

## Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial

Jeremy Fairbank, Helen Frost, James Wilson-MacDonald, Ly-Mee Yu, Karen Barker, Rory Collins for the Spine Stabilisation Trial Group

### Abstract

**Objectives** To assess the clinical effectiveness of surgical stabilisation (spinal fusion) compared with intensive rehabilitation for patients with chronic low back pain.

**Design** Multicentre randomised controlled trial.

**Setting** 15 secondary care orthopaedic and rehabilitation centres across the United Kingdom.

**Participants** 349 participants aged 18-55 with chronic low back pain of at least one year's duration who were considered candidates for spinal fusion.

**Intervention** Lumbar spine fusion or an intensive rehabilitation programme based on principles of cognitive behaviour therapy.

**Main outcome measure** The primary outcomes were the Oswestry disability index and the shuttle walking test measured at baseline and two years after randomisation. The SF-36 instrument was used as a secondary outcome measure.

**Results** 176 participants were assigned to surgery and 173 to rehabilitation. 284 (81%) provided follow-up data at 24 months. The mean Oswestry disability index changed favourably from 46.5 (SD 14.6) to 34.0 (SD 21.1) in the surgery group and from 44.8 (SD 14.8) to 36.1 (SD 20.6) in the rehabilitation group. The estimated mean difference between the groups was -4.1 (95% confidence interval -8.1 to -0.1,  $P=0.045$ ) in favour of surgery. No significant differences between the treatment groups were observed in the shuttle walking test or any of the other outcome measures.

**Conclusions** Both groups reported reductions in disability during two years of follow-up, possibly unrelated to the interventions. The statistical difference between treatment groups in one of the two primary outcome measures was marginal and only just reached the predefined minimal clinical difference, and the potential risk and additional cost of surgery also need to be considered. No clear evidence emerged that primary spinal fusion surgery was any more beneficial than intensive rehabilitation.

### Introduction

Chronic low back pain is a common cause of distress and results in considerable personal and public financial consequences. Management is mostly non-operative, but spinal fusion has been used for nearly 90 years. Spinal fusion rates vary between and within countries.<sup>1</sup> In England about 1000 lumbar fusions are performed per year.<sup>2</sup> An almost direct relation exists between the numbers of operations performed each year and of orthopaedic and neurosurgeons per head of population.<sup>3</sup> In the United States, spinal fusions for "degenerative changes" rose sharply from around 11 000 operations per year in 1996 to 37 000/year in 2001 (a 336% increase).<sup>4</sup> Both the rationale and the techniques used to fuse the spine have been questioned.<sup>5</sup> Multidisciplinary rehabilitation programmes that focus on physical, psychological, social, and occupational factors have been advocated for patients with chronic pain of the low back.<sup>6-8</sup>

This trial was conceived in response to the identification of weak evidence for surgery as a priority by the NHS standing group on health technology in 1994.<sup>9, 10</sup> The pragmatic trial was designed to compare two treatment strategies (spinal stabilisation surgery or intensive rehabilitation) for patients considered by surgeons to be candidates for surgical stabilisation of the lumbar spine.

### Methods

This multicentre, randomised trial was set in 15 hospitals in the United Kingdom. Only consultant surgeons with training and expertise in performing spinal fusions participated. We approached an additional 39 centres where either the surgeon was unwilling to recruit patients or implementation of the intensive rehabilitation programme was impossible.

Editorial by Koes and p 1239

Nuffield Orthopaedic Centre, Oxford OX3 7LD

Jeremy Fairbank  
consultant orthopaedic surgeon  
James Wilson-MacDonald  
consultant orthopaedic surgeon  
Karen Barker  
director of physiotherapy research

University of Warwick, Division of Health in the Community, Coventry CV4 7AL  
Helen Frost  
research fellow

Centre for Statistics in Medicine, Oxford OX3 7LF  
Ly-Mee Yu  
statistician

Clinical Trial Service Unit and Epidemiological Studies Unit, Radcliffe Infirmary, Oxford OX2 6HE  
Rory Collins  
professor

Correspondence to: J Fairbank  
jeremy.fairbank@ndos.ox.ac.uk

BMJ 2005;330:1233-9



Additional contributors and members of the MRC steering committee are on bmj.com

This article was posted on bmj.com on 23 May 2005: <http://bmj.com/cgi/doi/10.1136/bmj.38441.620417.8F>

**Table 1** Baseline characteristics of patients and clinical details at trial entry. Values are numbers of patients unless otherwise indicated

