Research Paper

PAIN®

OPEN

# Investigating generalizability of results from a randomized controlled trial of the management of chronic widespread pain: the MUSICIAN study

Gareth T. Jones[a,b], Elizabeth A. Jones[a,b], Marcus J. Beasley[a,b], Gary J. Macfarlane[a,b,*], On behalf of the MUSICIAN study team

**Abstract**

The generalisability of randomised controlled trials will be compromised if markers of treatment outcome also affect trial recruitment. In a large trial of chronic widespread pain, we aimed to determine the extent to which randomised participants represented eligible patients, and whether factors predicting randomisation also influenced trial outcome. Adults from 8 UK general practices were surveyed to determine eligibility for a trial of 2 interventions (exercise and cognitive behavioural therapy [CBT]). Amongst those eligible, logistic regression identified factors associated with reaching the randomisation step in the recruitment process. The main trial analysis was recomputed, weighting for the inverse of the likelihood of reaching the randomisation stage, and the numbers needed to treat were calculated for each treatment. Eight hundred eighty-four persons were identified as eligible for the trial, of whom 442 (50%) were randomised. Several factors were associated with the likelihood of reaching the randomisation stage: higher body mass index (odds ratio: 1.99; 0.85-4.61); more severe/disabling pain (1.90; 1.21-2.97); having a treatment preference (2.11; 1.48-3.00); and expressing positivity about interventions offered (exercise: 2.66; 1.95-3.62; CBT: 3.20; 2.15-4.76). Adjusting for this selection bias decreased the treatment effect associated with exercise and CBT but increased that observed for combined therapy. All were associated with changes in numbers needed to treat. This has important implications for the design and interpretation of pain trials generally.

**Keywords:** Chronic widespread pain, Fibromyalgia, RCTs, Methodology, External validity

## 1. Introduction

Randomised controlled trials (RCTs) remain the gold standard for assessing the efficacy and effectiveness of interventions. However, typically, they are conducted with highly selected patient populations and the results then generalised to wider patient populations.[9] The appropriateness of this generalisation is based, at least in part, on the extent to which the randomised patients resemble the entire eligible patient population, and the belief that the biological effect will be the same in other populations. A concern with the external validity of trials (including those concerned with pain) has led to renewed interest in "Real World Evidence" (ie, observational data) as perhaps providing more appropriate evidence on treatment effectiveness in settings in which they may be typically applied.[12]

These assumptions may not hold true. It is known that certain population groups are, generally, more willing to be randomised than others–these include the less well educated[6,11] and those with more severe symptoms[2,6]—and the generalisability of trial results may be compromised if certain patient characteristics that are associated with trial recruitment are also markers of the trial treatment outcome. However, the extent to which this is the case for individual trials is often impossible to gauge, as trial recruitment frequently occurs in such a way that detailed information on eligible but nonrandomised patients is not available.

Recent reviews and meta-analyses have shown that eligible individuals may be less likely to enter a trial if they have strong treatment preferences.[10,16] In addition, treatment preference may be associated with prognostic indicators in trial participants, such as anxiety,[15] and symptom severity.[2,10] There is also evidence that, among trial participants, treatment effect differs according to a priori treatment preferences.[10,16]

We conducted an RCT of the management of chronic widespread pain in primary care–the MUSICIAN study (Managing Unexplained Symptoms In primary Care: Involving traditional and Accessible New approaches).[13] The trial was a factorial 2 × 2 design and interventions were (1) prescribed exercise delivered by trained fitness instructors, and including access to a fitness facility; (2) cognitive behavioural therapy (CBT) delivered over the telephone by trained therapists; (3) both of the above; or (4) usual care. We found that both exercise and CBT were associated with important and statistically significant improvements in patient global assessment in both the medium and long terms, although no additional benefit was gained from receiving both treatments.[1,13] Trial patients were identified using a large

*Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.*

[a] *Epidemiology Group, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, United Kingdom,* [b] *Aberdeen Centre for Arthritis and Musculoskeletal Health, University of Aberdeen, Aberdeen, United Kingdom*

**Corresponding author. Address: Epidemiology Group, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, United Kingdom. Tel.: +44 (0) 1224 437 266. E-mail address: epidemiology@abdn.ac.uk (G.J. Macfarlane).*

*PAIN 158 (2017) 96–102*



*http://dx.doi.org/10.1097/j.pain.0000000000000732*

population-based survey. This gave rise to a unique opportunity to gather detailed information from a large pool of eligible individuals; to characterise those who did and those who did not consent to randomisation; and to determine the influence of treatment preference on the likelihood that an eligible individual would be randomised.

