

Professor Malcolm Hooper's further concerns about the PACE Trial article published in The Lancet

24th June 2011

Executive Summary

Scrutiny of the criteria used to determine which participants were “within the normal range” on the two primary outcomes in the PACE Trial -- physical function and fatigue -- reveals a manifest contradiction in the report published in The Lancet (PD White et al. Lancet 2011:377:823-836).

Ratings that would qualify a potential participant as sufficiently impaired to enter the trial were considered “within the normal range” when recorded on completion of the trial.

There is thus discordance between the designated entry criteria and the benchmarks of “the normal range” in assessing outcomes at the end of the Trial in respect of both physical function and fatigue.

It cannot be acceptable to describe a PACE Trial participant at the end of the trial as having attained levels of physical function and fatigue “within the normal range” and to consider the same participant sufficiently disabled and symptomatic, as judged by the same recorded levels of physical function and fatigue, to have qualified for entry into the PACE Trial in the first place.

This situation has arisen as a result of numerous changes and re-calculations by the Principal Investigators (PIs) in the relevant benchmarks, changes in the PIs' cited reference material as to what constitutes “the normal range”, and the PIs' use of inappropriate comparison groups.

It should be noted that the analysis refers to outcomes “within the normal range”. This is not necessarily the same as “normal”. It is a statistical concept, defined as the mean plus or minus one standard deviation from the mean. It may or may not equate well to what is typical in the population. In the case of physical functioning, the threshold of “the normal range” is far from what is “normal”. Due cognisance of this should have been taken in interpreting outcomes on physical function.

However, even with these factors mitigating in favour of positive reporting, only 30% of the CBT participants and 28% of the GET participants recorded outcomes “within the normal range” in respect of physical functioning and fatigue on conclusion of the PACE Trial.

The Trial Protocol sets out two “primary efficacy measures”. These consist of specific parameters delineating what is to be considered “a positive outcome” on physical functioning and fatigue, respectively. In combination, these were to have been used to identify “overall improvers”, but this analysis has been dropped by the PIs.

Professor Hooper is of the view that the PACE Trial fails on a fundamental aspect of clinical research in that there is no attempt to apply the pre-determined primary efficacy measures to the outcome data, and furthermore the benchmarks used to judge suitability for recruitment and outcomes are patently contradictory.

Together with others who have expressed concerns, Professor Hooper continues to believe that the need for an independent statistical re-evaluation of the raw data is overwhelming as, without such an independent assessment, doubts over the veracity of the claims made by Professor White et al cannot be resolved.

Replying to Professor Hooper’s complaint, Professor White et al state: *“The PACE trial paper...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)”*. If The Lancet accepts this, Professor Hooper asks that it publish an immediate and unequivocal clarification about this key issue, since during the 8-year life of the PACE Trial, virtually all the documents refer to “CFS/ME” and the published results are being applied to people with the distinct nosological disorder myalgic encephalomyelitis (ME).

Such clarification would serve to protect people with ME from implicit or explicit pressure to engage in exercise programmes (continuance of welfare benefits as well as medical support and basic civility being contingent upon compliance). ME patients have consistently reported that even graded exercise results in deterioration that is often long-lasting and severe.

Introduction

On 28th March 2011 Professor Hooper submitted his detailed concerns in the document “REPORT: COMPLAINT TO THE RELEVANT EXECUTIVE EDITOR OF THE LANCET ABOUT THE PACE TRIAL ARTICLES PUBLISHED BY THE LANCET” (<http://www.meactionuk.org.uk/COMPLAINT-to-Lancet-re-PACE.htm>).

Professor Peter White, the lead author of the PACE Trial article, was invited by The Lancet’s senior editorial staff to respond to it, which he did in an undated letter sent to Richard Horton, editor-in-chief of The Lancet (<http://www.meactionuk.org.uk/whitereply.htm>), as a result of which the complaint was rejected in its entirety by The Lancet’s senior editorial staff.

On 28th May 2011 Professor Hooper therefore responded to the failure of Professor White to address the important issues raised (<http://www.meactionuk.org.uk/Comments-on-PDW-letter-re-PACE.htm>).

On 3rd June 2011, Zoe Mullan, senior editor at The Lancet, indicated to a correspondent unconnected with Professor Hooper that if he had further concerns, she would welcome his contacting her about them. Having been made aware of this, he agreed to do so.

A specific and major concern is the focus of this present document, which relates to the PIs’ PACE Trial entry criteria and criteria for assessing outcomes on physical function and fatigue. The result is an overlap between the benchmarks of “the normal range” on these measures as applied to PACE participants’ outcomes and the benchmarks (on these same measures) denoting impairment at the outset of the Trial. Furthermore, the PIs have failed to report on the pre-defined criteria delineating “a positive outcome” that are specified in the Trial Protocol.

