research activities focus on the implementation of evidence-based care for patients with LBP. Although guideline implementation (GI) strategies have been studied with respect to provider adherence and patient outcomes,[4–6] it is still unsure whether these activities actually alter subsequent health care cost. Evidence from economic evaluations is lacking despite the importance of such studies for decision and policy makers when deciding on the most efficient resource use.

The aim of our study was to perform a cost-effectiveness analysis on 2 LBP GI strategies that have been tested in a cluster randomized controlled trial in German general practices.

MATERIALS AND METHODS

Study Design and Data Sources

This study is a secondary analysis of a 3-armed cluster randomized controlled GI study in primary care.[7]

At the index visit, patients were asked to fill out 2 sets of questionnaires, one while waiting and another one at home (sociodemographic and disease-related data, for postal return in a prepaid envelope). One baseline telephone interview (within 4 weeks) and 2 follow-up interviews (after 6 and 12 months) were performed by specially trained study nurses. Data on LBP-related health care resource use were collected during all 3 interviews. To shorten interviews, we had to omit non–medical costs such as travel and time expenses or out-of-pocket costs.

Interventions

Practices were assigned to the 3 study arms by permuted block randomization. General practitioners (GPs) in both intervention groups (GI and GI plus motivational counseling [MC]) were trained in using the LBP guideline of the German College of General Practitioners and Family Physicians (DEGAM)[8]: The guideline consists of 4 modules: a detailed version and a pocket card for physicians, a prescription-like short form information, and a more detailed information flyer for patients. Three interactive seminars for participating GPs were held, including information on performance of the diagnostic triage and identification of red flags (first session), early identification of yellow flags, including general behavioral principles on management of chronic pain patients (second session), and informing and advising patients (third session). The third session gave room for discussion of implementation barriers and individual experiences. All doctors of the intervention groups received information about relevant local facilities for pain patients (self-help groups, fitness clubs, teaching sessions organized by health insurers, specialists, *etc*.). Individual educational visits by study nurses to participating GPs were done twice to hand over the guideline and, after 3 to 6 months, to discuss individual problems with GI.

During the third educational session, GPs of the GI plus MC group were introduced to MC strategies. Two nurses per practice received a 20-hour training (2 full-day workshops and 1–3 supervision sessions) designed to increase the nurses' skills to motivate LBP patients for regular physical activity. Practice nurses were asked to invite all identified patients for up to 3 counseling sessions (maximum 10–15 minutes each), the first session within 1 to 3 weeks after inclusion in the study. They were encouraged to use specifically designed brochures on motivational and behavior change and posters to communicate the key messages. Study coordinators contacted the practice nurses regularly to identify barriers and problems with regard to the implementation of this new counseling strategy.

The control group received the guideline *via* postal mail, which has been proven to be ineffective with respect to patient outcomes.[9]

Study Sample

All patients consulting for LBP were recruited consecutively. Inclusion criteria were LBP on the day of inclusion, aged 20 years and older, ability to read and to understand German, and written consent. Exclusion criteria were pregnancy and isolated thoracic or cervical pain.

Setting

The study was conducted in 2 German regions (Marburg and Göettingen) in 2003 to 2004. Overall, 126 GPs (73 men, mean age of 49 years) belonging to 118 practices (57 group practices and 69 single practices) participated.

Clinical Measures

For description of pain, we asked for the pain intensity (numeric analogue scale), days in pain during the previous year, and a possible radiation of pain. For classification of the natural history of LBP, we used the von Korff scale to grade the severity of chronic pain.[10]

Effect Measures

The main outcome to assess the implementation effectiveness was functional capacity measured with the 12-item Hannover Functional Ability Questionnaire for Measuring Back Pain–Related Functional Limitations at baseline (questionnaire) and after a 6- and 12-month period (interview).[11,12] Normal function shows scores of 80% to 100%; scores around 70% equal a moderately limited function and scores below 60% show a severely limited function.

Secondary outcomes were physical activity during 1 week prior to the interview, days in pain and days of sick leave referring to the 6-month period prior to the 2 follow-up interviews, and quality of life measured with the EuroQol[13] after a 12-month period providing us with a single index value on a visual analogue scale (0–100).

Sociodemographic characteristics were assessed for description of the study sample. Furthermore, GPs reported for each patient the presence of complicating factors (generally unwell, neurological deficits, history of cancer, chronic inflammatory disease, osteoporosis with danger of fracture, immune deficiency, or severe trauma).

Measurement of Health Care Utilization

Consultation of health care providers (GP, specialists), diagnostic and therapeutic procedures, and auxiliaries were given in types and numbers. For pharmaceuticals, information given by patients was initially translated into drug codes. However, for practical reasons we had to change to free-text answers on type, doses, and package size of the medication during follow-up. Data on hospital and rehabilitation were given in days of care and reason for admission (*e.g.*, surgery, pain management).

Valuation of Direct and Indirect Costs

Cost calculations are based on LBP-related health care costs during months 1 to 6 and 7 to 12 during follow-up. All costs were valued for the year 2004 from the societal perspective.