Thus, using data from the MUSICIAN study, the aims of the current study were, firstly, to examine factors that may affect the generalisability of trial results and secondly, to examine the extent to which external validity may be compromised, by determining whether factors predicting randomisation also influenced trial outcome.

## 2. Methods

The MUSICIAN study was a 2 × 2 factorial RCT investigating the management of chronic widespread pain (registration number: ISRCTN67013851), the methods and main results (including CONSORT statement) of which have been described elsewhere.[1,13,14] In brief, potential trial participants were identified by means of a large-scale postal questionnaire survey, mailed to all 45,994 individuals aged 25 years and older registered with 8 general practitioners in the city of Aberdeen, Scotland, and North Cheshire, England. As over 95% of UK residents are registered at a GP practice, and these practices were located in areas of varying levels of socioeconomic status, this was considered to be suitably representative of the general population. Questionnaire respondents were potentially eligible to be randomised if they reported:

(1) Pain consistent with the American College of Rheumatology definition of chronic widespread pain in their 1990 classification criteria for fibromyalgia[20];
(2) Pain of some impact, defined as a score of ≥1 on the Chronic Pain Grade[19]; and
(3) Pain for which they had consulted their general practitioner at least once, within the previous 12 months.

In addition, trial inclusion criteria required patients to consent to be contacted again, and to have:

(4) No health condition identified as requiring an alternative treatment;
(5) Access to a land-line telephone; and
(6) No contra-indications to exercise. (Note: pain alone was not considered a contra-indication.)

The questionnaire provided brief information about the exercise and CBT treatments offered in the trial (although, at this stage, participants did not know that they might be invited to take part in a trial). It also elicited information about participants' familiarity with these treatments; how positive they would be about receiving the treatments (using a 0-10 visual analogue scale); and how effective they believed they would be, were they to receive them (on a 5 point Likert scale from "much improved," to "much worse"). Treatment preference was assessed by a single question asking participants which of the available treatments they would opt for, were they to have been given a choice.

Survey respondents who were potentially eligible for the RCT were then mailed information about the trial itself, after which they were contacted by a research nurse to confirm eligibility and arrange an initial assessment appointment in a local clinical research facility. At this appointment if eligibility was confirmed and consent was obtained, randomisation took place.

The primary outcome for the trial was a 7-point, patient global impression change score, assessed by self-completion questionnaire, at 6 and 9 months post-randomisation. Patients were asked to rate how they felt their health had changed since the period before entering the trial, ranging from 1 ("very much worse") to 7 ("very much better"). Questionnaire nonrespondents were asked the same question verbally, by telephone interview.

### 2.1. Analysis

Firstly, amongst individuals surveyed, responders and nonresponders were compared and among survey respondents eligible for randomisation, differences were examined between those individuals who were/were not subsequently randomised. This was done using $\chi^2$ tests and nonparametric tests for trend[5] and the magnitude of any differences characterised using logistic regression. Thus, differences are expressed as odds ratios with 95% confidence intervals (95% Confidence interval (CI)). Secondly, a forward stepwise regression model was constructed, to identify which variables independently predicted randomisation. If not already dichotomous, these variables were then dichotomized and N × 2 categories were created, where N represented the number of factors in the multivariable regression model. The primary trial analysis (presented elsewhere[13]) was then recomputed, weighting for the inverse of the likelihood of randomisation (ie, likelihood of reaching randomisation stage), for every given combination of N × 2 categories. Finally, the number needed to treat (NNT) was calculated for each of the treatments, based on the weighted odds ratios.