Professor Hooper cannot comprehend how The Lancet editors can accept such non-science as objective and reliable evidence of the success of the PACE Trial and he fails to understand how senior Lancet editors could be “*fully satisfied*” by the PIs’ illogical conclusion that the same requirement for admission to the trial has been judged by them to denote attainment within “the normal range” at the end of the trial, a situation that requires correction or clarification as a matter of urgency.

He believes that, as a UK custodian of valid science in medicine, The Lancet failed to recognise the very serious flaws in the PACE study itself and in the published article reporting the supposedly successful outcome.

On 18th April 2011 in his broadcast about the PACE Trial on Australian ABC Radio National, Richard Horton was disparaging about criticisms of the article, asserting that the PACE Trial was a well-designed and well-executed study; he also said: *“We will invite the critics to submit versions of their criticism for publication and we will try as best as we can to conduct a reasonable scientific debate about this paper. This will be a test I think of this particular section of the patient community to engage in a proper scientific discussion”* (<http://www.abc.net.au/rn/healthreport/stories/2011/3192571.htm>).

Professor Hooper asks that The Lancet honour Richard Horton’s call and that, as part of that process, this present submission be afforded due scrutiny by The Lancet’s independent statisticians.

For the avoidance of doubt, relevant extracts from the PACE Trial protocol and the published article are here provided:

PACE TRIAL PROTOCOL (extract)

“10.1 Primary outcome measures

10.1.1 Primary efficacy measures

Since we are interested in changes in both symptoms and disability we have chosen to designate both the symptoms of fatigue and physical function as primary outcomes. This is because it is possible that a specific treatment may relieve symptoms without reducing disability, or vice versa. Both these measures will be self-rated.

The 11 item Chalder Fatigue Questionnaire measures the severity of symptomatic fatigue, and has been the most frequently used measure of fatigue in most previous trials of these interventions. We will use the 0,0,1,1 item scores to allow a possible score of between 0 and 11. A positive outcome will be a 50 % reduction in fatigue score, or a score of 3 or less, this threshold having been previously shown to indicate normal fatigue.

The SF-36 physical function sub-scale measures physical function, and has often been used as a primary outcome measure in trials of CBT and GET. We will count a score of 75 (out of a maximum of

100) or more, or a 50 % increase from baseline in SF-36 sub-scale score as a positive outcome. A score of 70 is about one standard deviation below the mean score (about 85, depending on the study) for the UK adult population. Those participants who improve in both outcome measures will be regarded as overall improvers.

10.2 Secondary outcome measures

10.2.1 Secondary efficacy measures

LANCET ARTICLE REPORTING THE PACE TRIAL FINDINGS (extracts)

Study Design & Participants

“Other eligibility criteria consisted of a bimodal score of 6 of 11 or more on the Chalder fatigue questionnaire [ref. 15] and a score of 60 of 100 or less on the short form-36 physical function subscale. [ref. 16] 11 months after the trial began, this requirement was changed from a score of 60 to a score of 65 to increase recruitment”. (Professor White has now admitted in his letter to The Lancet that this “may affect generalisability”).

Outcomes

“The two participant-rated primary outcome measures were the Chalder fatigue questionnaire (Likert scoring 0,1, 2, 3; range 0–33; lowest score is least fatigue) [ref. 15] and the short form-36 physical function subscale (version 2; range 0–100; highest score is best function) [ref. 16]. Before outcome data were examined, we changed the original bimodal scoring of the Chalder fatigue questionnaire (range 0–11) to Likert scoring to more sensitively test our hypotheses of effectiveness”.

Statistical Analysis

“In another post-hoc analysis, we compared the proportions of participants who had scores of both primary outcomes within the normal range at 52 weeks. This range was defined as less than the mean plus 1 SD scores of adult attendees to UK general practice of 14.2 (+4.6) for fatigue (score of 18 or less) and equal to or above the mean minus 1 SD scores of the UK working age population of 84 (–24) for physical function (score of 60 or more) [refs. 32,33]”.

Results

25 (16%) of 153 participants in the APT group were within normal ranges for both primary outcomes at 52 weeks, compared with 44 (30%) of 148 participants for CBT, 43 (28%) of 154 participants for GET, and 22 (15%) of 152 participants for SMC”.

Failure to Report on “Positive Outcomes”

The PACE Trial Protocol sets out the criteria to be used to delineate “a positive outcome”. These criteria apply to the scores achieved on the two primary outcomes, physical function and fatigue, respectively (see box above).