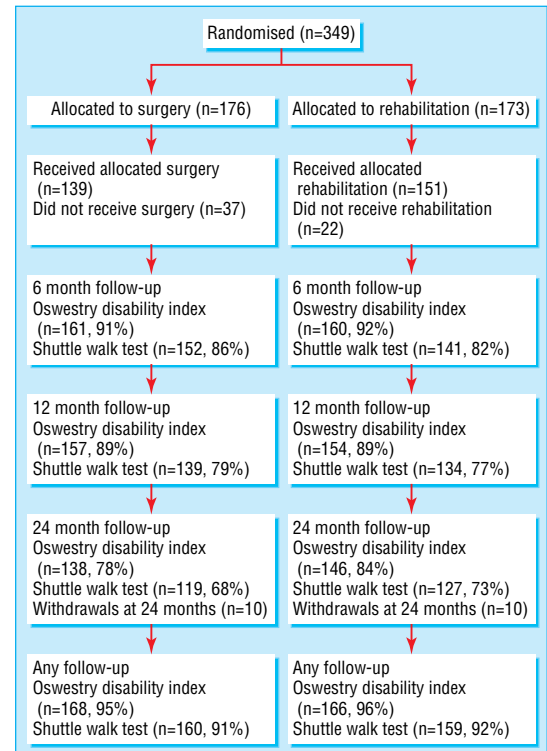
Characteristic	Surgery (n=176)	Rehabilitation (n=173)
Male	79 (44.9)	93 (53.8)
Female	97 (55.1)	80 (46.2)
Age:		
<30 years	24 (13.6)	20 (11.6)
30-39 years	63 (35.8)	67 (38.7)
40-49 years	56 (31.8)	66 (38.1)
≥50 years	33(18.8)	20 (11.6)
Centre:*		
A	55 (31.3)	54 (31.2)
B	28 (15.9)	27 (15.6)
C	45 (25.6)	43 (24.9)
D	48 (27.3)	49 (28.3)
Mean duration of back pain years (range) in years	8 (1-35)	8 (1-35)
Current smokers	76 (43.2)	74 (42.8)
Litigation	25 (14.2)	21 (12.1)
Currently in paid employment	88 (50.0)	94 (54.3)
Back pain interfered patient's ability to work:	149 (84.7)	149 (86.1)
Had to give up job	65 (43.6)	67 (45.0)
Had to change job	19 (12.7)	10 (6.7)
Had to reduce hours	17 (11.4)	12 (8.0)
Had to take sick leave	59 (39.6)	69 (46.3)
<b>Clinical details</b>		
Clinical classification:		
Spondylolisthesis	20 (11.4)	18 (10.4)
Post-laminectomy	14 (8.0)	14 (8.1)
Chronic low back pain	142 (80.6)	141 (81.5)
Planned surgery type:		
Graf	27 (15.3)	28 (16.2)
Fusion	149 (84.7)	144 (83.2)
Missing	0	1 (0.6)
Planned fused level:		
Single level	100 (56.8)	109 (63.0)
>1 level	70 (39.8)	62 (35.8)
Missing	6 (3.4)	2 (1.2)
<b>Mean score (SD)</b>		
Oswestry disability index	46.5 (14.6)	44.8 (14.8)
Shuttle walking test in metres	254 (209)	247 (185)
SF-36 physical component score	19.4 (8.8)	20 (9.7)
SF-36 mental component score	43.2 (10.9)	44.2 (12.6)
Modified somatic perception questionnaire	9.0 (6.4)	7.7 (5.7)
Zung self rating depression scale	31.8 (10.4)	31.2 (11.8)
Distress and risk assessment method:		
Normal	14 (8.0)	14 (8.1)
At risk	65 (36.9)	85 (49.1)
Distressed depressive	87 (49.4)	69 (39.9)
Distressed somatic	9 (5.1)	2 (1.2)
Missing	1 (0.6)	3 (1.7)

\*Refers to the three largest recruiting centres and a pool of the remaining centres.

**Eligibility criteria**

We used the uncertainty of outcome principle to define our entry criteria and therefore depended on the current practice of many experienced spine surgeons and their patients.<sup>11</sup> Patients who were candidates for surgical stabilisation of the spine were eligible if the clinician and patient were uncertain which of the study treatment strategies was best. Patients had to be aged between 18 and 55, with more than a 12 month history of chronic low back pain (with or without referred pain) and irrespective of whether they had had previous root decompression or discectomy.

Patients were ineligible if the surgeon considered that any medical or other reasons made one of the trial interventions unsuitable. These included infection or



Flow of participants through each stage of the spine stabilisation trial, showing numbers completing the primary outcome measures (Oswestry score and shuttle walking test) at each follow-up stage

other comorbidities (inflammatory disease, tumours, fractures), psychiatric disease, inability or unwillingness to complete the trial questionnaires, or pregnancy. If patients had had previous surgical stabilisation surgery of the spine they were also excluded.

**Objectives**

The aim was to determine whether surgical stabilisation of the spine (by fusion or flexible stabilisation) was more or less effective at achieving worthwhile relief of symptoms over a two year period than an intensive rehabilitation programme based on principles of cognitive behaviour therapy.

**Outcome measures**

We assessed outcomes at baseline and 6, 12, and 24 months from randomisation by a trial research therapist in each centre. If the patient was unable to attend the follow-up appointments we mailed the questionnaire. We approached non-responders by phone, through their family doctor, and via national databases.