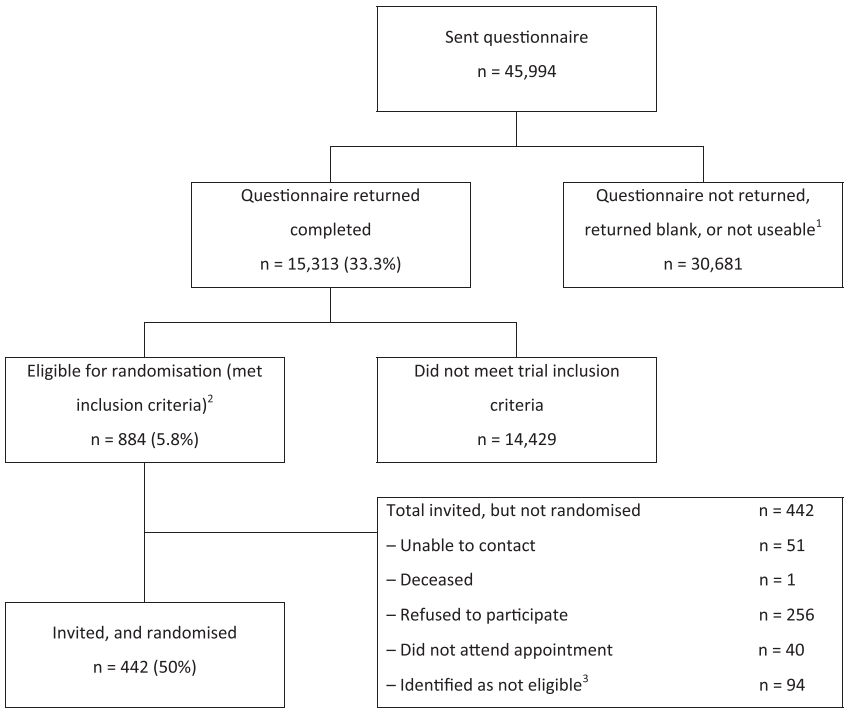
Statistical analysis was conducted using STATA 11.1 from STATACORP, Texas. Numbers needed to treat were calculated in Microsoft Excel, using published formula.[3]

## 3. Results

Of 45,994 individuals invited to participate in the survey, useable questionnaire responses were received from 15,313 (33%). Women were significantly more likely to respond than men (37% vs 29%; $\chi^2 = 328.1$, $P < 0.001$) and there was a significant increase in response rate with age (21% among those aged 25-40 years, increasing to 45% in those >60 years; nonparametric test for trend $P < 0.001$). Of the 15,313 responders, 1844 (12%) reported chronic widespread pain of whom 884 (48%) were eligible to take part in the trial and 442 (50%) were eventually randomised. Of the 442 responders not randomised, 94 were subsequently found to be ineligible, and one died before they attended the screening visit. Thus, there were 347 participants who met all trial inclusion criteria, but were not randomised. The flow of participants from initial survey invitation to subsequent randomisation is shown in **Figure 1**.

The median age of eligible participants was 57 years (interquartile range: 46-66 years) and 68% were females. Two-thirds (67%) rated their health as "good," or better; 28% had a body mass index >30 kg/m$^2$; and 51% were ex-smokers or current smokers. Of the eligible survey participants, those aged 41 to 60 years were significantly more likely to be randomised than younger respondents (odds ratio: 1.54; 95% confidence interval: 1.02-2.33). However, this effect was not linear and there was no further increase in the likelihood of randomisation among those aged >60 years (1.31; 0.87-1.98). Also, there was no difference in the likelihood of randomisation between men and women (odds ratio for women: 1.23; 0.91-1.66).

A significant trend existed, such that participants with higher BMI ($P = 0.03$) and higher Chronic Pain Grade (signifying more severe and/or disabling pain) ($P = 0.002$) were more likely to be randomised than other individuals (**Table 1**). Participants already taking some exercise (1-2 times/wk) were more likely to be




1 Includes one (eligible) person who returned a questionnaire but died before being invited.

2 Includes six people marked as not eligible, but invited due to error, one of whom was subsequently randomised.

3 Identified as ineligible either on the invitation phone call, or at the screening appointment.

**Figure 1.** Flow of participants in the study.

randomised in comparison with those not currently exercising, but those undertaking frequent exercise (>5 times/wk) were not more likely to be randomised than those not exercising. Participants with a treatment preference were twice as likely to be randomised as those without (2.11; 1.48-3.00), and this effect existed irrespective of whether the preference was for exercise, CBT, or both (**Table 2**). Positivity about receiving either exercise (2.66; 1.95-3.62) or CBT (3.20; 2.15-4.76) was associated with an increase in the likelihood of randomisation, although no such effect was observed with participant expectations of outcome, for either treatment (**Table 2**).