Analysis of these “primary efficacy measures” (there were no others) does not appear in the article published in The Lancet.

This omission may be viewed in the context of the prior reporting of disappointing results from the PACE Trial’s sibling, the MRC-funded FINE (Fatigue Intervention by Nurses Evaluation) Trial (AJ Wearden et al. BMJ 2010; 340; c1777). It is notable that the criteria specified in the PACE Trial Protocol to denote “a positive outcome” are identical to the criteria that were used to gauge outcomes in the FINE Trial, with the exception that a threshold of 70 (as opposed to 75) was used on physical functioning in the FINE Trial.

Given the close links between the PACE and FINE Trials, it is inconceivable that the PACE Trial Investigators would have been unaware that criteria differing little from their own pre-designated “positive outcome” measures in the PACE Trial had produced disappointing results when applied to the FINE data.

(The poor FINE Trial results may also have influenced the PACE Trial PIs' decision to change the method approved in respect of assessing outcomes on fatigue as recorded via the Chalder Fatigue Questionnaire, thus departing from the Trial Protocol – see below).

No other measure of “a positive outcome” is presented in The Lancet article. Instead, the analysis focuses on inter-group differences in scores recorded in respect of physical function and fatigue. These are described as “primary outcome measures”. However, without having a (pre)specified parameter on the relevant variables as to what is to be deemed a “primary outcome measure”, this description is meaningless.

The Lancet article does, however, present a secondary analysis of outcomes in respect of these variables, assessed against “the normal range” (see box above). It is this analysis that contains an inherent contradiction ie. it was possible for participants to be deemed to have attained levels of physical function and fatigue “within the normal range” when they had actually deteriorated on these parameters over the course of the PACE Trial.

Assessing Physical Function

Physical function was assessed using the Physical Function subscale of the Short Form 36 Health Survey Questionnaire (usually abbreviated to SF-36), with higher scores indicating better function (McHorney CA et al; Med Care 1993;31:247-263). The raw score range is over a 20 point range. However for purposes of analysis this is converted to a scale of 0-100, rising in increments of 5.

What is “Normal” Physical Function?

-

The situation whereby it was possible for a person to deteriorate on this measure over the course of the PACE Trial yet still be deemed to have attained physical function “within the normal range” on completion of the trial arose in part in consequence of the PIs' various revisions of the relevant benchmarks in respect of recruitment criteria and the assessment of outcomes.

The problem also resides in the standard practice of using the mean plus and minus one standard deviation (SD) from the mean to denote the “range of normal” on a variable. When data is “normally

distributed” (in statistical terms) around a mean, the concept relates well to what is the norm. In respect of physical function in general, and SF-36 scores in particular, data is skewed. In these circumstances, there is a difference between what is normal in the sense of being most frequently found, and “the normal range”. This should have been flagged up by the PIs in interpreting the reported outcomes on physical function.

The two problems are delineated below.

The Threshold of the “Range of Normal” as a Benchmark on Physical Function

The paper referenced in respect of the threshold of “the normal range” that has been applied in the PACE Trial (Bowling A et al; J Publ Health Med 1999:21:255-270) reviews normative data from a range of sources and concludes: *“These results confirm the highly skewed nature of the distributions (see Fig 1), which is a problematic feature of all health status scales.”*

This “problematic” feature is that the data are highly skewed towards the high end of the scale. Indeed, scrutiny of the relevant histogram in Fig 1 of the Bowling et al. paper suggests that there are more people who score the maximum 100 on the SF-36 physical functioning scale than the combined total of people who score anything other than 100.

In such circumstances, applying a benchmark of the mean minus one standard deviation to general population data on the SF-36 physical function subscale to denote the threshold of “the normal range”, while technically correct, does not equate to what would be understood as “normal” in respect of physical functioning in the general population.

Because of the skewed nature of distributions on health status scales, the use of a “reference range” may be more appropriate for comparative purposes. This describes the variations of a measurement or value in healthy individuals and is a basis for a [physician](#) or other [health professional](#) to interpret a set of results for a particular patient. The standard definition of a reference range originates in **what is most prevalent** in a control group taken from the population.

The PIs’ Definitions and Re-definitions of “Normal” and “The Range of Normal” on Physical Function

In the PACE Trial documents obtained under the FOIA it is recorded that the PIs' intention was to set the recruitment ceiling at a maximum of 70 and to define normal physical function as an SF-36 score of at least 75.

In his application dated 12th September 2002 to the West Midlands Multicentre Ethics Committee (MREC), Professor White described the derivation of this threshold of "normal" as follows: *"We will count a score of 75 [out of a maximum of 100] or more as indicating normal function, this score being one standard deviation below the mean score [90] for the UK working age population"*, citing Jenkinson C et al. Short form 36 (SF-36) Health Survey questionnaire: normative data from a large random sample of working age adults; BMJ:1993:306:1437-1440.

