www.margaretwilliams.me

Brief summary of key concerns about the PACE trial

(Detailed evidence can be found in "Key concerns about the PACE Trial" by Professor Malcolm Hooper, September 2013)

The Principal Investigators (and others involved in the trial):

1. failed to fully declare their competing interests, thus informed consent may not have been obtained
2. failed to comply with essential ethical guidance and Codes of Practice in many crucial domains
3. chose (their own) entry criteria that did not define the population purportedly being studied
4. failed to subgroup the cohort according to the presence of neurological signs and symptoms, without which CFS/ME cannot be diagnosed
5. failed to adhere to the published protocol (Dr Ben Goldacre of "Bad Science" says of such practice: *"in a trial… you have to say which is the 'primary outcome' before you start: you can't change your mind about what you're counting as your main outcome…. It's not just dodgy, it also messes with the statistics ….You cannot change the rules after the game has started. You cannot even be seen to do that*" (The Guardian: 5th January 2008). The fact is that the PACE Investigators did change the rules after the game had started and they have been seen to do that
6. changed the required entry score on the SF-36 physical function scale after the trial had started
7. changed their scoring methods and thresholds once they had obtained their data
8. carried out the one objective test (6 MWT) in a meaningless way in the context of CFS/ME (after one year of therapy, PACE participants – average age 38 years -- did not achieve a one-off result achievable by healthy people aged 85)
9. admitted they were not studying "CFS/ME" after all (merely "fatigue")
10. reported that there was no reduction in State or insurance benefits claimed by participants (in fact there was an increase in benefit uptake)
11. refused to supply the return-to-work figures despite FOIA requests
12. reported that at the end of the trial, the Clinical Global Impression results showed that 60% of participants in the GET group and 58% of participants in the CBT group reported negative or minimal change in overall health
13. proclaimed the trial as successful, whilst ignoring the fact that 70-72% were not helped.

The "normal range": the PIs used an inappropriate comparator to define their own "normal range", which enhanced the claimed efficacy of the interventions; furthermore, **their *post-hoc* changes, revisions and recalculations resulted in the illogical situation whereby participants could score worse on completion than on entry but still be classed as being "within the normal range" due to the (alleged) success of the interventions** (NB. The "normal range" is a statistical term and does not equate to "normal" health, and certainly not to "recovery").

"Recovery" scores: the PIs' *post-hoc* metric for physical function warrants close scrutiny because its derivation contains a significant statistical error.

In Psychological Medicine White et al wrote: *"We changed our original protocol's threshold score for being within a normal range on this measure from a score of >=85 to a lower score as that threshold would mean that approximately half the general working age population would fall outside the normal range. The mean (SD) scores for a demographically representative English adult population were 86.3 (22.5) for males and 81.8 (25.7) for females (Bowling et al 1999). We derived a mean (SD) score of 84 (24) for the whole sample, giving a normal range of 60 or above for physical function"* (Psychological Medicine 2013: Oct; 43(10):2227-35: Epub ahead of print). **This statement proved to be inaccurate**.

It is clear that from the start of the trial Professor White et al had two distinct concepts in mind: "positive outcome" (defined as the mean SF-36 PF score minus 1 SD or above) and "recovery" (a higher threshold defined as an SF-36 PF score of 85 or above). It is instructive to note the progressive widening of these thresholds over time:

Year	Source	Mean minus 1 SD	Positive Outcome	Recovery
2002	Trial protocol	75 [1]	75	not specified
2007	Trial protocol	70 [2]	75	>=85
2011	Lancet	60	60	not specified
2013	Psych Med	60	>=60	>=60

[1] 2002: *"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"*

[2] 2007: *"A score of 70 is about one standard deviation below the mean score (about 85, depending on the study) for the UK adult population".*

Therefore it can be seen that between 2002 and 2011-2013 the Investigators' derivation of the mean SF-36 PF score minus 1SD fell from a score of 75 or above to a score of 60 or above. Similarly, their definition of recovery fell from a score of 85 or above to a score of only 60 or above.

Consequently, by publication, there was no difference between a positive outcome and recovery, both of which fell under the common rubric of the Investigators' chosen "normal range".

Not only do the published results lack conceptual clarity, they also contain an important statistical error. The Investigators' stated justification for reducing the SF-36 physical function threshold of the "normal range" from 85 to 60 (namely that approximately half the general working age population would fall below an SF-36 physical function threshold of 85) is not supported by any cited reference and specifically not by Bowling et al (J Pub Health Med 1999:21:255-270), **although it appears possible that the Investigators intended readers to assume that they were relying Bowling et al for that statement.**

Independent re-analysis of Bowling's raw data shows that just 18% (not approximately 50% as claimed by the PACE Investigators) fall below an SF-36 physical function threshold of 85, and once those with long-term health issues are excluded, the figure falls to 8%. These figures are nowhere near the figure of approximately 50% upon which the Investigators relied.

This vitiates the Investigators' stated reason for lowering the score from 85 to 60 and consequently invalidates the conclusion of their published paper on "recovery" (Psychological Medicine 2013: Oct; 43(10):2227-35: Epub ahead of print).

The Investigators' new threshold of 60 is noteworthy because it is lower than the score of 65 required for entry to the trial, so a participant could deteriorate or stay the same but still be counted as recovered in the published results.

This has resulted in an explicit contradiction by the Investigators because, having set the lower bound for recovery at 60, they also state <u>in the same paper</u> that any SF-36 score of less than or equal to 65 represents abnormal physical function, therefore, <u>in the same paper</u>, scores of 60 and 65 represent both abnormal physical function and recovery.

This is not just a theoretical concern, as an FOIA request revealed that nearly 13% of participants had scores of 60 or 65 when they entered the trial: if 13% entered the trial with "normal" function, why were they treated?

It is important to be aware that the figure of 60 for "recovery" was used by the Investigators <u>specifically for the PACE trial</u> and it contradicts how they themselves previously defined markers of recovery in the same disorder using the same measure: in 2007 they stated: *"A patient had to score 80 or higher to be considered as recovered"* (Psychother Psychosom 2007:76:171-176) and in 2009 their Dutch colleagues asserted: "*A cut-off of less than or equal to 65 was considered to reflect severe problems with physical functioning"* (European Journal of Public Health 2009:20:3:251-257).

Common sense would suggest that a mathematically-derived recovery threshold which allows a participant to deteriorate and still be described as recovered must contain a mistake. Yet common sense has not prevailed in this instance and the co-editor-in-chief of Psychological Medicine (Professor Sir Robin Murray, Professor of Psychiatric Research at The Institute of Psychiatry; Fellow of the Royal College of Psychiatrists; elected a Fellow of the Royal Society in 2010 and knighted in 2011 for his services to medicine), has declined to correct obvious errors when they were pointed out to him.