Five factors were found to be independently associated with randomisation (ie, reaching the randomisation step in the recruitment process): age, positivity about exercise, positivity about CBT, more severe disabling Chronic Pain Grade, and taking regular exercise. Weighting the analysis by the inverse of the likelihood of randomisation (essentially, simulating the effect of all eligible nonparticipants actually being randomised) resulted in slight difference in the treatment effect estimates at both 6 and 9 months. For the single therapies, at 6 months, the weighted model resulted in an 11% decrease in the magnitude of treatment effect for CBT (from an odds ratio of 6.45; 2.42-17.2 to 5.72; 1.92-17.0) and a 25% decrease in the treatment effect associated with exercise (from 7.28; 2.79-19.0 to 5.49; 1.89-16.0). In contrast, the weighted model gave a 16% increase in the estimate of treatment effect of the combined therapy (**Table 3**). The same pattern was true at 9 months, although the magnitude of the changes in effect estimates was less (5% decrease, 11% decrease, and 19% increase, respectively). For CBT, the weighted model produced no change in the number needed to treat. However, for exercise, there was an increase in the NNT from 4 to 5, for improvement at 6 months, and from 7 to 8 for improvement at 9 months. For the combined therapy, NNT fell from 5 to 4 for improvement at 9 months.

## 4. Discussion

In the context of a large randomised controlled trial examining the effectiveness of exercise therapy and CBT for chronic widespread musculoskeletal pain, we have shown that individuals who were randomised were different, in a number of ways, from the entire eligible patient population that was originally identified. Randomised individuals had a higher BMI, and more severe and/or disabling pain. They were also more likely to have a treatment preference, for either or both of available trial treatments, and be more positive about receiving either of the treatments available in the trial. We have demonstrated that this selection bias resulted in a change in treatment effect estimation, and in the associated NNT, although the changes noted were modest.

The design of the MUSICIAN study and, specifically, the opportunity to collect a large amount of data on individuals who were eligible to participate in the trial, but who were not ultimately randomised, allowed an assessment of potential selection bias which is rare in trials. This notwithstanding, there are a number of methodological issues to discuss, in interpreting these findings. The first issue is the timing of data collection. All predictors of randomisation were collected by population survey typically 1 to 2 weeks before randomisation. Although this has the advantage that participants completed these questions naïve to their eligibility for the trial, it may be that participants report different treatment preferences, positivity and expectations in what they believe to be a hypothetical situation, than they would if actually faced with the possibility of receiving either therapy. Secondly, only one-

**Table 1**

**Differences in demographics and health, between eligible survey participants who were/were not randomised.**

| | Randomised | | Odds ratio (95% CI) for randomisation | |
|---|---|---|---|---|
| | Yes | No | Crude | Age adjusted |
| Age, y | | | | |
| 25-40 | 61 | 64 | 1.00 | — |
| 41-60 | 197 | 134 | 1.54 (1.02-2.33) | — |
| >60 | 184 | 147 | 1.31 (0.87-1.98) | — |
| Gender | | | | |
| Male | 135 | 120 | 1.00 | 1.00 |
| Female | 307 | 225 | 1.21 (0.90-1.64) | 1.23 (0.91-1.66) |
| Self-rated health | | | | |
| Excellent | 7 | 10 | 1.00 | 1.00 |
| Very good | 76 | 63 | 1.72 (0.62-4.79) | 1.70 (0.61-5.47) |
| Good | 210 | 158 | 1.90 (0.71-5.10) | 1.84 (0.68-4.96) |
| Fair | 127 | 96 | 1.89 (0.69-5.15) | 1.82 (0.66-4.98) |
| Poor | 20 | 16 | 1.79 (0.55-5.74) | 1.71 (0.53-5.54) |
| BMI, kg/m$^2$ | | | | |
| ≤20 | 15 | 15 | 1.00 | 1.00 |
| 20.1-25.0 | 133 | 119 | 1.13 (0.53-2.40) | 1.10 (0.51-2.36) |
| 25.1-30.0 | 157 | 128 | 1.23 (0.58-2.60) | 1.16 (0.54-2.49) |
| 30.1-35.0 | 74 | 53 | 1.40 (0.63-3.10) | 1.30 (0.58-2.91) |
| >35.0 | 62 | 31 | 2.00 (0.87-4.61) | 1.99 (0.85-4.61)* |
| Smoking status | | | | |
| Never | 219 | 161 | 1.00 | 1.00 |
| Ex-smoker | 154 | 111 | 1.02 (0.74-1.40) | 1.02 (0.74-1.40) |
| Current smoker | 63 | 67 | 0.69 (0.46-1.03) | 0.70 (0.47-1.05) |
| Regular exercise† | | | | |
| None | 84 | 82 | 1.00 | 1.00 |
| 1-2 times per week | 177 | 113 | 1.53 (1.04-2.25) | 1.55 (1.05-2.28) |
| 3-4 times per week | 100 | 75 | 1.30 (0.85-1.99) | 1.33 (0.86-2.04) |
| ≥5 times per week | 79 | 72 | 1.07 (0.69-1.66) | 1.08 (0.69-1.69) |
| Chronic Pain Grade‡ | | | | |
| I | 86 | 100 | 1.00 | 1.00 |
| II | 152 | 117 | 1.51 (1.04-2.20) | 1.53 (1.05-2.23) |
| III | 85 | 53 | 1.86 (1.19-2.92) | 1.90 (1.21-2.99) |
| IV | 86 | 53 | 1.89 (1.21-2.95) | 1.90 (1.21-2.97)§ |

