Taylor & Francis
Taylor & Francis Group

Fatigue: Biomedicine, Health & Behavior

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rftg20

Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial

Carolyn Wilshire, Tom Kindlon, Alem Matthees & Simon McGrath

To cite this article: Carolyn Wilshire, Tom Kindlon, Alem Matthees & Simon McGrath (2017) Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial, Fatigue: Biomedicine, Health & Behavior, 5:1, 43-56, DOI: 10.1080/21641846.2017.1259724

To link to this article: https://doi.org/10.1080/21641846.2017.1259724

Published online: 14 Dec 2016.

FATIGUE: BIOMEDICINE, HEALTH & BEHAVIOR, 2017
VOL. 5, NO. 1, 43–56
http://dx.doi.org/10.1080/21641846.2017.1259724

Taylor & Francis
Taylor & Francis Group

Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial

Carolyn Wilshire[a], Tom Kindlon[b], Alem Matthees[c] and Simon McGrath[d]

[a]School of Psychology, Victoria University of Wellington, Wellington, New Zealand; [b]Irish ME/CFS Association, Dublin, Republic of Ireland; [c]Perth, Australia; [d]Monmouth, UK

ABSTRACT

BACKGROUND: Publications from the PACE trial reported that 22% of chronic fatigue syndrome patients recovered following graded exercise therapy (GET), and 22% following a specialised form of CBT. Only 7% recovered in a control, no-therapy group. These figures were based on a definition of recovery that differed markedly from that specified in the trial protocol.

PURPOSE: To evaluate whether these recovery claims are justified by the evidence.

METHODS: Drawing on relevant normative data and other research, we critically examine the researchers' definition of recovery, and whether the late changes they made to this definition were justified. Finally, we calculate recovery rates based on the original protocol-specified definition.

RESULTS: None of the changes made to PACE recovery criteria were adequately justified. Further, the final definition was so lax that on some criteria, it was possible to score below the level required for trial entry, yet still be counted as 'recovered'. When recovery was defined according to the original protocol, recovery rates in the GET and CBT groups were low and not significantly higher than in the control group (4%, 7% and 3%, respectively).

CONCLUSIONS: The claim that patients can recover as a result of CBT and GET is not justified by the data, and is highly misleading to clinicians and patients considering these treatments.

ARTICLE HISTORY
Received 7 July 2016
Accepted 6 November 2016

KEYWORDS
Chronic fatigue syndrome; myalgic encephalomyelitis; graded exercise therapy; cognitive behavioural therapy

In chronic fatigue syndrome (CFS), one of the major presenting symptoms is long-term, disabling fatigue (often accompanied by neurocognitive impairment, post-exertional malaise, pain, flu-like symptoms or other symptoms [1]). CFS patients' activity levels can also be greatly reduced. Consequently, the question arises as to whether physical deconditioning contributes significantly to the clinical picture in CFS. Indeed, according to one theoretical account – the behavioural/deconditioning model [2] – physical deconditioning plays a primary role in perpetuating the disease, and is (directly or indirectly) responsible for

CONTACT Carolyn Wilshire carolyn.wilshire@vuw.ac.nz School of Psychology, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

many of its most common symptoms, ranging from fatigue and sleep disturbance to orthostatic intolerance [2]. According to this model, these deconditioned CFS patients experience fatigue, stiffness and other symptoms when they attempt to increase their activity, and they misinterpret these as signs of ongoing disease. They then become more focused on their symptoms, and fearful of further activity, creating a self-perpetuating cycle [2].

The behavioural/deconditioning model predicts that if patients' cognitions and behaviours can be effectively modified, full recovery from CFS is possible. Two forms of intervention have been developed based on this model. The first is graded exercise therapy (GET): patients receive support and guidance on how to gradually increase their activity so as to minimise unpleasant symptoms. The second is a special form of cognitive behavioural therapy (CBT) which encourages patients to re-examine how they interpret their symptoms, and also to experiment with very gradually increasing their activity.

The PACE trial was designed to examine the efficacy of these two treatments [2]. Reports of the trial's findings concluded that both GET and CBT led to recovery in over a fifth of patients [3]. Such a claim, if true, would have significant consequences for the understanding and treatment of CFS. In this commentary, we critically evaluate the evidence upon which this claim is based.

Summary of the PACE trial

The PACE trial was an open-label, randomised trial designed to examine the effectiveness of various behavioural interventions for CFS [2]. The largest trial of its kind, it involved more than 600 adult participants. All were diagnosed with CFS according to the Oxford case definition, which focuses exclusively on fatigue, and in which fatigue must be the principal symptom [4]. The fatigue must be of more than six months' duration, have a definite onset, and result in significant disability. Participants in the trial also had to meet additional minimum criteria with respect to their scores on self-report measures of fatigue and disability.[1] All participants were aged 18 or over; the vast majority (97%) were under 60 [7].

All patients in the trial received at least three specialist medical care consultations, where they were prescribed medication as necessary for symptoms such as pain and insomnia. One group (Control) received no further treatment. The remaining three groups received up to 15 additional sessions of one of three therapies: CBT; GET; or a novel treatment, adaptive pacing therapy. The CBT programme was designed to address what were seen as patients' 'unhelpful cognitions' about their illness and their fears about exercise [2, p. 825]. As part of the programme, patients were encouraged to plan gradual increases in their physical activity and assess their effects. They were told that CBT was 'a powerful and safe treatment which has been shown to be effective in … CFS/ME' and that 'many people have successfully overcome CFS/ME using cognitive behaviour therapy, and have maintained and consolidated their improvement once treatment has ended' (CBT participants' manual [8, p. 123]).