It should be noted that the comparative data related to the UK working age population.

A ceiling of 70 in respect of recruitment and a threshold of 75 to denote "normal function" on the SF-36 physical function subscale was accordingly presented in the PACE Trial Identifier. As the SF-36 Physical Function subscale proceeds *in increments of 5* this meant that there was the narrowest of margins between the ceiling on physical function in respect of entry to PACE, and the threshold of "normal" on conclusion.

The proposed threshold for entry was discussed at the Trial Steering Committee held on 22nd April 2004 (at which Professor White was present) and those discussions are minuted as follows: *"7. The outcome measures were discussed. It was noted that there may need to be an adjustment of the threshold needed for entry **to ensure improvements were more than trivial** (emphasis added). For instance a participant with a Chalder score of 4 would enter the trial and be judged improved with an outcome score of 3. The TSC (Trial Steering Committee) suggested one solution would be that the entry criteria for the Chalder scale score should be 6 or above, so that a 50% reduction would be consistent with an outcome score of 3. **A similar adjustment should be made for the SF-36 physical function subscale**"* (emphasis added).

Consequently, when the PACE Trial began (the first participant having been randomised on 18th March 2005), the ceiling in respect of SF-36 at entry was a score of 60.

In the Trial Protocol, an SF-36 threshold of 75 remains in respect of assessment of outcomes and plays a part in the identification of "a positive outcome": *"We will count a score of 75 (out of a*

maximum of 100) or more, or a 50% increase from baseline in SF-36 subscale score as a positive outcome”.

This applies in both the full 226 page final version (unpublished by the PIs but obtained under the FOIA and available at <http://www.meactionuk.org.uk/FULL-Protocol-SEARCHABLE-version.pdf>) and the shortened 20-page version of the Protocol that was published in 2007 (www.biomedcentral.com/1471-2377/7/6 -- which was not peer-reviewed by the journal because it had already received ethical and funding approval by the time it was submitted, the Editor commenting: *“We strongly advise readers to contact the authors or compare with any published result(s) articles to ensure that no deviations from the protocol occurred during the study”*).

Curiously, although the SF-36 score threshold remains at 75, the threshold of “normal” cited by the PIs in the PACE Trial protocol has been lowered to 70: *“A score of 70 is about one standard deviation below the mean score (about 85, depending on the study) for the UK adult population”*. The PIs cite two references in support (Jenkinson C et al; BMJ 1993;306:1437-1440 – ie. the same reference as in the application to the MREC -- and Bowling A et al; J Publ Health Med 1999;21:255-270). It is notable that the normative group identified now relates to **the adult population as a whole** (ie. it includes elderly people, whereas the normative group previously cited was **the working age population**).

Because of continued problems attaining recruitment targets, on 9th February 2006 Professor White wrote to Mrs Anne McCullough, Administrator at the West Midlands MREC, requesting a substantial amendment to the trial's entry criteria as **he wished to raise the SF-36 threshold required for inclusion criteria for the trial from 60 to 65**. He stated: *“Increasing the threshold [from 60 to 65] will improve generalisation.... The TMG (Trial Management Group) and TSC (Trial Steering Committee) believe this will also make a significant impact on recruitment”*.

(It is notable that in this request for a substantive amendment dated 9th February 2006, Professor White assured the MREC that *“Increasing the threshold [from 60 to 65] will improve generalisation”* but in his response to Professor Hooper’s complaint on this point, Professor White admitted that: *“Such a change may affect the generalisability...of the results”*. The context of this statement is such that a stricture rather than an improvement is implicit. In effect, this change meant that, in the midst of the recruitment period, the pool of potential candidates was increased by relaxing the entry criteria to allow people with better physical capacity to take part).

Furthermore, it narrowed the gap between how physically impaired a person had to be in order to be recruited, and how well they had to function to be deemed to have a positive outcome, leading to the following approach to the MREC from Professor White: ***“This would mean the entry criterion on this measure was only 5 points less than the categorical positive outcome of 70 on this scale. We therefore propose an increase of the categorical positive outcome from 70 to 75, reasserting a ten point score gap between entry criterion and positive outcome”*** (emphasis added).

Given that the threshold of **positive outcome** is stated as 75 in the Trial Protocol, this is baffling. (It was the **threshold of normal function in the population** that is cited as 70.) Unless there is an as-yet unidentified document reducing the SF-36 threshold denoting a positive outcome from 75 to 70, it would appear that Professor White was confused as to the existing benchmark.