* Test for trend across categories $P = 0.03$.

† Number of times per week doing 30minutes of moderate physical activity or walking that increased the heart rate or increased breathing.

‡ Due to trial eligibility criteria, there were no participants with a Chronic Pain Grade = 0.

§ Test for trend across categories $P = 0.002$.

BMI, body mass index; CI, confidence interval.

third of the survey questionnaires were returned. Population survey questionnaire response rates are falling over time[8] and participation rates of 33% are not uncommon. The current study aimed to determine whether trial participants were different from eligible but nonrandomised participants. By definition, individuals who failed to complete the initial survey questionnaire were not eligible for the trial. This study looked at how refusal to participate after the identification of eligible patients affected representativeness; a separate source of selection bias (not under examination in the current study) comes from not being able to identify eligible patients in the first place. Although the prevalence of chronic widespread pain in the current study was very similar to other large population studies,[14] we know that responders and nonresponders differ with respect to age and gender. The differences were 24% and 8.0%, respectively, with older individuals and women significantly more likely to respond than other individuals, and among all respondents, these individuals were also significantly more likely to be randomised. This illustrates further that trial participants are different from the wider eligible patient population.

Our findings concur with other studies which have shown that trial participants differ from the wider eligible population in a number of ways and that participants with severe or disabling pain were more likely to be randomised is perhaps no surprise. These individuals may be more willing than other participants to try novel or hard-to-access treatments. It is also plausible that those with a higher BMI may have been more willing to enter the trial, to benefit (potentially) from the exercise therapy. What is particularly pertinent, however, is not why randomised and nonrandomised participants are different, but the fact that they are different with respect to a number of important prognostic markers. Increasing the likelihood that persons agree to take part in trials for which they are eligible is key to reducing this selection bias. A systematic review of factors which could potentially increase the chance of an approached person agreeing to take part in a trial for which they are eligible showed the following to be effective: strategies to increase awareness of the health problem being studied (including an interactive computer programme, education session, or video about the health problem being studied). In contrast increasing patients'

**Table 2**

**Differences in treatment preference and expectation, between eligible survey participants who were/were not randomised.**

| | Randomised | | Odds ratio (95% CI) for randomisation | |
|---|---|---|---|---|
| | Yes | No | Crude | Age adjusted |
| Treatment preference | | | | |
| No* | 68 | 95 | 1.00 | 1.00 |
| Yes | 362 | 245 | 2.06 (1.45-2.93) | 2.11 (1.48-3.00) |
| Treatment preference | | | | |
| None* | 68 | 95 | 1.00 | 1.00 |
| Exercise | 170 | 151 | 1.57 (1.07-2.30) | 1.60 (1.09-2.34) |
| CBT | 27 | 16 | 2.36 (1.18-4.71) | 2.38 (1.18-4.76) |
| Both treatments | 165 | 78 | 2.96 (1.96-4.46) | 2.10 (2.04-4.70) |
| Expectations of exercise† | | | | |
| Improve | 347 | 236 | 1.00 | 1.00 |
| No change | 58 | 64 | 1.62 (1.10-2.40) | 1.67 (1.12-2.48) |
| Worsen | 21 | 17 | 1.36 (0.66-2.83) | 1.34 (0.64-2.79) |
| Expectations of CBT† | | | | |
| Improve | 228 | 129 | 1.00 | 1.00 |
| No change | 175 | 168 | 1.70 (1.25-2.30) | 1.74 (1.28-2.37) |
| Worsen | 3 | 5 | 0.58 (0.14-2.45) | 0.55 (0.13-2.33) |
| Positivity re: exercise‡§ | | | | |
| Low | 113 | 159 | 1.00 | 1.00 |
| Moderate/high | 325 | 182 | 2.51 (1.86-3.40) | 2.66 (1.95-3.62) |
| Positivity re: CBT‡ | | | | |
| Low | 125 | 154 | 1.00 | 1.00 |
| Moderate | 165 | 116 | 1.75 (1.25-2.45) | 1.85 (1.31-2.60) |
| High | 141 | 60 | 2.90 (1.97-4.25) | 3.20 (2.15-4.76) |

