



# Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial

Jonathan C Hill, David G T Whitehurst, Martyn Lewis, Stirling Bryan, Kate M Dunn, Nadine E Foster, Kika Konstantinou, Chris J Main, Elizabeth Mason, Simon Somerville, Gail Sowden, Kanchan Vohora, Elaine M Hay

## Summary

Lancet 2011; 378: 1560–71

Published Online  
September 29, 2011  
DOI:10.1016/S0140-6736(11)60937-9

See [Comment](#) page 1530

Arthritis Research UK Primary Care Centre, Primary Care Sciences, Keele University, Stoke-on-Trent, UK (J C Hill PhD, D G T Whitehurst PhD, M Lewis PhD, K M Dunn PhD, Prof N E Foster DPhil, K Konstantinou PhD, Prof C J Main PhD, E Mason MSc, S Somerville MSc, G Sowden MSc, K Vohora BSc, Prof E M Hay MD); School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada (Prof S Bryan PhD, D G T Whitehurst); and Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada (Prof S Bryan, D G T Whitehurst)

Correspondence to: Dr Jonathan C Hill, Arthritis Research UK Primary Care Centre, Primary Care Sciences, Keele University, Keele, ST5 5BG, UK [j.hill@cphc.keele.ac.uk](mailto:j.hill@cphc.keele.ac.uk)

For the Keele STarT Back Screening Tool see <http://www.keele.ac.uk/startback>

See Online for webappendix

**Background** Back pain remains a challenge for primary care internationally. One model that has not been tested is stratification of the management according to the patient's prognosis (low, medium, or high risk). We compared the clinical effectiveness and cost-effectiveness of stratified primary care (intervention) with non-stratified current best practice (control).

**Methods** 1573 adults (aged  $\geq 18$  years) with back pain (with or without radiculopathy) consultations at ten general practices in England responded to invitations to attend an assessment clinic. Eligible participants were randomly assigned by use of computer-generated stratified blocks with a 2:1 ratio to intervention or control group. Primary outcome was the effect of treatment on the Roland Morris Disability Questionnaire (RMDQ) score at 12 months. In the economic evaluation, we focused on estimating incremental quality-adjusted life years (QALYs) and health-care costs related to back pain. Analysis was by intention to treat. This study is registered, number ISRCTN37113406.

**Findings** 851 patients were assigned to the intervention ( $n=568$ ) and control groups ( $n=283$ ). Overall, adjusted mean changes in RMDQ scores were significantly higher in the intervention group than in the control group at 4 months ( $4.7$  [SD  $5.9$ ] vs  $3.0$  [ $5.9$ ]), between-group difference  $1.81$  [95% CI  $1.06$ – $2.57$ ] and at 12 months ( $4.3$  [ $6.4$ ] vs  $3.3$  [ $6.2$ ]),  $1.06$  [ $0.25$ – $1.86$ ]), equating to effect sizes of  $0.32$  ( $0.19$ – $0.45$ ) and  $0.19$  ( $0.04$ – $0.33$ ), respectively. At 12 months, stratified care was associated with a mean increase in generic health benefit ( $0.039$  additional QALYs) and cost savings ( $\pounds 240.01$  vs  $\pounds 274.40$ ) compared with the control group.

**Interpretation** The results show that a stratified approach, by use of prognostic screening with matched pathways, will have important implications for the future management of back pain in primary care.

**Funding** Arthritis Research UK.

## Introduction

Back pain remains a major international health problem, with a lifetime prevalence of 80–85%<sup>1</sup> that poses substantial challenges for clinical management.<sup>2</sup> For example, in the UK, each year 6–9% of adults consult their general practitioner about back pain,<sup>3</sup> with only 20–40% no longer reporting pain or disability a year later.<sup>4</sup> Therefore, improvement of the primary care management of low back pain has the potential to reduce the long-term effects of back pain, including persistent disabling symptoms, low quality of life, and reduced capacity to work.<sup>5</sup>

Results of primary care trials<sup>6–9</sup> show that more sophisticated treatments (such as manual treatments, exercise, and cognitive behavioural approaches) are more effective than usual or minimal care of back pain. However, because of insufficient evidence, guideline recommendations<sup>10</sup> are not clear about the clinical selection of patients who are likely to benefit from additional interventions. Generally, for most patients with non-specific back pain, initial referral decisions are based on clinical intuition despite evidence to suggest that this provides inefficient and inconsistent access to treatment.<sup>11</sup> The alternative option, referral of all patients

with back pain for treatment, is generally thought to be unnecessary, impractical, and inefficient because of the high numbers and costs.<sup>5,10–12</sup>

A one-size-fits-all primary care strategy<sup>13</sup> is suboptimum because it ignores the heterogeneity in patients.<sup>14</sup> A novel approach, gaining interest in other medical specialties,<sup>15</sup> but not yet tested in the management of back pain, is to test whether stratified care according to the estimated risk of poor prognosis (defined here as persistent disability because of back pain) improves clinical outcomes while remaining cost effective. We developed a stratified model of primary care management of back pain,<sup>16</sup> which consists of two complementary components. First, a previously validated, simple-to-use prognostic screening method (the Keele STarT Back Screening Tool),<sup>17–19</sup> to allocate patients into one of three risk-defined groups—low, medium, and high (webappendix p 1). Second, three treatment pathways, developed with clinical experts (webappendix p 2), were matched to these risk groups.<sup>16</sup>

In this trial, we test the main hypothesis that a stratified approach to primary care management for low back pain results in clinical and economic benefits compared with current best practice. Our other aims were to test

differences within each of the three risk groups—for low-risk patients, whether minimum treatment provided non-inferior clinical outcomes to current best care; for medium-risk patients, whether systematic referral to physiotherapy led to better clinical outcomes than did current best care; and for high-risk patients, whether systematic referral to psychologically augmented physiotherapy led to better clinical outcomes than did current best care.

## Methods

### Participants

The methods are reported in full in the protocol.<sup>16</sup> In ten general practices within the Keele General Practice Research Partnership, England, adults who had consulted their doctor about back pain during June, 2007, to November, 2008, were identified through weekly searches of electronic patients' records for morbidity codes for back pain.<sup>16</sup> Individuals who were identified were sent a letter from back pain referral services in two National Health Service (NHS) centres, with an invitation to telephone to make an appointment at the initial assessment clinic, information about the trial, and baseline questionnaires. At the clinic, a research nurse assessed patients for eligibility, obtained written informed consent, and checked completion of the questionnaire, including the STarT Back Screening Tool (webappendix p 1).<sup>17</sup> Patients were included in the study if they were at least 18 years old, could speak and understand English, and had back pain of any duration, with or without associated radiculopathy. We excluded patients with potentially serious disorders (eg, cauda equina compression, inflammatory arthritis, and malignancy), serious illness or comorbidity (including those undergoing treatment for a prevalent axis 1 or 2 mental health disorder according to the Diagnostic and Statistical Manual of Mental Disorders, fourth edition [DSM-IV] criteria), who had spinal surgery in the past 6 months, who were pregnant, who were receiving back treatments (except primary care), and who were unable or unwilling to attend (figure 1).

A trial steering and independent data monitoring committee oversaw the trial. The North Staffordshire Local Research Ethics Committee approved the protocol.

### Randomisation and masking

A clinic administrator telephoned a remote clinical trials unit (Keele University), which assigned participants to intervention and control groups by use of computer-generated stratified block randomisation (block sizes of three) in a 2:1 ratio to enable future secondary analysis of targeted treatment mechanisms. Stratification was according to the NHS centre (n=2) and STarT Back Screening Tool risk subgroup. Participants, administrator, or physiotherapists could not be masked to randomisation because therapists were administering the active intervention; however, we ensured that the therapists administering the control treatment were not made aware of the

details of the stratified model of care during the trial, and that patients were made aware that they would be treated according to one of two primary care management models. The research nurse who retrieved outcome data was masked to randomisation, and the concealment strategies were that the nurse was in a separate office during clinic, the administrator randomly assigned the patients, and the use of a masked follow-up database.

### Baseline clinical assessment and treatment

On the same day and in the same NHS centre as the nurse assessment, baseline clinical and treatment sessions were delivered to intervention and control groups by study physiotherapists. To guard against learning and contamination effects, different physiotherapists (13 in the intervention group and 40 in the control group) delivered initial clinic and ongoing physiotherapy. Irrespective of the treatment group, participants were not restricted from using health care elsewhere or seeing their general practitioner during the follow-up.