In any event, the gap proposed was ten points – representing a minimum increment of two stages on the SF-36 scale.

Professor White further assured the MREC that this change would bring the PACE Trial into line with its “sister study”, the FINE Trial, and that it would not affect the analysis of the trial data: *“The other advantage of changing to 75 is that it would bring the PACE trial into line with the FINE trial, an MRC funded trial for CFS/ME and the sister study to PACE. This small change is unlikely to influence power calculations or analysis”*.

The presentation of trial data in The Lancet demonstrates that Professor White did not observe the assurances he provided to the ethics committee.

In The Lancet article reporting the results of the PACE Trial, the primary efficacy measures as set out in the trial protocol have been abandoned altogether. There is no reference to any measure of “a positive outcome”.

However, a “post hoc” analysis is presented, which entails comparing PACE participants’ outcomes against a threshold of “the normal range” in respect of physical function. Defined as the mean minus one standard deviation in respect of a normative population and having been specified as 75 in the application to the MREC and reduced to 70 in the protocol, in the analysis published in The Lancet the threshold of “the normal range” is further reduced to an SF-36 score of 60.

This was based on *“the mean minus 1 SD scores of the UK working age population of 84 (–24) for physical function”*.

One reference is cited in respect of this threshold, this being the Bowling et al paper that was one of two cited at the PACE Trial Protocol stage. That paper reviews normative data from a range of sources, none of which appears to provide the figures cited (see Table 4: *“Comparison of SF-36 dimension norms in Britain”* in the Bowling et al paper).

Following Professor Hooper's complaint, Professor White responded in his letter to The Lancet: *"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' ".* Such a comparator is inappropriate because, by definition, the English adult population includes elderly people. The appropriate comparison would be with the SF-36 physical function scores for **age-matched healthy people**. However this would have raised the threshold of the normal range to a higher level, thus making it more difficult – if not impossible – for the PIs to claim even moderate success for the PACE Trial.

Furthermore, the data analysis published in The Lancet is at odds with one of the reasons given by Professor White to the MREC for previously setting the "categorical positive outcome" at 75, namely to put PACE into line with the FINE Trial.

It is notable that, when the FINE Trial results were reported (in the spring of 2010), only 17 of the 81 participants assessed had met the relevant parameter in respect of physical function at the primary outcome point -- a score of at least 75 or an improvement of 50% from baseline.

Remarkably, in view of the abstruse complexity of much of the analysis presented in The Lancet article, the PACE Trial PIs have stated: *"Changes to the original published protocol were made to improve either recruitment or interpretability"* (The Lancet: doi:10.1016/S0140-6736(11)60651-X).

In summary, since it was possible to score 65 on the SF-36 and still be recruited to the PACE Trial, setting the threshold of "the normal range" at 60 on completion meant that there was a **negative** five point score gap, meaning that a participant could actually **deteriorate** during the course of the trial and leave the trial **more disabled** than before treatment, but still fall within the PIs' new definition of "normal" (ie. attainment of "normality" was set **lower** than the entry criteria, which by any standards is illogical).

Assessing "Normal" Fatigue

In the PACE Trial, fatigue was assessed using the Chalder Fatigue Questionnaire or CFQ (Chalder T, Wessely S et al; J Psychosom Res 1993:37:147-153).

The Chalder Fatigue Questionnaire comprises eleven questions. Respondents are asked to indicate their situation in respect of each of these on a four-point scale: “less than usual”; “no more than usual”; “more than usual”; “much more than usual”.

The fatigue score is the sum total of the scores obtained in respect of the eleven items in the Chalder Fatigue Questionnaire. The higher the score, the greater the impact of fatigue. However, there are two methods of scoring responses.

Change by the PIs in Method of Scoring Outcomes

One method of producing a fatigue score involves scoring these respective responses on a scale from 0 to 3 and summing the total. This method (known as Likert scoring, which has a possible range of 0 - 33) was used to assess outcomes on fatigue.

However, for the purposes of screening for entry to PACE, a different method of scoring the responses was adopted. Known as bimodal analysis, this entails placing each response into one of two categories: any item rated “less than usual” or “no more than usual” is allocated a score of 0; any item rated “more than usual” or “much more than usual” is allocated a score of 1. The possible range is therefore 0 -11.

It is notable that the original proposal – as set out in the MREC application, the Trial Identifier, and the Trial Protocol - was to analyse results bimodally. This was to feed into one of two “primary efficacy measures”: “A positive outcome will be a 50 % reduction in fatigue score, or a score of 3 or less, **this threshold having been previously shown to indicate normal fatigue** (Trial Protocol, citing Chalder T, Berelowitz G, Hirsch S, Pawlikowska T, Wallace P, Wessely S and Wright D: Development of a fatigue scale. J Psychosom Res 1993, 37:147-153.)