The GET programme focused on gradually increasing patients' activity levels. Participants were told that 'in previous research studies, most people with CFS/ME felt either "much better" or "very much better" with GET', and that GET was 'one of the most effective therapy strategies currently known' (GET participants' manual [9, p. 28]). In contrast, the adaptive pacing therapy programme encouraged patients to restrict their activity levels to no more than 70% of what they felt they could safely do, with a view to avoiding

an exacerbation of symptoms. No specific claims were made to patients as to its effectiveness [10].

The primary outcome measures, obtained 52 weeks after treatment allocation, were: (a) self-rated physical function using the Physical Function subscale of the SF-36 Health Survey (SF-36; [5]) and (b) self-rated fatigue, measured using the Chalder Fatigue Questionnaire [6]. Self-rated fatigue and physical function improved in all groups, but significantly more so for the CBT and GET groups. Recovery results appeared in a subsequent paper ([3]; henceforth referred to as *the Recovery paper*). It was reported that 22% of patients in the CBT group and 22% in the GET group recovered after treatment. In contrast, only 8% and 7% recovered in the adaptive pacing therapy and Control (no therapy) groups respectively. The authors concluded that CBT and GET are effective interventions for CFS *that can often lead to recovery*, and that these should be widely recommended, at least to ambulatory patients.

Definition of recovery

In the Recovery paper, four criteria were used to define recovery. To be classed as 'recovered', the patient had to meet a specified threshold score on each of the following three self-report scales: (a) The SF-36 Physical Function subscale; (b) the Chalder Fatigue Questionnaire and (c) the Clinical Global Impression (CGI) scale, a self-rated measure of overall health change (based on Guy [11]). The fourth criterion was that participants should no longer meet a specified case definition for CFS. However, the specific thresholds used to define recovery with respect to all four of these criteria were substantially modified from those specified in the original trial protocol [12]. In each instance, the changes to the criteria made 'recovery' easier to achieve. Table 1 summarises the original and revised criteria. Recovery rates according to the original protocol-specified definition were not published.

Current study: aims and methods

In this paper, we critically examine the PACE researchers' definition of recovery, drawing on normative data for the outcome measures used in their definition, as well as other

Table 1. Summary of protocol-specified and revised recovery criteria.

Outcome measure	Original protocol recovery criterion [12]	Recovery paper criterion [3]
SF-36 Physical Function subscale [5]	≥85	≥60
Chalder Fatigue Questionnaire [6]	≤3 on 11-point, bimodal-scored scale	≤18 on 33-point, Likert-scored scale
Clinical Global Impression 7-point scale [11]	Score of 1 ('very much better')	Score of 1 or 2 ('very much better' or 'much better')
Clinical case definition of CFS	Patients had to fail to meet the Oxford case definition for CFS [4], as well as the CDC and London case definitions [13,14].	For the primary analysis of recovery rates ('trial recovery'), patients had to fail to meet only one case definition: the modified Oxford case definition used at trial entry. In this modified version, the participant's SF-36 Physical Function score had to be at or below 65 and their Chalder fatigue score (bimodal scoring method) had to be six or above

relevant research. We also present some simple analyses of recovery rates using the PACE trial dataset that was recently made available to the public.[2] First, for each of the four individual recovery criteria, we describe the final definition used in the published findings, and how it differed from that specified in the original trial protocol. We discuss whether these changes were justified, referring where appropriate to evidence from normative studies. Using the publicly available data set, we also calculate the proportions of patients that would qualify as 'recovered' on that criterion according to both the original protocol-specified definition, and the final definition. This enabled us to assess the actual impact of the changes on recovery outcomes.

Second, we compare the overall recovery rates reported by the researchers with our own calculations of the recovery rates based on a definition that is almost identical to that specified in the original trial protocol. We used an intention-to-treat approach, as specified in the protocol: any missing outcome measure was replaced with the most recent available score for that person on that outcome. We compared the rates of recovery obtained in this way for the GET and Control (no therapy) groups using Fisher's exact test; the same approach was applied for the comparison of the CBT and Control (no therapy) groups. In the final sections of the article, we consider the wider question as to how recovery should best be defined, and more specifically, whether any of the other outcomes that were measured in the PACE trial may provide further insights into the rates of recovery.

Criterion 1. The SF-36 physical function score

The SF-36 Physical Function subscale [5] asks participants to rate how limited they are in doing 10 daily activities, ranging from running/lifting heavy objects/doing strenuous sports, to walking 100 yards and bathing/dressing. Respondents can answer, 'Yes, limited a lot' (score 0), 'Yes, limited a little' (score 1), or 'No, not limited at all' (score 2). The total score is multiplied by five. The minimum score of 0 indicates major limitations in all activities; the maximum score of 100 indicates no impairment on any activity. To be eligible to enter the PACE trial, patients had to score 65 or lower on this scale.