Because we had no data on insurance status, we postulated a 10% rate of privately insured patients for all direct cost categories as this percentage corresponds to the average of privately insured patients in Germany. Physician consultations and diagnostic or therapeutic procedures were priced using provider-specific charges. The costs for drugs were based on package prices according to the official German price list of drugs. Expenditure for hospital care is based on diagnosis-related groups or on department-specific daily charges. Inpatient rehabilitation was valued by sector-specific charges, all data inflated for 2004 by the sector- and state-specific inflation rate. If necessary, we accounted for patient copayment in all cost categories. Cost estimations for auxiliaries are based on average prices (recommended by Krauth *et al*[14]; inflated for 2004) or by personal information from medical supply stores.

For an estimation of indirect costs, we used the human capital approach multiplying the number of missed work hours with the average daily labor cost in Germany.[15] A more detailed description of valuation in costs as well as sources of information may be seen in Becker *et al*.[16]

GI costs were calculated from the study perspective as well as from a societal perspective (Table 1). Costs for guideline development were not accounted for because they refer to all 3 study arms.

Statistical Analysis

Basic statistical analysis was performed with SPSS 17.0 (SPSS, Inc., Chicago, IL) and R (R Development Core Team, http://www.R-project.org/) .[17]

Our sample size calculation was based on changes in the primary outcome functional capacity: Expecting small effects ($f = 0.1$) and a dropout rate of 25%, we aimed for 1874 patients ($\alpha = 0.05$, power $1 - \beta = 80\%$, intracluster correlation $\rho = 0.03$, expected cluster size $n = 16$).

Nonparametric bootstrapping using the program R (with 100,000 simulated replicates) was applied to present means and confidence intervals (CIs) of cost data for each study group. Incremental cost-effectiveness ratios are based on the differences in total costs between the intervention groups and the control group divided by the difference in effects for functional capacity, days in pain, or quality of life. They are (BUGS project, http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml) calculated with bootstrapping (5000 simulated replicates) and presented on a cost-effectiveness plane. Acceptability curves were drawn to show the statistical confidence of cost-effectiveness as a function of willingness to pay, measured in Euro (€) per point on the effectiveness scale.

For sensitivity analysis, we performed Markov-chain Monte Carlo simulations with the WinBUGS software[18] to adjust for direct and indirect costs during the 6 months prior to recruitment as well as for clustering of data. This was done using a Bayesian hierarchical model developed by Grieve *et al*[19] that jointly models a cost and an effectiveness variable as well as their correlation and includes practice clusters by means of a random effect. According to Nixon and Thompson,[20] we further extended this model to take into account costs during the 6 months prior to recruitment as a covariable for both effectiveness and cost. The effectiveness variable, that is, functional capacity, was assumed to be normally distributed. To account for the skewness of cost (here, during the period after recruitment), the model involved a gamma distribution for this variable. Three Markov chains were generated (in total 3×6667, that is, 20,001 replicates that were actually analyzed). Two-dimensional confidence ellipses were derived from the joint distribution of the cost and functional capacity for the replicates.

RESULTS

We finally included 1378 patients from 76 practices (1–20 per practice) in the main study (participation rate: 44%, dropout: 12.1% in 12 months). Reasons for exclusion from the main study are given in Figure 1. For cost analysis, we had to further exclude 55 patients who refused to participate in the baseline interview on health care utilization and 1 patient who showed implausible data on days of sick leave. This left us with 1322 eligible patients for the cost-effectiveness analysis. Baseline and sociodemographic characteristics stratified for study groups are shown in Table 2 and Table 3 and in more detail in Becker *et al*.[7,16]

Absence From Work and Health Care Costs

After 6-month follow-up, 93 patients of the GI group (mean: 23 days; 95% CI, 16–32; median: 10 days) and 90 patients of the GI plus MC group (mean: 25 days; 95% CI, 16–33; median: 12 days) reported absence from work because of LBP during the previous 6-month period. Less patients (n = 81) in the control group showed on average more days off work (mean: 40 days; 95% CI, 29–52; median: 14 days) than those in the intervention groups. At 12-month follow-up, 62 patients of the GI group (mean: 19 days; 95% CI, 11–28; median: 10 days), 51 patients of the GI plus MC group (mean: 22 days; 95% CI, 12–31; median: 10 days), and 53 controls (mean: 29 days; 95% CI, 16–42; median: 10 days) reported absence from work again during the previous 6-month period.

Table 4 shows differences in means of health care costs between pairs of the 3 study groups during follow-up. Results are shown for months 1 to 6 and 7 to 12 after recruitment. Direct, indirect, and total costs during follow-up are markedly lower in both intervention arms compared with controls. However, this difference was already present at baseline (Table 5) and has been accounted for in additional sensitivity analyses.

Intervention Effects

Table 6 presents the results that contribute to the calculation of cost-effectiveness ratios for both intervention groups compared with controls. With respect to intervention effects after 6 and 12 months, improvement of functional capacity was more pronounced (but not significant) in the intervention groups compared with controls. Tendencies in effect remain visible after 12 months. There are no significant effects regarding the other secondary outcomes: physical activity or quality of life.