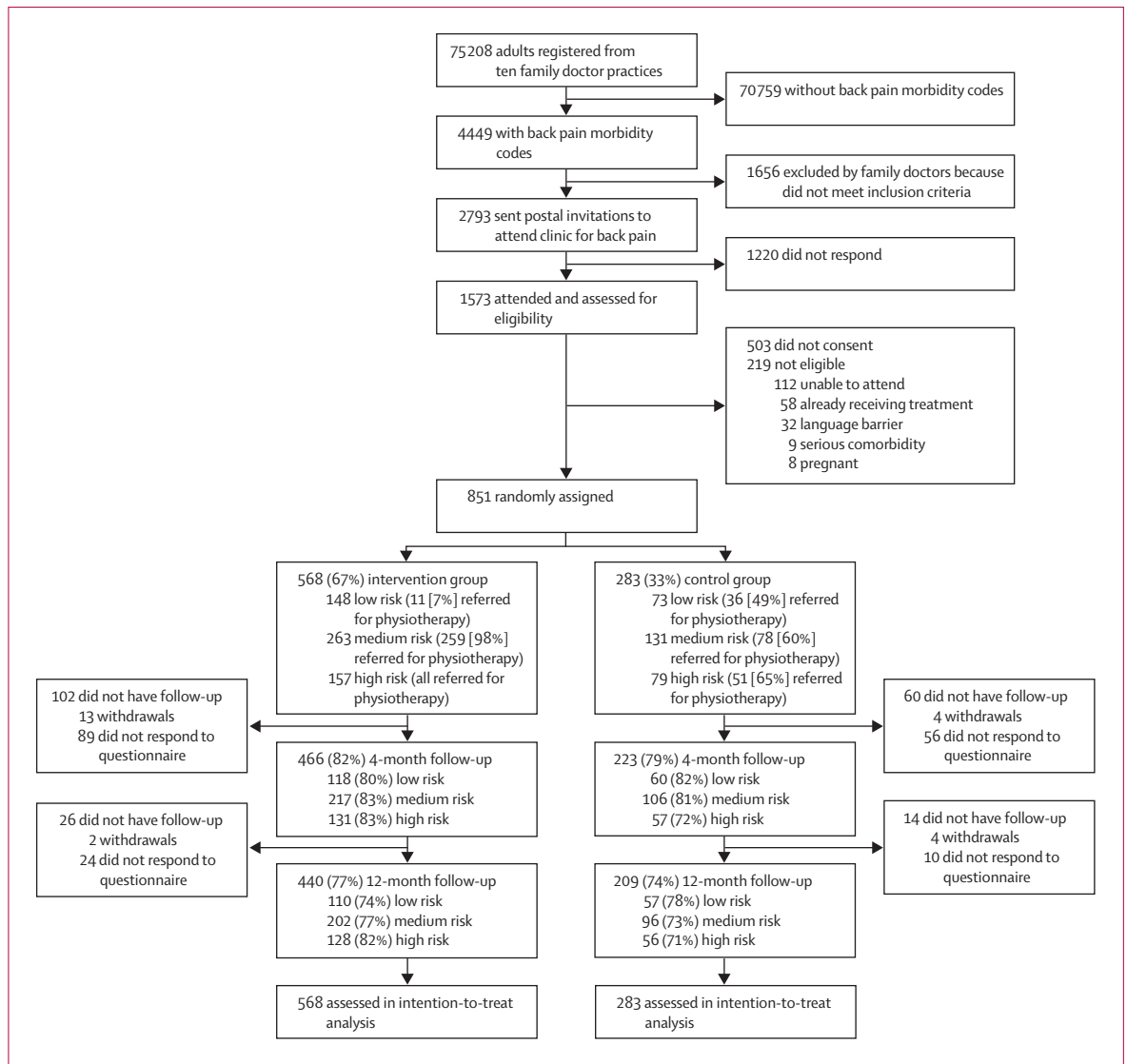
In the control group, during the baseline clinical assessment and treatment session, decisions about referral were made on the basis of the physiotherapists' clinical judgment, without knowledge of a participant's STarT Back Tool classification. Participants received a 30-min physiotherapy assessment and initial treatment including advice and exercises, with the option of onward referral to further physiotherapy. Control physiotherapists received a half day training to familiarise them with study procedures.

In the intervention group, during the baseline clinical assessment and treatment session, decisions about referral were made by use of the STarT Back Screening Tool classification. The 30-min assessment and initial treatment were delivered according to an agreed protocol, with advice focusing on promotion of appropriate levels of activity, including return to work, and a pamphlet about local exercise venues and self-help groups. Participants were shown a 15-min educational video entitled *Get Back Active*<sup>20</sup> and given the *Back Book*.<sup>21</sup> Low-risk patients were only given this clinic session; medium-risk and high-risk patients were referred for further physiotherapy-led treatment sessions. Clinic physiotherapists doing the assessments in the intervention group were given 1 day of training about the STarT Back Screening Tool, and to standardise advice and data gathering from the case report forms.

### Follow-up treatment sessions

In the control group, referral for further physiotherapy was entirely at the clinical discretion of the physiotherapists delivering the baseline session. The physiotherapists to whom they could refer were in usual NHS practice settings and did not overlap with the physiotherapists providing follow-up sessions in the intervention group. Such physiotherapists have general training in physical therapies and some training in more complex psychologically

For the STarT Back trial protocol see <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2377248/pdf/1471-2474-9-58.pdf>



**Figure 1: Trial profile**

Two patients in the control group died during the study (one before 4-month follow-up and one before 12-month follow-up), one developed a serious comorbidity in the intervention group (before 4-month follow-up), and 20 contacted the National Health Service centre to withdraw from the study (14 in intervention group [12 before 4-month follow-up and two before 12-month follow-up] and six in control group [three before 4-month follow-up and three before 12-month follow-up]). At 4 months, 622 (90%) of 689 responders sent their responses in the post (419 [90%] of 466 in intervention group and 203 [91%] of 223 in control group); the remaining 67 (10%) responders supplied key outcome responses by telephone (47 [10%] in intervention group and 20 [9%] in control group). At 12 months, 567 (87%) of 649 responders sent their responses in the post (386 [88%] of 440 in intervention group and 181 [87%] of 209 in control group); 82 (13%) of 649 patients responded by telephone (54 [12%] in intervention group and 28 [13%] in control group). Of 689 responders at the 4-month follow-up, one patient in the control group did not complete the primary outcome measure (Roland and Morris Disability Questionnaire).

informed treatments, but none underwent special training or instruction related to this study.

In the intervention group, medium-risk patients, according to the STarT Back Screening Tool, were referred for standardised physiotherapy to address symptoms and function. High-risk patients were referred for psychologically informed physiotherapy to address physical symptoms and function, and also psychosocial obstacles to recovery. Physiotherapists delivering the medium-risk intervention were given 3 days of additional training, and

those delivering the high-risk intervention had 6 days of additional training (ie, 9 days in total).

**Outcomes**

Demographic data and clinical outcomes were gathered before randomisation and 4 months and 12 months later by use of postal questionnaires. Case report forms were used to gather data about treatment content from the physiotherapists. The primary clinical outcome was the effect of treatment on the Roland and Morris Disability

Questionnaire<sup>22</sup> (RMDQ; scale 0–24; high scores indicate severe disability) score at 12 months.

Descriptions and psychometric properties of the secondary outcome measures are reported in detail elsewhere.<sup>16</sup> Secondary outcome measures were referral for further physiotherapy, back pain intensity, the Pain Catastrophizing Scale<sup>23</sup> (measures the extent to which someone has a pessimistic outlook of back pain), fear-avoidance beliefs (Tampa Scale of Kinesiophobia<sup>24</sup>), Hospital Anxiety and Depression Scale,<sup>25</sup> health-related quality of life (EuroQol EQ-5D;<sup>26</sup> Short Form 12<sup>27</sup> physical and mental component scores), S'TarT Back Screening Tool risk-subgroup reduction, perception of overall change in back pain (global change), number of physiotherapy treatment sessions, attendance at initial physiotherapy treatment, adverse events, health-care resource use and costs over 12 months, number of days off work because of back pain, and satisfaction with care.

We did a telephone follow-up of non-responders at both timepoints for disability, and catastrophizing and global changes. Adverse events were defined as any serious morbidity or events causing unwarranted distress to a participant that were potentially related to either intervention; information was gathered by the physiotherapists and from patients' self-completed questionnaires.

### Statistical analysis

The trial hypotheses were tested by use of pretreatment randomisation allocation to low-risk, medium-risk, and high-risk groups in the intervention and control groups. Since the secondary hypotheses to test differences at a risk-group level required a larger sample size than did the primary hypothesis, we used this for our power calculation. Hence, the sample size calculation was based on the ability to detect, for high-risk and medium-risk groups, a between-treatment mean difference of 2.5 RMDQ points at the 12-month primary endpoint with a 5% two-tailed significance level; and for the low-risk group, an equivalent one-tailed 2.5% level non-inferiority test. Allowing for a 20% loss to follow-up, we aimed to recruit 800 participants. However, during recruitment, on the basis of the recommendation by our trial data monitoring committee, this number was revised to 850 participants to compensate for a larger than anticipated loss at the 12-month follow-up of 25%. Although powered to detect differences in the low-risk, medium-risk, and high-risk groups, the revised total sample size provided 80% power to detect an overall treatment effect size of 0.2, equivalent to a mean difference in RMDQ scores between the intervention and control groups of about 1.

Analysis was by intention to treat. For the primary analyses, imputed datasets were used for all descriptive and inferential assessments to address attrition bias, generated through multiple imputation (pooled estimates of five imputed datasets) by use of simulation based on a multivariate normal model (numerical variables) and a logistic regression model (categorical outcomes).<sup>28</sup>

Estimates of treatment effect (mean difference for numerical outcomes, odds ratios for categorical outcomes, and incidence rate ratios for lost work days), with 95% CI, were obtained by use of linear, binary logistic, ordinal logistic, and Poisson regression models respectively, with adjustment for baseline score, age, sex, RMDQ, and back pain duration (as agreed a priori with our trial steering group). Standardised effect sizes were reported, alongside numbers needed to treat (NNT) by use of at least 30% change in RMDQ;<sup>29</sup> and calculation of 95% CI for NNT by use of Stang and colleagues' recommendations.<sup>30</sup> Sensitivity analyses were done by use of complete-case analysis (ie, non-imputed dataset) and further adjustment for therapist's effects with random-effects modelling of main therapist. All analyses were done with SPSS (version 17.0.1) and STATA (11.0).