In the original study protocol, recovery on this measure was defined as a score of 85 or above [12], a similar figure to that used in previous studies of behavioural interventions for CFS (e.g. [15,16]). This was a reasonable, albeit arguably low threshold: according to reference data from a large British community sample, the vast majority (90%) of people aged 18–59 without a long-term illness or disability actually score 90 or higher ([17]: for summary report, see Bowling et al. [18]). However, in the Recovery paper, the minimum score for a recovery outcome was lowered to 60. This late post-protocol change increased the proportion of participants who met this criterion from 14% to 45%, a more than threefold increase. The justification given for this change was that the original threshold of 85 was so high that 'approximately half the general working age population would not meet it' [3, p. 2229]. This claim is clearly incorrect: it seems to have been based on summary statistics from a large British reference sample reported in Bowling et al. [18], in which almost a third of participants were aged 60 or over, and one-fifth reported a long-term illness or disability that limited their daily activities or the work they could do. Since the PACE trial participants were screened and excluded for the presence of fatiguing illnesses other than CFS, any normative dataset used to define recovery should have also excluded such illnesses. In addition, the authors seem to have derived their 'approximately half'

figure from calculations based on the sample mean and standard deviation, a method which was inappropriate, given the highly skewed distribution of scores (see Figure 1 for illustration).

Indeed, as Figure 1 shows, if we look just at those participants from the British reference sample who were aged 18–59 *and* did not have a long-term illness or disability, the median (and modal) score for this highly skewed normative sample was 100 and 93% scored at or above the original recovery threshold of 85 [17]. Their arguments do not therefore justify the lowering of the SF-36 physical function threshold score from the originally specified minimum level.

The revised recovery threshold score is so low that it is close to the mean score of patients with osteoarthritis of the hip, rheumatoid arthritis, and Class II congestive heart failure [19,20], as shown in Figure 1. This is a serious concern. But perhaps even more worrying, the new, lowered threshold of 60 meant that a patient could qualify as 'recovered' on this criterion with a lower score than was required to *enter* the trial (the trial entry threshold was 65, a level described by the study authors as representing 'abnormal levels of physical function' [3, p. 2229]). An analysis of the publicly available PACE trial data reveals that 13% of participants qualified as recovered on this revised criterion *before the trial even began*, and for three of these cases, physical function scores were actually *lower* at the end of the trial than they were at its commencement.

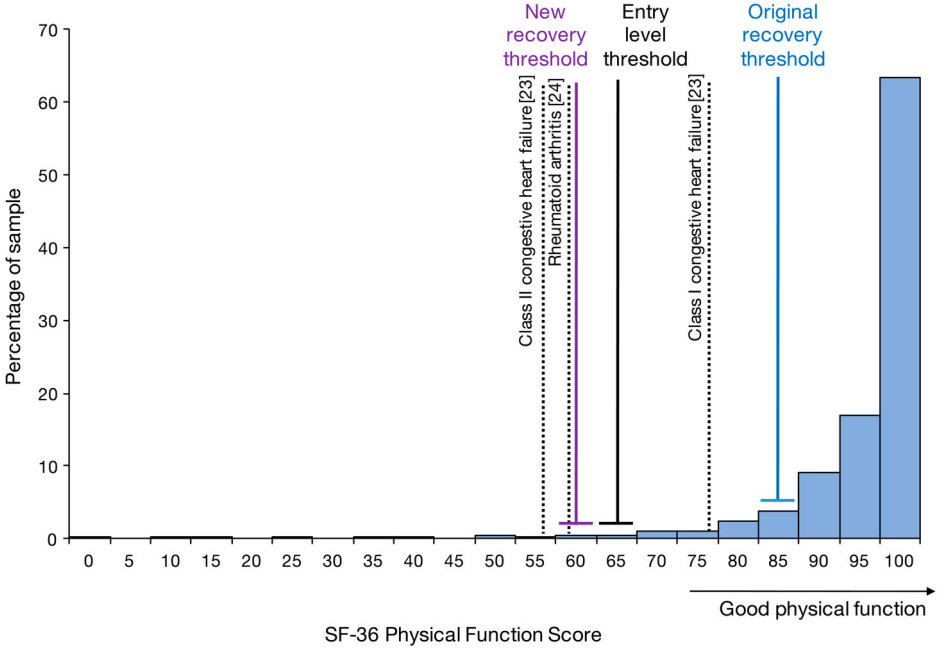

Figure 1. Distribution of SF-36 physical function scores from the subsample of Bowling et al's. [18] British community participants who were aged 18–59 and without long-term illness or disability (as calculated from the published data [17]). Also presented for comparison are: (a) the recovery criteria specified in the original protocol [12]; (b) the revised criteria used in the recovery paper [3] and (c) mean scores for several chronic illness populations.

Criterion 2. The Chalder Fatigue Questionnaire

The Chalder Fatigue Questionnaire [6] contains 11 questions related to fatigue, such as 'Do you have problems with tiredness?', 'Do you need to rest more?' and 'Do you have difficulty concentrating?' When responding, PACE participants were asked to compare themselves over the previous month to when they were last well. In the original trial protocol, each response of 'less than' or 'no more than' usual scored a 0, whereas a response of 'more than' or 'much more than' usual scored a 1, giving a minimum score of 0 (indicating no more fatigue than premorbidly on all questions) and a maximum possible score of 11 (severe fatigue). To be eligible to enter the PACE trial, patients had to score 6 or above on this 11-point scale [2]. To be considered recovered according to the original trial protocol, patients could score no more than 3 – the same threshold recommended by the authors of the questionnaire [6], and used in a previous study of CBT for CFS [15].

