

CONFIDENTIAL

Dr Richard Horton
Editor
The Lancet
London

Dear Editor

Response to the complaint to The Lancet of March 2011

We respond to scientific questions and ethical concerns where they relate to the PACE Trial, and not to ad hominem criticisms. The criticisms of PACE trial investigators and clinicians were included in a much longer letter of complaint to the Medical Research Council in 2010, and were not upheld; the complaints being judged to be “groundless and without substance” (letters from MRC Head of Corporate Governance and Policy available if requested). Since criticisms mentioned in the introduction are repeated in individual sections, we respond within those sections as numbered within the complaint.

1. *Terminology and Classification* (pages 4 to 6)

We did not use the ICD-10 classification of myalgic encephalomyelitis (ME) because it does not describe how to diagnose the condition using standardised criteria, so cannot be used as reliable eligibility criteria. The PACE trial paper refers to chronic fatigue syndrome (CFS) which is operationally defined; it does not purport to be studying CFS/ME but CFS defined simply as a principal complaint of fatigue that is disabling, having lasted six months, with no alternative medical explanation (Oxford criteria). We also used the best available (operationalised) alternative criteria for CFS and ME (International [Centers for Disease Control] and London criteria) and determined which participants met these.

We did not ask for ethical approval for doctors to refer anyone “whose main problem is fatigue (or a synonym)” to enter the trial; they also had to be definitely or provisionally diagnosed as having CFS before being screened for eligibility. The full substantial amendment clarifying this is available on request.

2. *Fast track publication* (page 6) - It is not for us to comment on the editorial practices of a highly respected international journal.

3. *Competing interests* (pages 7-12)

Authors declared possible competing interests to the Lancet prior to acceptance and publication. All the treatment manuals have been published and are available to download on www.pacetrial.org at no cost.

Withdrawals/drop-outs and missing data

These were handled using accepted methods, which are unlikely to have introduced bias, especially given the uncommon frequency of these occurrences and their similarity across treatment arms in the trial.

4. *The trial did not study ME/CFS* (pages 12-18)

The selection of patients was for CFS operationalised using the broadest criteria (the Oxford criteria). No sensible neurologist would apply the diagnosis of CFS (or indeed ME) to patients who had “proven organic brain disease”, such as Parkinson’s disease. For the purposes of this trial ME was not regarded as a “proven organic brain disease”. In order to ensure balance between the trial arms in those participants who met alternative criteria for CFS and ME, randomisation was stratified by the International (Centers for Disease Control) criteria (which require additional symptoms) and by the London ME criteria (based on Melvin Ramsay’s original description, and which excludes co-existing “primary” psychiatric disorders [which we interpreted as any psychiatric disorder] and emphasises post-exertional fatigue). We were provided with the second revised version of the London ME criteria; we did not invent our own. We considered use of the Canadian criteria for ME but we found it impossible to operationalise them adequately for research purposes; to our knowledge they have not been used in a major research trial. We studied the results for differently defined subgroups and they were similar to those in the entire group.

Biomarkers (page 13)

Possible biomarker data were not ignored, but were irrelevant to the main aims of the trial since knowledge of their reported associations with CFS did not alter the need to do the trial. We did apply for a grant to study associations between treatment response and candidate genes, but were not funded.

Entry inducements (page 15)

At no time was anyone offered money to persuade a patient to enter the PACE trial.

5. *Failure to comply with ethics* (page 19)

All participants received a standardised CFS clinic leaflet explaining current understanding of the causes of CFS, including immune, endocrine, and viral aetiologies and possible treatments. This is and was available on the trial website (www.pacetrials.org).

Standardised specialist medical care was designed to reflect current specialist medical care. As in any clinical service some patients were seen by consultants and some by trainees under consultant supervision. All participants were told that they would be offered three outpatient sessions with their doctor during their treatment (see patient information sheet, available on the trial website).

Adaptive pacing therapy was designed in collaboration with a national ME charity, was led by a clinician who is an expert in pacing as well as activity management, and was piloted with patients to optimise its efficacy.

The aim of CBT and GET was to improve function and symptoms, with the potential for recovery, although the information about potential for recovery was not included in either the patient information sheet or the patient clinic leaflet. This is the model of the treatment, based on at least two studies that showed recovery is possible (references available).

The right for individuals to decline to participate was respected – 564 people did not consent to either research assessments or randomisation.

Most importantly, patients who declined either research assessment or randomisation were offered continuing medical care and therapies at the Royal Free hospital Fatigue service throughout the time of the trial.

The Fatigue service at the Royal Free hospital is not closed and still assesses and manages patients.