The analysis of cost-effectiveness focused on estimation of mean incremental quality-adjusted life years (QALYs) and back-pain-related health-care costs for the overall stratified management approach by use of a within-trial analysis. QALYs were calculated with the EQ-5D. Details of the numbers of physiotherapy sessions attended by each participant were obtained through case report forms and an audit of clinical notes for the participating physiotherapy services. Other health-care costs were estimated from responses to the resource-use items contained within the 12 month self-report questionnaire. Details of the unit costs applied to units of resource use are provided in webappendix p 5. Like the clinical analysis, the economic evaluation was replicated in the complete-case dataset.

To assess the economic consequences of the stratified management intervention beyond health-care resources, costs were also assigned to self-reported work absence by use of the human capital approach; self-reported work absence was weighted by respondent-specific wage rates identified from data for yearly earnings and UK Standard Occupational Classification codes.<sup>31,32</sup> Because of the 12-month follow-up during the study, costs or health benefits were not discounted.

This study is registered, number ISRCTN37113406.

### Role of the funding source

Keele University sponsored the trial, approved the design, and appointed the trial steering and data monitoring committee. The funder was not involved in the preparation of the study protocol, management of the trial, analysis of the data, or preparation of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### Results

Figure 1 shows the trial profile; 851 patients were randomly assigned to the intervention and control groups. The S'TarT Back Screening Tool classified 221 (26%) patients as low risk, 394 (46%) as medium risk, and 236 (28%) as high risk in the intervention and

For the UK Standard Occupation Classification see <http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/standard-occupational-classification-2000/about-soc-2000/index.html>

|                                 | All participants           |                       | Low-risk participants      |                      | Medium-risk participants   |                       | High-risk participants     |                      |
|---------------------------------|----------------------------|-----------------------|----------------------------|----------------------|----------------------------|-----------------------|----------------------------|----------------------|
|                                 | Intervention group (n=568) | Control group (n=283) | Intervention group (n=148) | Control group (n=73) | Intervention group (n=263) | Control group (n=131) | Intervention group (n=157) | Control group (n=79) |
| <b>Demographics</b>             |                            |                       |                            |                      |                            |                       |                            |                      |
| Age (years)                     | 50.1 (15.0)                | 49.1 (14.3)           | 46.5 (14.3)                | 47.6 (14.7)          | 50.5 (15.3)                | 49.3 (13.5)           | 52.7 (14.5)                | 50.1 (15.3)          |
| Sex, female                     | 330 (58%)                  | 170 (60%)             | 82 (55%)                   | 42 (58%)             | 160 (61%)                  | 83 (63%)              | 88 (56%)                   | 45 (57%)             |
| Routine and manual occupations* | 287 (51%)                  | 149 (53%)             | 61 (41%)                   | 26 (36%)             | 137 (52%)                  | 65 (50%)              | 89 (57%)                   | 58 (73%)             |
| Currently in paid employment    | 350 (62%)                  | 174 (61%)             | 112 (76%)                  | 50 (68%)             | 158 (60%)                  | 83 (63%)              | 80 (51%)                   | 41 (52%)             |
| Time off work for back pain†    | 185 (53%)                  | 90 (52%)              | 40 (36%)                   | 12 (24%)             | 95 (60%)                   | 51 (61%)              | 50 (63%)                   | 27 (66%)             |
| <b>Back pain and function</b>   |                            |                       |                            |                      |                            |                       |                            |                      |
| RMDQ disability score           | 9.8 (5.6)                  | 9.7 (5.8)             | 4.6 (3.5)                  | 4.2 (3.3)            | 9.9 (4.5)                  | 9.8 (4.8)             | 14.4 (4.6)                 | 14.7 (4.4)           |
| Back pain intensity             | 5.3 (2.2)                  | 5.2 (2.2)             | 3.4 (1.6)                  | 3.5 (1.7)            | 5.5 (1.7)                  | 5.3 (1.8)             | 7.0 (1.8)                  | 6.8 (2.0)            |
| <b>Duration of back pain</b>    |                            |                       |                            |                      |                            |                       |                            |                      |
| <1 month                        | 97 (17%)                   | 54 (19%)              | 25 (17%)                   | 14 (19%)             | 44 (17%)                   | 24 (18%)              | 28 (18%)                   | 16 (20%)             |
| 1–3 months                      | 124 (22%)                  | 66 (23%)              | 31 (21%)                   | 15 (21%)             | 61 (23%)                   | 39 (30%)              | 32 (20%)                   | 12 (15%)             |
| 3–6 months                      | 81 (14%)                   | 42 (15%)              | 19 (13%)                   | 12 (16%)             | 38 (14%)                   | 16 (12%)              | 24 (15%)                   | 14 (18%)             |
| 6 months to 3 years             | 144 (25%)                  | 65 (23%)              | 45 (30%)                   | 21 (29%)             | 65 (25%)                   | 27 (21%)              | 34 (22%)                   | 17 (22%)             |
| >3 years                        | 122 (21%)                  | 56 (20%)              | 28 (19%)                   | 11 (15%)             | 55 (21%)                   | 25 (19%)              | 39 (25%)                   | 20 (25%)             |
| Radiating leg pain              | 352 (62%)                  | 178 (63%)             | 61 (41%)                   | 28 (38%)             | 176 (67%)                  | 89 (68%)              | 115 (73%)                  | 61 (77%)             |
| Radiating pain below knee       | 179 (32%)                  | 93 (33%)              | 24 (16%)                   | 10 (14%)             | 75 (29%)                   | 47 (36%)              | 80 (51%)                   | 36 (46%)             |
| <b>Psychological measures</b>   |                            |                       |                            |                      |                            |                       |                            |                      |
| PCS, catastrophising score      | 16.3 (10.8)                | 15.9 (11.1)           | 8.6 (5.8)                  | 8.1 (6.5)            | 14.6 (8.1)                 | 13.6 (8.1)            | 26.4 (10.6)                | 26.9 (10.3)          |
| PCS ≥20                         | 186 (33%)                  | 89 (31%)              | 8 (5%)                     | 4 (5%)               | 60 (23%)                   | 28 (21%)              | 118 (75%)                  | 57 (72%)             |
| TSK, fear avoidance score       | 40.3 (6.1)                 | 40.6 (6.3)            | 36.5 (4.9)                 | 36.5 (5.8)           | 39.2 (5.0)                 | 39.7 (4.7)            | 45.8 (5.0)                 | 46.0 (5.7)           |
| HADS, anxiety subscale          | 7.4 (4.1)                  | 7.6 (4.0)             | 5.2 (2.9)                  | 5.4 (3.3)            | 7.0 (3.7)                  | 7.4 (3.7)             | 10.1 (4.2)                 | 10.1 (3.8)           |
| HADS, depression subscale       | 5.8 (4.1)                  | 6.0 (4.1)             | 3.1 (2.7)                  | 3.0 (2.5)            | 5.5 (3.3)                  | 6.0 (3.8)             | 8.9 (4.3)                  | 8.9 (3.7)            |
| <b>General health</b>           |                            |                       |                            |                      |                            |                       |                            |                      |
| Widespread pain‡                | 186 (33%)                  | 104 (37%)             | 33 (22%)                   | 16 (22%)             | 93 (35%)                   | 56 (43%)              | 60 (38%)                   | 32 (41%)             |
| SF12, physical component        | 37.0 (10.4)                | 36.4 (10.6)           | 45.9 (8.6)                 | 46.0 (9.1)           | 35.7 (9.5)                 | 35.1 (8.6)            | 30.8 (7.6)                 | 29.6 (8.2)           |
| SF12, mental component          | 48.2 (11.9)                | 47.8 (11.7)           | 53.7 (7.3)                 | 53.0 (8.1)           | 49.6 (11.5)                | 48.6 (11.3)           | 40.6 (12.5)                | 41.5 (12.3)          |

Data are mean score (SD) or number (%). RMDQ=Roland and Morris Disability Questionnaire. PCS=Pain Catastrophising Scale. TSK=Tampa Scale of Kinesiophobia. HADS=Hospital Anxiety and Depression Scale. SF12=Short Form 12. \*Based on major groups 5–9 of the UK Standard Occupation Classification (2000) for current or most recent paid employment (inclusive of 36 missing data [26 in intervention group and ten in control group]). †Responders who reported being currently in paid employment at baseline and taking time off during the 12 months before baseline. ‡Based on the American College of Rheumatology's definition.<sup>33</sup>

**Table 1: Baseline characteristics of participants randomly assigned according to treatment group in total and stratified according to risk group**