* Includes participants with no preference, and those who responded "don't know."
† The imagined effect of 6 months of treatment, on participants' chronic pain.
‡ How positive participants would be about receiving the treatment, one a 0 to 10 scale; divided into tertiles for analysis.
§ Due to the skewed distribution of positivity regarding exercise, the moderate and high tertiles form one category.
CBT, cognitive behavioural therapy; CI, confidence interval.

understanding of the trial process, recruiter differences, and various methods of randomisation and consent design were not associated with improved recruitment.[4]

Our findings also show that eligible individuals with a preference for one or both of the investigative treatments in the MUSICIAN trial were more likely to be randomised than those with no preference. This is likely to be at least partially explained by the nature of the interventions offered in the MUSICIAN trial. In the UK, neither prescribed exercise (including free gym membership for 6 months, and complimentary access to a fitness instructor) nor CBT are routinely available for chronic widespread pain in primary care. Previous trials have reported that a strong treatment preference was a key reason for refusing randomisation[7,10,17,18] and this also has important implications for the generalisability of findings. A recent meta-analysis of 11 musculoskeletal trials found that, among participants, treatment preference was an important determinant of outcome.[16]

We have also shown that the factors that influence whether a potential participant is likely to be randomised into a trial also influence trial outcome. Re-computing the main trial analysis, to adjust for the fact that the randomised participants are different from the total eligible patient population, gave intriguing findings. For the single therapies, our weighted model resulted in a decrease in treatment effect, suggesting that any selection bias (in the original analysis) acted to overestimate treatment effects. Whereas, for combined therapy, the opposite was true, suggesting that any selection bias led to an underestimate of the effect of treatment. In the context of the current trial, where the treatment effect sizes were large ($OR_{range}$: 6.45-7.28 at 6 months, and 3.41-5.57 at 9 months)

**Table 3**

**The influence of factors associated with randomisation, on trial outcome.**

| Treatment group | Improvement* at 6 mo post randomisation, odds ratio (95% CI) [NNT] | | Improvement* at 9 mo post randomisation, odds ratio (95%CI) [NNT] | |
|---|---|---|---|---|
| | Original findings† | Weighted model | Original findings† | Weighted model |
| Treatment as usual | 1.00 | 1.00 | 1.00 | 1.00 |
| CBT | 6.45 (2.42-17.2) [NNT = 4] | 5.72 (1.92-17.0) [NNT = 4] | 5.57 (2.34-13.3) [NNT = 5] | 5.31 (2.06-13.7) [NNT = 5] |
| Exercise | 7.28 (2.79-19.0) [NNT = 4] | 5.49 (1.89-16.0) [NNT = 5] | 3.41 (1.42-8.15) [NNT = 7] | 3.02 (1.18-7.76) [NNT = 8] |
| CBT + Exercise | 6.76 (2.56-17.8) [NNT = 4] | 7.86 (2.69-23.0) [NNT = 4] | 5.18 (2.19-12.3) [NNT = 5] | 6.19 (2.41-15.9) [NNT = 4] |

* "Much better" or "very much better" on patient global change score on how patients felt their health had changed since entering the trial, from 1 ("very much worse") to 7 ("very much better").
† Effect estimates and NNTs differ slightly from those in the original manuscript 7 because we have excluded individuals with missing values for variables used in the weighting calculation.
NNT, numbers needed to treat; CBT, cognitive behavioural therapy; CI, confidence interval.

an over- or under-estimate of the magnitude observed in the current study makes little difference to the overall conclusions of the trial. However, many trials have smaller effect sizes and, while it is impossible to predict what the results would be, over- or under-estimates of between 10% and 24% may have important implications in interpretation of trial findings. As in the current study, even minor changes in effect size, may result in changes in NNT, and this may have potentially important implications for estimates of the cost-effectiveness of treatments. In the original MUSICIAN trial for the primary outcome[14] exercise was not cost effective, and the cost effectiveness of CBT was marginal. In this context, even minor errors in estimation of effect measures are important.