The rationale provided for the change to Likert scoring in the consideration of outcomes in The Lancet article was: “Before outcome data were examined, we changed the original bimodal scoring of the Chalder fatigue questionnaire (range 0-11) to Likert scoring to more sensitively test our hypothesis of effectiveness”.

However, one consequence of adopting a Likert approach to processing responses is that it becomes easier to demonstrate differences between the groups when such differences are relatively small.

This had been demonstrated in respect of the fatigue outcome data in the FINE Trial: analysed using bimodal scoring as set out in the FINE Trial protocol, there was no statistically significant improvement in fatigue between the FINE interventions and the “treatment as usual” control group at the primary outcome point (Wearden AJ et al; BMJ 2010:340:c1777).

However, following publication of those results, the FINE Trial Investigator (Dr Alison Wearden PhD, an observer on the PACE Trial Steering Committee) reappraised the FINE Trial data according to Likert scoring and produced a “*clinically modest, but statistically significant effect...at both outcome points*” (http://www.bmj.com/cgi/eletters/340/apr22_3/c1777#236235), a fact of which the PACE Trial PIs would have been well aware.

What is “Normal” Fatigue? What is “Abnormal” Fatigue?

As with the physical function scores, there is an overlap between the level of fatigue deemed sufficiently significant to qualify a person to participate in the PACE Trial and the level of fatigue deemed to denote a positive outcome.

This means that identical responses on the Chalder Fatigue Questionnaire could qualify a person as sufficiently “fatigued” for entry to the PACE trial and later allow them to be deemed to have attained “normality” in terms of their level of fatigue at the outcomes assessment stage.

This absurdity is somewhat opaque owing to the use of a different method of processing responses to the Chalder Fatigue Questionnaire at entry stage (bimodal) and outcomes assessment (Likert) stage (see above). Nonetheless it is possible to demonstrate a manifest contradiction and flaws in the definitions used.

Qualifying threshold re: Fatigue for Entry to the PACE Trial

As with physical function, the criterion that was used to recruit participants to PACE in respect of fatigue differed from what was originally specified.

In his application dated 12th September 2002 to the MREC, Professor White stated: *“We will operationalise CFS in terms of fatigue severity ... as follows: a Chalder fatigue score of four or more.”* He also referred to: *“a score of 4 having been previously shown to indicate abnormal fatigue.”*

The PACE Trial Identifier repeated the requirement for a fatigue score of 4 or more to indicate caseness at entry.

However, following discussion at the Trial Steering Committee (on 22nd April 2004) this was revised upward to 6 in order to allow for a more appropriate gap to appear between the required level of fatigue on entry and the threshold of an outcome denoting improvement (at that point, a score of 3 or less, or a 50% improvement from baseline -- however, the consideration of this “primary efficacy measure” was later dropped). PACE participants were recruited on this basis.

Ceiling of “Normal” Fatigue on **Completion** of the PACE Trial

The commitments given before the PACE Trial interventions began were consistently for a ceiling score of 3 on the Chalder Fatigue Questionnaire, rated bimodally, to represent “normal” fatigue on completion of the trial. The rationale for treating bimodally rated scores of 4 and above as representing abnormal levels of fatigue is repeatedly cited as Chalder T et al. J Psychosom Res 1993;37:147-153. That paper is the work of the lead author of the Chalder Fatigue Questionnaire, PACE Trial Principal Investigator Professor Trudie Chalder, a co-author being the Director of the PACE Trial Clinical Unit and member of the Trial Management Group Professor Simon Wessley. For example:

In his application to the MREC, under the heading “What is the primary end point?” Professor White stated: *“We will use the 0,0,1,1 item scores to allow a categorical threshold measure of “abnormal” fatigue with a score of 4 having been previously shown to indicate abnormal fatigue.”*

In the PACE Trial Identifier, under the heading “3.9 What are the proposed outcome measures? Primary efficacy measures” Professor White stated: *“We will use the 0,0,1,1 item scores to allow a categorical threshold measure of “abnormal” fatigue with a score of 4 having been previously shown to indicate abnormal fatigue [ref 23]”* (Chalder T et al. J Psychosom Res 1993; 37: 147-153.)

In the PACE Trial protocol, under the heading “10.1 Primary outcome measures; 10.1.1 Primary efficacy measures” Professor White stated: *A positive outcome will be a 50 % reduction in fatigue score, or a score of 3 or less, this threshold having been previously shown to indicate normal fatigue*” (Chalder T et al. J Psychosom Res 1993, 37:147-153).