However, in the 2013 Recovery paper, the PACE authors changed this requirement in two ways. First, they adopted a different scoring method, where each response to each question was scored on a scale from 0 to 3: 'less than usual' scored 0, 'no more than usual' scored 1, 'more than usual' scored 2 and 'much more than usual' scored 3. This method, which they referred to as the Likert method, gives a maximum possible score of 33. A score of 11 or below indicates no greater overall fatigue than premorbidly.

Second, the authors lowered the threshold for defining recovery on this measure. A participant now had to score 18 or below on the 'Likert' scoring method to be considered recovered. This threshold is so generous that patients officially diagnosed with CFS – even within the centres associated with the PACE trial – commonly scored within the 'recovery' range.[3] It is also much less stringent than the original threshold. A score of 18 translates into between 4 and 9 using the original binary scoring method (depending upon each respondent's scoring pattern: see Figure 2). Our analyses show that in PACE, changing this threshold doubled the number of patients who qualified as 'recovered' on this criterion (the total recovered rose from 15% to 29%). 16 of the new qualifying cases reported continuing fatigue on seven out of the 11 CFQ items, and one case even reported fatigue on 8 of the 11 items. These scores indicate considerably greater levels of fatigue than the maximum score of 3 specified on the original protocol. Finally, and perhaps most worryingly, seven of the PACE participants themselves fulfilled this new recovery criterion upon trial entry [21].

According to the authors, their change to this criterion was made '… following the publication of a much larger study of fatigue in adults in a representative population sample of patients registered with a GP from South East England' [3 p. 2229]. The mean score in this sample (using the Likert-style scoring method) was 14.2, and 18 was chosen as a threshold because it was the highest integer score within one standard deviation of the mean [22]. This explanation is puzzling. First, the data analysed in this recent study [22] were not in fact new: they had been included in a report of a much larger sample published some 16 years earlier [23], whose authors included one of the PACE co-authors (Chalder). Second, the group described in this new paper [22] consisted of patients who had visited their GP within a year of being sampled in the earlier study, and a substantial proportion had compromised health. In the parent population of over 15,000 participants from which this sample was drawn [23], 34% of participants reported a significant health problem that could cause significant fatigue. These problems included

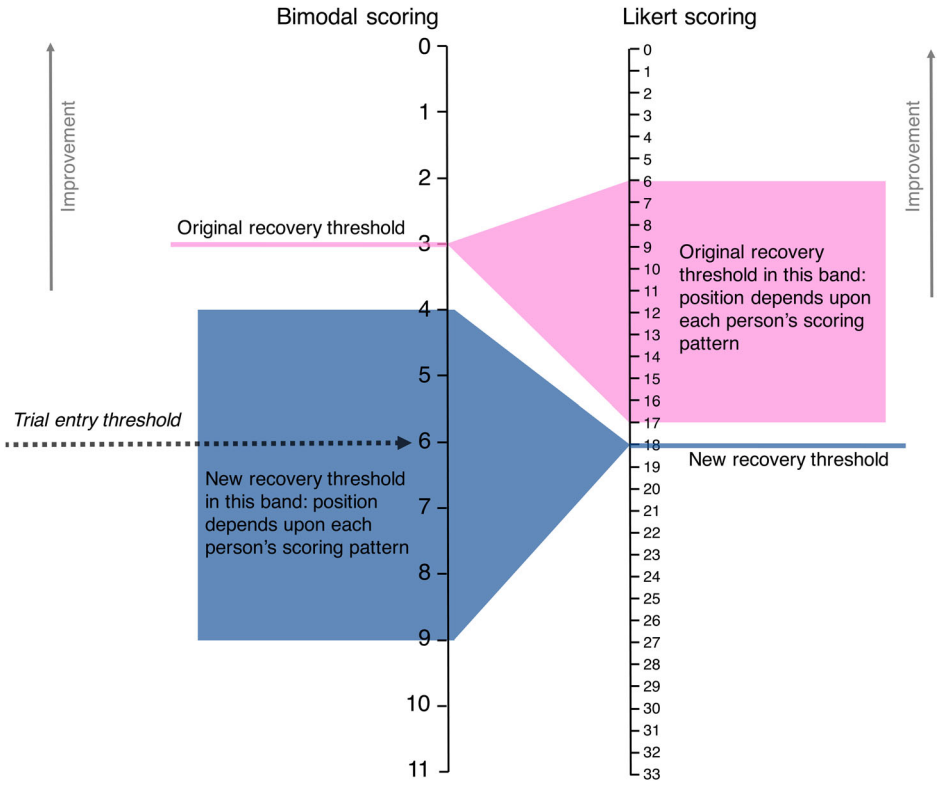

Figure 2. The new Chalder Fatigue Questionnaire recovery threshold in context.

surgery, anaemia, pregnancy and psychiatric problems such as depression or anxiety, and even CFS. A more appropriate sample for benchmarking recovery would have been the one reported by Loge et al. [24], which was randomly selected from the Norwegian National Register, and presented data separately for those who did not report a current disease or health problem. This study had also been published long before the PACE trial. The sample contained over 2000 respondents, and the mean score for healthy working age participants was 11.2 on the Likert-scoring method (a score of 11 indicates no more or less fatigue than when a person was last well, as would be expected from a healthy subset).