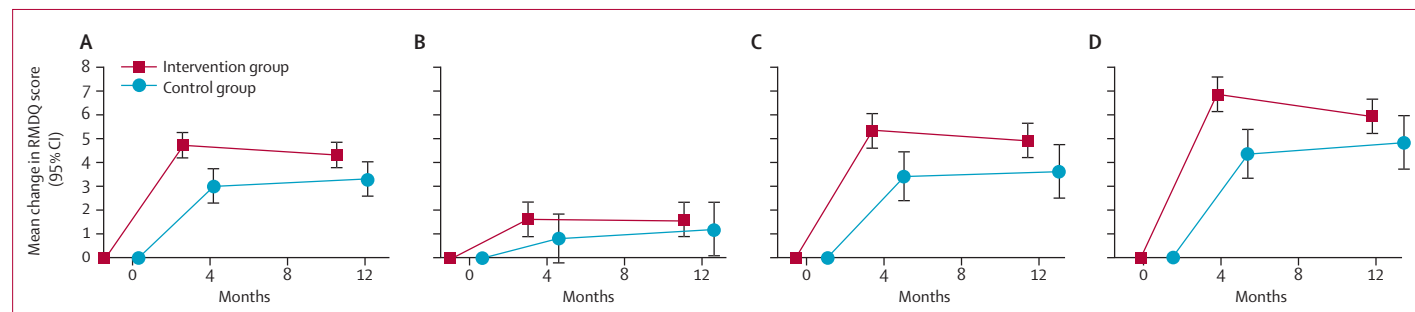
control groups. The mean age of participants was 50 years (range 18–87), and 500 (59%) of 851 patients were women. Table 1 shows the baseline characteristics of the participants. Baseline sex, disability, and duration of pain were similar for participants with data at 12 months and those lost to follow-up. However, those lost to follow-up were younger: mean age of the individuals who responded at 12 months was 52.0 years [SD 14.2] compared with 42.3 years [14.0] for those who did not respond. Responders at 12 months had a mean baseline RMDQ score of 9.8 [5.7] and non-responders had a score of 9.8 [5.6]. 386 (59%) of 649 responders and 114 (56%) of 202 non-responders at 12 months were women. 95 (47%) non-responders versus 292 (45%) responders at 12 months had their pain for longer than 6 months.

The mean number of total treatment sessions was similar in the intervention and control groups (3.9 [SD 2.6] vs 3.8 [3.6], respectively), although a

higher proportion of patients in the intervention group were referred for further physiotherapy (427 [75%] of 568 vs 165 [58%] of 283). In the intervention group, referral decisions were in agreement with those predicated by use of the STarT Back Screening Tool in 553 cases, but not in 15 cases—11 low-risk patients were referred and four medium-risk patients were not referred (figure 1). This difference arose because therapists were allowed to overrule the STarT Back Screening Tool recommendation if the clinician thought that the decision to overrule was appropriate. Referral patterns were substantially different in controls, with more than a third of medium-risk and high-risk patients not referred for physiotherapy, and about half of all low-risk patients referred for treatment (figure 1). Among the participants referred, initial treatment attendance was 528 (93%) in the intervention group and 263 (93%) in the control group, with those in the intervention group receiving fewer physiotherapy sessions (mean

4.2 sessions [SD 2.1] vs 5.1 sessions [3.5]) during a shorter time span (47.7 days [33.0] vs 69.4 days [50.3]) than did participants in the control group. No serious or non-serious adverse events were reported.

Overall, the reduction in RMDQ score (primary outcome) was larger in the intervention group than in the control group (figure 2), even with conservative sensitivity analyses (table 2). For the primary outcome,



**Figure 2:** Mean change from baseline in RMDQ (primary outcome measure) scores at 4-month and 12-month follow-ups in all participants (A), low-risk participants (B), medium-risk participants (C), and high-risk participants (D)  
RMDQ=Roland and Morris Disability Questionnaire.

|                                   | All participants    | p value | Low-risk participants | p value | Medium-risk participants | p value | High-risk participants | p value |
|-----------------------------------|---------------------|---------|-----------------------|---------|--------------------------|---------|------------------------|---------|
| Intervention group                | 568                 | ..      | 148                   | ..      | 263                      | ..      | 157                    | ..      |
| Control group                     | 283                 | ..      | 73                    | ..      | 131                      | ..      | 79                     | ..      |
| <b>4 months</b>                   |                     |         |                       |         |                          |         |                        |         |
| Mean change (SD)                  |                     |         |                       |         |                          |         |                        |         |
| Intervention group                | 4.7 (5.9)           | ..      | 1.6 (4.4)             | ..      | 5.3 (6.0)                | ..      | 6.8 (6.9)              | ..      |
| Control group                     | 3.0 (5.9)           | ..      | 0.8 (4.3)             | ..      | 3.4 (6.1)                | ..      | 4.4 (6.1)              | ..      |
| Mean difference* (95% CI)         | 1.81 (1.06 to 2.57) | <0.0001 | 0.66 (-0.41 to 1.74)  | 0.2211  | 1.99 (0.75 to 3.22)      | 0.0012  | 2.53 (0.90 to 4.16)    | 0.0024  |
| Sensitivity analysis†             | 1.54 (0.77 to 2.30) | <0.0001 | 0.66 (-0.42 to 1.74)  | 0.2275  | 1.59 (0.49 to 2.69)      | 0.0048  | 2.28 (0.49 to 4.07)    | 0.0129  |
| Sensitivity analysis‡             | 1.83 (1.01 to 2.65) | <0.0001 | 0.74 (-0.50 to 1.98)  | 0.2447  | 1.99 (0.73 to 3.25)      | 0.0021  | 2.57 (0.81 to 4.32)    | 0.0045  |
| Effect size§ (95% CI)             | 0.32 (0.19 to 0.45) | <0.0001 | 0.19 (-0.12 to 0.51)  | 0.2211  | 0.43 (0.16 to 0.70)      | 0.0012  | 0.56 (0.20 to 0.92)    | 0.0024  |
| Participants with good outcome¶   |                     |         |                       |         |                          |         |                        |         |
| Intervention group                | 392 (69%)           | ..      | 100 (68%)             | ..      | 186 (71%)                | ..      | 106 (68%)              | ..      |
| Control group                     | 158 (56%)           | ..      | 42 (58%)              | ..      | 75 (57%)                 | ..      | 40 (51%)               | ..      |
| Odds ratio   (95% CI)             | 1.85 (1.36 to 2.51) | <0.0001 | 1.58 (0.83 to 3.00)   | 0.1618  | 2.00 (1.20 to 3.33)      | 0.0073  | 1.96 (1.03 to 3.71)    | 0.0390  |
| Number needed to treat** (95% CI) | 7.0 (13.5 to 4.9)   | <0.0001 | 9.5 (-21.8 to 4.4)    | 0.1618  | 6.5 (22.8 to 4.1)        | 0.0073  | 6.2 (135.0 to 3.5)     | 0.0390  |
| <b>12 months</b>                  |                     |         |                       |         |                          |         |                        |         |
| Mean change (SD)                  |                     |         |                       |         |                          |         |                        |         |
| Intervention group                | 4.3 (6.4)           | ..      | 1.6 (4.5)             | ..      | 4.9 (5.9)                | ..      | 5.9 (7.2)              | ..      |
| Control group                     | 3.3 (6.2)           | ..      | 1.2 (4.8)             | ..      | 3.6 (6.3)                | ..      | 4.8 (6.3)              | ..      |
| Mean difference* (95% CI)         | 1.06 (0.25 to 1.86) | 0.0095  | 0.12 (-1.13 to 1.38)  | 0.8456  | 1.33 (0.15 to 2.52)      | 0.0253  | 1.22 (-0.47 to 2.91)   | 0.1547  |
| Sensitivity analysis†             | 1.06 (0.26 to 1.87) | 0.0099  | 0.29 (-0.86 to 1.44)  | 0.6177  | 1.18 (0.01 to 2.34)      | 0.0482  | 1.30 (-0.54 to 3.13)   | 0.1653  |
| Sensitivity analysis‡             | 1.08 (0.18 to 1.98) | 0.0186  | 0.06 (-1.37 to 1.50)  | 0.9298  | 1.35 (0.12 to 2.58)      | 0.0317  | 1.56 (-0.72 to 3.84)   | 0.1710  |
| Effect size§ (95% CI)             | 0.19 (0.04 to 0.33) | 0.0095  | 0.04 (-0.33 to 0.41)  | 0.8456  | 0.29 (0.03 to 0.55)      | 0.0253  | 0.27 (-0.10 to 0.65)   | 0.1547  |
| Participants with good outcome¶   |                     |         |                       |         |                          |         |                        |         |
| Intervention group                | 367 (65%)           | ..      | 98 (66%)              | ..      | 176 (67%)                | ..      | 93 (59%)               | ..      |
| Control group                     | 160 (57%)           | ..      | 44 (60%)              | ..      | 77 (59%)                 | ..      | 39 (49%)               | ..      |
| Odds ratio   (95% CI)             | 1.48 (1.02 to 2.15) | 0.0344  | 1.22 (0.60 to 2.47)   | 0.5754  | 1.55 (0.93 to 2.57)      | 0.0874  | 1.57 (0.82 to 3.00)    | 0.1723  |
| Number needed to treat** (95% CI) | 10.8 (206.0 to 5.8) | 0.0344  | 21.4 (-7.9 to 5.4)    | 0.5754  | 10.0 (-56.5 to 5.1)      | 0.0874  | 9.0 (-20.2 to 4.0)     | 0.1723  |

Data are number (%), unless otherwise indicated. RMDQ=Roland and Morris Disability Questionnaire. \*Mean for the intervention group minus mean for the control group (by use of linear regression adjusted for age, sex, baseline RMDQ, and duration of back pain). †Based on observed or available case data. ‡Based on imputed datasets with adjustment for clustering by therapist (by use of random-effects linear regression modelling). §Mean difference relative to the pooled SD of baseline scores. ¶Defined as at least 30% change in the RMDQ score compared with baseline; the numbers represent average rounded counts of five imputed datasets. ||Odds of a good outcome in the intervention group relative to the control group (by use of binary logistic regression with adjustment for age, sex, baseline RMDQ, and duration of back pain). \*\*Based on the point estimate for the proportion of patients with good outcomes in the control group and the odds ratio for good outcomes in the intervention group relative to the control group; smaller positive numbers for the 95% CI convey a stronger association than do larger positive numbers and smaller negative numbers indicate a greater advantage towards the control group, and therefore the number needed to treat does not necessarily lie within the 95% CI.