TABLE 1. Costs Referring to Guideline Implementation

	Study Perspective		Societal Perspective	
	Center I	Center II	Center I	Center II
Guideline implementation				
Educational training for physicians/ sessions: Physician reimbursement (126 physicians, 3 sessions each)	3 sessions/physician (75€ physician/each) 3 × 126 × 75€ = 28,350€		0.78€/physician minute (5), 1.5 h per session • 126 × 3 × 0,78€ × 90 min = 26,535.60€	
Educational training for physicians/ session: Moderators (3 sessions of 1.5 h, 4 times/each center)	1 research assistant, 1 study nurse • Research assistant*: 36 h × 28.53€ = 1027.08€ • Practice nurse†: 36 h × 22.08€ = 794.88€		Trainer (moderator) 183.33€‡ × 3 sessions × 4 times = 2200€	91.67€‡ × 3 sessions × 4 times = 1100€
Educational training for physicians and PN: Rental fee for teaching location	None		2€ per person (126 physicians: 3 × 4 times; 72 PN: 2 sessions) 2€ × (126 × 12 + 72 × 2) = 3312€	
Educational training for physicians and PN: Catering (physicians: 12 sessions; PN: 4 dates in each center)	Overall: 1826.40€	• Physicians (40.90€/session): 40.90€ × 12 = 490.80€ • PN: 127.82€/session 4 × 127.82€ = 511.28€	1826.40€	• Physicians: 40.90€/session, 40.90€ × 12 = 490.80€ • PN: 127. 82€/session 4 × 127.82€ = 511.28€
Purchase fee for guidelines	No cost differences compared with controls who received the guidelines *via* post			
Additional costs for motivational counseling				
Educational training for PN: PN reimbursement (72 PN, 20 h in total)	Overall 75€ per PN 72 × 75€ = 5400€		On the basis of an hourly rate 11.57 €§ × 72 × 20 h = 16660.80€	
Educational training PN: Personnel 10 groups, each 4 times 2 sessions	1 Group moderator (research assistant) 80 h × 28.53€ = 2282.40€		Moderator 183.33‡ × 2 × 4 = 1466.64€	1.67€‡ × 2 × 4 = 733.36€
Patient counseling: Reimbursement PN 585 sessions (489 patients), each 20 min	Overall 7.50€/session/patient 585 × 7.50€ = 4387.50€		On the basis of an hourly rate 585 × 20 min × (11.57€§/60 min) = 2256.15€	
Patient counseling: Patient reimbursement	None			
Implementation cost study arm I (GI)	33000.44€		35976.08€	
Implementation cost study arm II (GI plus MC)	45070.34€		57093.03€	

**Calculation on the basis of 4394€/mo, assuming 38.5 h/wk.*

†Calculation on the basis of 3401€/mo, assuming 38.5 h/wk.

‡AQUA-Institute (Institute for Applied Quality Improvement and Research in Health Care), Göttingen: personal information: 200€ in Hessen, 100€ in Niedersachsen. Prices were discounted by the German average inflation rate.

§Calculation on the basis of 1782.39€/mo, assuming 38.5 h/wk.

PN indicates practice nurses; GL, guideline implementation; MC, guideline implementation plus motivational counseling.

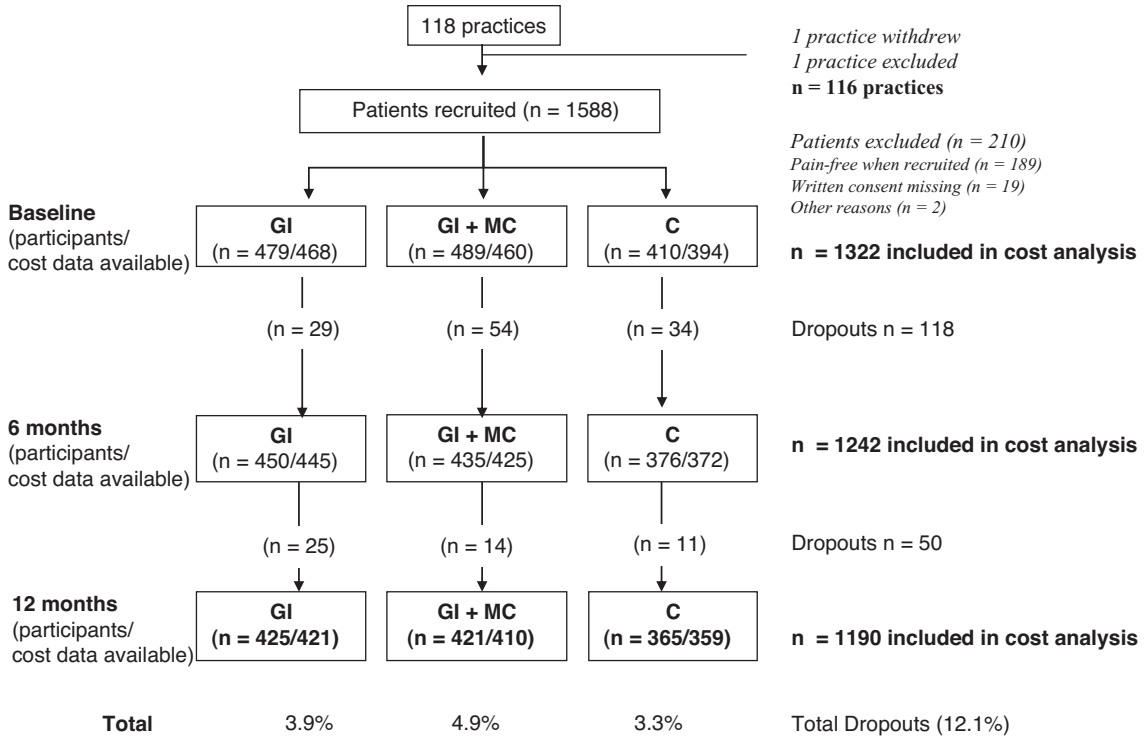

Figure 1. Patient flowchart. Dropouts are given in italics. GI indicates guideline implementation group; GI + MC, guideline implementation plus motivational counseling; C, controls (postal dissemination of the guideline).