**Primary outcome**

The two primary measures at 24 months included a back pain specific questionnaire and a standardised walking test. The Oswestry low back pain disability index is scored from 0% (no disability) to 100% (totally disabled or bedridden) and designed to assess limitations of various activities of daily living.<sup>12-13</sup> The shuttle walking test is a standardised, progressive, maximal test of walking speed and endurance.<sup>14-16</sup>

**Secondary outcomes**

The short form 36 general health questionnaire (SF-36) includes 35 items summarised in two measures related to physical and mental health. Each scale

ranges from 0 (worst health state) to 100 (best health state). The summary measures are transformed to give a population mean of 50 (SD 10). The SF-36 is recommended as an outcome assessment for spinal disorders because it provides strong psychometric support and extensive normative data.

**Psychological assessment**—We used the distress and risk assessment method (DRAM), which includes the modified Zung depression index and somatic perception questionnaire, to assess anxiety and depression.<sup>17</sup>

**Complications**—We recorded the intraoperative use of anaesthetic agents, implants, and radiological investigations; complications of surgery and any adverse effects of rehabilitation; postoperative complications, implant failure and repeat surgery; and personal items and devices purchased by the patient because of lower back pain. Work status was monitored. We recorded “obvious pseudoarthrosis” only where it was clear to the treating surgeon that fusion had failed and that this was a problem to the patient.

### Sample size

We used the Oswestry disability index to determine the sample size. The trial was designed to be able to detect a difference in mean score between the intervention groups of as little as 4 points.<sup>12 13</sup> We estimated that 133 subjects would be required in each group to detect such a difference at the  $\alpha = 0.05$  level with 80% power. We initially planned to recruit at least this number of patients in each of three separate clinical groups to allow reliable subgroup analysis, but most of the patients were recruited in one clinical category.

### Interventions

**Spinal stabilisation surgery**—The particular technique used for spinal fusion was left to the discretion of the operating surgeon. This allowed choice of the most appropriate surgical approach, implant (if any), interbody cages, and bone graft material for that patient. A small number of surgeons used flexible stabilisation of the spine (the Graf or Global technique). This was recorded for each patient before randomisation.

**Intensive rehabilitation programme**—Each centre was modelled on a daily outpatient programme of education and exercise running on five days per week for three weeks continuously. Further details of the programme are reported elsewhere.<sup>15</sup> Most centres offered 75 hours of intervention (range 60–110 hours), with one day of follow-up sessions at one, three, six, or 12 months after treatment. The rehabilitation programmes were led by physiotherapists but included clinical psychologists in all but one centre, as well as medical support. The daily exercises were individually tailored and paced to increase repetitions and duration, aiming to build on the participants’ baseline ability. They included stretching of major muscle groups, spinal flexibility exercises, general muscle strengthening, spine stabilisation exercises, and cardiovascular endurance exercise using any mode of aerobic exercise (treadmill walking, step-ups, cycling, rowing). All but one centre included daily sessions of hydrotherapy. We used principles of cognitive behaviour therapy to identify and overcome fears and unhelpful beliefs that many patients develop when in pain.

**Table 2** Compliance with allocated intervention and further treatment

	Surgery (n=176)	Rehabilitation (n=173)
<b>No (%) patients who received allocated intervention</b>	<b>139 (79)</b>	<b>151 (87)</b>
Time from randomisation to intervention received per No of patients:		
<1 week	1	10
1-2 weeks	8	25
2 weeks-1 month	20	43
1-3 months	86	63
4-6 months	16	9
7-12 months	5	1
>12 months	3	0
Low attendance (1-49% available time)*	N/A	12
High attendance (50-100% available time)*	N/A	139
Median time from randomisation to intervention in months (range)	1.6 (0.2-15.4)	0.9 (0.1-10.2)
<b>No (%) of patients who did not receive allocated treatment</b>	<b>37 (21)</b>	<b>22 (13)</b>
Switched to the other treatment group	7	10
Physiotherapy	3	0
Medical treatment	11	2
No recorded treatment	16	10
<b>No (%) of patients who required further treatment after allocated treatment</b>	<b>97 (55.1%)</b>	<b>68 (39.3%)</b>
Further surgery or surgery*	11	38
Additional rehabilitation programme	7	0
Additional physiotherapy treatment	47	8
Additional medical treatment	32	22
No recorded additional treatment	42	83

\*Rehabilitation only.

### Treatment allocation and recruitment

Surgeons approached patients who were candidates for spinal fusion. Each centre employed a trial research therapist to organise the trial locally, recruit patients, book treatment appointments, and carry out assessments. Patients were given verbal, written, and videotape (OMI, Oxford) explanations of the background and nature of the trial. The trial research therapists obtained written consent and carried out baseline assessments before randomisation.