**Table 2:** Mean change in scores with the Roland and Morris Disability Questionnaire (primary outcome measure) during 4-month and 12-month follow-ups

the between-group adjusted mean differences in change in RMDQ scores were significant at 4 months and 12 months, equating to standardised effect sizes of 0·32 and 0·19, respectively (table 2).

Differences in secondary outcome measures in favour of the intervention for all participants were significant at 4 months for pain intensity, catastrophising, fear, anxiety, depression, general health (physical component), STarT Back Screening Tool risk reduction, and global change (table 3); and at 12 months for catastrophising, fear, depression, general health (physical component), and risk reduction (table 3). The patients in the intervention group were significantly more likely to be satisfied with treatment (data available for 4 months follow-up), and took fewer days off work because of back pain (data available for 12 months follow-up).

Details of health-care resource use, health-care costs, days off work, and health-related quality of life (EQ-5D scores and QALY estimates) are provided in webappendix p 7 for each group. The stratified management intervention resulted in greater mean health benefit (0·039 additional QALYs), achieved at a lower mean health-care cost (cost saving £34·39; webappendix p 8), than the control. Similar inferences were drawn from the complete-case analysis (0·033 additional QALYs and a

mean cost saving of £41·93; webappendix p 8). Figure 3 shows the cost-effectiveness plane for the primary analysis, generated from 25 000 bootstrap samples.<sup>16</sup> The dominance of the stratified intervention (ie, greater benefit at lower cost) was shown in 92% of replications (ie, 92% of bootstrapped cost-effect pairs were in the southeast quadrant).

The societal benefit from fewer work days lost because of back pain corresponded to a mean indirect (productivity) cost saving of £675 over the 12-month follow-up for the intervention group compared with the control group (webappendix p 8).

For the primary and secondary clinical outcomes, differences between the intervention and control groups within the low-risk group indicated non-inferiority, although significantly fewer work days were lost in the intervention group (table 2, table 3). For the medium-risk and high-risk groups, the adjusted between-group mean differences in RMDQ scores were significant at 4 months (table 2), but the difference was only significant for the medium-risk group and not the high-risk group at 12 months (table 2). One notable finding was that the mean number of work days lost at 12 months was far fewer in the medium-risk participants in the intervention group than in the control group.

|                              | All participants                      |                             |                               |         | Low-risk participants                 |                            |                               |         | Medium-risk participants              |                             |                               |         | High-risk participants                |                            |                               |         |
|------------------------------|---------------------------------------|-----------------------------|-------------------------------|---------|---------------------------------------|----------------------------|-------------------------------|---------|---------------------------------------|-----------------------------|-------------------------------|---------|---------------------------------------|----------------------------|-------------------------------|---------|
|                              | Inter-<br>vention<br>group<br>(n=568) | Control<br>group<br>(n=283) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=148) | Control<br>group<br>(n=73) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=263) | Control<br>group<br>(n=131) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=157) | Control<br>group<br>(n=79) | Point<br>estimate<br>(95% CI) | p value |
| <b>Back pain intensity</b>   |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months, mean change (SD)   | 3·2 (2·5)                             | 2·6 (2·4)                   | 0·55* (0·23 to 0·86)          | 0·0005  | 1·7 (2·2)                             | 1·5 (2·1)                  | 0·28* (-0·29 to 0·84)         | 0·3182  | 3·5 (2·6)                             | 2·8 (2·1)                   | 0·58* (0·06 to 1·10)          | 0·0200  | 4·2 (2·3)                             | 3·4 (2·9)                  | 0·73* (0·04 to 1·42)          | 0·0316  |
| 12 months, mean change (SD)  | 3·0 (2·8)                             | 2·8 (2·6)                   | 0·11* (-0·21 to 0·43)         | 0·5002  | 1·7 (2·3)                             | 1·7 (2·4)                  | 0·10* (-0·51 to 0·70)         | 0·7461  | 3·3 (2·6)                             | 3·0 (2·8)                   | 0·16* (-0·44 to 0·76)         | 0·5802  | 3·7 (2·7)                             | 3·6 (3·2)                  | -0·01* (-0·72 to 0·70)        | 0·9755  |
| <b>STarT Back risk group</b> |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months                     | ..                                    | ..                          | 2·09† (1·40 to 3·12)          | 0·0003  | ..                                    | ..                         | 1·16† (0·36 to 3·67)          | 0·8021  | ..                                    | ..                          | 1·70† (0·95 to 3·03)          | 0·0678  | ..                                    | ..                         | 3·14† (1·63 to 6·06)          | 0·0006  |
| Low risk                     | 439 (77%)                             | 186 (66%)                   | ..                            | ..      | 134 (91%)                             | 65 (89%)                   | ..                            | ..      | 201 (76%)                             | 89 (68%)                    | ..                            | ..      | 104 (66%)                             | 32 (41%)                   | ..                            | ..      |
| Medium risk                  | 103 (18%)                             | 70 (25%)                    | ..                            | ..      | 12 (8%)                               | 7 (10%)                    | ..                            | ..      | 55 (21%)                              | 36 (27%)                    | ..                            | ..      | 36 (23%)                              | 27 (34%)                   | ..                            | ..      |
| High risk                    | 26 (5%)                               | 27 (10%)                    | ..                            | ..      | 2 (1%)                                | 1 (1%)                     | ..                            | ..      | 7 (3%)                                | 6 (5%)                      | ..                            | ..      | 17 (11%)                              | 20 (25%)                   | ..                            | ..      |
| 12 months                    | ..                                    | ..                          | 1·51† (1·01 to 2·25)          | 0·0405  | ..                                    | ..                         | 0·91† (0·31 to 2·65)          | 0·8562  | ..                                    | ..                          | 1·51† (0·85 to 2·66)          | 0·1525  | ..                                    | ..                         | 1·62† (0·82 to 3·23)          | 0·1566  |
| Low risk                     | 430 (76%)                             | 195 (69%)                   | ..                            | ..      | 131 (89%)                             | 65 (89%)                   | ..                            | ..      | 204 (78%)                             | 94 (72%)                    | ..                            | ..      | 95 (61%)                              | 36 (46%)                   | ..                            | ..      |
| Medium risk                  | 111 (20%)                             | 71 (25%)                    | ..                            | ..      | 17 (11%)                              | 7 (10%)                    | ..                            | ..      | 56 (21%)                              | 35 (27%)                    | ..                            | ..      | 38 (24%)                              | 29 (37%)                   | ..                            | ..      |
| High risk                    | 27 (5%)                               | 17 (6%)                     | ..                            | ..      | 0 (1%)                                | 1 (1%)                     | ..                            | ..      | 3 (1%)                                | 2 (2%)                      | ..                            | ..      | 24 (15%)                              | 14 (18%)                   | ..                            | ..      |

(Continues on next page)