Cost-Effectiveness

Taking into account the health care costs during follow-up (months 1–6) both interventions seem less expensive and more effective than the control group concerning the primary outcome functional capacity as well as for days in pain. Figures 2 and 3 show the cost-effectiveness plane for functional capacity comparing GI or GI plus MC with the control group (C). Most bootstrap ratios (84%) appear in the bottom right quadrant meaning lower costs and better functional capacity for GI *versus* C (95% for GI plus MC *vs.* C). The acceptability curves show a maximum statistical confidence of 0.98 for a willingness to pay of 115€ per point on the functional capacity scale (GI plus MC *vs.* C: 0.998 for 134€).

At 12-month follow-up, trends are still visible for all comparisons but less prominent. The maximum confidence of 0.97 for the intervention (GI group *vs.* controls) being

TABLE 2. Sociodemographic Characteristics of Patients by Intervention and Control Groups (n = 1322)

	Intervention Groups		Control Groups
Variables	**GI (n = 468)**	**GI plus MC (n = 460)**	**Postal Dissemination (n = 394)**
Age (mean ± SD, range)*	49.2 yr ± 13.3, 21–83	47.6 yr ± 13.5, 20–91	50.2 yr ± 14.1, 20–81
Sex (males)*	191 (41)	177 (39)	181 (46)
Employment status			
Working full- or part-time	260 (55.6)	256 (55.7)	208 (52.8)
Keeping house	37 (7.9)	46 (10.0)	35 (8.9)
Retired	81 (17.3)	67 (14.6)	77 (19.5)
Unemployed	19 (4.1)	16 (3.5)	17 (4.3)
Other	14 (3.0)	22 (4.8)	13 (3.3)

Values are numbers (column percentage), unless otherwise indicated.

**Significant difference between study groups, $\alpha = 0.05$.*

GI indicates guideline implementation; MC, motivational counseling.

TABLE 3. Baseline Characteristics by Study Arms (n = 1322 Patients)

Variables		Guideline Implementation (n = 468)	Guideline Implementation Plus Motivational Counseling (n = 460)	Postal Dissemination (n = 394)
Functional capacity	Mean (SD)	67.33 (21.37)	69.00 (21.20)	65.64 (22.02)
Pain intensity (NRS 0–10)	Mean (SD)	5.4 (1.74)	5.3 (1.70)	5.27 (2.12)
Days of pain in the previous year	Mean (SD)	101 (132.32)	103 (123.57)	114 (131.71)
Chronic pain grade*, n (%)				
Low disability/low intensity)		98 (29.6)	112 (33.4)	82 (28.9)
Low disability/high intensity		95 (28.7)	83 (24.8)	71 (25.0)
High disability/moderately limiting		88 (26.6)	89 (26.6)	75 (26.4)
High disability/severely limiting		50 (15.1)	51 (15.2)	56 (19.7)
Depression score	Mean (SD)	14.98 (9.37)	15.80 (9.34)	15.15 (9.29)
Quality of life (VAS score, 0–100)	Mean (SD)	56.96 (19.9)	58.57 (18.94)	55.38 (19.03)
Days of sick leave	Mean (SD)	6.08 (18.0)	8.10 (26.39)	10.83 (31.64)
Red flags, n (%)				
Generally unwell		11 (2.3)	8 (1.7)	9 (2.3)
Neurological deficits		4 (0.9)	9 (2.0)	6 (1.5)
History of cancer		4 (0.9)	10 (2.2)	7 (1.8)
Chronic inflammatory disease		9	7 (1.5)	14 (3.5)
Osteoporosis		9 (1.9)	4 (0.9)	16 (4.1)
Fever		1 (0.2)	0	1 (0.3)
Immune deficiency		0	0	0
Severe trauma		1 (0.2)	4 (0.9)	2 (0.5)

More than 20% missing. Percentages refer to patients per group who completed the respective questions.

NRS indicates numeric rating scale; VAS, visual analogue scale.

cost-effective is obtained at a willingness to pay of 67€ per point on the functional capacity scale. At 99€ per point, a maximum cost-effectiveness confidence of 0.99 is reached when comparing the GI plus MC group with controls.

For reasons of simplicity, implementation costs have not been included in the analysis because they occur per practice rather than per patient. They account for about 5% of total costs (Table 1).

TABLE 4. Mean Differences in Health Care Costs After 6 and 12 Months for All Patients

	1–6 mo Follow-up			7–12 mo Follow-up		
	GI to C (n = 825)	GI plus MC to C (n = 810)	GI plus MC to GI (n = 885)	GI to C (n = 789)	GI plus MC to C (n = 785)	GI plus MC to GI (n = 846)
Direct cost	−65.55 (−355; 200)	−179.76 (−467; 31)	−114.21 (−357; 81)	−97.95 (−285; 68)	−97.21 (−281; 58)	0.74 (−145; 139)
Indirect cost	**−332.51 (−650; −44)**	**−302.83 (−621; −6)**	29.68 (−204; 265)	−122.92 (−363; 79)	−114.71 (−354; 90)	8.21 (−165; 179)
Total cost	−398.06 (−901; 60)	**−482.59 (−983; −54)**	−84.53 (−464; 274)	−220.87 (−569; 78)	−211.93 (−560; 92)	8.94 (−244; 268)

GI indicates guideline implementation; C, control group; MC, motivational counseling.

