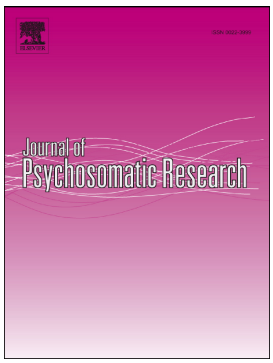


Guided graded exercise self-help for chronic fatigue syndrome: Long term follow up and cost-effectiveness following the GETSET trial

Lucy V. Clark, Paul McCrone, Francesca Pesola, Mario Vergara-Williamson, Peter D. White

PII: S0022-3999(21)00129-X

DOI: https://doi.org/10.1016/j.jpsychores.2021.110484

Reference: PSR 110484

To appear in: *Journal of Psychosomatic Research*

Received date: 23 June 2020

Revised date: 30 March 2021

Accepted date: 31 March 2021

Please cite this article as: L.V. Clark, P. McCrone, F. Pesola, et al., Guided graded exercise self-help for chronic fatigue syndrome: Long term follow up and cost-effectiveness following the GETSET trial, *Journal of Psychosomatic Research* (2021), https://doi.org/10.1016/j.jpsychores.2021.110484

Guided graded exercise self-help for chronic fatigue syndrome: long term follow up and cost-effectiveness following the GETSET trial

Lucy V Clark,[a] Paul McCrone[b], Francesca Pesola[c], Mario Vergara-Williamson[d], Peter D White[a]

[a] Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Bart's and the London School of Medicine and Dentistry, Queen Mary University, London, UK.

[b] King's Health Economics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. E-mail: p.mccrone@greenwich.ac.uk

[c] Cancer Prevention Group, School of Cancer & Pharmaceutical Sciences, Faculty of Medicine and Life Sciences, King's College London, London, UK. E-mail: francesca.pesola@kcl.ac.uk

[d] Kent & Medway NHS and Social Care Partnership Trust, Kent & Medway CFS/ME Service, Maidstone, Kent, UK. E-mail: m.vergara-williamson@nhs.net

Corresponding author at: Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Bart's and the London School of Medicine and Dentistry, Queen Mary University, London, EC1A 7BE, UK. E-mail address: l.clark@qmul.ac.uk

Abstract

Objective: The GETSET trial found that guided graded exercise self-help (GES) improved fatigue and physical functioning more than specialist medical care (SMC) alone in adults with chronic fatigue syndrome (CFS) 12-weeks after randomisation. In this paper, we assess the longer-term clinical and health economic outcomes.

Methods: GETSET was a randomised controlled trial of 211 UK secondary care patients with CFS. Primary outcomes were the Chalder fatigue questionnaire and the physical functioning subscale of the short-form-36 survey. Postal questionnaires assessed the primary outcomes and cost-effectiveness of the intervention 12-months after randomisation. Service costs and quality-adjusted life years (QALYs) were combined in a cost-effectiveness analysis.

Results: Between January 2014 and March 2016, 164 (78%) participants returned questionnaires 15-months after randomisation. Results showed no main effect of intervention arm on fatigue ($chi^2(1)=4.8$, p=0.03) or physical functioning ($chi^2(1)=1.3$, p=0.25), adjusting for multiplicity. No other intervention arm or time*arm effect was significant. The short-term fatigue reduction was maintained at long-term follow-up for participants assigned to GES, with improved fatigue from short- to long-term follow up after SMC, such that the groups no

longer differed. Healthcare costs were £85 higher for GES and produced more QALYs. The incremental cost-effectiveness ratio was £4,802 per QALY.

Conclusions: The short-term improvements after GES were maintained at long-term follow-up, with further improvement in the SMC group such that the groups no longer differed at long-term follow-up. The cost per QALY for GES compared to SMC alone was below the usual threshold indicating cost-effectiveness, but with uncertainty around the result.

Keywords: Chronic fatigue syndrome, myalgic encephalomyelitis, graded exercise therapy, guided self-help, cost-effectiveness, follow up

Abbreviations

CFQ, Chalder Fatigue questionnaire; CFS, chronic fatigue syndrome; GES, guided graded exercise self-help; GETSET, guided graded exercise therapy self-help trial; SF-36 PF, short form-36 physical function subscale; SMC, standard medical care

1. Introduction

Chronic fatigue syndrome (CFS), also known as myalgic encephalomyelitis (ME), is a relatively common illness characterised by chronic and disabling fatigue, plus a range of additional symptoms, which cause severe impairment in daily functioning [1]. The prevalence of CFS in the population is estimated to be between 0.2% and 2.6% [1], but prevalence rates vary depending on the definition used. If untreated, the prognosis of CFS is poor, and whilst duration and severity may impact on improvement and recovery, best estimates from a systematic review suggest that 5% of those diagnosed reported recovery and 40% reported symptom improvement [2].

When we planned the graded exercise therapy guided self-help trial (GETSET), a large trial had suggested that graded exercise therapy (GET), delivered by specialist therapists over 15 sessions, could improve CFS patient outcomes [3]. The GETSET trial was designed to compare the efficacy and safety of GET, delivered as a guided graded exercise self-help (GES) intervention added to specialist medical care (SMC), with SMC alone, for adult patients with chronic fatigue syndrome (CFS) [4-5]. The trial found that 12 weeks after randomisation, patients that had been assigned to GES had significantly greater improvements in their fatigue, and to a lesser extent physical functioning than those that had been assigned to SMC alone [4].

Long-term studies of participants in trials of therapist-delivered graded exercise therapy for CFS suggest that improvements in fatigue and functioning are maintained in the long-term [6-7], although a systematic review of exercise interventions (of any sort, including Qigong) for CFS reported low certainty of evidence for longer-term benefit [8]. Health economic analysis has suggested that therapist-delivered GET is cost-effective [9]. In this paper we report the findings following a longer-term follow-up of GETSET trial participants [3]. We aim to compare primary outcomes, the main

secondary outcome, and health economic outcomes of participants both within and between the two randomised intervention groups, one year after randomisation.

2. Methods

2.1. The GETSET Trial

The GETSET trial protocol and main results have been published previously, and describe the trial methods, including details of the interventions (guided graded exercise self-help (GES) added to specialist medical care (SMC) and SMC alone) [3-4]. The GETSET trial was a pragmatic two-arm, randomised controlled trial undertaken at two UK National Health Service (NHS) secondary care clinics for CFS in London and Kent. The trial tested a guided graded exercise self-help intervention (GES) for adult patients meeting NICE criteria for chronic fatigue syndrome (CFS) [10]. We collected, and therefore report, information on Oxford and Centers for Disease Control and Prevention (CDC) criteria [11-12] for reasons of generalisation. Between May 15, 2012 and Dec 24, 2014 a total of 211 participants were recruited and randomly allocated to unstandardised SMC alone or SMC plus GES. Participants in the SMC group remained on a waiting list for therapy whilst GES participants were given a self-help booklet [13] describing a manualised [14] six step programme of graded exercise self-management, and offered up to four sessions of Skype™ or telephone guidance (the first could be face-to face), totalling 90 minutes over 8 weeks, from a physiotherapist. The primary outcome measures of the trial were fatigue (CFQ) and physical functioning (SF-36 PF), measured 12 weeks after randomisation using self-report scales described below [15-16].