Randomisation was generated centrally by computer program, with minimisation for various potential confounding factors: age, smoking, litigation, Oswestry score, clinical classification, and planned use of the Graf procedure.

### Statistical methods

We carried out an intention to treat analysis. We used analysis of covariance (ANCOVA) to analyse quantita-

**Table 3** Complication due to surgery (each subject could have more than one complication)

Complication	No of patients
At treatment site:	
Dural tear	5
Excessive bleeding	3
Implant problems	5
Bone fracture	1
Vascular injury	1
Loss of purchase or fixation	3
Broken drain	1
Associated with surgical approach:	
Vascular injury	1
Other (loss of swab 1, peritoneal tear 2)	3
Systemic:	
Haemorrhage	1
Further surgery (up to 2 years’ follow-up)	11

A total of 19 patients had complications as a result of surgery.

**Table 4** Mean(SD) outcome values at 24 months, and differences in changes from baseline to 24 months

	Surgery (n=176)	Rehabilitation (n=173)	Difference in change 0-24 months (95% CI)*	P value†
<b>Oswestry disability index</b>	(n=138)	(n=146)		
24 months	34.0 (21.1)	36.1 (20.6)	-4.1 (-8.1 to -0.1)	0.045
<b>Shuttle walking test</b>	(n=118)	(n=126)		
24 months	352 (244)	310 (202)	34 (-8 to 77)	0.12
<b>SF-36 physical component score</b>	(n=115)	(n=131)		
24 months	28.8 (14.9)	27.6 (14.6)	2.0 (-1.2 to 5.3)	0.21
<b>SF-36 mental component score</b>				
24 months	47.4 (12.2)	48.1 (12.6)	-0.2 (-2.9 to 2.6)	0.90
<b>Domains of SF-36</b>				
General health perception:				
Baseline	47.6 (20.5)	46.5 (22.0)		
24 months	57.7 (23.6)	53.8 (24.5)	3.2 (-1.9 to 8.2)	0.22
Physical functioning:				
Baseline	33.6 (19.0)	39.5 (22.1)		
24 months	50.0 (28.2)	49.8 (28.7)	4.8 (-1.2 to 10.8)	0.11
Role limitation (physical):				
Baseline	15.0 (27.1)	17.6 (30.5)		
24 months	39.6 (42.1)	38.6 (42.7)	2.4 (-7.5 to 12.3)	0.63
Role limitation (emotional):				
Baseline	43.2 (41.4)	51.2 (44.0)		
24 months	65.2 (42.7)	65.4 (43.4)	2.9 (-7.1 to 13.0)	0.57
Pain:				
Baseline	28.6 (17.3)	30.0 (16.0)		
24 months	48.1 (26.4)	44.9 (25.1)	4.1 (-1.67 to 10.0)	0.16
Social functioning:				
Baseline	41.1 (23.3)	42.8 (22.9)		
24 months	53.6 (26.2)	55.6 (26.2)	-0.9 (-6.5 to 4.7)	0.76
Mental health:				
Baseline	60.1 (19.9)	60.3 (21.6)		
24 months	66.5 (21.5)	68.4 (23.1)	-1.8 (-6.4 to 2.8)	0.45
Energy and vitality:				
Baseline	35.4 (20.0)	37.4 (21.7)		
24 months	46.7 (22.8)	46.4 (24.9)	1.5 (-3.8 to 6.7)	0.58

\*Adjusted for baseline measures. Rehabilitation group is the reference group.

†Analysis of covariance adjusted for baseline measure.

tive outcomes at 24 months, with corresponding baseline values and treatment group as covariates.

We used multiple imputation to handle missing data. To impute the missing data we constructed multiple regression models including variables potentially related to the fact that the data were missing and also variables correlated with that outcome. We used Stata (StataCorp, College Station, Texas, USA)<sup>18</sup> and PROC MI in SAS (SAS Institute, Cary, NC, USA) to obtain similar answers, and only the former are presented.

## Results

A total of 349 patients were randomised between June 1996 and February 2002 from 15 centres in the UK (176 allocated to surgery and 173 to rehabilitation).

**Table 5** Summary of results from available cases and from multiple imputation analyses

Outcome	Available cases		Imputed analyses	
	Estimated difference (95% CI)	P value*	Estimated difference (% CI)	P value*
Oswestry disability index	-4.1 (-8.1 to -0.1)	0.04	-4.5 (-8.2 to -0.8)	0.02
Shuttle walking test	34 (-8 to 77)	0.12	25 (-16 to 66)	0.23
SF-36 (physical component score)	2.0 (-1.2 to 5.3)	0.21	2.5 (-0.4 to 5.5)	0.09
SF-36 (mental component score)	-0.2 (-2.9 to 2.6)	0.90	0.4 (-2.0 to 2.9)	0.73

\*Analysis of covariance adjusted for baseline measure.

The figure shows the progression through the trial. Table 1 shows the baseline characteristics of patients who entered the trial.