TABLE 5. Cost Data in Euro for the 6-Month Period Before Recruitment (Means and Bootstrapped Confidence Intervals)

	Guideline Implementation (n = 468)	Guideline Implementation Plus Motivational Counseling (n = 460)	Postal Dissemination/ Controls (n = 394)
Physician consultations	66.84 (59–81)	76.28 (63–100)	86.45 (73–118)
Drugs	32.64 (25–52)	35.18 (27–50)	65.40 (47–103)
Diagnostic procedures	59.22 (45–79)	86.74 (65–118)	119.03 (91–154)
Therapeutic procedures	171.42 (140–242)	172.42 (148–204)	213.89 (184–250)
Auxiliaries	18.03 (12–26)	15.80 (10–26)	20.55 (14–29)
Hospital care	103.82 (50–197)	114.04 (55–26)	170.76 (93–294)
Rehabilitational care	105.10 (64–167)	69.28 (35–125)	154.36 (97–237)
Direct costs	557.13 (447–712)	571.78 (462–734)	829.54 (680–1028)
Indirect costs	553.07 (435–726)	672.55 (507–909)	826.74 (615–1121)
Total costs	1110.20 (918–1372)	1244.33 (1010–1581)	1656.29 (1347–2063)

Sensitivity Analysis

Figure 4 shows the results on incremental functional capacity at 6-month follow-up without adjustment for clustering (solid circles), taking into account cluster randomization (dotted circles), as well as allowing for baseline differences in costs in addition to clustering (dashed circles). There still remains a superior effect in functional capacity but no reduction in costs.

DISCUSSION

In this cost-effectiveness analysis, patients in both intervention groups showed lower costs compared with controls at 6- and 12-month follow-up. MC in addition to physician education seems slightly more cost-effective in most outcomes than physician education alone. However, any differences in cost-effectiveness diminish when adjusting for clustering of data and costs during a 6-month period prior to patient recruitment.

Limitations

Our results are based on a secondary analysis of a randomized controlled trial and therefore have to be interpreted with caution. With respect to external validity, we have to consider selection bias caused by physicians participating in research studies and by study patients who are more likely to follow an activity-enhancing therapy approach. However, a possible

TABLE 6. Incremental Cost-Effectiveness Ratios

	Guideline Implementation *Versus* Controls			Guideline Implementation Plus Motivational Counseling *Versus* Controls		
1- to 6-mo Follow-up	Incremental Cost	Incremental Effect	CER	Incremental Costs	Incremental Effect	CER
Functional capacity	−403.51	1.75	−230.58	−484.97	2.62	−185.10
Physical activity	−391.22	−6.03	64.88	−507.57	5.03	−100.91
Quality of life	−406.92	0.58	−701.59	−518.89	1.77	293.16
1- to 12-mo follow-up						
Functional capacity	−608.56	0.80	−760.70	−694.37	2.01	−345.46
Physical activity	−571.27	4.77	−119.76	−694.72	3.64	−190.86
Days in pain	−472.28	−27.14*	17.40	−712.46	−26.61	−26.77
Quality of life	−615.82	−0.43	1432.14	−659.45	0.74	−891.15

Differences in costs of the same study arms and time period derived from differences of patient numbers (only patients without missing in the respective variables were included); implementation costs are not included; days in pain refer to only a 12-month period.

**Significant with $P < 0.05$.*

CER indicates cost-effectiveness ratios.