As is detailed in our main trial paper [3] the original protocol had only one primary outcome measure, the SF-36 PF. However, when a significant minority of participants scored close to the mean of the general population (i.e., normal physical function) at randomisation despite reductions in functioning in other domains, such as mental or social activity levels, we decided to also include fatigue, using the CFQ, as a co-primary outcome. This decision was approved by the Research Ethics Committee, the Data Monitoring and Ethics Committee, and the Trial Steering Committee and was made mid-way through trial recruitment (on June 20, 2013, after recruitment of 99 [47%] patients), before any outcome data had been examined.

Ethical approval for the trial and follow up study was given by the UK National Research Ethics Service Committee London – London Bridge (reference 11/LO/1572). Written consent to be contacted for the follow up was obtained in the consent form signed by participants at the time of the original trial recruitment.

2.2. Long-term follow-up

Postal questionnaires were sent to participants at least one year after the date they were randomised, with one further postal questionnaire being sent to non-responders. Those who still did not respond

were contacted by telephone or e-mail as described in the protocol for the main trial [4]. The follow-up questionnaires included: the primary outcomes of both fatigue severity using the Chalder fatigue questionnaire (CFQ) [15] and physical functioning using the SF-36 physical functioning subscale (SF-36 PF) [16]; the main secondary outcomes of perceived change in overall health (CGI-health) and CFS (CGI-CFS), using the Clinical Global Impression (CGI) change score [17], since this also provides a measure of deterioration; the cost-effectiveness outcomes of perceived quality of life using the EQ-5D-3L (EQ-5D), which includes five domains (mobility, self-care, usual activities, pain and discomfort, and anxiety and depression), each of which is scored as 1, 2 or 3 depending on the severity of problems [18], and service use and informal care using an adapted version of the Client Services Receipt Inventory (CSRI) [19].

The questionnaire items and administration were identical to those used in the main part of the trial, except that during the trial some participant questionnaires were handed-in by participants at their assessment visit and for this follow-up they were all self-rated at home and returned by post. We chose to economically evaluate the intervention at one year, rather than at 12 weeks, as it was felt that 12 weeks was not a long enough period to establish cost-effectiveness.

2.3 Statistical methods

Primary outcomes

As was pre-specified in the protocol [4], missing information at the 12 week and one year follow-ups was pro-rated for patients who had less than 20% missing information on a scale. Descriptive statistics were used to summarise primary outcomes, using means and standard deviations (SDs), medians and interquartile ranges (IQRs) or frequencies and proportions, as appropriate. We constructed profile plots for CFQ and SF-36 PF to represent scores at baseline, 12 weeks and one year broken down by study arm.

Differences between study arms on CFQ and SF-36 PF were explored using mixed-effects regression models with robust standard errors. We used patients as the cluster, and random slopes and intercepts to account for repeated measures at 12 weeks and one year. As in the main trial analysis [3], we used a Bonferroni adjustment to account for having two primary outcomes (significance at $p< 0.025$).

We regressed each outcome on the intervention arm (GES versus SMC), time (12 weeks versus one year) and an interaction term between arm and time (arm*time) and these effects were tested using ANOVA-style tests. As in the main analysis of the trial [3] we adjusted the model for stratifiers and baseline scores as failing to do so has been found to lead to biased results [20].

Moreover, the analyses were adjusted for covariates associated with the outcome and missing data, based on evidence that shows that failing to do so may lead to biased results in complete case analysis [21, 22]. Relevant variables were identified by exploring the association between outcomes and baseline characteristics as well the association between missing status (yes/no) and baseline

characteristics, using t-test, chi-square test and correlations. The following variables were therefore included in our models: education level, days from randomisation (log transformed), International Physical Activity Questionnaire (IPAQ) baseline scores [23], and the Oxford criteria for CFS [11].

To assess differences between arms on CGI scores on CFS and overall health, we compared the rate of participants reporting an improvement between 12 weeks and long-term follow-up (CGI scores of 1 or 2 corresponding to "very much" or "much" better) compared to those who reported either no change or deterioration (scores of 3-7; i.e. positive change) at one year. We also compared the frequency of participants reporting a deterioration between 12 weeks and long-term follow-up (CGI scores of 6 or 7 corresponding to "much worse" or "very much worse") compared to those who reported either no change or an improvement (scores of 1 through 5; i.e. negative change). To do so, we used a binary logistic regression with positive/negative change regressed onto intervention arm and adjusted for change score at 12 week follow-up.

All analyses were adjusted using the study stratification variables (high depression score, high SF-36 PF score, and recruitment centre) [3] and covariates associated with missing outcome information as described above. In the analyses the SMC-alone arm was the reference category so that results are for the GES arm in relation to the SMC arm. For the primary outcomes we also undertook sensitivity analyses in which we adjusted only for a) stratifiers and b) stratifiers and baseline scores. In our sensitivity analyses, we adjusted the analysis to account for the number of service therapy sessions attended after randomisation. This variable was positively skewed and could not be successfully transformed; thus, we categorised it using quartiles. Analyses were conducted using Stata 15.

Cost-effectiveness analyses

The cost-effectiveness analyses took an NHS/social care perspective and combined costs with quality-adjusted life years (QALYs). In secondary analyses, the costs of unpaid care from family/friends was also considered. The time horizon for the analyses was the one year follow-up period. Using the five domains of the EQ-5D, 243 different health states were used to derive a utility weight, anchored by 1 representing full health and 0 representing death [24]. These utility scores were used to generate QALYs using area under the curve methods [25]. QALYs were compared between the two groups using a regression model with baseline utility controlled for.

The cost of the intervention was based on the GES physiotherapist time spent in contact with the participant. Given the training and supervision required, we based the unit cost of these clinicians on the cost used in the PACE trial [5] which, inflated to 2015/16 prices, was £113 per hour. Other service use was measured with the Client Service Receipt Inventory (CSRI) [19]. This included primary care, secondary care and social care services used in the 12 weeks prior to each of the data collection points: baseline, 12 week and one year follow-up. Participants were asked whether they had used a specific service and if so how often and (where appropriate) for how long. These data were combined

with appropriate unit cost information from the University of Kent and Department of Health (NHS reference cost PSSRU) [26]. Informal care costs were calculated by combining data on the number of care hours received per week with average hourly wages, multiplied by 12. Given that costs were not directly measured for the period between 12 week follow-up and 12 weeks before the one year follow-up, we extrapolated for this period on the basis of costs data that were available, excluding intervention costs which were assumed to be confined to the first 12 week trial period. Costs were compared between the two groups with baseline costs controlled for. A boot-strapped regression model was used due to the likely skewed cost distribution and 95% confidence intervals generated using the percentile method.