In conclusion, the justification given for changing the recovery threshold on the fatigue score criterion was inadequate. The revised threshold does not indicate recovery in any common-sense meaning of the word.

Criterion 3. The CGI score

The CGI scale used in the Recovery paper is a 7-point, single-item, self-report rating of change in health status (based on Guy [11]). PACE participants rated their health change following the intervention on a scale from 1 ('very much better') to 7 ('very much worse'). In the original trial protocol, a score of 1 ('very much better') was defined as indicating recovery on this measure [12]. However, in the Recovery paper, this threshold

was also changed, so that a score of 2 ('much better') was now also considered sufficient. This change increased the proportion of participants who met this criterion from 12% to 34%. The justification for this change, in the words of the authors, was that they 'considered that participants rating their overall health as "much better" represented the process of recovery' [3, p. 2230].

However, this is not the sense in which the term 'recovery' is used elsewhere in the paper, including the Conclusions section (see, e.g. [3, p. 2233]). Also, the phrase 'process of recovery' implies that improvement was very likely to continue in the future, and it is far from clear that this was the case. A follow-up study conducted a median of 31 months after each patient entered the trial found that there had been no reliable further improvement on the primary outcome measures in the GET arm, and only very small improvements in the CBT arm [25.] In the context of the PACE trial, lowering the CGI threshold was inappropriate, in our view, and artificially inflated the recovery figures.

Criterion 4. CFS caseness

A fourth criterion for defining recovery was that participants should no longer meet a case definition of CFS. In the Recovery paper, this case definition was a strict one. It consisted of the standard Oxford criteria for CFS caseness (the principal symptom must be fatigue, and it must be of more than six months' duration, have a definite onset, and result in significant disability) plus two additional requirements: (a) that the patient respond to at least 6 of the 11 fatigue questions on the Chalder Fatigue Questionnaire with 'more than usual' or 'much more than usual' and (b) that they score no higher than 65 on the SF-36 Physical Function scale. Since the public dataset identifies which participants failed to meet the Oxford CFS case definition at 52 weeks, and also includes the relevant CFQ and SF36 scores, it is relatively simple to calculate the proportion of participants meeting this recovery criterion: it is 48%. That is, nearly half of all participants counted as 'recovered' on this caseness criterion.

As was the case for the first three recovery criteria, this caseness criterion had also been modified substantially since the publication of the trial protocol. According to the protocol, participants had to fail to meet the *standard* Oxford case definition (without the added thresholds for fatigue and physical function), as well as the CDC and London case definitions [12].[4] Again, the impact of these late modifications was not trivial. We can see from the publicly released data that only 24% of participants failed to meet the standard Oxford CFS definition at 52 weeks. Therefore, the proportion of patients who qualified as recovered on the original caseness criterion cannot have exceeded 24%. (If we were to take into account the CDC and London case definitions – for which data is not provided in the publicly available dataset – the final proportion may be even lower than 24%.) Again, it is worrying that the seemingly minor changes to this criterion resulted in a substantial increase in the number qualifying as recovered, just as was the case for the previous three criteria.

Recovery outcomes according to the original criteria

The lack of adequate justification for the changes to the original, protocol-specified recovery criteria, and the ease with which patients could qualify as 'recovered' according to the new thresholds, raises the question of how many would have recovered according to the

original definition. Applying a definition of recovery that was almost identical to the protocol-specified one, we found that recovery rates dropped from 22% to 7% in the CBT treatment group, from 22% to 4% for the GET treatment group, and from 7% to 3% for the Control (no therapy) group. The complete analysis may be found in Matthees et al. [26]. The rates for the CBT and GET groups were not significantly higher than those for the Control group (*p*-values for Fisher's exact test were 0.199 and 0.770 for the 2 pairwise comparisons, respectively).[5] Consequently, it can be seen that the changes the authors made to the definition of recovery dramatically increased the overall proportion of patients that could be counted as 'recovered', and also enabled them to report statistically significant group differences, which would not have been possible using the original definition.

Other considerations

The definition of recovery used in the PACE trial was heavily reliant on self-report measures: indeed, all four recovery criteria involved self-report. This is a problem for two reasons. First, given that the trial was non-blinded, and that the CBT and GET treatments were promoted to participants as 'highly effective', such self-report measures could be vulnerable to expectancy effects [27–30]. Second, CBT (and to a lesser extent GET) were designed to reduce 'symptom focusing', and could therefore have altered symptom-reporting behaviour (for example, self-reported fatigue) in the absence of real change. Of course, blinding is not possible in behavioural intervention trials, but the issue of bias needs to be addressed in any definitive study of treatment efficacy. One simple check is to examine whether self-reported improvements are mirrored by improvements on more objectively measured outcome variables, which appear to be less vulnerable to expectancy effects ([28], see also [31–33] for discussion).[6]