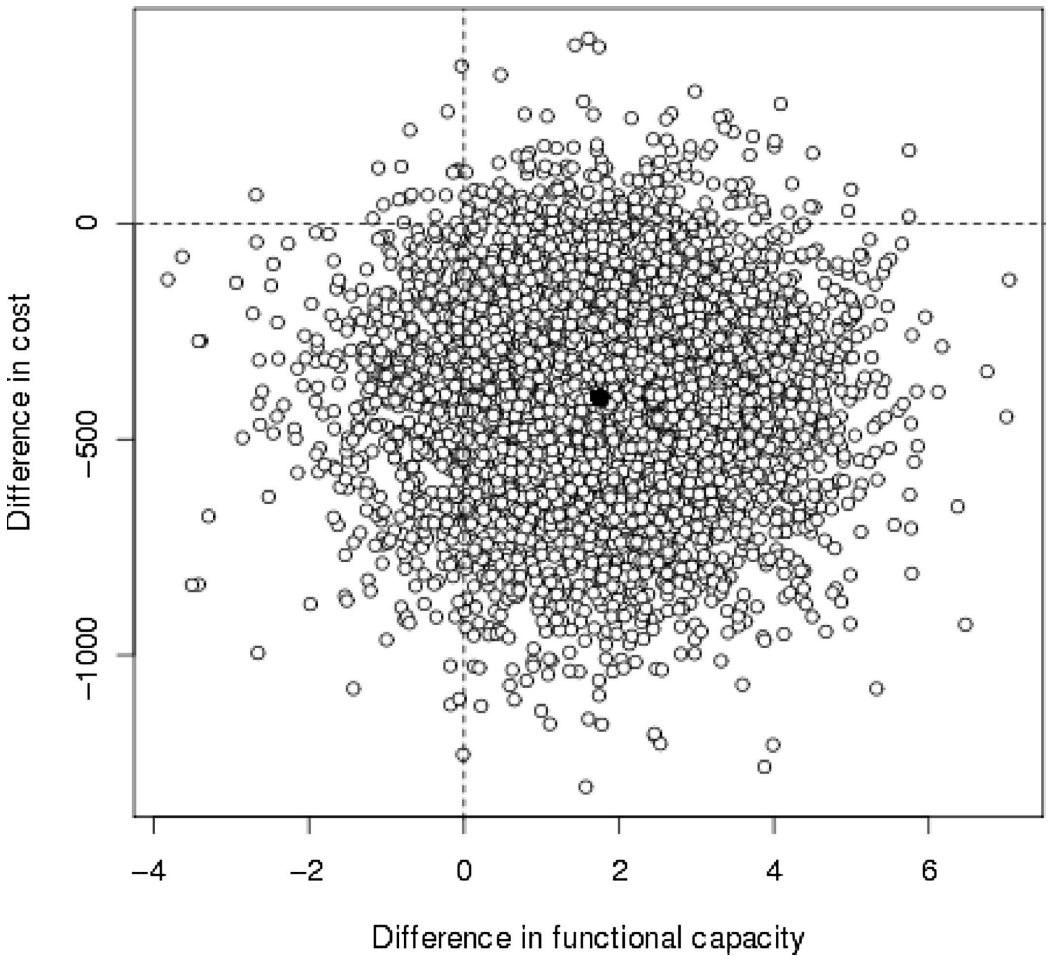

Figure 2. Incremental cost-effectiveness plane for functional capacity for a 6-month follow-up period, comparing guideline implementation *versus* controls (postal dissemination).

overestimation of cost differences between intervention groups and controls will be attenuated, because the control physicians and patients are also likely to be more enthusiastic about participating in a GI trial than a comparable population sample.

Cost data in our study are gained by retrospective interviews. We cannot exclude recall bias, information bias (*e.g.*, by patients who are unable to distinguish between LBP-related procedures and others), or social desirability bias. Under- or overestimation of costs is possible. However, these refer to all study groups and rather affect the generalizability of the study results than any group differences. Total costs are most likely underestimated, because we had to restrict interviews to key issues from the societal perspective. For a valuation of physician contacts, we followed the recommendations from Krauth *et al*,[14] which allow for only an approximation of costs.

Our cost-effectiveness calculation is based on cost consequences during follow-up and does not include expenses regarding guideline development or the implementation itself, which accounts for 5% of total costs in the intervention arms. Nevertheless, implementation benefits are expected to last longer than the study period and to be of use for more patients than our study participants. This makes implementation cost negligible and supports judgment on cost-effectiveness on the basis of follow-up health care utilization only.

We decided to perform a sensitivity analysis with respect to differences in costs prior to the beginning of the study in order to adjust for health care utilization bias. Possibly, these differences are caused by some sort of selection bias as by imbalances of patient comorbidities. We cannot comment on these, but we found no differences in quality of life which may serve as a surrogate for general health status.

Bias due to insufficient blinding is also conceivable. The baseline data collection took place after randomization on practice level. Patients could have noticed which study groups they were in and answered the questions accordingly. However, this would be the same concerning data on health care utilization as well as for data on depression, activity, and self-efficacy, in which we found no baseline imbalances. Possibly due to the skewed distribution in health care utilization data and costs, such that few patients account for the majority of total expenses, imbalances are much more likely to occur for costs than in different outcomes. This would mean that higher sample sizes than ours are needed to achieve equal distributions in health care utilization and costs, even in a randomized setting. Whether health care utilization prior to the study should be considered at all is up for discussion, especially when there are no imbalances in clinical parameters such as back pain severity. This has not been recommended in current recommendations on cost-effectiveness studies on