However, in The Lancet article reporting the results of the PACE Trial, when “normal” levels of fatigue were judged on completion of the trial, the analysis conducted related to: *“the proportions of participants who had scores of both primary outcomes within the normal range at 52 weeks. This range was defined as less than the mean plus 1 SD scores of **adult attendees to UK general practice** of 14.2 (+4.6) for fatigue (**score of 18 or less**) ... ”*: (32: Cella M, Chalder T et al: J Psychosom Res 2010:69:17-22).

A Likert score of 18 can translate to a bimodal score of between 4 and 9, depending on the specific responses that combine to produce the Likert score. According to the PIs, a bimodal score of 4 or more indicates abnormal fatigue (see above). Hence a Likert score of 18 always represents a state of abnormal fatigue.

In order to allow for a sufficient gap between the positive outcome criterion then proposed – a bimodal score of 3 or less -- the threshold of fatigue at entry to the PACE Trial had been set at 6. However, it is possible to record responses producing a Likert score of 18 (ie. the ceiling of “the range of normal” fatigue on conclusion of the PACE Trial) which translates to bimodal scores of 6, 7, 8, and 9.

Either the threshold of “normal” denoting a positive outcome should have been lower than the measure used in the analysis in The Lancet (Likert 18), and/or the threshold of caseness at recruitment (bimodal score of 6) should have been higher.

The net result of the analysis conducted is that identical responses could both qualify a person as sufficiently “fatigued” for entry to the PACE trial and at completion of the trial allow them to be deemed to be within “the range of normal” in terms of their level of fatigue.

What’s more, as with physical function, it would be possible for a person to record poorer responses in respect of fatigue on completion of the trial than at the outset, yet still be deemed by the PIs to be within “the range of normal” on this subjective primary outcome.

Further Issues Regarding the Assessment of Fatigue in the PACE Trial

Several further points are relevant in this regard.

First, the cited reference for the benchmark chosen to assess PACE outcomes, co-authored by the PACE Trial Principal Investigator Trudie Chalder, also provides bimodal scores for the same population: “*community sample: mean fatigue 3.27 (S.D. 3.21)*”. This places the ceiling at which a person can have fatigue and still be considered within the normal range at a bimodal score of 6.

This is inconsistent with the PACE Trial literature, which repeatedly refers to “*a score of 4 having been previously shown to indicate abnormal fatigue*” (see above), citing a paper lead-authored by Trudie Chalder and co-authored by Director of the PACE Trial Clinical Unit and member of the Trial Management Group Professor Simon Wessely.

Secondly, the Lancet article states that the benchmark employed was derived from fatigue scores from “*adult attendees to UK general practice*”. That study was part of a long-term longitudinal scrutiny of a cohort group but, notably, “*only completed data from those who went to see their general practitioner **the following year**... were used in this study*” (emphasis added). The Chalder Fatigue Questionnaires therefore related to the year prior to the selected cohort becoming “attendees to UK general practice”.

This is a curious and convoluted selection of a comparison population from which to derive normative data. Moreover the nature of this comparison group is by no means obvious from the PIs’ description (“*adult attendees to general practice*”) that is set out in The Lancet article on the PACE Trial results. Again, Trudie Chalder was an author of both papers.

Finally, it is possible that fatigue - unlike physical function - is “normally distributed” in the general population, as asserted by (then Dr) Simon Wessely. Referring to the findings of a study based on data from over 15,000 people (Pawlikowska T, Chalder T, Wessely S et al. BMJ 1994:308:743-746), he stated: “*18% had experienced substantial fatigue for six months or longer. **Fatigue, however, was ‘normally’ distributed...***” (Epidemiology of CFS: in “A Research Portfolio on Chronic Fatigue”; edited by Robin Fox for The Linbury Trust; RSM Press 1998).

If fatigue is normally distributed, then the method of equating “the range of normal” (a statistical concept) with “normality” (what is widespread in the population), as in reporting the PACE Trial

results, is acceptable. However, this would differentiate attempts to measure fatigue (ie. by using the Chalder Fatigue Questionnaire) from ***“all health status scales”*** in respect of which distributions are ***“highly skewed”*** (Bowling A et al. J Publ Health Med 1999:21:255-270) as referenced in the PACE Trial documentation.

The implications of this are profound, suggesting as it does that fatigue has a uniquely different relationship to health status.

It would, however, be in keeping with Wessely’s own findings (as published in his 1998 article on the Epidemiology of CFS referenced above) that ***“the world could not be divided into those with chronic fatigue (the ill group) and those without (the well)”*** (emphasis added).