The PACE trial included several such objective measures. These included a six-minute walking test and a step fitness test [2,38]. The researchers also recorded work days lost to illness in the period prior to, and in the 12 months following, treatment allocation [39]. Data for one of these measures was recently made publicly available: distance walked on the six-minute walking test. PACE's CBT programme encouraged patients to overcome their fear of activity, and to also experiment with gradual increases in activity, so patients who recovered after CBT would be expected to score within the normal range on this test, especially after the passage of an entire year. The same would be expected for the GET programme, which set a goal of five exercise sessions a week, with walking being the most popular choice of exercise. We can estimate the normal range for this test from recently published norms based on a comparable version of this task (like PACE, it used a 10 m track length) [40]. Taking into account the PACE participants' gender composition, average age and body mass index, and adopting the formula derived from the published norms, the lower bound of normal for this test is 589 m. None of the patients in the CBT, GET or Control groups who qualified as 'recovered' achieved a walking distance that approached this lower bound, even after a whole year – irrespective of whether the protocol-specified or the revised definition of recovery is used. Unfortunately, individual patient data for the other objective measures have not yet been made available, so we cannot evaluate how the 'recovered' patients fared on these.

However, we do know that overall, treatment with CBT or GET did not have any significant effects on these other outcomes [2, 38,39].

General discussion

The PACE trial is the largest clinical trial ever conducted on patients with CFS, and was regarded by its authors as 'definitive' [41, p. 2]. Its influence on the treatment of patients in the UK and elsewhere has been considerable. There were many strengths in its design, including the large sample size (determined *a priori* using power analysis), the inclusion of a no-therapy group, the random allocation of patients to treatment arms, and the use of procedures to minimise drop-outs. Another strength was that the active groups received a substantial dose of therapy, and that standardised manuals ensured comparability of treatments across centres and therapists.

However, when it comes to claims regarding recovery rates, there were some major problems. The term 'recovery' implies more than mere improvement; it implies a return to good health. Any operational definition needs to reflect this core meaning (see [42] for discussion). In the Recovery paper, the authors described their own definition of recovery as 'conservative' [3, p. 2228 and 2231]. Contrary to this claim, we have shown here that their definition does not represent recovery in any commonsense meaning of the word - that is, full restoration of health. Of even greater concern, their definition was significantly altered from the original protocol, making 'recovery' much easier to achieve. All four of the criteria used to define recovery were substantially relaxed. In fact, on the SF-36 physical function criterion, a patient could qualify as 'recovered' with a score below that required for trial *entry*, and 13% of patients already met this criterion *when they entered the trial*.

These changes appear to have been made at a very late stage in the project, and are not mentioned in the final version of the statistical analysis plan [41]. This is a particular concern given that the changes operated to favour the researchers' predictions. There may be instances where changing a trial protocol is appropriate, but the onus is on the researchers to provide independent justification [43,44]. In no case were these changes adequately justified on independent grounds, and results based on the original protocol-specified definition were never reported. In this paper, we estimated the rates of recovery based on the original protocol-specified definition; these rates were far lower than in the trial reports. Perhaps most important, recovery rates in the CBT and GET groups were not significantly higher than those in the Control, no-therapy group. This raises significant doubts and concerns about the findings reported in the Recovery paper.

Of course, any complete definition of recovery would also include objective evidence of a return to normal functionality – for example, normal scores on tests of physical performance (especially for those treatments that explicitly incorporated activity increases), and a return to normal hours at work, school, or other former duties. It is therefore surprising that none of these aspects of recovery was considered at all in the PACE papers, even as secondary measures. Our exploratory analysis indicates that, if the definition of recovery had also required objective indicators of a return to normal function, the estimates of recovery rates would most probably have been even lower.

We conclude that the Recovery paper reported inflated and misleading estimates of recovery rates in the PACE trial. There is no evidence from this trial that individuals

can recover from CFS as a result of these treatments. Patients and their clinicians should be made aware of this information when considering whether to enrol in such programmes.

Notes

1. Specifically, patients had to score 65 or less on the Short-Form Health Survey Physical Function subscale [5]; and 6 or more out of 11 on the Chalder Fatigue Questionnaire [6].
2. The data set on which these analyses are based is publicly available from: https://sites.google.com/site/pacefoir/pace-ipd_foia-qmul-2014-f73.xlsx Readme file: https://sites.google.com/site/pacefoir/pace-ipd-readme.txt.
3. For example, 18% of patients diagnosed with CFS at the Chronic Fatigue Unit, South London and Maudsley NHS Trust had scores of 18 or less on the Chalder fatigue questionnaire (Likert scoring) before treatment for their fatigue [22].
4. The CDC definition of CFS requires at least four of: post-exertional malaise, impaired memory/concentration, unrefreshing sleep, headaches of a new kind, muscle or joint pain, tender lymph nodes or sore throat [13]. The London definition requires: exercise-induced fatigue precipitated by trivially small exertion, impaired short-term memory and loss of concentration, fluctuation of symptoms, and no comorbid mood disorder [14]. Data based on an alternative criterion were also presented. The 'clinical recovery' criterion required that in addition to no longer meeting the modified Oxford definition of CFS, the participants should also no longer have sufficient additional non-fatigue symptomatology to qualify them as meeting the CDC or London definitions. This alternative definition yielded trivially poorer recovery rates for CBT and GET.
5. The publicly available data provides only one case definition of CFS – the Oxford definition – so we counted any person who failed to meet that definition as having met this criterion for recovery. This approach may provide a slightly generous estimate of the proportion recovered according to the original criteria: it is hypothetically possible that some of the people classified as recovered may still have met one of the other case definitions, and therefore may not have been defined as recovered according to the original criterion.
6. A reviewer of this paper drew our attention to a recent study arguing that the 'placebo response' may be low in this population [34]. However, the study in question operationalized the 'placebo response' as simply the rates of study-defined improvement in a wide variety of control conditions included in CFS treatment studies, ranging from passive waitlist through to fully blinded control arms. Curiously, if we were to apply the same definition to the PACE trial, the 'placebo response' in this study would be estimated as high (45% of participants in the control group improved following treatment by the authors' definition). Of course, neither conclusion is safe; this definition of the 'placebo response' does not distinguish amongst the various factors that affect baseline improvement rates, which may include pre-trial patient expectations, but also the definition of improvement used, the rates of spontaneous remission, the degree of fluctuation in symptoms, and the presence/absence of treatment blinding [27, 35, 36, 37].