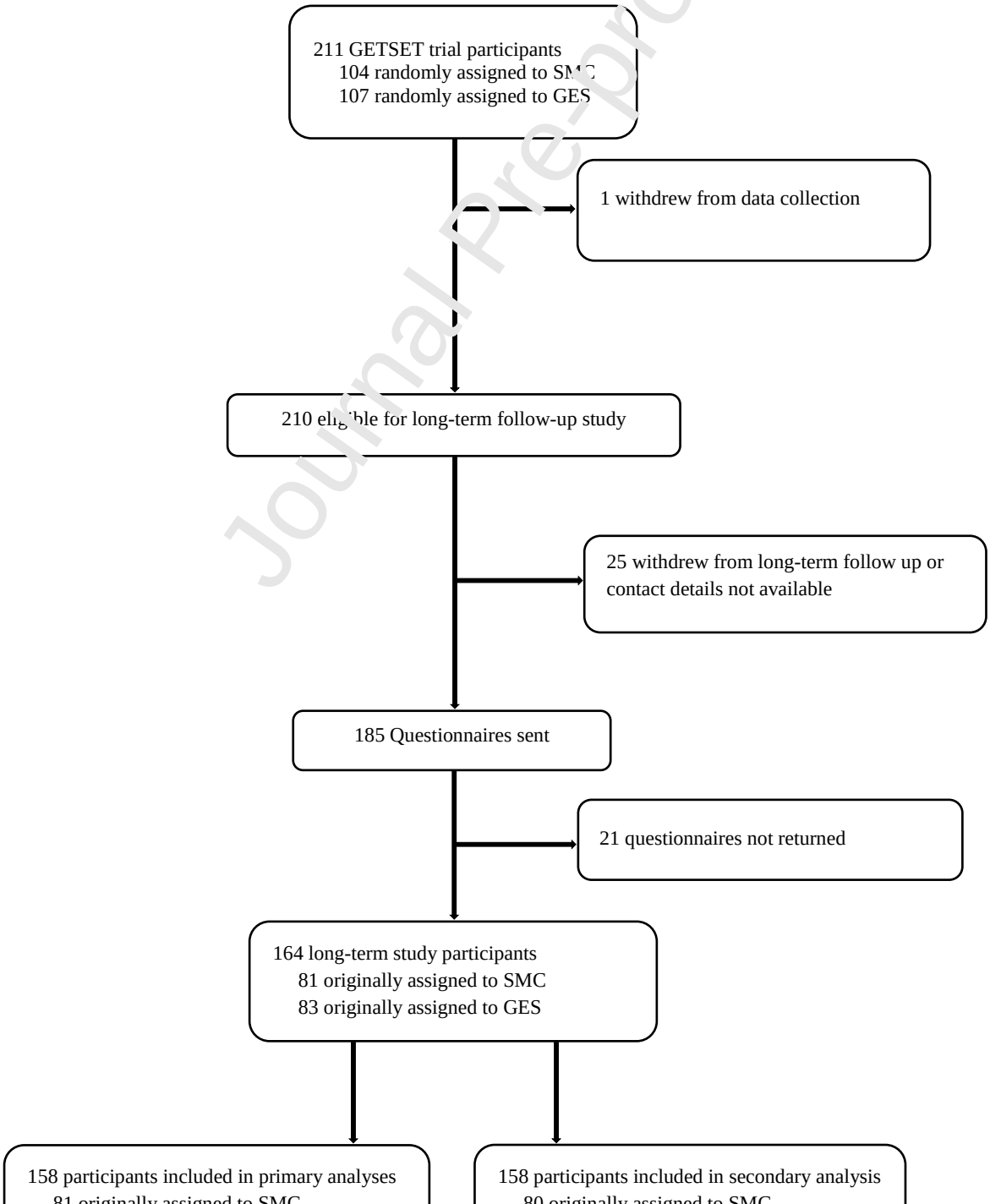


Figure 1: Long-term follow up study profile

The QALY and cost differences derived from the regression models were entered into the cost-effectiveness analyses. If GES resulted in lower (higher) costs and more (fewer) QALYs than usual care, then it would be 'dominant' ('dominated'). If GES had higher (lower) costs and resulted in more (fewer) QALYs then it would be a value judgement as to whether the extra (reduced) costs are justified in order to obtain more (fewer) QALYs. To inform this judgement an incremental cost-effectiveness ratio was used; this was the incremental cost divided by the incremental QALY gain. In England, NICE favours interventions with an incremental cost per QALY of below £20,000. To address uncertainty in the results we plotted 1,000 cost and QALY differences obtained from boot-strapped resamples on a cost-effectiveness plane. This allowed us to estimate the probability that GES results in higher costs and more QALYs, lower costs and more QALYs, lower costs and fewer QALYs, or higher costs and fewer QALYs than usual care. The probability that GES is more cost-effective than SMC alone was assessed using a cost-effectiveness acceptability curve where the extra QALYs gained were multiplied by different threshold values for a QALY (a range of £0 to £40,000 in steps of £5000), with costs then subtracted. This was done for the 1,000 boot-strapped sample and the proportion of these that were above zero gave us the required probability for that threshold value.

As with many studies, the sample size was determined by power calculations to show significant differences in clinical outcome measures rather than differences in cost-effectiveness.

3. Results

We obtained consent from all 211 trial participants to contact them for a long-term follow-up assessment; this was obtained at the time of randomisation into the GETSET trial, (Figure 1). We were unable to confirm current contact details for 25 (12%) trial participants, and one (<1%) participant subsequently withdrew their consent for further data collection. We therefore sent questionnaires to 185 of the original trial participants. Between January 27, 2014 and March 10, 2016, we received 164 usable questionnaires which is 78% of the full cohort and 89% of participants who

were sent questionnaires. Analyses were restricted to participants with complete information on baseline and both the short- and long-term follow-up for the primary (n = 158; GES = 77 vs. SMC = 81) and secondary (n = 158; GES = 78 vs. SMC = 80) outcomes. A similar proportion of participants returned questionnaires in each of the randomised groups (p = 1.0) and in each recruitment centre (p = 0.74). The median time from randomisation to long-term follow-up was 15 months (IQR 13-19) with a range of 11 to 36 months.

The baseline characteristics were similar between participants who did and did not take part in this follow-up (Table 1). Within the long-term follow-up study sample, baseline characteristics of the randomised intervention groups were similar, apart from for the number of minutes of physical activity undertaken (Appendix; Table A), which was significantly higher in those allocated to SMC, a finding that was already present between groups at baseline [3].

Most (91%; 149 of 164) of the long-term follow-up study participants reported receiving a service therapy in their respective centre after their final 12 week outcome assessment. The number of participants who received at least one session of therapy was similar between the original randomised

Table 1: Baseline characteristics of GETSET trial participants who did and did not participate in the follow-up

	Participated (n = 164)	**Did not participate (n = 47)**	***p* value***
Age (years) median (IQR)	38 (30-48)	35 (26-45)	0.12
Female	127 (77%)	40 (85%)	0.25
White	147 (90%)	41 (87%)	0.64
CDC criteria	120 (73%)	30 (64%)	0.21
Oxford Criteria	137 (84%)	33 (70%)	0.04
Duration of illness	42.0 (24.5 to 95.4)	54.0 (23.4 to 148.6)	0.33
Physical activity (min per week)**	152.5 (60.0 to 540.0)	155 (25.0 to 235.0)	0.11

Data are mean (SD), n(%), or median (IQR). CDC = US Centers for Disease Control and Prevention
* Chi-Squared test (categorical variables), independent *t* test (continuous variables, or Mann-Whitney *U* test; p-value for difference between groups.
** Data for this physical activity (min per week) were available for 206 participants (participated = 162 and did not participate = 46).

groups (GES = 75 [90%] of 83; SMC = 74 [91%] of 81; p = 0.82). A minority of participants in each of the original randomised groups received what has previously been defined as 'adequate' therapy (10 sessions or more) [27]. There was no significant difference between the groups in the proportions receiving at least this number of therapy sessions (GES = 19 [23%] of 83; SMC = 27 [33%] of 81; p = 0.14).