|                                | All participants                      |                             |                               |         | Low-risk participants                 |                            |                               |         | Medium-risk participants              |                             |                               |         | High-risk participants                |                            |                               |         |
|--------------------------------|---------------------------------------|-----------------------------|-------------------------------|---------|---------------------------------------|----------------------------|-------------------------------|---------|---------------------------------------|-----------------------------|-------------------------------|---------|---------------------------------------|----------------------------|-------------------------------|---------|
|                                | Inter-<br>vention<br>group<br>(n=568) | Control<br>group<br>(n=283) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=148) | Control<br>group<br>(n=73) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=263) | Control<br>group<br>(n=131) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=157) | Control<br>group<br>(n=79) | Point<br>estimate<br>(95% CI) | p value |
| (Continued from previous page) |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| Global change                  |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months                       | ..                                    | ..                          | 1.55†<br>(1.16 to<br>2.07)    | 0.0032  | ..                                    | ..                         | 1.45†<br>(0.80 to<br>2.60)    | 0.2153  | ..                                    | ..                          | 1.57†<br>(1.04 to<br>2.36)    | 0.0304  | ..                                    | ..                         | 1.70†<br>(0.98 to<br>2.96)    | 0.0595  |
| Much better                    | 252<br>(44%)                          | 99<br>(35%)                 | ..                            | ..      | 70<br>(47%)                           | 28<br>(38%)                | ..                            | ..      | 125<br>(48%)                          | 50<br>(38%)                 | ..                            | ..      | 57<br>(36%)                           | 21<br>(27%)                | ..                            | ..      |
| Better                         | 151<br>(27%)                          | 82<br>(29%)                 | ..                            | ..      | 38<br>(26%)                           | 23<br>(32%)                | ..                            | ..      | 70<br>(27%)                           | 40<br>(31%)                 | ..                            | ..      | 43<br>(27%)                           | 19<br>(24%)                | ..                            | ..      |
| Not better                     | 165<br>(29%)                          | 102<br>(36%)                | ..                            | ..      | 40<br>(27%)                           | 22<br>(30%)                | ..                            | ..      | 68<br>(26%)                           | 41<br>(31%)                 | ..                            | ..      | 57<br>(36%)                           | 39<br>(49%)                | ..                            | ..      |
| 12 months                      | ..                                    | ..                          | 1.23†<br>(0.95 to<br>1.60)    | 0.1104  | ..                                    | ..                         | 0.80†<br>(0.41 to<br>1.54)    | 0.4812  | ..                                    | ..                          | 1.54†<br>(0.99 to<br>2.39)    | 0.0520  | ..                                    | ..                         | 1.26†<br>(0.72 to<br>2.21)    | 0.4072  |
| Much better                    | 229<br>(40%)                          | 99<br>(35%)                 | ..                            | ..      | 57<br>(39%)                           | 32<br>(44%)                | ..                            | ..      | 116<br>(44%)                          | 46<br>(35%)                 | ..                            | ..      | 56<br>(36%)                           | 21<br>(27%)                | ..                            | ..      |
| Better                         | 121<br>(21%)                          | 63<br>(22%)                 | ..                            | ..      | 36<br>(24%)                           | 17<br>(23%)                | ..                            | ..      | 59<br>(22%)                           | 31<br>(24%)                 | ..                            | ..      | 26<br>(17%)                           | 15<br>(19%)                | ..                            | ..      |
| Not better                     | 218<br>(38%)                          | 121<br>(43%)                | ..                            | ..      | 55<br>(37%)                           | 24<br>(33%)                | ..                            | ..      | 88<br>(33%)                           | 54<br>(41%)                 | ..                            | ..      | 75<br>(48%)                           | 43<br>(54%)                | ..                            | ..      |
| PCS-catastrophising            |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months, mean change (SD)     | 6.4<br>(10.8)                         | 3.9<br>(10.4)               | 2.35*<br>(0.91 to<br>3.79)    | 0.0010  | 2.6<br>(7.3)                          | 1.3<br>(8.0)               | 1.10*<br>(-0.82 to<br>3.02)   | 0.2590  | 5.9<br>(10.1)                         | 3.6<br>(10.1)               | 1.88*<br>(-0.09 to<br>3.84)   | 0.0531  | 10.8<br>(12.1)                        | 6.9<br>(13.3)              | 3.91*<br>(0.41 to<br>7.41)    | 0.0242  |
| 12 months, mean change (SD)    | 6.1<br>(10.6)                         | 4.4<br>(11.4)               | 1.69*<br>(0.37 to<br>3.01)    | 0.0113  | 3.2<br>(8.6)                          | 1.6<br>(7.7)               | 1.47*<br>(-0.60 to<br>3.54)   | 0.1516  | 5.7<br>(9.0)                          | 3.6<br>(10.5)               | 1.74*<br>(-0.11 to<br>3.60)   | 0.0630  | 9.4<br>(12.1)                         | 8.1<br>(12.7)              | 1.35*<br>(-1.70 to<br>4.40)   | 0.3816  |
| TSK-fear avoidance             |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months, mean change (SD)     | 5.5<br>(7.0)                          | 3.2<br>(6.0)                | 2.52*<br>(1.61 to<br>3.43)    | <0.0001 | 3.5<br>(7.1)                          | 2.8<br>(7.3)               | 0.86*<br>(-1.01 to<br>2.73)   | 0.3540  | 5.2<br>(7.9)                          | 2.8<br>(5.7)                | 2.72*<br>(1.45 to<br>3.98)    | <0.0001 | 7.9<br>(8.0)                          | 4.3<br>(6.0)               | 3.70*<br>(1.70 to<br>5.70)    | 0.0003  |
| 12 months, mean change (SD)    | 5.2<br>(7.3)                          | 3.3<br>(10.3)               | 2.18*<br>(0.73 to<br>3.63)    | 0.0013  | 3.5<br>(7.2)                          | 2.4<br>(9.3)               | 1.22*<br>(-1.26 to<br>3.70)   | 0.3030  | 4.9<br>(7.6)                          | 3.2<br>(8.2)                | 2.14*<br>(0.23 to<br>4.04)    | 0.0181  | 7.3<br>(7.7)                          | 4.3<br>(7.2)               | 3.12*<br>(1.18 to<br>5.06)    | 0.0016  |
| HADS, anxiety subscale         |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months, mean change (SD)     | 1.7<br>(3.6)                          | 1.2<br>(4.0)                | 0.63*<br>(0.08 to<br>1.18)    | 0.0208  | 0.6<br>(3.3)                          | 0.9<br>(3.5)               | -0.17*<br>(-0.97 to<br>0.62)  | 0.6666  | 1.7<br>(3.8)                          | 0.8<br>(3.7)                | 1.10*<br>(0.33 to<br>1.86)    | 0.0039  | 2.8<br>(4.3)                          | 2.2<br>(4.5)               | 0.56*<br>(-0.57 to<br>1.69)   | 0.3155  |
| 12 months, mean change (SD)    | 1.3<br>(3.9)                          | 1.0<br>(4.4)                | 0.45*<br>(-0.10 to<br>1.01)   | 0.1035  | 0.5<br>(3.2)                          | 0.8<br>(4.0)               | -0.13*<br>(-1.22 to<br>0.96)  | 0.8023  | 1.3<br>(4.2)                          | 0.6<br>(4.2)                | 0.88*<br>(0.08 to<br>1.68)    | 0.0283  | 2.1<br>(4.5)                          | 1.7<br>(5.0)               | 0.28*<br>(-1.14 to<br>1.71)   | 0.6783  |
| HADS, depression subscale      |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months, mean change (SD)     | 1.7<br>(3.7)                          | 1.1<br>(3.3)                | 0.71*<br>(0.26 to<br>1.15)    | 0.0018  | 0.3<br>(3.2)                          | 0.2<br>(3.3)               | 0.09*<br>(-0.69 to<br>0.88)   | 0.8086  | 1.7<br>(3.6)                          | 1.2<br>(3.5)                | 0.79*<br>(0.13 to<br>1.45)    | 0.0193  | 3.0<br>(4.3)                          | 1.9<br>(3.8)               | 1.13*<br>(0.16 to<br>2.09)    | 0.0223  |
| 12 months, mean change (SD)    | 1.4<br>(4.1)                          | 0.9<br>(4.0)                | 0.62*<br>(0.07 to<br>1.17)    | 0.0227  | 0.2<br>(3.3)                          | 0.2<br>(3.5)               | -0.03*<br>(-0.92 to<br>0.86)  | 0.9494  | 1.3<br>(3.7)                          | 1.0<br>(3.8)                | 0.63*<br>(-0.10 to<br>1.36)   | 0.0863  | 2.7<br>(4.7)                          | 1.5<br>(4.5)               | 1.21*<br>(0.02 to<br>2.40)    | 0.0413  |
| SF12, physical component       |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months, mean change (SD)     | -7.5<br>(13.0)                        | -5.2<br>(13.3)              | -2.83*<br>(-4.62 to<br>-1.05) | 0.0010  | -3.2<br>(9.6)                         | -1.8<br>(9.7)              | -1.50*<br>(-3.89 to<br>0.89)  | 0.2137  | -9.1<br>(11.7)                        | -6.4<br>(10.7)              | -3.27*<br>(-5.51 to<br>-1.03) | 0.0037  | -8.9<br>(15.1)                        | -6.4<br>(15.8)             | -3.13*<br>(-7.50 to<br>1.23)  | 0.1215  |
| 12 months, mean change (SD)    | -7.5<br>(11.3)                        | -5.2<br>(10.9)              | -2.93*<br>(-4.31 to<br>-1.56) | <0.0001 | -4.0<br>(9.7)                         | -2.4<br>(10.1)             | -1.70*<br>(-4.25 to<br>0.85)  | 0.1825  | -8.8<br>(11.5)                        | -5.7<br>(11.7)              | -3.74*<br>(-6.02 to<br>-1.46) | 0.0012  | -8.6<br>(12.2)                        | -6.8<br>(13.1)             | -2.46*<br>(-5.90 to<br>0.97)  | 0.1464  |
| SF12, mental component         |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |

(Continues on next page)