Professor White has never been “in overall charge” of this clinic; he has never worked at the Royal Free hospital in any capacity.

In the interests of transparency, we would like to add the following information. At five of the PACE trial centres, there was a pre-trial management service, which continued to offer potential participants the alternative of specific treatment for their CFS outside of the trial. At a sixth centre, there had never been, and continues not to be, a management service. Patients at this centre were provided with a diagnostic service, as was the case before the trial. Patients with CFS who were either ineligible for the trial, or declined either research assessment or randomisation were offered one of the following:

1. Referral to a community group CFS rehabilitation programme (This was funded and set up specifically to offer patients an alternative to the PACE trial.)

2. GP referral for cognitive behaviour therapy via the local clinical psychology service.

6. *Failure to “control”* (page 24)

A control condition in an experiment or trial means an appropriate comparator. Both the paper and protocol explain that this trial was designed to compare effectiveness across treatment arms, with particular comparisons being prespecified; for each comparison of two treatments, one functioned as the control.

Matching of groups (page 25)

Participants in randomised controlled trials are not matched; they are randomised so important characteristics are balanced between the groups; the paper shows that balance was achieved in this trial. The stratification errors were consequences of human error in applying complicated multiple criteria. The paper gives details of both actual stratification factors and as randomised, and clearly shows where we applied actual stratification factors and those as randomised. These errors were of little practical importance – stratification was used to ensure balance of important prognostic factors across arms. The results show that the true status in each case was balanced across the arms. Errors in assigning stratification status do not mean that the trial was poorly controlled and they did not affect the differences that were found between the trial arms.

7. *Adverse events* (pages 25-28)

The PACE trial reported five separate safety outcomes, including serious adverse events and reactions (all of which were reported individually in the web appendix). We used the definitions of adverse events, serious adverse events and serious adverse reactions of the European Union Clinical Trials Directive for medicinal products, with significantly more robust definitions and standard

operating procedures than are normally used for a trial of therapies.¹ None of the safety results gave cause for concern. We cannot comment on individuals who may or may not have been trial participants. The number of non-serious adverse events reported by patients in response to direct questioning was indeed high. We have examined adverse events in other trials of treatments for CFS, and found similar high rates (eg 89.6% of participants) when assessed in a similar way to this trial.² The important point is that the non-serious adverse events were similar in number between the groups (apart from CBT being associated with fewer than other groups) indicating that they most probably reflected the illness and not the effect of specific treatments. An independent group of three CFS specialist doctors determined which adverse events were serious and which serious adverse events were possible or probable adverse reactions to treatment. There were no definite serious adverse reactions to treatment. All adverse events were reported up until participants completed or dropped out of trial follow up.

The numbers of participants withdrawing from treatment due to worsening is plainly stated in table 4. There were 6 participants who withdrew for this reason, with no statistically significant difference between the trial arms in this outcome.

8. *Changes to entry criteria* (page 28)

A change was made in the eligibility score on the SF-36 physical function scale to enhance recruitment, as stated in the paper. This change was made by the Trial Management Group after approval by the independent Trial Steering Committee (TSC). It is common for entry criteria to be amended when they pose an unacceptable barrier to recruitment that was not fully anticipated at the start of a trial. Such a change may affect generalisability but not the validity of the results. The change to the required SF36 score did not have an effect on treatment differences because participants recruited both before and after the change were balanced across arms. The mean SF36 scores were very similar in all trial arms at baseline, as would be expected from randomisation.

The change in eligibility regarding participants' previous experience of trial treatments was because we found it hard to ascertain the nature and content of previous treatments provided at non-PACE clinics; again stated in the paper, and approved by the TSC.

9. *Outcome results* (page 29)

Statistical significance and confidence intervals - We had difficulty understanding many of the comments about standard deviations, error bars and confidence intervals. We are confident that the analyses were properly presented in the paper. Standard deviations would be expected to increase with time relative to baseline; the groups were less variable at baseline because of the entry criteria. Larger standard deviations decrease the likelihood of finding a difference between the groups, so the larger amount of variation would have made it more difficult to find differences between the

groups at 52 weeks. It is true that, with large numbers, small differences can be found to be statistically significant. However, we defined clinically useful differences before the analysis of outcomes and 7 out of 8 treatment differences in the primary outcomes were in excess of these. The description “almost always exceeded” was used because all except the comparison of CBT and SMC exceeded the clinically useful difference. It is incorrect to describe confidence intervals as showing poor confidence in results. It is difficult to argue both that there were so many people that we found statistically, but not clinically, significant results, but that our confidence intervals are wide – confidence intervals will become small when there are large amounts of data.