A number of variables were found to be associated with the outcomes and their missingness: education level, Oxford CFS classification, days to randomisation and activity level measured by the IPAQ. The latter also captures differences in activity measured in minutes observed between participants in the two study arms at baseline (median (IQR): GES = 150 (60-420) vs. SMC = 202.5 (65- 690.5). Indeed, IPAQ activity levels and physical activity (minutes) were highly correlated (polychoric rho = 0.88, se = 0.02).

Results showed no main effect (allowing for Bonferroni correction) of the intervention arm on fatigue ($chi^2(1) = 4.8$, p = 0.03; marginal mean differences: 2.2, 95%CIs: 0.2- 4.1) or functioning ($chi^2(1) = 1.3$, p = 0.25; marginal mean difference: -3.5, 95%CI: -9.5- 2.5). Results did not change when adjusting for number of post-trial therapy sessions attended (CFQ: $chi^2(1) = 3.8$, p = 0.052; SF-36 PF: $chi^2(1) = 1.1$, p = 0.30). We also ran unadjusted analyses, and found that the marginal mean difference was weakened but aligned with the one observed in the primary analysis and their CIs overlap (see Appendix 1, Table B).

Table 2: Outcomes by original intervention assignments

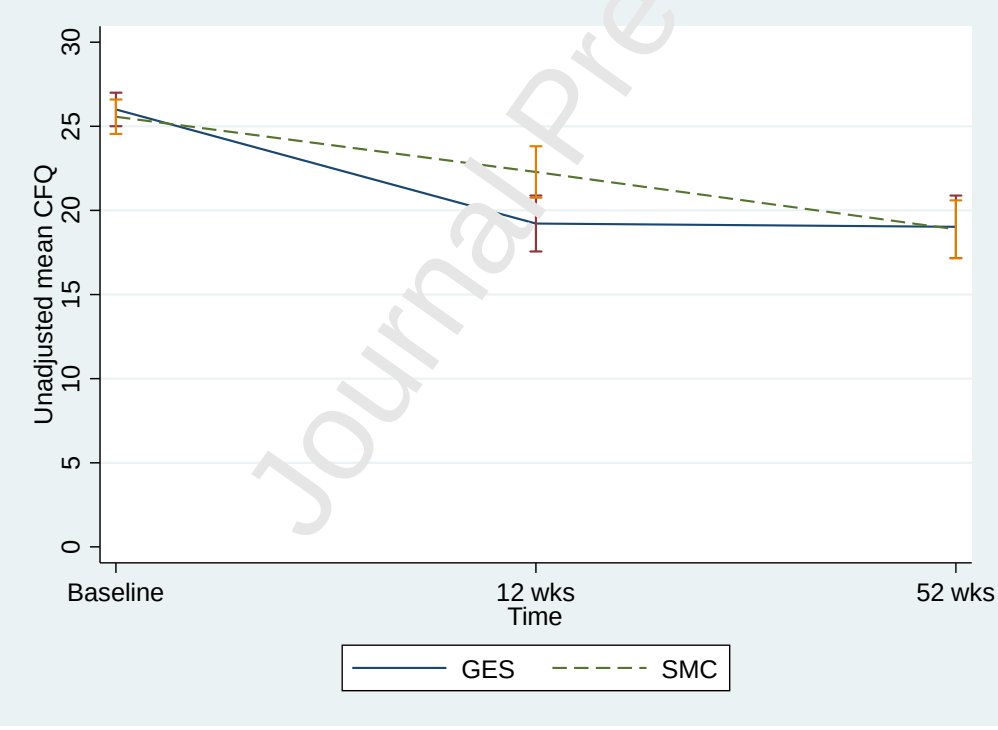
	GES (n = 77)	SMC (n = 81)	Difference between arms
Fatigue (CFQ)			
Mean score (SD) at baseline	26.0 (4.5)	25.6 (4.7)	0.43 (-1.0 to 1.9)
Mean score (SD) at 12 weeks	19.2 (7.4)	22.3 (7.0)	-3.1 (-5.3 to -0.8)
Mean score (SD) at long-term follow-up	19.0 (8.3)	18.9 (7.9)	0.2 (-2.4 to 2.7)
Physical functioning (SF-36 PF)			
Mean score (SD)at baseline	50.1 (21.9)	53.4 (22.6)	-3.3 (-10.3 to 3.7)
Mean score (SD) at 12 weeks	55.6 (24.3)	53.6 (26.3)	-1.9 (-6.0 to 9.9)
Mean score (SD) at long-term follow up	58.2 (29.6)	60.4 (26.2)	-2.2 (-11.0 to 6.6)

Data as mean (SD), or median difference (95%CI). Lower scores are better for fatigue, higher scores are better for physical functioning. Differences in unadjusted mean between-group differences are obtained from independent t-tests comparing GES and SMC. CFQ = Chalder Fatigue questionnaire. SF-36 PF = short form-36 physical function subscale. SMC = specialist medical care alone. GES = guided graded exercise self-help.

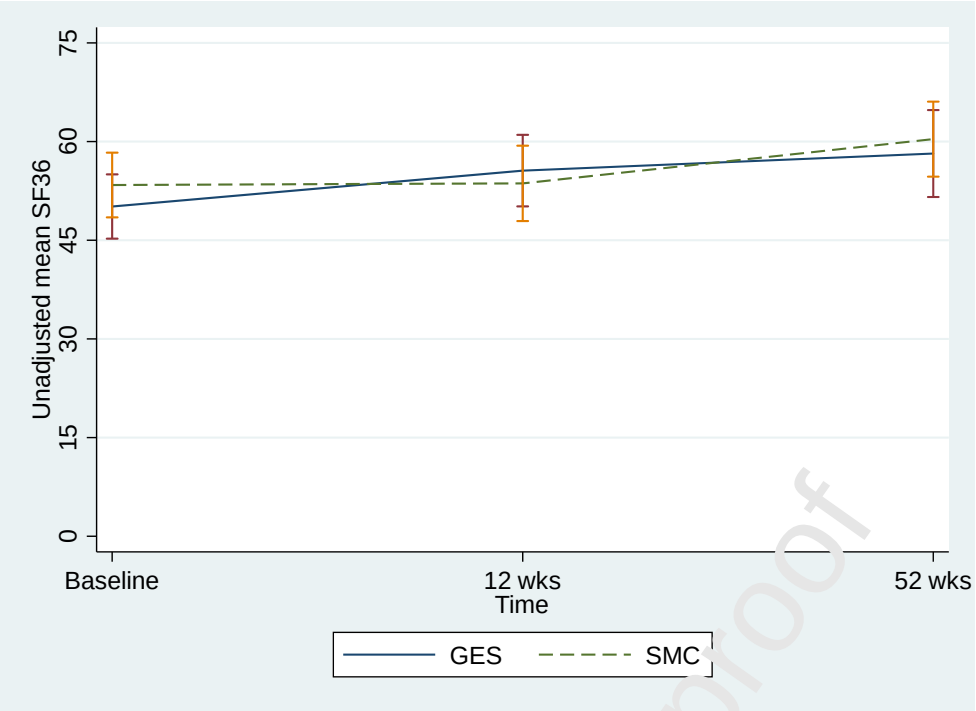
There was a main effect of time on functioning ($chi^2(1) = 7.3$, p = 0.007), indicating an overall increase in functioning scores between short-term and long-term follow up. There was also a significant effect of time on fatigue ($chi^2(1) = 6.3$, p = 0.001). Figure 2 indicates an overall decrease in fatigue between short-term and long-term follow-up. Results did not change when further adjusting for number of post-trial therapy sessions attended (CFQ: $chi^2(1) = 6.3$, p = 0.001; SF-36 PF:$chi^2(1) = 7.3$, p = 0.007).