## Compliance with treatment and follow-up

Table 2 shows data on participants' compliance with their treatment and follow-up. Forty eight (28%) patients randomised to rehabilitation had surgery by two years. Seven (4%) patients randomised to surgery had rehabilitation instead of surgery.

## Complications

Intraoperative complications occurred in 19 surgical cases (table 3). Eleven patients required further operations on their lumbar spine during the two year follow-up. We did not identify any specific complications of the rehabilitation programmes.

## Clinical outcomes

Oswestry scores improved slightly more in favour of surgery (-4.1, 95% confidence interval -8.1 to -0.1, P=0.045). After imputation for missing follow-up data the mean difference was -4.5 (-8.2 to -0.8, P=0.02) (tables 4 and 5). No significant heterogeneity in the effect on the Oswestry score was observed between the predefined groups of patient (table 6). No other difference between groups in any of the other outcomes at 24 months reached significance, even when we used imputed values (tables 4 and 5).

## Discussion

Patients with low back pain who are considered by surgeons to be candidates for spinal fusion may obtain similar benefits from an intensive rehabilitation programme as they do from surgery. Our large randomised controlled trial of spinal fusion surgery compared with intensive rehabilitation was limited by recruitment difficulties, some crossover between intervention groups, and incomplete follow-up at 24 months, but the results should help clinicians and service providers make decisions about the management of chronic low back pain. Both groups improved over time, but this effect may reflect a natural resolution of chronic low back pain or regression to the mean. The Oswestry scores improved significantly more in patients allocated to surgery than in those allocated to rehabilitation. Although this difference just exceeds the 4 points specified in the sample size calculation, clinically this difference is small considering the potential risks and additional costs of surgery. Analyses adjusting for baseline variations or per protocol analysis do not change this interpretation (data not shown). Overall, since the other primary outcome of the shuttle walking test and the other measures did not differ (even after imputation for missing values), the small difference observed in Oswestry scores should be interpreted cautiously. Furthermore, the confidence intervals can be used to rule out differences in Oswestry scores of more than 10 points in favour of surgery and of more than 2 points in favour of rehabilitation. Consequently, they narrow substantially the range of plausible estimates for any benefit of surgery.

## Comparison with related research

A Cochrane review in 1999 found a complete absence of randomised controlled evidence for spinal fusion.<sup>5</sup>

**Table 6** Mean Oswestry disability score (with standard deviations) at baseline and 24 months by different subgroups

Subgroup	Surgery	Rehabilitation
<b>Centre</b>		
A:		
Baseline	42.2 (14.9)	40.3 (12.1)
24 months	24.6 (20.7)	28.9 (17.3)
B:		
Baseline	46.7 (13.1)	46.9 (15.1)
24 months	31.4 (18.5)	43.7 (20.9)
C:		
Baseline	48.7 (14.0)	47.8 (15.8)
24 months	42.0 (18.8)	43.3 (19.7)
D:		
Baseline	49.3 (14.8)	45.8 (15.8)
24 months	39.7 (20.9)	33.6 (22.1)
<b>Clinical classification</b>		
Spondylolisthesis:		
Baseline	42.1 (15.1)	38.0 (13.8)
24 months	35.7 (25.4)	30.1 (19.9)
Post-laminectomy:		
Baseline	50.7 (14.2)	47.3 (10.2)
24 months	38.0 (18.4)	33.7 (14.8)
Chronic back pain:		
Baseline	46.7 (14.5)	45.4 (15.2)
24 months	33.3 (20.8)	37.2 (21.1)
<b>Litigation</b>		
No:		
Baseline	45.6 (14.9)	44.2 (15.1)
24 months	34.0 (21.7)	35.4 (20.9)
Yes:		
Baseline	52.0 (11.3)	49.1 (12.4)
24 months	34.1 (17.7)	41.5 (17.0)
<b>Smoking</b>		
No:		
Baseline	45.5 (14.6)	43.1 (14.9)
24 months	29.5 (19.1)	34.6 (19.3)
Yes:		
Baseline	47.8 (14.5)	46.9 (14.6)
24 months	40.6 (22.2)	38.4 (22.2)
<b>Planned type of surgery</b>		
Graf:		
Baseline	45.2 (12.6)	42.5 (17.2)
24 months	35.1 (18.9)	35.6 (24.3)
Fusion:		
Baseline	46.8 (14.9)	45.3 (14.4)
24 months	33.9 (21.5)	36.4 (19.8)

Three randomised controlled trials have been reported subsequently. Möller and Hedlund reported a trial in isthmic spondylolisthesis, with 77 patients randomised to different forms of surgery and 34 patients randomised to an exercise programme.<sup>19</sup> The patients allocated to surgery reported greater benefits at two years in terms of Oswestry scores compared with those allocated to exercise, but instrumentation and bone grafting was not found to produce an advantage over bone grafting alone. A Swedish trial randomised 222 patients to three surgical groups of equal size and 72 patients to physiotherapy.<sup>20</sup> They reported decreased pain and disability in the surgical group compared with physiotherapy at two years but no difference in outcomes between the different surgical techniques. Little effect of physiotherapy was apparent at two years, although this may have been because of the type or intensity of treatment. Routine physiotherapy and intensive rehabilitation are not the same and should not be considered as such. Brox et al reported no

differences between groups in a small trial of 64 patients comparing instrumented posterior fusion with rehabilitation followed to 12 months.<sup>21</sup> Improvements in outcomes were comparable with those in both arms of the present trial and in the surgical arm of the Swedish trial.