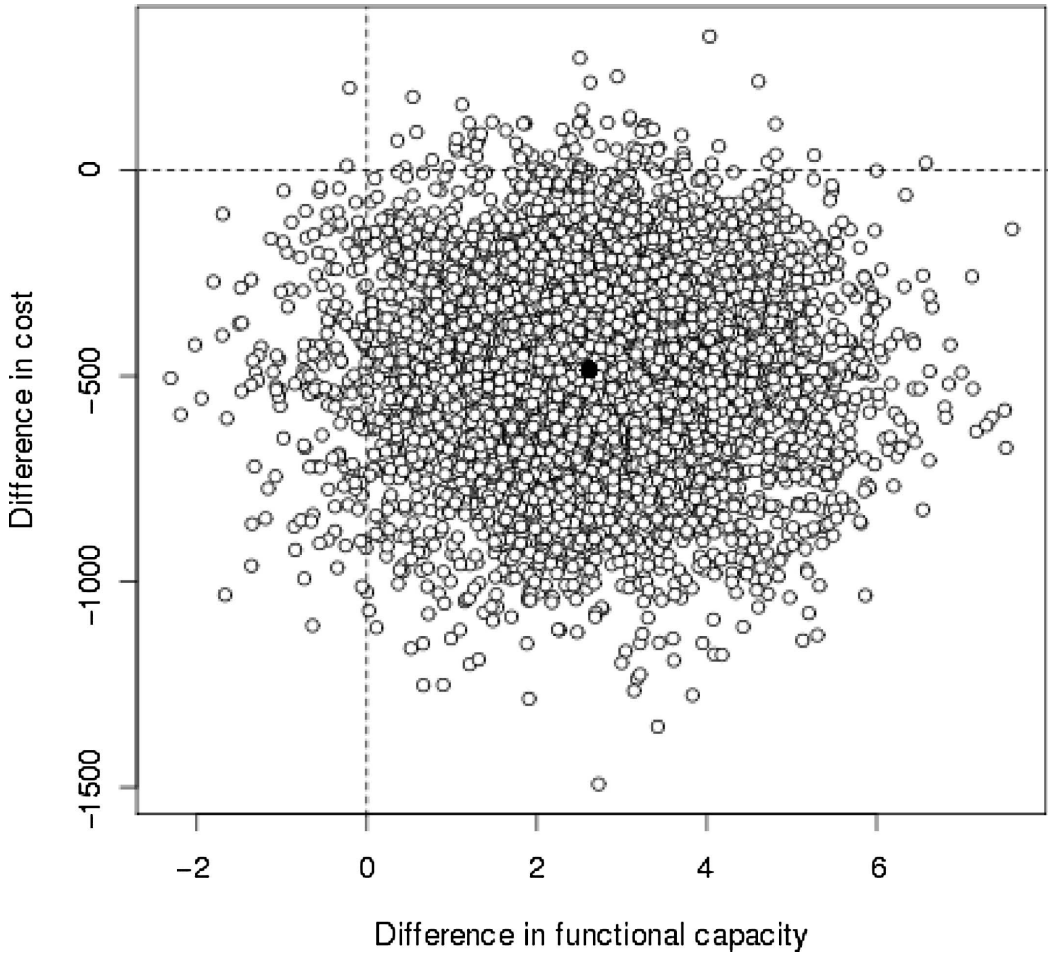

Figure 3. Incremental cost-effectiveness plane for functional capacity for a 6-month follow-up period, comparing guideline implementation plus motivational counseling *versus* controls (postal dissemination).

LBP.[21] However, we think that these baseline costs are a surrogate for patients' health care utilization pattern. Results from sensitivity analysis underline the need not to merely rely on the evaluation of randomization success by reviewing non–cost parameters such as clinical data or sociodemographic characteristics but to include some sort of health care utilization measurement prior to the study period.

International Literature

There are various reasons why the implementation strategies did not prove to be cost-effective after allowing for baseline costs. Recent reviews on the effectiveness of implementation strategies have highlighted the evidence for the effectiveness of multifaceted intervention, interactive education, and clinical reminder systems.[22] Although our interventions comply with this evidence, the intervention may have been too weak. Minor effects or just tendencies as study results are common in primary care intervention studies, referring partly to the complexity of these studies.[23] A systematic review by Grimshaw *et al*[24] found modest to moderate improvement of care in the majority of studies on GI strategies. Only about a third of the included studies (n = 235) reported economic data with rather poor methodology, which makes it difficult to draw conclusions regarding cost-effectiveness. Hoeijenbos *et al*[25] studied the cost-effectiveness of an active implemen-

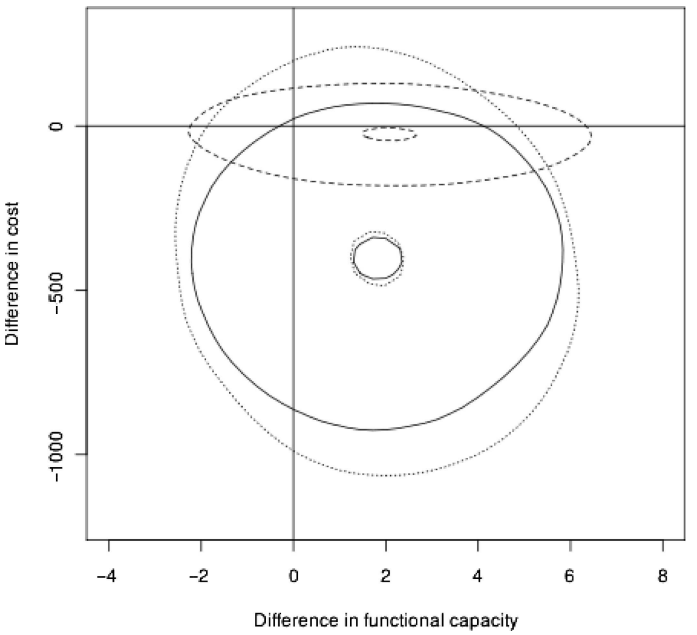

Figure 4. Sensitivity analysis for the incremental cost-effectiveness plane (CEP) shown in Figure 2 (change in functional capacity during 6 months comparing guideline implementation *versus* controls). Solid ellipses indicate raw CEP without adjustment; dotted ellipses, CEP adjusted for cluster effects; dashed ellipses, CEP adjusted for cluster effects and baseline cost differences. The inner and outer ellipses refer to the 5% and 95% confidence levels, respectively.

tation strategy compared with the standard dissemination strategy of the Dutch physiotherapy guideline. The authors found no differences in the quality of life, direct medical costs, and productivity costs. This was different in a recent study by Lambeek *et al*[26] who did a secondary cost-effectiveness analysis on a study evaluating the effects of integrated care of patients listed as sick with chronic LBP. The authors found lower total costs in the intervention group compared with controls concluding its cost-effectiveness.