|                                | All participants                      |                             |                               |         | Low-risk participants                 |                            |                               |         | Medium-risk participants              |                             |                               |         | High-risk participants                |                            |                               |         |
|--------------------------------|---------------------------------------|-----------------------------|-------------------------------|---------|---------------------------------------|----------------------------|-------------------------------|---------|---------------------------------------|-----------------------------|-------------------------------|---------|---------------------------------------|----------------------------|-------------------------------|---------|
|                                | Inter-<br>vention<br>group<br>(n=568) | Control<br>group<br>(n=283) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=148) | Control<br>group<br>(n=73) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=263) | Control<br>group<br>(n=131) | Point<br>estimate<br>(95% CI) | p value | Inter-<br>vention<br>group<br>(n=157) | Control<br>group<br>(n=79) | Point<br>estimate<br>(95% CI) | p value |
| (Continued from previous page) |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| 4 months, mean change (SD)     | -2.1 (11.3)                           | -2.1 (11.0)                 | -0.16* (-1.63 to 1.32)        | 0.8327  | 0.5 (12.4)                            | -1.0 (10.2)                | 0.96* (-1.68 to 3.60)         | 0.4640  | -1.5 (10.4)                           | -1.1 (12.0)                 | -0.80* (-2.92 to 1.31)        | 0.4497  | -5.5 (12.5)                           | -4.8 (14.3)                | -0.27* (-3.98 to 3.45)        | 0.8811  |
| 12 months, mean change (SD)    | -1.7 (13.0)                           | -1.2 (13.4)                 | -0.69* (-2.39 to 1.01)        | 0.4050  | 1.3 (10.7)                            | -0.4 (9.9)                 | 1.18* (-1.12 to 3.49)         | 0.3124  | -1.2 (12.3)                           | -0.1 (12.7)                 | -1.53* (-3.62 to 0.55)        | 0.1468  | -5.5 (13.8)                           | -3.6 (13.8)                | -1.31* (-5.04 to 2.43)        | 0.4682  |
| <b>Work loss‡</b>              |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
| Days off work, mean (SD)       | 4.4 (21.2)                            | 12.2 (35.1)                 | 2.76§ (2.52 to 3.01)          | <0.0001 | 0.4 (1.2)                             | 3.0 (11.9)                 | 7.06§ (4.28 to 11.6)          | <0.0001 | 4.1 (15.8)                            | 18.4 (47.2)                 | 5.17§ (4.52 to 5.92)          | <0.0001 | 9.9 (35.4)                            | 10.6 (18.2)                | 1.46§ (1.24 to 1.72)          | <0.0001 |
| <b>Satisfaction with care¶</b> |                                       |                             |                               |         |                                       |                            |                               |         |                                       |                             |                               |         |                                       |                            |                               |         |
|                                | ..                                    | ..                          | 1.98† (1.45 to 2.70)          | <0.0001 | ..                                    | ..                         | 1.24† (0.68 to 2.25)          | 0.4784  | ..                                    | ..                          | 2.15† (1.35 to 3.40)          | 0.0013  | ..                                    | ..                         | 3.18† (1.68 to 6.00)          | 0.0005  |
| Very satisfied                 | 166/416 (40%)                         | 46/197 (23%)                | ..                            | ..      | 30/106 (28%)                          | 14/56 (25%)                | ..                            | ..      | 87/197 (44%)                          | 22/91 (24%)                 | ..                            | ..      | 49/113 (43%)                          | 10/50 (20%)                | ..                            | ..      |
| Quite satisfied                | 139/416 (33%)                         | 81/197 (41%)                | ..                            | ..      | 33/106 (31%)                          | 21/56 (38%)                | ..                            | ..      | 63/197 (32%)                          | 39/91 (43%)                 | ..                            | ..      | 43/113 (38%)                          | 21/50 (42%)                | ..                            | ..      |
| Not satisfied                  | 111/416 (27%)                         | 70/197 (36%)                | ..                            | ..      | 43/106 (41%)                          | 21/56 (38%)                | ..                            | ..      | 47/197 (24%)                          | 30/91 (33%)                 | ..                            | ..      | 21/113 (19%)                          | 19/50 (38%)                | ..                            | ..      |

Data are number (%), unless otherwise indicated. PCS=Pain Catastrophizing Scale. TSK=Tampa Scale of Kinesiophobia. HADS=Hospital Anxiety and Depression Scale. SF12=Short Form 12. RMDQ=Roland and Morris Disability Questionnaire. \*Mean difference—ie, mean for the intervention group minus mean for the control group (by use of linear regression adjusted for age, sex, baseline score, baseline RMDQ, and duration of back pain). †Odds ratio represents the proportional odds in terms of ordered response categories for the intervention group relative to the control group (by use of ordinal logistic regression adjusted for age, sex, baseline RMDQ, and duration of back pain). ‡Analysis of time off work (days) is based on a total subsample of 298 of 567 responders who reported being currently employed at 12 months follow-up, and relates to leave due to low back pain during the period between baseline and follow-up at 12 months. §Incidence rate ratio for the extra work days lost in the control group relative to the intervention group (by use of Poisson regression adjusted for age, sex, baseline RMDQ, and duration of back pain). ¶Analysis was based on 613 responders to the question about satisfaction with care at 4 months follow-up: 416 in the intervention group and 197 in the control group.

Table 3: Summary data for secondary outcome measures at follow-ups of 4 months and 12 months

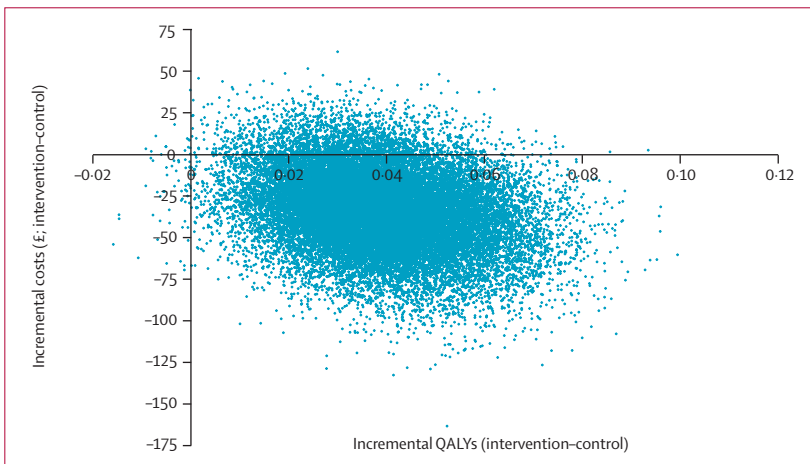


Figure 3: Cost-effectiveness plane for the comparison of the stratified management approach (intervention group) with current best practice (control group), based on 25 000 bootstrapped cost-effect pairs. QALYs=quality-adjusted life years.

### Discussion

A stratified management approach in which prognostic screening and treatment targeting were combined resulted in improved primary care efficiency, leading to higher health gains for patients with back pain than did

existing non-stratified best care. Significant improvements were not only noted in the primary outcome measure (disability) at both 4-month and 12-month follow-ups, but also for a range of secondary outcome measures, including physical and emotional functioning, pain intensity, quality of life, days off work, global improvement ratings, and treatment satisfaction. Although the effect sizes for the targeted intervention were similar to other primary care trials,<sup>6-9</sup> the additional benefit provided by stratified care is noteworthy when the size of the disability reduction in the control group is considered. Mean RMDQ change scores in the control group were larger at 4 months and 12 months than the within-group change of 2.5 points generally judged to be a clinically meaningful change,<sup>34</sup> and also larger than reductions noted in the active intervention groups of other trials (panel).<sup>6,7,9</sup>

From an economics perspective, the stratified management approach was associated with improvements in health-related quality of life (QALYs), a reduction in health-care use, and fewer days off work related to back pain. Health benefits and cost savings attributable to the stratified intervention might have been underestimated; resource use outside the study clinic sessions was less common and the between-group difference in quality of

life was greater at 12 months than at 4 months. The potential for long-term economic benefits lends further support that a stratified management approach provides value for money.

Use of the screening method resulted in important differences in the pattern of treatment referral between the groups. Outcomes in the low-risk group were non-inferior despite far less low-risk intervention patients (7%) having a referral for further treatment than the low-risk controls (49%) referred for an average of five physiotherapy sessions. This finding was in keeping with our theory that a substantial proportion of referrals based on clinical judgement alone (controls) might be unnecessary and that many low-risk patients are receiving unnecessary treatment in current practice.<sup>35</sup> By contrast, 113 (40%) medium-risk and 91 (32%) high-risk patients in the control group were not offered further treatments, which is likely to have contributed to their significantly smaller reductions in disability than in the medium-risk and high-risk patients in the intervention group at

4 months and 12 months. Our interpretation of these findings is that, without systematic prognostic screening to assist treatment referral, many medium-risk and high-risk patients are potentially being denied access to more sophisticated treatments that are likely to improve their clinical outcomes. This issue is perhaps most important for patients classified as high risk (28%) who showed significant, substantial reductions in disability (RMDQ) at 4 months (mean change 6·8; table 2). Although these results lend some support to the use of psychologically informed physiotherapy for patients with psychological distress,<sup>7</sup> further research is needed to establish if these short-term benefits can be sustained, because the difference between the high-risk groups at 12 months (5·9) was not significant. Noteworthy is that, with the design used for this pragmatic clinical trial, we are not able to ascertain whether risk-group level benefits were the result of improvements in referral patterns, or to improvements in the content or quality of the follow-up physiotherapy sessions. Nevertheless, our findings support the clinical effectiveness and cost-effectiveness of a combined stratified management approach.

The strengths of this trial include high internal validity, with remote randomisation, treatments delivered according to protocols, effective masking of assessors, consistent findings for several disparate outcomes, and a sample size large enough to enable a separate examination of the effectiveness within the low-risk, medium-risk, and high-risk groups. Another strength is that the sensitivity analyses, including a conservative approach with adjustment for any potentially unequal therapist expertise or skills (therapist effects),<sup>36</sup> did not change the clinical or economic findings. Limitations included a greater than anticipated loss to follow-up and a slight imbalance in attrition between the groups, addressed through extended recruitment and imputation. Also because delivery of physiotherapy was complex, with a variable number of sessions by a variable number of different physiotherapists, the random-effect adjustment for the therapist could only be applied in the adjustment for differences with a designated main therapist. Additional between-group variability might have remained unadjusted.