Figure 2. Unadjusted mean scores on (a) fatigue and (b) physical functioning by intervention group for participants with complete information (n = 158). GES= guided graded exercise self-help. SMC=specialist medical care alone.

(a)



(b)



The arm*time interaction on CFQ was just significant at our adjusted alpha level ($chi^2(1) = 5.0$, p = 0.0247), which indicates that scores over time were moderated by intervention arm. Figure 2 suggests that improvements in fatigue were sustained between short and long term follow up for participants in the GES arm while those in SMC showed a reduction in fatigue between short and long term follow up (Figure 2 and Table 2).

The intervention arm*time interaction was not significant for SF-36 PF ($chi^2(1) = 1.4$, p = 0.23), which indicates that scores over time were not moderated by intervention arm (Figure 2). In our sensitivity analysis, we adjusted the model to account for the number of therapy sessions attended during the 15 months following randomisation, this adjustment did not change the results (CFQ: $chi^2(1) = 5.0$, p=0.0247; SF-36 PF: $chi^2(1) = 1.4$, p = 0.23).

The proportion of participants reporting a positive change in both overall health (CGI-health) and CFS (CGI-CFS) increased significantly from 12 weeks to long-term follow-up in both groups (see Table 3). The logistic regression showed there was no main effect of intervention arm on difference in positive change on the CGI-health scale (OR = 1.0, 95%CI: 0.4 to 2.3, p = .98) or CGI-CFS scale (OR = 1.0, 95%CI: 0.4 to 2.4, p = 0.95). Indeed, the proportion of participants with a positive change at long-term follow-up did not differ between arms for CGI-Health (GES = 23; 30% vs SMC = 19; 24%) or CGI-CFS (GES = 20; 26% vs. SMC: 17; 21%; Table 3). Results did not change after adjusting for the number of post-trial therapy sessions attended (CGI-CFS: OR = 1.1, 95%CI: 0.5 to 2.7, p = 0.83; CGI-Health: OR = 1.1, 95%CI: 0.5 to 2.7, p = 0.78).

Table 3: Within-group comparisons of long-term CGI-Health and CGI-CFS follow-up with final trial outcome

Change in perceived health (CGI-Health)	**GES (n = 78)**	**SMC (n = 80)**
12 weeks		
Positive change n (%)	13 (17)	5 (6)
No change	64 (82)	70 (88)
Negative change	1 (1)	5 (6)
Long-term n		
Positive change n (%)	23 (30)	19 (24)
No change	48 (62)	56 (70)
Negative change	7 (9)	5 (6)
Comparison		
Difference in positive change*^	5.0, p = 0.03	12.3, p = 0.001
Difference in negative change*^	4.5, p = 0.03	0, p = 1.0
Change in perceived CFS (CGI-CFS)	**GES (n = 78)**	**SMC (n = 80)**
12 weeks		
Positive change n (%)	10 (13)	6 (8)
No change	68 (87)	70 (88)
Negative change	0 (0)	4 (9)
Long-term		
Positive change n (%)	20 (26)	17 (21)
No change	54 (69)	57 (71)
Negative change	4 (5)	6 (8)
Comparison		
Difference in positive change*^	6.3, p= 0.01	7.1, p= 0.01
Difference in negative change*^	4.0, p = 0.05	0.5, p = 0.48

^ McNemar's Chi square test; * All differences between 12 week outcome and long-term follow-up. Difference in proportions with positive change and 95% confidence intervals for change in perceived health as compared to baseline.
Positive change was defined as' very much better' or 'much better'. No change was defined as 'a little better', 'no change', or 'a little worse'. Negative change was defined as 'much worse' or 'very much worse'. p-values for patient rated clinical global impression from McNemar test.
SMC = specialist medical care alone, GES = guided graded exercise self-help. CI = confidence interval.

There was no main effect of intervention arm on difference in negative change in the CGI-Health scale (GES: n = 7; 9% vs SMC 5; 6%; OR: 2.1, 95%CIs: 0.5 to 8.6, p = 0.29) or CGI-CFS (GES: 4; 5% vs. SMC: 6; 8%; OR: 0.8, 95%CIs: 0.2 to 3.5, p = 0.80). The proportion of participants reporting a negative change increased from 12 week to long-term follow-up in the GES group for both CGI-health (n = 1;1% to n = 7; 9%; McNemar's $chi^2(1) = 4.5$; p = 0.03) and CGI-CFS (n = 0; 0% to n = 4; 5%; McNemar's $chi^2(1) = 4.0$; p = 0.05) (Table 3). Results did not change after adjusting for the number of post-trial therapy sessions attended (CGI-CFS: OR = 0.9, 95% CI: 0.2 to 3.7, p = 0.84; CGI-Health: OR = 1.9, 95% CI: 0.4 to 7.9, p = 0.40).

Cost-effectiveness

Service use at baseline was similar between the two groups (Appendix 1, Table C). Around three-quarters of each group had GP contacts and half had contacts with other doctors. About one-fifth had contacts with complementary healthcare services. Total healthcare costs were just under £400 in each group over the previous three months. The level of informal care was high and for those who received this, the average was over ten hours per week. By the 12 week follow-up there was a slight decline in GP and other doctor use in both groups. Almost all of the GES group received the intervention and did so for a mean of 86 minutes (sd = 17.7). In the 12 weeks prior to long-term follow-up there was little change in GP or other doctor use compared to the 12 weeks after randomisation. Use of other healthcare professionals was somewhat higher in the SMC alone group at long-term follow-up. There were few other differences of note between the groups. Informal care declined from 12 weeks to long-term follow-up in both groups but was still received by most participants.

The average service costs were relatively high for GPs, other doctors and psychologists at all time points. Inpatient costs were very variable (Table 4). Total health and social care costs at baseline were very similar between the two groups. At both follow-up points the total healthcare costs were highest for the GES group. At 12 weeks follow-up this was due to the trial intervention and at long-term follow-up it was largely due to inpatient costs, even though this only pertained to a small number of participants (as can be seen from Table 3). Over the long-term follow-up period, the mean costs were £1,552 for GES and £1,524 for SMC alone. Adjusting for baseline and focusing only on those with QALYs recorded, showed that GES had costs that were on average higher by £85 (bootstrapped 95% CI, -£303 to £542).

Table 4. Mean (SD) service costs by trial arm and time point*.