CONCLUSION

Our study shows superiority by trend of the intervention arms with respect to functional capacity, days in pain, and follow-up cost. Furthermore, it underlines the importance of including health care utilization data at baseline when performing economic analyses. Future research is needed to further clarify the cost-effectiveness of different implementation strategies and the cost consequences of evidence-based LBP management.

➢ Key Points

- ❑ International guidelines have been developed to improve evidence-based care for LBP patients. To find a cost-effective implementation strategy seems crucial to give support to decision makers in health care.
- ❑ We present a secondary cost-effectiveness study alongside a cluster randomized controlled trial evaluating 2 GI strategies (physician education alone and in combination with MC by practice nurses).
- ❑ At first glance, both intervention groups show lower total costs and better outcomes compared with controls. However, effects attenuate when adjusting for clustering of data and health care costs 6 months prior to the study period.
- ❑ Future cost-effectiveness studies should include information on health care utilization behavior prior to the study period instead of merely relying on randomization success shown by non–cost parameters such as clinical data and sociodemographic characteristics.

Acknowledgments

The authors thank all patients, practice nurses, and general practitioners who participated in the study.

References

1. Papageorgiou AC, Croft PR, Ferry S, et al. Estimating the prevalence of low back pain in the general population. Evidence from the South Manchester Back Pain Survey. *Spine (Phila Pa 1976)* 1995;20:1889–94.
2. Raspe HH, Kohlmann T. The current backache epidemic [in German]. *Ther Umsch* 1994;51:367–74.
3. Von Korff M. Studying the natural history of back pain. *Spine (Phila Pa 1976)* 1994;19:2041S–6S.
4. Engers AJ, Wensing M, van Tulder MW, et al. Implementation of the Dutch low back pain guideline for general practitioners: a cluster randomized controlled trial. *Spine (Phila Pa 1976)* 2005;30:559–600.
5. van Tulder M, Becker A, Bekkering T, et al. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J* 2006;15(suppl 2):S169–91.
6. Burton AK, Balague F, Cardon G, et al. Chapter 2. European guidelines for prevention in low back pain: November 2004. *Eur Spine J* 2006;15(suppl 2):S136–68.
7. Becker A, Leonhardt C, Kochen MM, et al. Effects of two guideline implementation strategies on patient outcomes in primary care: a cluster randomized controlled trial. *Spine (Phila Pa 1976)* 2008;33:473–80.
8. Becker A, Chenot JF, Niebling W, et al. *DEGAM Guideline Low Back Pain* [in German]. Düsseldorf, Germany: Omikron Publishing; 2003.
9. Henrotin YE, Cedraschi C, Duplan B, et al. Information and low back pain management: a systematic review. *Spine (Phila Pa 1976)* 2006;31:E326–34.
10. Von Korff M, Ormel J, Keefe FJ, et al. Grading the severity of chronic pain. *Pain* 1992;50:133–49.
11. Kohlmann T, Raspe H. [Hannover Functional Questionnaire in ambulatory diagnosis of functional disability caused by backache]. *Rehabilitation (Stuttg)* 1996;35:I–VIII.
12. Roese I, Kohlmann T, Raspe H. [Measuring functional capacity in backache patients in rehabilitation: a comparison of standardized questionnaires]. *Rehabilitation (Stuttg)* 1996;35:103–8.
13. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
14. Krauth C, Hessel F, Hansmeier T, et al. [Empirical standard costs for health economic evaluation in Germany—a proposal by the working group methods in health economic evaluation]. *Gesundheitswesen* 2005;67:736–46.
15. Boss A, Elendner T. [Tax reform and wage tax in Germany. Simulations based on wage tax statistics.]. *Die Weltwirtschaft* 2003;4:368–87.
16. Becker A, Held H, Redaelli M, et al. Low back pain in primary care: costs of care and prediction of future health care utilization. *Spine (Phila Pa 1976)* 2010;35:1714–20.
17. R Development Core Team. *R: A language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2010.
18. Lunn DJ, Thomas A, Best N, et al. WinBUGS—A Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000;10:325–37.
19. Grieve R, Nixon R, Thompson SG. Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized trials. *Med Decis Making* 2010;30:163–75.
20. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Econ* 2005;14:1217–29.
21. Korthals-de Bos I, van Tulder M, van Dieten H, et al. Economic evaluations and randomized trials in spinal disorders: principles and methods. *Spine (Phila Pa 1976)* 2004;29:442–8.
22. Prior M, Guerin M, Grimmer-Somers K. The effectiveness of clinical guideline implementation strategies—a synthesis of systematic review findings. *J Eval Clin Pract* 2008;14:888–97.
23. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694–6.
24. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8:iii–iv, 1–72.
25. Hoeijenbos M, Bekkering T, Lamers L, et al. Cost-effectiveness of an active implementation strategy for the Dutch physiotherapy guideline for low back pain. *Health Policy* 2005;75:85–98.
26. Lambeek LC, Bosmans JE, Van Royen BJ, et al. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. *BMJ* 2010;341:c6414.