Acknowledgements

This paper represents a collaborative effort between psychology researchers and patients. In addition to those named as authors, several others also contributed important ideas, suggestions and information. This paper could not have been written without their help. We also wish to thank Keith Laws for his helpful comments, and David Tuller and Sam Carter for drawing our attention to some of the issues noted in this paper. Open Access publication has been made possible thanks to funding from The ME Association Medical Education Programme.

Disclosure statement

Tom Kindlon works in a voluntary capacity for the Irish ME/CFS Association.

Notes on contributors

Carolyn Wilshire is a University academic. She has published numerous articles in cognitive psychology, neuropsychology and more recently, health psychology. She recently co-authored a major critical review of the concept of psychological causation in medicine for the journal Perspect Psychol Sci.

Tom Kindlon is a patient with chronic fatigue syndrome (CFS), and also Assistant Chairperson of the Irish ME/CFS Association. He has published numerous articles and letters on the topic of CFS.

Alem Matthees is a patient with CFS who has previously published commentaries and analyses of the research in this field. AM was instrumental in obtaining and analysing the PACE dataset that is referred to in this article.

Simon McGrath is a patient with CFS and blogs about CFS research. The work reported represents a collaborative effort amongst patients and researchers. It illustrates the powerful contribution patients can make to health research.

References

[1] Komaroff AL, Fagioli LR, Doolittle TH, et al. Health status in patients with chronic fatigue syndrome and in general population and disease comparison groups. Am J Med. 1996;101 (3):281–290.
[2] White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. Lancet. 2011;377(9768):823–836
[3] White PD, Goldsmith K, Johnson AL, et al. Recovery from chronic fatigue syndrome after treatments given in the PACE trial. Psychol Med. 2013;43(10):2227–2235.
[4] Sharpe MC, Archard LC, Banatvala JE, et al. A report-chronic fatigue syndrome: guidelines for research. J Roy Soc Med. 1991;84(2):118.
[5] Ware JE, Sherbourne CD. The MOS 36-Item short-form health survey (SF-36). Med Care. 1992;30 (6):473–483.
[6] Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. J Psychosom Res. 1993;37(2):147–153.
[7] Couch IC. Re: PACE trial authors' reply to letter by Kindlon. BMJ. 2013;347:f5963.
[8] Burgess M, Chalder T. PACE Manual for participants: cognitive behavioural therapy. 2004. Available from: http://www.wolfson.qmul.ac.uk/images/pdfs/4.cbt-participant-manual.pdf
[9] Bavinton J, Dyer N, White PD, PACE Manual for participants: graded exercise therapy. 2004. Available from: http://www.wolfson.qmul.ac.uk/images/pdfs/6.get-participant-manual.pdf
[10] Cox D, Ludlam S, Mason L, et al. PACE Manual for participants: adaptive pacing therapy. 2004. Available from: http://www.wolfson.qmul.ac.uk/images/pdfs/2.apt-participant-manual.pdf
[11] Guy W. Clinical Global Impression scale. The ECDEU assessment manual for psychopharmacology-revised Volume DHEW Publ No ADM. 1976;76(338):218–212.
[12] White PD, Sharpe MC, Chalder T, et al. Protocol for the PACE trial: a randomised controlled trial of adaptive pacing, cognitive behaviour therapy, and graded exercise as supplements to standardised specialist medical care versus standardised specialist medical care alone for patients with the chronic fatigue syndrome/myalgic encephalomyelitis or encephalopathy. BMC Neurol. 2007;7(1):118.
[13] Reeves WC, Lloyd A, Vernon SD, et al. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. BMC Health Serv Res. 2003;3:3.