The wider implications of the results of this trial are that patients' outcomes can be improved with a stratified approach to primary care management of low back pain. Results from primary care cohort studies<sup>18</sup> suggest that the low-risk group might represent as many as 56% of all back pain consultations with the family doctor. Hence, substantial reductions could have occurred in health-care use and time off work if stratified primary care were implemented. We recognise the challenges in implementing these findings in clinical services, and that our results might not be generalisable to other health-care settings or musculoskeletal complaints. Our research group and others are investigating implementing stratified management into routine primary care in the

#### Panel: Research in context

##### Systematic review

Previous reports of relevant large randomised controlled trials, international guidelines, and meeting abstracts were searched with Ovid Medline up to May, 2011, using exploded Medline Medical Subject Headings and free text terms for "back pain" in combination with "primary health care". This search was restricted to English language papers and the few Medline clinical queries were used to restrict the search to studies of treatments only (to achieve maximum specificity). The evidence suggests that sophisticated treatments (such as manual treatments and exercise) are more effective than are minimal packages of care but similar to each other.<sup>10</sup> There are no trials of methods for tailoring treatment to individual patients or investigation of the clinical effectiveness and cost-effectiveness of using prognostic screening protocols to help identify patients to be targeted with referral beyond minimum treatment.<sup>10</sup>

##### Interpretation

This trial is unique in that the results show the role of system-wide changes with a stratified approach to improve outcomes in patients with low back pain and reduce health-care costs. Although the effect sizes were similar to those of other trials,<sup>6-9</sup> these improvements are noteworthy because improvements in the control group were substantial and similar to those of the active interventions in other clinical trials,<sup>6-9</sup> the control included sophisticated treatments, not just minimum care; and the stratified intervention was highly cost effective. Although referral rates with stratified management were higher, these health-sector costs were outweighed by savings due to reductions in referral of low-risk patients and overall use of health-care resources during the follow-up. The results of this trial have important implications for commissioners and providers of services for back pain.

UK,<sup>37</sup> with further collaborative research in the USA, Denmark, and the Netherlands.

For many years, the potential for targeting treatment has been emphasised as a research priority for back pain.<sup>38</sup> The results of this trial provide the first evidence that a stratified management approach to target the provision of primary care significantly improves patient outcomes and is associated with substantial economic benefits compared with current best practice. As such, the findings of this study represent an important advance in primary care management of back pain, and have important implications for commissioners and providers of services for back pain.

#### Contributors

JCH led the development of the Keele STarT Back Screening Tool and matched treatment pathways; contributed to the project management and interpretation of the statistical analyses, and was responsible for drafting the final report. DGTW is the health economist and was responsible for the analysis and reporting cost-effectiveness, and contributed to the final report. ML was the study statistician and was responsible for statistical analysis and reporting. SB provided additional health economic support for the study. KMD participated in the design and conduct of the trial, and contributed to the final report. NEF provided her expertise in primary care, low back pain, and trials, participated in data interpretation, the conception of the protocol, and contributed to the acquisition of funding and to the discussion section of the final report. KK, CJM, GS, and SS were responsible for developing and overseeing the low-risk, medium-risk, and high-risk interventions. EM was the study manager and was responsible for the set up and management of the data. KV was the head research nurse and was responsible for managing recruitment and follow-up. EMH was the chief investigator and grant holder of the trial, responsible for the original idea and conception of the protocol, and the first author of the protocol paper; she contributed to the design of the interventions, oversaw the project conduct, analysis, and reporting, and had overall responsibility for the final report.

#### STarT Back Trial investigators

*Clinicians:* S Baker, A Bishop, H Bradbury, C Doyle, J Fleet, J French, K Gorsky, J Hill, M Holden, L Huckfield, A Jones, K Konstantinou, K Major, E Mason, S McKenna, L Noble, C Powys, G Sowden, E Stuart, L Wood.

*Trial support:* J Bailey, C Calverley, C Clements, W Clowe, P Croft, K Dziedzic, J Handy, R Hughes, J Jordon, R Mullis, F Rawlings, I Thomas, D van der Windt, K Vohora, C Warlow, J Young.

*Trial Steering Committee and Data Monitoring Committee (independent members):* C Jerosch-Herold, R Mcmanus, C Roberts, J Selfe.

#### Conflicts of interest

We declare that we have no conflicts of interest.

#### Acknowledgments

Arthritis Research UK funded the study (grant code I7741). We thank all the participants and general practices who participated in the STarT Back trial; and NHS Stoke-on-Trent and North Staffordshire for hosting the conduct of the study, and the North Staffordshire Primary Care Research Consortium and the primary care delivery arm of West Midlands North Comprehensive Local Research Network for their service support funding.

#### References

- WHO. The burden of musculoskeletal conditions at the start of the new millennium. *World Health Organ Tech Rep Ser* 2003; **919**: 1–218.
- Costa-Black KM, Loisel P, Anema JR, Pransky G. Back pain and work. *Best Pract Res Clin Rheumatol* 2010; **24**: 227–40.
- Dunn KM, Croft PR. Classification of low back pain in primary care: using “bothersomeness” to identify the most severe patients. *Spine* 2005; **130**: 1887–92.
- Hestback L, Leboeuf YC, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. *Eur Spine J* 2003; **12**: 149–65.
- McGrail MP Jr, Lohman WH, Gorman R. Disability prevention principles in the primary care office. *Am Fam Phys* 2001; **63**: 679–84.
- Lamb SE, Lall R, Hansen Z, et al. A multi-centred randomised controlled trial of a primary care-based cognitive behavioural programme for low back pain. The Back Skills Training (BeST) trial. *Health Technol Assess* 2010; **14**: 1–iv.
- Hollingshurst S, Sharp D, Ballard K, et al. Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain: economic evaluation. *BMJ* 2008; **337**: a2656.
- Hay EM, Mullis R, Lewis M, et al. Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet* 2005; **365**: 2024–30.
- UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *BMJ* 2004; **329**: 1377.
- Savigny P, Kuntze S, Watson P, et al. Low back pain: early management of persistent non-specific low back pain. London: National Institute of Clinical Evidence, 2009. <http://www.nice.org.uk/CG88> (accessed March 19, 2010).
- Department of Health (UK). The musculoskeletal services framework. A joint responsibility: doing it differently. Leeds: Department of Health, 2006.
- Imison C, Naylor C. Referral management: lessons for success. London: King's Fund, 2010. [http://www.kingsfund.org.uk/publications/referral\\_management.html](http://www.kingsfund.org.uk/publications/referral_management.html) (accessed on Dec 14, 2010).
- Foster NE, Hill JC, Hay EM. Subgrouping patients with low back pain in primary care: are we getting any better at it? *Man Ther* 2010; June 30, 2010. DOI:10.1016/j.math.2010.05.013.
- van der Windt D, Hay EM, Jellema P, Main CJ. Psychosocial interventions for low back pain in primary care: lessons learned from recent trials. *Spine* 2008; **33**: 81–89.
- Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009; **338**: b606.
- Hay EM, Dunn KM, Hill JC, et al. A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. *BMC Musculoskelet Disord* 2008; **9**: 58.
- Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum* 2008; **59**: 632–41.
- Hill JC, Vohora K, Dunn KM, Main CJ, Hay EM. Comparing the STarT Back Screening Tool's subgroup allocation of individual patients with that of independent clinical experts. *Clin J Pain* 2010; **26**: 783–87.
- Hill JC, Dunn KM, Main CJ, Hay EM. Subgrouping low back pain: a comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 2010; **14**: 83–89.
- Royal College of General Practitioners and NHS Executive. Get back active—the back book video. London: Stationery Office, 2002.
- Royal College of General Practitioners and NHS Executive. The back book. London: Stationery Office, 2002.
- Roland M, Morris R. A study of the natural-history of back pain: development of a reliable and sensitive measure of disability in low-back-pain. *Spine* 1983; **8**: 141–44.
- Sullivan M, Bishop S, Pivik J. The Pain Catastrophizing Scale: development and validation. *Psychol Assess* 1995; **7**: 524–32.
- Kori SH, Miller RP, Todd DD. Kinisophobia: a new view of chronic pain behavior. *Pain Manag* 1990; 35–43.
- Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**: 361–70.
- Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. Discussion paper number 138. York: Centre for Health Economics, University of York, 1995.
- Ware J Jr, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996; **34**: 220–33.
- Little R, Rubin D. Statistical analysis with missing data. New York: John Wiley and Sons, 1981: 97–124.

- 29 Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. *J Clin Epidemiol* 2006; **59**: 45–52.
- 30 Stang A, Pooleb C, Bender R. Common problems related to the use of number needed to treat. *J Clin Epidemiol* 2010; **63**: 820–25.
- 31 Annual survey of hours and earnings (ASHE). London: Office for National Statistics, 2008. <http://www.ons.gov.uk/ons/rel/ashe/annual-survey-of-hours-and-earnings/2008-results/index.html> (accessed Sept 6, 2011).
- 32 Office for National Statistics. Standard occupational classification 2000. Volume 2. The coding index. London: Stationery Office, 2000.
- 33 Wolfe F, Smythe H, Yunus M, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. *Arthritis Rheum* 1990; **33**: 160–71.
- 34 Beurskens AJHM, de Vet HCW, Koke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996; **65**: 71–76.
- 35 Deyo RA, Mirza SK, Turner JA, Martin IB. Over treating chronic back pain: time to back off? *J Am Board Fam Med* 2009; **22**: 62–68.
- 36 Lewis M, Morley S, van der Windt DAWM, et al. Measuring practitioner/therapist effects in randomised trials of low back pain and neck pain interventions in primary care settings. *Eur J Pain* 2010; **14**: 1033–39.
- 37 Foster NE, Mullis R, Young J, et al, IMPaCT Back Study team. IMPaCT Back study protocol. Implementation of subgrouping for targeted treatment systems for low back pain patients in primary care: a prospective population-based sequential comparison. *BMC Musculoskelet Disord* 2010; **20**: 186.
- 38 Guccione A, Goldstein M, Elliott S. Clinical research agenda for physical therapy. *Phys Ther* 2000; **80**: 499–513.