	Baseline		12 week follow-up		Long-term follow-up	
	GES	**SMC**	**GES**	**SMC**	**GES**	**SMC**
Intervention	0 (0)	0 (0)	158 (40) [a]	0 (0)	0 (0)	0 (0)
GP	53 (46)	60 (50)	44 (45)	44 (42)	39 (46)	49 (50)
Other doctor	95 (120)	96 (129)	85 (139)	87 (122)	74 (123)	68 (106)
Other healthcare professional	17 (38)	24 (46)	15 (35)	17 (40)	17 (35)	39 (83)
Nurse	9 (23)	10(25)	9 (23)	6 (16)	11 (32)	13 (33)

Complementary healthcare	32 (99)	24 (64)	34 (92)	34 (70)	47 (126)	27 (66)
Psychologist	57 (254)	108 (529)	47 (166)	83 (466)	116 (237)	149 (552)
Occupational therapist	6 (23)	7 (24)	4 (20)	2 (13)	19 (56)	16 (66)
Physiotherapist	7 (28)	7 (31)	17 (73)	9 (45)	9 (39)	19 (76)
Counsellor	17 (102)	6 (34)	7 (53)	7 (40)	1 (5)	6 (39)
Inpatient	44 (327)	0 (0)	15 (88)	17 (129)	239 (1496)	36 (232)
Accident & emergency	12 (40)	6 (27)	5 (31)	9 (36)	15 (56)	8 (30)
Total healthcare cost	**348 (502)**	**347 (610)**	**439 (357)**	**316 (518)**	**587 (1576)**	**428 (693)**
Informal care	2057 (3138)	1388 (1684)	1652 (2137)	1513 (2008)	1255 (1444)	1675 (2918)
*Costs in 2016/17 £s						

The baseline mean EQ-5D utility scores for the GES group at baseline, 12 week follow-up and long-term follow-up were 0.5147, 0.5895 and 0.5882 respectively. The figures for SMC alone were 0.5498, 0.5616 and 0.5923. GES resulted in 0.5911 QALYs and standard care 0.5812. Adjusting for baseline, the difference was 0.0177 in favour of GES. The incremental cost-effectiveness ratio (ICER) was £85 divided by 0.0177. This came to £4802 per QALY. However, both the cost and QALY differences were small and there was much uncertainty around them, as shown in the cost-effectiveness plane (Figure 3). Whilst the most likely outcome would be GES resulting in higher costs and producing more QALYs (56% of replications), there was also a 19% likelihood of GES having lower costs and producing more QALYs and 18% likelihood of higher costs and fewer QALYs. The least likely result is GES saving money and having fewer QALYs. Figure 4 shows that if a QALY is not valued at all then there is less than 30% likelihood that that the intervention is cost-effective. With higher values given to a QALY the likelihood increases and at a threshold of £20,000 per QALY there is a 63% likelihood that GES was the most cost-effective option.

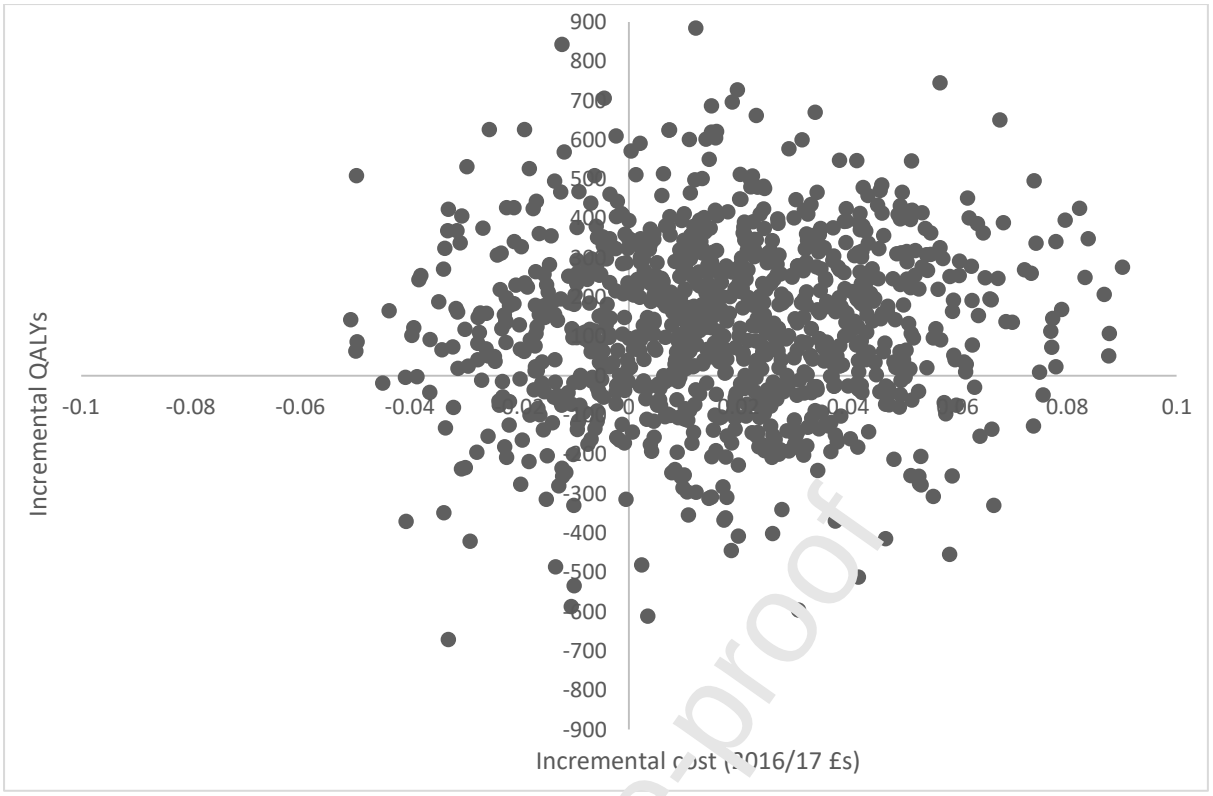


Figure 3. Cost-effectiveness plane.