Evidence is moderate to strong that multidisciplinary rehabilitation including general exercise programmes of muscle strengthening, flexibility training, and cardiovascular endurance along with a cognitive behaviour approach improves function, reduces pain, and work loss in patients with chronic pain of the low back compared with usual care or non-multidisciplinary treatment.<sup>8, 22, 23</sup> This type of treatment was difficult to implement in the trial and, although recommended in recent European guidelines,<sup>24</sup> is not routinely available in the NHS.

### Strengths and limitations of the study

The uncertainty principle had initially been expected to aid trial accrual by bringing the process of informed consent closer to standard medical practice. However, recruitment was slow and numbers enrolled smaller than planned. Eligibility was based on the uncertainty of outcome principle, but uncertainty does not come easily to surgeons when patients are demanding clear direction and advice. Factors influencing recruitment will be presented elsewhere. This pragmatic trial reflects current practice across the UK of experienced spine surgeons selecting patients for fusion. Surgeons may argue that we excluded the best candidates for surgery through “certainty” of outcome, but this certainty varied between surgeons. Evidence from the Swedish trial<sup>25</sup> shows that patients with low neurotism, narrow discs, and a short time off work do best with surgery.

### Surgical issues

Surgeons were allowed their own choice of operation to improve the chance of clinical success. The Swedish trial showed no difference between three surgical techniques of fusion.<sup>24</sup> These results call into question what lumbar fusion is actually doing to patients with chronic back pain. Elucidation of this question was not the objective of this study. The results are highly relevant to spinal fusion surgery, as well as the new techniques of flexible stabilisation and disc replacement that are being applied to this group of patients.

### Loss to follow-up

Loss to follow-up at 24 months (20%) limits the internal validity of the trial. We used multiple imputation as a sensitivity analysis to tackle potential bias resulting from the poor response rate. Overall estimates of the treatment effect were very similar with all methods of statistical analysis.

### Blinding

The pre-randomisation outcomes were scored by the trial research therapists and later checked by computer. All subsequent outcomes were scored centrally. We were not able to blind the trial research therapists to patient allocation after the baseline assessment.

### Limitation of outcomes

The available outcome measures are blunt instruments for assessing a complex condition. The minimum clini-

### What is already known on this topic

Limited evidence shows that patients with severe chronic low back pain treated with spinal stabilisation surgery have a better outcome in terms of pain and disability than with traditional conservative management

The results of spinal stabilisation surgery seem to be similar whatever surgical technique is used

Intensive multidisciplinary rehabilitation including a biopsychosocial approach improves pain and function in severe chronic low back pain compared with usual care or traditional conservative treatment

### What this study adds

No clear evidence emerged that primary spinal fusion surgery was more beneficial than intensive rehabilitation using principles of cognitive behaviour therapy

Evidence exists to support intensive rehabilitation with cognitive behaviour principles as an alternative to spinal fusion surgery in the management of chronic low back pain

cally important change in the Oswestry scores has been estimated by different observers as being somewhere between 4 and 17.<sup>26</sup> Debate continues among back pain experts over the question of what represents a clinically important change. Functional measures are difficult to apply in a multicentre setting, and although the use of muscle measurement techniques may be useful, it was not possible to use them in this trial because of financial limitations. Walking capacity was chosen as it is simple and cheap to measure and often a limitation for people with chronic low back pain.

#### Compliance with treatment protocol

The 48 (28%) patients who were randomised to rehabilitation and then had additional surgery by two years should be considered as an additional outcome of the trial and taken into account in the interpretation of the results. Although some patients and surgeons were clearly not satisfied with the results of rehabilitation, many more seem to have benefited and avoided surgical intervention.

#### Conclusion

Nearly three quarters of those patients allocated to rehabilitation avoided surgery by two years. Rehabilitation including a cognitive behaviour approach is not routinely or widely available to patients with chronic pain of the low back, and this trial implies that it should be. Rehabilitation programmes require finance, space, and training, but above all they need the strong support of all clinicians involved in the care of these patients.

We thank the patients, who permitted a difficult decision to be made for them; referees, physiotherapist, and surgeons, inside and outside the trial ([www.ndos.ox.ac.uk/SST](http://www.ndos.ox.ac.uk/SST)), who helped

develop the protocol; the Medical Research Council for supporting the study; NHS R&D (especially Richard Lilford) for supporting and promoting the study.