Overlap in confidence intervals at 24 weeks is not relevant as the pre-specified primary end-point was 52 weeks and our primary analysis used data from all follow up times. Analyses were guided by a pre-specified analysis plan, which we plan to publish. We report both unadjusted results and adjusted results in our models. Figure 2 shows unadjusted differences. The final results, shown in figure 3, are adjusted for baseline value of the outcome, amongst other things. The final results are not directly comparable to a simple comparison because they incorporate outcomes from all time points, adjust for stratification factors and baseline values (recommended approaches), and for clustering within therapists.

Clinically useful difference (page 30) - The figures of 7.4 and 6.9 come from unadjusted figures – the adjusted difference between GET and SMC was 9.4, not 6.9, which exceeds the prespecified clinically useful difference. Comparisons with APT were pre-specified and were not introduced simply because the APT group had a lower mean. In addition, the comparisons were made when the study group were blinded to the trial arms, so these numbers were obtained before we knew which group was which.

Normal ranges - The primary analysis compared the mean differences in the primary outcome scores across treatment arms, which are in the paper. The normal range analysis was plainly stated as post hoc, given in response to a reviewer’s request. We give the results of the proportions with both primary outcomes within *normal ranges*, described a priori, using population derived anchors.

SF-36 scores (page 31) - The definition of a “normal range” for the SF36 in the paper is different from that given in the protocol for “recovery”. Firstly, being within a “normal range” is not necessarily the same as being “recovered”. Secondly, the *normal range* we gave in the paper was directly taken from a study of the most representative sample of the adult population of England (mean - 1 SD = 84 – 24 = 60). The threshold SF36 score given in the protocol for *recovery* (85) was an estimated mean (without a standard deviation) derived from several population studies. We are planning to publish a paper comparing proportions meeting various criteria for *recovery or remission*, so more results pertinent to this concern will be available in the future. We did however make a descriptive error in referring to the sample we referred to in the paper as a “UK working age population”, whereas it

should have read “English adult population”, and have made this clear in our response to correspondence.

Fatigue measure (page 32) - We explained in the paper why we changed our scoring of the fatigue measure from bimodal to Likert scoring, in order to improve sensitivity to change to better test our hypotheses, and did this before outcome data were examined. This was included in our pre-specified analysis plan approved by the TSC.

Walking test (page 33) - The interpretation of the walking test results seems to be one of scientific debate. Statistical testing takes into account variability. The GET group were still significantly different than the SMC and APT groups despite a large amount of variability in the measure. In addition, one cannot focus solely on absolute metres walked for individual trial arms as these may or may not be influenced by treatment. The valid comparisons are between trial arms. We did not ask participants to undertake a practice walking test for the reason mentioned in the complaint; post-exertional fatigue being a characteristic feature of CFS.

10. *Data not reported (page 35)*

Not all the measures listed in the protocol are described in the paper. That is because it was impossible to present all the data collected in a single paper of limited words. The measures reported in the main paper were specified before analysis. Future papers that will include these additional measures are in preparation including reports of economic outcomes, different definitions of recovery and remission, mediators and moderators, and long-term follow up.

We used patients’ self ratings to measure outcome. Given that the illness is defined by patient reports, we argue that patient reports are the most important outcomes.

Actigraphy was dropped as an outcome measure before the trial started, not afterwards. This measure was dropped mainly in response to the MRC Board and reviewers of the grant suggesting that the outcome load was excessive for participants. We agreed that asking participants to wear an actometer around their ankle for a week might increase the number trial drop-outs at our primary end-point. This change, like all others made, was approved by the TSC.

11. *Overview (page 38)* - On one hand, the complaint suggests we do not present sufficient results, whereas here the complaint is that the results were too complex. We believe the complexity, such as it was, was at an appropriate level for the research questions we sought to answer.

12. *Science media centre* (pages 38-40) - This appears to be a complaint about the Science Media Centre.

13. *Summary* (page 40) - We do not comment on these complaints which extend far beyond the PACE trial.

Yours sincerely,

PD White, KA Goldsmith, AL Johnson, R Walwyn, HL Baber, T Chalder, M Sharpe, on behalf of all the co-authors

References

1. White PD, Sharpe MC, Chalder T, DeCesare JC, Walwyn R; on behalf of the PACE trial group. 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. *BioMed Cent Neurol* 2007; **7**: 6.

2. Blacker CVR, Greenwood DT, Wesnes KA, et al. Effect of galantamine hydrobromide in chronic fatigue syndrome: A randomized controlled trial. *JAMA* 2004; **292**: 1195-1204.