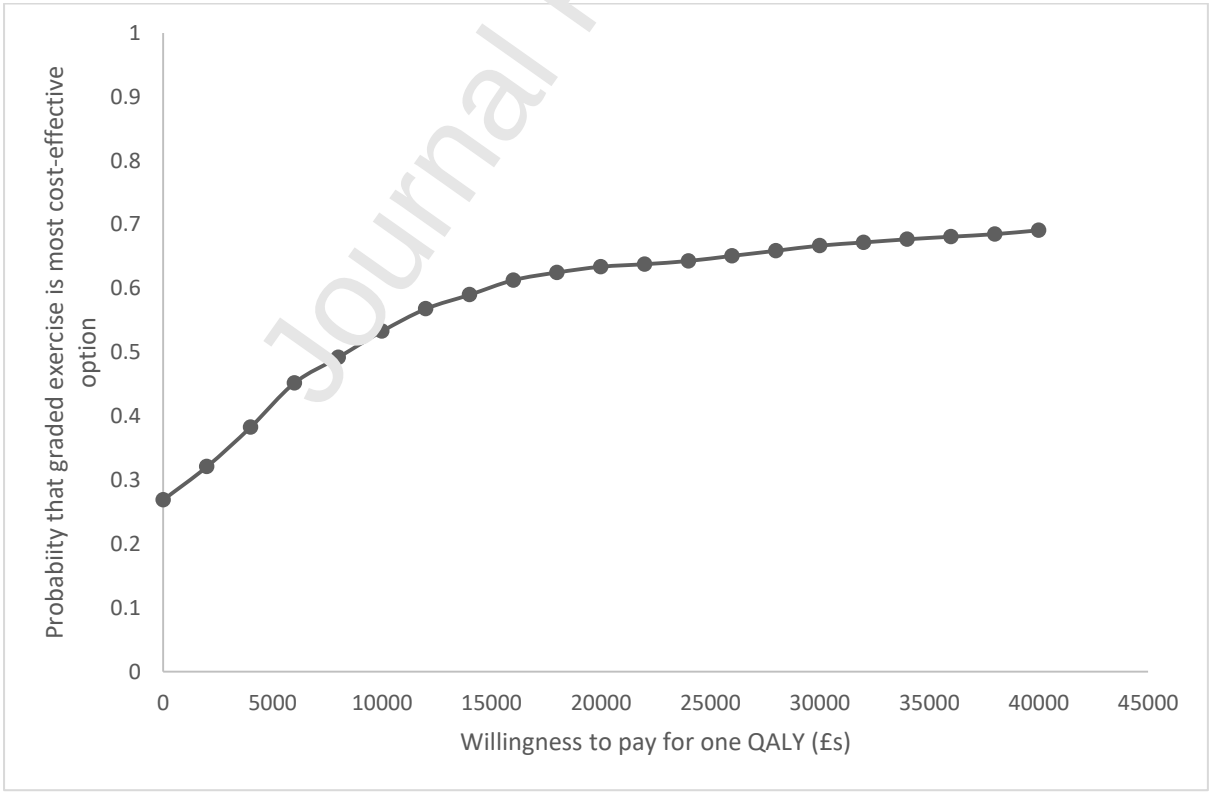


Figure 4. Cost-effectiveness acceptability curve.

4. Discussion

In the main GETSET trial we found a beneficial effect of GES on fatigue and physical functioning at 12 weeks after randomisation [3]. By 15 months' follow up there were no significant differences in either of the primary outcomes between the two interventions. This long term follow-up shows that at 15 months after randomisation the short-term improvements after GES, which we reported in the main paper [4], were maintained, but with no additional improvement, and no difference from the comparison group. There is no evidence that the improvements observed in the SMC group were due to them having received more exposure to therapy than the GES group after trial completion. As randomisation ceased after the 12 week follow-up, we are unable to speculate further as to why outcomes might have changed or not changed in either group in the 12 months since the end of randomisation.

In support of previous research in this field [7,8], GES does not appear to be associated with deterioration in overall health or CFS in the longer-term as these outcomes did not significantly differ between the originally allocated SMC and GES groups at long term follow up. There was an increase in the proportion reporting a deterioration from 12 weeks to long term follow up in those originally allocated to GES. This pattern found in the GES group may reflect that only one participant reported a deterioration at 12 weeks, having just participated in the intervention [3]. The fact that the deterioration occurred following rather than during the intervention period suggests that it was not likely to be the result of the intervention itself.

The health economic analyses indicated that GES added only slightly to healthcare costs and produced more QALYs. Although differences were small, the incremental cost-effectiveness ratio of £4,802 per QALY was substantially below the threshold often used in England to determine cost-effectiveness (£20,000 to £30,000 per QALY). However, there was still only a 63% likelihood that the intervention was cost-effective at £20,000 per QALY.

A systematic review has reported on outcomes without intervention in patients with CFS. This review concluded that prognosis was generally poor with improvement reported by 40% of participants [2]. A recent systematic review looking at outcomes following exercise interventions (of any sort, such as Qigong) reported on eight trials. Two trials of graded exercise therapy reported on long-term follow up, and found that benefits of GET were maintained at 2 years [6] and 2.5 years [5], but the review reports that that there was low certainty of evidence longer-term [8]. Our findings are consistent with results from both of those studies - that improvements from graded exercise therapy interventions are maintained in the long-term, and provides greater certainty of maintained benefits.

Limitations of this study include the issue of incomplete response rate. However, of note, we adjusted our analysis to account for the fact that some baseline characteristics were associated with missing

information on the outcome measure to improve our model estimates and avoid bias. The duration of follow-up varied from 1 to 3 years, although this is unlikely to have caused bias as we controlled for it in our sensitivity analysis for CGI and it did not change the findings. As is common in this type of research, our outcomes were all self-rated. This could be considered a limitation, although they are potentially the most important outcomes in conditions such as CFS, which are defined by symptoms which are described by the individual. It could be considered a limitation that we added fatigue as a second primary outcome mid-way through trial recruitment for reasons given in the methods. However this was before any outcome data were formally analysed, sample size was recalculated using an adjusted significance level accordingly and both primary outcomes are reported. The number of minutes of physical activity undertaken at baseline was significantly higher in those allocated to SMC, but this was a finding that was already present between groups at baseline and a sensitivity analysis shows it made no difference to our findings. Finally, interpretation of these results is limited by participants having had the opportunity to have a service therapy between 12 weeks and long-term follow-up, and apart from the collection of appointment data from service records, participants were not monitored during this period.

The strength of this study is that the results provide first estimates of long-term outcome following a graded exercise guided self-help intervention. In this trial we were able to offer the GES intervention by telephone and Skype, mitigating the fatiguing effects of travelling for the intervention and enabling those with more severe illness to be included. Our qualitative study suggested that those who had been ill for longer made less improvement [26]. Some of those using GES, and probably including those with a longer duration or more severe illness, will need additional support, and we suggest that GES is considered as a first step in a stepped care approach for people with CFS, if these results are confirmed by others.

In summary, a year after the guided intervention ended, there were no significant differences between the two groups in fatigue or physical functioning, and no evidence of a difference between groups in deterioration in overall health. The short-term improvements after GES on fatigue and functioning were maintained at long-term follow-up although primary outcomes no longer differed from those in the SMC group. The addition of GES to SMC may be a cost-effective intervention, but based on the present findings we cannot say this with confidence. Better treatments are still needed for patients with this chronically disabling illness.

Contributors

PDW & LVC conceived the follow up study with PMc and FP contributing to its design. PDW was the PI, LVC was the trial manager, PMc was the Health Economist and FP was the statistician. LVC, PMc, PDW and FP drafted the manuscript. PMc designed the methodology for the health economic outcomes and undertook the health economic analysis. FP designed the statistical methodology for the clinical outcomes and undertook the analyses. MVW participated in the trial design and coordination as centre leader. All authors contributed to the final manuscript.

Conflicts of interest

PDW reports grants from the National Institute of Health Research and the Sue Estermann Charity during the conduct of the study; personal fees from Swiss Re-insurance company, outside the submitted work; and PDW is an appointed member of the Independent Medical Experts' Group, a non-department public body, which advises the UK Ministry of Defence about its Armed Forces Compensation Scheme.

Funding sources and acknowledgements

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0610-22060). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. This study was also supported by funding through the Sue Estermann Fund. We thank Michelle Beynon for facilitating the long-term follow-up data collection and entry.

Data sharing

Suitably anonymised data are available for external researchers to request through the Yale University Open Data Access (YODA) Project which is prepared to share these data. The YODA project is an independent organisation that advocates for the responsible sharing of clinical research data. Data from available clinical trials are shared through the YODA Project with registered users with approved proposals for scientific research. For the YODA project see http://yoda.yale.edu

Funding: UK National Institute for Health Research (RfPB Programme) and the Sue Estermann Fund.

Trial details: The trial is registered at http://istctn.org, ISRCTN22975026.