**Contributors:** JF was responsible for the overall study design, the organisation of the study, recruiting and operating on patients in the study, data analysis, and wrote the first draft of this report. HF was responsible for overall study design, the organisation of the study, the design and implementation of the rehabilitation programme, data analysis, and writing the first draft of the report. JWM was responsible for overall study design, the organisation of the study, recruiting and operating on patients in the study, data analysis, and editing the report. RC was responsible for overall study design, statistical advice and data analysis, and presented analyses to the data monitoring committee. LMY was responsible for statistical analysis. JF and HF are guarantors. KB was responsible for recruiting patients and data analysis. Of the contributors who are not listed as authors, Douglas Altman was responsible for statistical analysis and sat on the data monitoring committee. Alastair Gray was responsible for study design, the organisation of the study and economic analysis. Nicolas Maniadakis was responsible for economic data collection and analysis. Kate Johnston, Helen Campbell, and Oliver Rivero were responsible for economic analysis. Patricia Carver was responsible for data collection and analysis. L Morgan was responsible for data collection and database design. Kate Stevens, Victoria Erlanger, Rebecca Bale collected and entered data. Peter Smith developed and maintained the database.

**Funding:** The Medical Research Council supported the trial financially and was represented on the steering committee. The NHS (326) or private patient insurance (23) funded the treatment of patients. MRC grant number G94431172.

**Competing interests:** JF and JWM have received funding from Synthes for a spinal fellow.

**Ethical approval:** The trial was approved by a multicentre research ethics committee (twice; references 98/5/14 for original and 03/05/034 for long term follow-up) and 15 local research ethics committees.

- Katz J. Lumbar spinal fusion: surgical rates, costs, and complications. *Spine* 1995;20:78S-83S.
- Department of Health (England). Hospital episode statistics, 1998-2003. London: DoH, 1998-2003.
- Cherkin D, Deyo R, Loeser J, Bush T, Waddell G. An international comparison of back surgery rates. *Spine* 1994;19:1201-6.
- Deyo RA, Gray DT, Kreuter W, Mirza S, Martin, BI. United States trends in lumbar fusion surgery for degenerative conditions. *Spine* 2005 (in press).
- Gibson J, Grant I, Waddell G. The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. *Spine* 1999;24:1820-32.
- Mayer T, Gatchel R, Kishino N. Objective assessment of spine function following industrial injury. A prospective study with comparison group and one year follow up. *Spine* 1985;10:482-93.
- Lindstrom I, Ohlund C, Eek C. The effect of graded activity on patients with sub acute low back pain: A randomised prospective clinical study with an operant conditioning behavioural approach. *Phys Ther* 1992;72:279-93.
- Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. *BMJ* 2001;322:1511-6.
- NHS Standing Group. *Health technology (SGHT) report*. London: Department of Health, 1994.
- Turner J, Ersek M, Herron L, Hesekorn J, Kent D, Ciol M, et al. Patient outcomes after lumbar spinal fusions. *JAMA* 1992;268:907-11.
- Wejter C, Shapiro S, Cranley Glass K. For and against: clinical equipoise and not the uncertainty principle is the moral underpinning of the randomised controlled trial. *BMJ* 2000;321:756-8.
- Fairbank J, Pynsent P. The Oswestry Disability Index. *Spine* 2000;25:2940-53.
- Roland M, Fairbank J. The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. *Spine* 2000;25:3115-24.
- Taylor S, Frost H, Taylor A, Barker K. Reliability and responsiveness of the shuttle walking test in patients with chronic low back pain. *Physiother Res Int* 2001;6:170-8.
- Frost H, Lamb S, Shackleton C. A functional restoration programme for chronic low back pain: a prospective outcome study. *Physiotherapy* 2000;86:285-93.
- Pratt R, Fairbank J, Virr A. The reliability of the shuttle walking test, the Swiss spinal stenosis questionnaire, the Oxford spinal stenosis Score, and the Oswestry disability index in the assessment of patients with lumbar spinal stenosis. *Spine* 2002;27:84-91.
- Main CJ, Wood PLR, Hollis S, Spanswick CC, Waddell G. The distress and risk assessment method: A simple patient classification to identify distress and evaluate the risk of poor outcome. *Spine* 1992;17:42-52.
- Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227-41.