It is worth reiterating that in response to Professor Hooper’s complaint to The Lancet, Peter White, writing on behalf of all contributors to The Lancet article, stated: ***“The PACE trial paper 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)”***.

Why would The Lancet fast-track an article concerning a spurious disorder defined ***“simply as a principal complaint of fatigue”***?

Against this background, what was the purpose of the PACE Trial, given that the Director of the PACE Clinical Trial Unit, Professor Simon Wessely, is on record -- long before the PACE Trial began -- stating his empirically-based conclusion that the world cannot be divided into ***“the ill”*** and ***“the well”*** on the basis of the degree of fatigue experienced?

Conclusion

Reporting on the results of the PACE Trial, The Lancet article states: ***“25 (16%) of 153 participants in the APT group were within normal ranges for both primary outcomes at 52 weeks, compared with 44 (30%) of 148 participants for CBT, 43 (28%) of 154 participants for GET, and 22 (15%) of 152 participants for SMC.”***

In the light of the contradictions and other considerations outlined above, it would appear that these figures, modest as they are, inflate the proportions who may be deemed to be within “the normal range” on conclusion of the PACE Trial (**but being within “the normal range” does not necessarily equate to what would be considered “normal” in the typical sense of the word**).

“The normal range” is a statistical term; “normality” is the usual/regular/common/typical value of a variable in respect of an appropriate control population. Where a measure is “normally distributed” in the general population, the method chosen to identify the “normal range” – ie. the mean plus or minus one standard deviation from the mean – equates well to what is “normal”. **Where the distribution is skewed, as it is in respect of physical function, then the application of this formula fails to deliver a meaningful threshold in terms of what is “normal” in the population.**

Furthermore, there were numerous changes to the chosen thresholds and cut off points, both in terms of entry to the PACE Trial and in respect of the assessment of outcomes.

Manipulation of the benchmarks used to recruit to the PACE Trial and to judge whether or not participants were “within the normal range” at its conclusion has produced an absurd situation whereby the same requirement for admission to the trial is deemed by the PIs to denote success at the end of the trial. With regard to these issues:

- the PIs’ chosen thresholds of the “normal range” on the two “primary outcomes” are contrived, unrepresentative, and unduly low in respect of physical function and high in respect of fatigue
- the nature of the comparison group in respect of physical function is misrepresented in the article published in The Lancet, which refers to a “working age population”. The threshold of the range of normal is now said to have been derived from figures relating to the “adult population as a whole” ie. including elderly people. This affords a lower threshold of the “normal range”, thus boosting the proportion of PACE participants who could be deemed to have attained the benchmark level of physical functioning
- the reference cited in respect of the chosen threshold of the range of normal physical functioning does not appear to provide the figures cited by the PIs (ie. Bowling A et al. *Publ Health Med* 1999;21:255-270)

- the benchmark chosen in respect of “fatigue” is at odds with the threshold of “abnormal” fatigue as “demonstrated” in previously published work by the PIs, as cited in the Trial Protocol.

These factors makes the PACE Trial outcomes appear more favourable than is warranted; this in turn misrepresents the claimed efficacy of the interventions CBT and GET.

At the same time, the two “**primary outcome measures**” that were specified to delineate “**a positive outcome**” are not reported. No alternative “**primary efficacy measures**” are proposed, nor is there any reference to parameters of “**a positive outcome**” in The Lancet article.

The analysis given greatest prominence simply compares mean scores between the various intervention and control groups on physical function and fatigue and, having identified some statistically significant differences between these, concludes that CBT and GET “**moderately improve outcomes**”.

On behalf of all of the contributors to the PACE Trial article published in The Lancet, Peter White has agreed with something that people with myalgic encephalomyelitis have been pointing out, ie. the article does not relate to people with ME but to “Oxford”-defined chronic fatigue syndrome: “a principal complaint of fatigue that is disabling, having lasted six months, with no alternative medical explanation.”

Consequently, there should be immediate, high profile, unequivocal clarification specifying to which patients the PACE Trial findings can legitimately be applied.

The PACE Trial Protocol states that the main aim of the trial was to “provide high quality evidence to inform choices made by patients, patient organisations, health services and health professionals about the relative benefits, cost-effectiveness, and cost-utility, as well as adverse effects, of the most widely advocated treatments for CFS/ME”.

The problematic analysis and presentation of data means that the PACE Trial has failed to provide “high quality evidence”, which is an unacceptable outcome for an eight-year project involving 641 participants that cost £5 million to execute.

Patients, clinicians and tax-payers have a right to expect higher scientific exactitude from The Lancet, and the PIs have an ethical and fiscal duty to allow an independent re-evaluation of the data.