[14] Tyrrell DAJ. Report from the National Task Force on Chronic Fatigue Syndrome (CFS), Post Viral Fatigue Syndrome (PVFS) and Myalgic Encephalomyelitis (ME). 1994. Westcare: Bristol.
[15] Deale A, Husain K, Chalder T, et al. Long-term outcome of cognitive behavior therapy versus relaxation therapy for chronic fatigue syndrome: a 5-year follow-up study. Am J Psychiat. 2001;158(12):2038–2042.
[16] Knoop H, Bleijenberg G, Gielissen MFM, et al. Is a full recovery possible after cognitive behavioural therapy for chronic fatigue syndrome? Psychother Psychosom. 2007;76(3):171–176.
[17] Office of Population Censuses and Surveys. Social Survey Division. OPCS Omnibus Survey, November 1992. UK Data Service. 1997. DOI:10.5255/UKDA-SN-3660-1
[18] Bowling A, Bond M, Jenkinson C, et al. Short Form 36 (SF-36) Health Survey Questionnaire: which normative data should be used? Comparisons between the norms provided by the omnibus survey in Britain, the health survey for England and the Oxford healthy life survey. J Public Health. 1999;21(3):255–270.
[19] Picavet HSJ. Health related quality of life in multiple musculoskeletal diseases: SF-36 and EQ-5D in the DMC3 study. Ann Rheum Dis. 2004;63(6):723–729.
[20] Juenger J. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. Heart. 2002;87(3):235–241.
[21] Queen Mary University of London. PACE Trial normal range at entry. FOI Request: 2013/F42 [cited 2015 Aug 26]. Available from: https://docs.google.com/document/d/17H8AQsMKkXnKDdm0JbpAWQXWWTNEXvaKWNROBWhKW40/pub
[22] Cella M, Chalder T. Measuring fatigue in clinical and community settings. J Psychosom Res. 2010;69(1):17–22.
[23] Pawlikowska T, Chalder T, Hirsch SR, et al. Population based study of fatigue and psychological distress. BMJ. 1994;308(6931):763–766.
[24] Loge JH, Ekeberg Ø, Kaasa S. Fatigue in the general Norwegian population: normative data and associations. J Psychosom Res. 1998;45(1):53–65.
[25] Sharpe M, Goldsmith KA, Johnson AL, et al. Rehabilitative treatments for chronic fatigue syndrome: long-term follow-up from the PACE trial. Lancet Psychiatry. 2015;2(12):1067–1074.
[26] Matthees A, Kindlon T, Maryhew C, et al. A preliminary analysis of 'recovery' from chronic fatigue syndrome in the PACE trial using individual participant data. 2016 Available from: http://www.virology.ws/2016/09/21/no-recovery-in-pace-trial-new-analysis-finds
[27] Lilienfeld SO, Ritschel LA, Lynn SJ, et al. Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. Perspect Psychol Sci. 2014;9(4):355–387.
[28] Hróbjartsson A, Emanuelsson F, Thomsen ASS, et al. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. Int J Epidemiol. 2014;43(4):1272–1283.
[29] Boot WR, Simons DJ, Stothart C, et al. The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. Perspect Psychol Sci. 2013;8(4):445–454.
[30] Page MJ, Higgins JP, Clayton G, et al. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. PLoS One. 2016;11(7):e0159267.
[31] Kewley AJ. Does cognitive behavioural therapy or graded exercise therapy reduce disability in chronic fatigue syndrome patients? Objective measures are necessary. Clin Psychol Sci Prac. 2013;20:321–322. doi:10.1111/cpsp.12042.
[32] Kindlon TP. Objective measures found a lack of improvement for CBT & GET in the PACE trial: subjective improvements may simply represent response biases or placebo effects in this non-blinded trial. BMJ. 2015;350:h227. Available from: http://www.bmj.com/content/350/bmj.h227/rr-10
[33] Twisk FN. A definition of recovery in myalgic encephalomyelitis and chronic fatigue syndrome should be based upon objective measures. Qual Life Res. 2014;23(9):2417–2418.
[34] Cho HJ, Hotopf M, Wessely S. The placebo response in the treatment of chronic fatigue syndrome: a systematic review and meta-analysis. Psychosom Med. 2005;67(2):301–313.

[35] Artus M, van der Windt DA, Jordan KP, et al. Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. Rheumatology. 2010;49(12):2346.
[36] Allan LG, Siegel S. A signal detection theory analysis of the placebo effect. Eval Health Prof. 2002;25(4):410–420.
[37] Hróbjartsson A, Kaptchuk TJ, Miller FG. Placebo effect studies are susceptible to response bias and to other types of biases. J Clin Epidemiol. 2011;64(11):1223–1229.
[38] Chalder T, Goldsmith KA, White PD, et al. Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. Lancet Psychiatry. 2015;2(2):141–152.
[39] McCrone P, Sharpe M, Chalder T, et al. Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: a cost-effectiveness analysis. PLoS One. 2012;7:e40808.
[40] Beekman E, Mesters I, Gosselink R, et al. The first reference equations for the 6-minute walk distance over a 10 m course. Thorax. 2014;69(9):867.
[41] Walwyn R, Potts L, McCrone P, et al. 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. 2013;14(1):1.
[42] Adamowicz JL, Caikauskaite I, Friedberg F. Defining recovery in chronic fatigue syndrome: a critical review. Qual Life Res. 2014;23(9):2407–2416.
[43] Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011;22(11):1359–1366. doi:10.1177/0956797611417632. Epub 2011 Oct 17. PMID: 22006061. Available from: http://pss.sagepub.com/content/22/11/1359.long
[44] Evans S. When and how can endpoints be changed after initiation of a randomized clinical trial? PLoS Clin Trials. 2007;2(4):e18. PMID 17443237. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1852589