- 19 Möller H, Hedlund R. Surgery versus conservative treatment in adult isthmic spondylolisthesis. A prospective randomized study part 1. *Spine* 2000;25:1171-5.
- 20 Fritzell P, Hagg O, Wessburg P, Nordwall A, Group SLSS. Chronic back pain and fusion: a comparison of three surgical techniques: a prospective multicentre randomized study from the Swedish Lumbar Spine Study Group. *Spine* 2002;27:1131-41.
- 21 Brox J, Sørensen R, Friis A, Nygaard Ø, Indahl A, Keller A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003;28:1913-21.
- 22 Schonstein E, Kenny D, Keating J, Koes B, Herbert R. Physical conditioning programs for workers with back and neck pain: a Cochrane systematic review. *Spine* 2003;28(19):E391-5.
- 23 Liddle S, Baxter G, Gracey J. Exercise and chronic low back pain: what works? *Pain* 2004;107:176-90.
- 24 European Commission Research Directorate General. COST B13 Management Committee. *European guidelines for the management of low back pain*. www.backpainurope.org, 2005 (accessed 17 May 2005).
- 25 Hägg O, Fritzell P, Elkselius L, Nordwall A. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish lumbar spine study. *Eur Spine J* 2003;12:22-33.
- 26 Taylor S, Taylor A, Foy M, Fogg A. Responsiveness of common outcome measures for patients with low back pain. *Spine* 1999;24:1805-12.

(Accepted 24 March 2005)

doi 10.1136/bmj.38441.620417.8F

## Surgical stabilisation of the spine compared with a programme of intensive rehabilitation for the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial

Oliver Rivero-Arias, Helen Campbell, Alastair Gray, Jeremy Fairbank, Helen Frost, James Wilson-MacDonald for the Spine Stabilisation Trial Group

### Abstract

**Objective** To determine whether from a health provider and patient perspective, surgical stabilisation of the spine is cost effective when compared with an intensive programme of rehabilitation in patients with chronic low back pain.

**Design** Economic evaluation alongside a pragmatic randomised controlled trial.

**Setting** Secondary care.

**Participants** 349 patients randomised to surgery (n = 176) or to an intensive rehabilitation programme (n = 173) from 15 centres across the United Kingdom between June 1996 and February 2002.

**Main outcome measures** Costs related to back pain and incurred by the NHS and patients up to 24 months after randomisation. Return to paid employment and total hours worked. Patient utility as estimated by using the EuroQol EQ-5D questionnaire at several time points and used to calculate quality adjusted life years (QALYs). Cost effectiveness was expressed as an incremental cost per QALY.

**Results** At two years, 38 patients randomised to rehabilitation had received rehabilitation and surgery whereas just seven surgery patients received had both treatments. The mean total cost per patient was estimated to be £7830 (SD £5202) in the surgery group and £4526 (SD £4155) in the intensive rehabilitation arm, a significant difference of £3304 (95% confidence interval £2317 to £4291). Mean QALYs over the trial period were 1.004 (SD 0.405) in the surgery group and 0.936 (SD 0.431) in the intensive rehabilitation group, giving a non-significant difference of 0.068 (–0.020 to 0.156). The incremental cost effectiveness ratio was estimated to be £48 588 per QALY gained (–£279 883 to £372 406).

**Conclusion** Two year follow-up data show that surgical stabilisation of the spine may not be a cost effective use of scarce healthcare resources. However, sensitivity analyses show that this could change—for

example, if the proportion of rehabilitation patients requiring subsequent surgery continues to increase.

### Introduction

The first UK based trial comparing surgical stabilisation of the spine with a programme of intensive rehabilitation, has shown a significant difference in the Oswestry disability index,<sup>1</sup> at two years in patients randomised to spinal fusion surgery compared with intensive rehabilitation.<sup>2</sup> This statistical difference between treatment groups in only one of the two primary outcome measures was marginal and only just reached the predefined minimal clinical difference. The potential risk and additional cost of surgery also need to be considered. No clear evidence emerged that primary fusion was any more beneficial than intensive rehabilitation. We report an economic evaluation of this trial. We employ a cost utility framework to determine whether any net health gain from using surgery would be sufficient to justify a likely increase in the costs of treatment. The chosen form of analysis will facilitate comparisons between the cost effectiveness of surgery and that of other healthcare interventions competing for healthcare resources.

### Methods

Full details of the randomised controlled trial are published in parallel with this paper.<sup>2</sup> We recruited 349 patients who met trial eligibility criteria from 15 centres around the UK between June 1996 and February 2002. Of these patients, 176 were ran-

See also *Editorial* by Koes and p 1233

Health Economics Research Centre, Department of Public Health, University of Oxford, Oxford OX3 7LF

Oliver Rivero-Arias  
*research officer*

Helen Campbell  
*research officer*  
Alastair Gray  
*professor of health economics*

Nuffield Orthopaedic Centre, Oxford OX3 7LD

Jeremy Fairbank  
*consultant orthopaedic surgeon*  
James Wilson-MacDonald  
*consultant orthopaedic surgeon*

Division of Health in the Community, University of Warwick, Warwick CV4 7AL

Helen Frost  
*research fellow*

Correspondence to: H Campbell  
helen.campbell@dphpc.ox.ac.uk

BMJ 2005;330:1239–43



Additional details of the costing method, the participating units, and the members of the MRC steering committee are on [bmj.com](http://bmj.com)



This is the abridged version of an article that was posted on [bmj.com](http://bmj.com) on 23 May 2005: <http://bmj.com/cgi/doi/10.1136/bmj.38441.429618.8F>