In summary, the status of randomised controlled trials as the gold standard method for determining the effectiveness of healthcare interventions is based on their inherent internal validity and the ability to control potential confounding variables, but they are commonly conducted on highly selected patient groups. Their real world value, therefore, depends on the assumption that these patient groups adequately represent the entire eligible patient population, yet rarely is information available to test this assumption. Capitalising on a unique opportunity to collect data on a wider eligible population we have shown, firstly, that trial participants differ not only in terms of clinical variables, but also in terms of treatment preference; and, secondly, that the factors associated with trial participation also influence trial outcome. This has important implications for trials generally and emphasises that, where possible, collecting information on eligible but nonrandomised patients allows a better estimate of treatment effectiveness.

## Conflict of interest statement

The authors have no conflicts of interest to declare.

## Acknowledgements

The MUSICIAN trial was supported by an award from Arthritis Research UK, Chesterfield, UK. Grant number: 17292. Ethical approval for the study was granted by Cheshire NHS Research Ethics Committee; reference number: 07/Q1506/61. All participants provided written consent.

The MUSICIAN study team is Professor G. J. Macfarlane (chief investigator), Professor Phil Hannaford, Dr Philip Keeley, Professor Karina Lovell, Dr John McBeth, Professor Paul McNamee, Professor Deborah Symmons and Dr Steve Woby (investigators), M. J. Beasley (research assistant), Chrysa Gkazinou (trial manager), Dr E. A. Jones (PhD student), Dr Gordon Prescott (statistician) and Dr Graham Scotland (health economist).

Additional contributions: we are grateful to the following GP practices and their patients for participating in the study in Aberdeen: Carden Medical Centre, Elmbank Medical Practice, Great Western Medical Practice, Garthdee Medical Group; and in Macclesfield: Readesmoor Medical Group Practice, Lawton House Surgery, Bollington Medical Centre, and Park Lane Surgery. The Scottish Primary Care Research Network facilitated access to patient information at the practices in Aberdeen city. At the University of Aberdeen John Norrie (while director of CHaRT) and Ashraf El-Metwally, PhD (while Lecturer in Epidemiology) were originally investigators on the MUSICIAN study. Alison MacDonald and Gladys McPherson of the Health Services Research Unit (HSRU) at the University of Aberdeen provided input regarding the conduct of the study. Dev Acharya, Jennifer Banister, Gertrude Chikwekwe, Rowan Jasper, Flora Joyce, Karen Kane and Michelle Rein, were project assistants on the study. Alison Littlewood performed the study management at the Manchester site and Charlie Stockton was the study manager during the setting up and part of study conduct. Research nurses Daniel Barlow, Roslyn Campbell and Vivien Vaughan conducted the pre-randomisation clinic assessments. Wesley Bramley, Julie Carney and Richard Paxton delivered the exercise intervention, and Jayne Fox, Nicola McConnell, Marie Pope and Lindsay Rigby delivered the CBT. We are grateful to Professor Shaun Treweek (University of Aberdeen) for providing comments on the manuscript.

In addition, we are grateful to the independent members of the Trial Steering Committee: Professor Matthew Hotopf, Professor Tracey Howe, and Professor Martin Underwood; and the Data Monitoring Committee: Dr Marwan Bukhari, Professor Hazel Inskip and Dr Chris Edwards.

Author contributions: G. T. Jones, E. A. Jones, M. J. Beasley, and G. J. Macfarlane were involved in the conception and design, analysis and interpretation of data. G. T. Jones drafted the article, G. J. Macfarlane revised it and prepared it for submission, M. J. Beasley and E. A. Jones critically reviewed the manuscript.

**Article history:**
Received 21 April 2016
Received in revised form 31 August 2016
Accepted 21 September 2016
Available online 29 September 2016

## References

[1] Beasley M, Prescott GJ, Scotland G, McBeth J, Lovell K, Keeley P, Hannaford PC, Symmons DP, MacDonald RI, Woby S, Macfarlane GJ. Patient-reported improvements in health are maintained 2 years after completing a short course of cognitive behaviour therapy, exercise or both treatments for chronic widespread pain: long-term results from the MUSICIAN randomised controlled trial. RMD Open 2015;1:e000026.

[2] Bedi N, Chilvers C, Churchill R, Dewey M, Duggan C, Fielding K, Gretton V, Miller P, Harrison G, Lee A, Williams I. Assessing effectiveness of treatment of depression in primary care. Partially randomised preference trial. Br J Psychiatry 2000;177:312–18.