References

[1] J.B. Prins, J.W. van der Meer, G. Bleijenberg, Chronic fatigue syndrome, Lancet 367 (2006) 346-55. https://doi.org/10.1016/S0140-6736(06)68073-2.
[2] R. Cairns, M. Hotopf, A systematic review describing the prognosis of chronic fatigue syndrome, Occup Med 55 (2005) 20-31. https://doi.org/10.1093/occmed/kqi013.
[3] P.D. White, K.A. Goldsmith, A.L. Johnson, L. Potts, R. Walwyn, J.C. DeCesare, H.L. Baber, M. Burgess, L.V. Clark, D.L. Cox, J. Bavinton, B.J. Angus, G. Murphy, M. Murphy, H. O'Dowd, D. Wilks, P. McCrone, T. Chalder, M. Sharpe on behalf of the PACE trial management group, Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial, The Lancet 377 (2011) 823-836. https://doi.org/10.1016/S0140-6736(11)60096-2.
[4] L.V. Clark, F. Pesola, J.M. Thomas, M. Vergara-Williamson, M. Beynon, P.D. White, Guided graded exercise self-help plus specialist medical care versus specialist medical care alone for chronic fatigue syndrome (GETSET): a pragmatic randomised controlled trial, The Lancet 390 (2017) 363-73. https://doi.org/10.1016/S0140-6736(16)32589-2.
[5] L.V. Clark, P. McCrone, D. Ridge, A. Cheshire, M. Vergara-Williamson, F. Pesola, P.D. White, 2016. Graded exercise therapy guided self-help trial for patients with chronic fatigue syndrome (GETSET): protocol for a randomized controlled trial and interview study, JMIR Res Protoc. 5, e70. https://doi.org/10.2196/resprot.5395.
[6] P. Powell, R.P. Bentall, F.J. Nye, R.H.T. Edwards, Patient education to encourage graded exercise in chronic fatigue syndrome: 2-year follow-up of randomised controlled trial, BJPsych 184 (2004) 142-146. https://doi.org/10.1192/bjp.184.2.142.
[7] M. Sharpe, K.A. Goldsmith, A.L. Johnson, T. Chalder, J. Walker, P.D. White, Rehabilitative treatments for chronic fatigue syndrome: long-term follow-up from the PACE trial. The Lancet Psychiatry 2 (2015) 1067-74. https://doi.org/10.1016/S2215-0366(15)00317-X
[8] L. Larun, J. Brurberg, J. Odgaard-Jensen, J.K. Price J,. Exercise therapy for chronic fatigue syndrome. Cochrane Systematic Review, 2019. https://doi.org/10.1002/14651858.CD003200.pub8.
[9] P. Mc Crone, M. Sharpe, T. Chalder, M. Knapp, A.L. Johnson, K.A. Goldsmith, P.D. White, 2012. Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: a cost-effectiveness analysis. PLOSONE. 7, e40808. https://doi.org/10.1371/journal.pone.0040808.
[10] National Institute for Health and Clinical Excellence. Clinical Guideline CG53. Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy): diagnosis and management of CFS/ME in adults and children. http://guidance.nice.org.uk/CG53 (accessed 27th Jan 2020).
[11] M.C. Sharpe, L.C. Archard, J.E. Banatvala, L.K. Borysiewicz, A.W. Clare, A. David, R.H. Edwards, K.E. Hawton, H.J. Lambert, R.J. Lane. A report – chronic fatigue syndrome: guidelines for research, J R Soc Med 84 (1991) 118-121.
[12] Reeves WC, Lloyd A, Vernon SD, Klimas N, Jason LA, Bleijenberg G, International Chronic Fatigue Syndrome Study Group. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. BMC Health Serv Res 2003 Dec 31;3(1):25
[13] L.V. Clark (ed), Graded exercise therapy: a self-help guide for those with chronic fatigue syndrome/myalgic encephalomyelitis, Bart's and the London NHS Trust, London, 2011. https://www.qmul.ac.uk/wolfson/media/wolfson/current-projects/GET-guide-booklet-version-1-22062010.pdf (accessed 27th Jan 2020).
[14] L. V. Clark, E. Tims & P.D. White. (2012) Guided Support Manual: Guided Self-help graded exercise therapy for CFS/ME. https://www.qmul.ac.uk/wolfson/media/wolfson/current-projects/GETSET-therapists-manual-with-appendices.pdf
[15] T. Chalder, G. Berelowitz, S. T. Pawlikowska, L. Watts, S. Wessely, D. Wright, E.P. Wallace, Development of a fatigue scale. J Psychosom Res 37 (1993) 147-53. https://doi.org/10.1016/0022-3999(93)90081-p.

[16] C.A. McHorney, J.E. Ware, A.E. Raczek, The MOS 36 item short form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 31 (1993) 247-253. https://doi.org/10.1097/00005650-199303000-00006.

[17] W. Guy, ECDEU assessment manual for psychopharmacology, National Institute of Mental Health Rockville, MD, 1976 218-22.

[18] The EuroQol Group, EuroQol: a facility for the measurement of health related quality of life, 16 Health Policy 199–208. https://doi.org/10.1016/0168-8510(90)90421-9.

[19] J. Beecham, M. Knapp, Costing psychiatric interventions, in: G. Thornicroft (Ed.), Measuring mental health needs, Gaskell, London, 2001, pp. 200–224.

[20] J. Twisk, L. Bosman, T. Hoekstra, J. Rijnhart, M. Welten, M. Heymans. Different ways to estimate treatment effects in randomised controlled trials. Contemporary Clinical Trials Communications 10 (2018) 80-85. https://doi.org/10.1016/j.conctc.2018.03.008

[21] M. Pampaka, G. Hutcheson, J. Williams. Handling missing data: analysis of a challenging data set using multiple imputation. International Journal of Research & Method in Education. 39 (2016). https://doi.org/10.1080/1743727X.2014.979146

[22] R.H.H. Groenwold, A. Rogier T. Donders, K.C.B. Roes, F.E. Harrell, K.G.M. Moons. Dealing With Missing Outcome Data in Randomized Trials and Observational Studies. American Journal of Epidemiology. 175 (2011) 210-217. https://doi/10.1093/aje/kwr302.

[23] C.L. Craig, A.L. Marshall, M. Sjostrom, A.E. Bauman, M.L. Booth, B.E.Ainsworth, M.Pratt, U.Ekelund, A. Yngve, J.F. Sallis, P. Oja, International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 35 (2003) 1381-1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB.

[24] P. Dolan, C. Gudex, P. Kind, A. Williams (1995) A social tariff for EuroQol: results from a UK population survey. York: University of York. https://www.york.ac.uk/che/pdf/DP138.pdf.

[25] A. Manca, N. Hawkins, M.J. Sculpher, Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility, Health Econ 14 (2005) 487–496. https://doi.org/10.1002/hec.94[illegible]

[26] L. Curtis, Unit costs of health and social care. Canterbury: PSSRU, 2010.

[27] R. Walwyn, L. Potts, P. McCrone, A. L. Johnson, J.C. DeCesare, H. Baber, K. Goldsmith, M. Sharpe, T. Chalder, P.D. White, A randomised trial of adaptive pacing therapy, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome (PACE): statistical analysis plan, Trials 14 (2013) 386. https://dx.doi.org/10.1186%2F1745-6215-14-386.

[28] A. Cheshire, D. Ridge, L. Clark, P.D. White, Guided graded Exercise Self-help for chronic fatigue syndrome: patient experiences and perceptions, Disabil Rehabil 2018. https://doi.org/10.1080/09638288.2018.1499822.

Highlights

- Guided graded exercise self-help (GES) can lead to sustained improvement in patients with chronic fatigue syndrome.
- There was no evidence of greater harm after GES compared to specialist medical care at long-term follow-up.
- The study showed that GES was probably cost-effective.
- Most patients remained unwell at follow up; more effective treatments are required.