[3] Bender R. Number needed to treat. In: Armitage P, Colton T, editors. Encyclopedia of Biostatistics. 2nd ed. Chichester: John Wiley & Sons Ltd, 2005. p. 3752–61.

[4] Caldwell PH, Hamilton S, Tan A, Craig JC. Strategies for increasing recruitment to randomised controlled trials: systematic review. PLoS Med 2010;7:e1000368.

[5] Cuzick JA. Wilcoxon-type test for trend. Stat Med 1985;4:87–90.

[6] Detre KM, Guo P, Holubkov R, Califf RM, Sopko G, Bach R, Brooks MM, Bourassa MG, Shemin RJ, Rosen AD, Krone RJ, Frye RL, Feit F. Coronary revascularization in diabetic patients: a comparison of the randomised and observational components of the Aypass Angioplasty Revascularization Investigation (BARI). Circulation 1999;99:633–40.

[7] Foster NE, Thomas E, Hill JC, Hay EM. The relationship between patient and practitioner expectations and preferences and clinical outcomes in a trial of exercise and acupuncture for knee osteoarthritis. Eur J Pain 2010;14:402–9.

[8] Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007;17:643–53.

[9] Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials 2015;3:495.

[10] King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, Sibbald B, Lai R. Impact of participant and physician intervention preferences on randomised trials: a systematic review. JAMA 2005; 293:1089–99.

[11] King SB III, Barnhart HX, Kosinski AS, Weintraub WS, Lembo NJ, Petersen JY, Douglas JS Jr, Jones EL, Craver JM, Guyton RA, Morris DC, Liberman HA. Angioplasty or surgery for multivessel coronary artery disease: comparison of eligible registry and randomised patients in the EAST trial and influence of treatment selection on outcomes. Emory

Angioplasty versus Surgery Trial Investigators. Am J Cardiol 1997;79: 1453–9.
[12] Li G, Sajobi TT, Menon BK, Korngut L, Lowerison M, James M, Wilton SB, Williamson T, Gill S, Drogos LL, Smith EE, Vohra S, Hill MD, Thabane L; Registry-Based Randomized Controlled Trials in Calgary. Registry-based randomized controlled trials: advantages, challenges and areas for future research. J Clin Epidemiol 2016. pii: S0895–4356(16) 30350-X.
[13] McBeth J, Prescott G, Scotland G, Lovell K, Keeley P, Hannaford P, McNamee P, Symmons DP, Woby S, Gkazinou C, Beasley M, Macfarlane GJ. Cognitive behavior therapy, exercise, or both for treating chronic widespread pain. Arch Intern Med 2012;172:48–57.
[14] Macfarlane GJ, Beasley M, Jones EA, Prescott GJ, Docking R, Keeley P, McBeth J, Jones GT; MUSICIAN Study Team. The prevalence and management of low back pain across adulthood: results from a population-based cross-sectional study (the MUSICIAN study). PAIN 2012;153:27–32.
[15] Mills N, Metcalfe C, Ronsmans C, Davis M, Lane JA, Sterne JA, Peters TJ, Hamdy FC, Neal DE, Donovan JL. A comparison of socio-demographic and psychological factors between patients consenting to randomisation and those selecting treatment (the ProtecT study). Contemp Clin Trials 2006;27:413–19.
[16] Preference Collaborative Review Group. Patients' preferences within randomised trials systematic review and patient level meta-analysis. Br Med J 2008;337:a1864.
[17] Raue PJ, Schulberg HC, Heo M, Klimstra S, Bruce ML. Patients' depression treatment preferences and initiation, adherence, and outcome: a randomised primary care study. Psychiatr Serv 2009;60: 337–43.
[18] Tincello DG, Kenyon S, Slack M, Toozs-Hobson P, Mayne C, Jones D, Taylor D. Colposuspension or TVT with anterior repair for urinary incontinence and prolapse: results of and lessons from a pilot randomised patient-preference study (CARPET 1). BJOG 2009;116:1809–14.
[19] Von Korrf M, Dworkin S, LeResche L. Graded chronic pain status: an epidemiologic evaluation. PAIN 1990;40:279–29121.
[20] Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, Tugwell P, Campbell SM, Abeles M, Clark P, Fam AG, Farber SJ, Fiechtner JJ, Franklin CM, Gatter RA, Hamaty D, Lessard J, Lichtbroun AS, Masi AT, Mccain GA, Reynolds WJ, Romano TJ, Russel IJ, Sheon RP. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum 1990;33: 160–72.