
A preliminary analysis of ‘recovery’ from chronic fatigue syndrome in the PACE trial using individual participant data

Wednesday 21 September 2016

Alem Matthees (1), Tom Kindlon (2), Carly Maryhew (3), Philip Stark (4), Bruce Levin (5).

1. Perth, Australia. alem.matthees@gmail.com
2. Information Officer, Irish ME/CFS Association, Dublin, Ireland.
3. Amersfoort, Netherlands.
4. Associate Dean, Mathematical and Physical Sciences; Professor, Department of Statistics; University of California, Berkeley, California, USA.
5. Professor of Biostatistics and Past Chair, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, USA.

Summary

The PACE trial tested interventions for chronic fatigue syndrome, but the published ‘recovery’ rates were based on thresholds that deviated substantially from the published trial protocol. Individual participant data on a selection of measures has recently been released under the Freedom of Information Act, enabling the re-analysis of recovery rates in accordance with the thresholds specified in the published trial protocol. The recovery rate using these thresholds is 3.1% for specialist medical care alone; for the adjunctive therapies it is 6.8% for cognitive behavioural therapy, 4.4% for graded exercise therapy, and 1.9% for adaptive pacing therapy. This re-analysis demonstrates that the previously reported recovery rates were inflated by an average of four-fold. Furthermore, in contrast with the published paper by the trial investigators, the recovery rates in the cognitive behavioural therapy and graded exercise therapy groups are not significantly higher than with specialist medical care alone. The implications of these findings are discussed.

Introduction

The PACE trial was a large multi-centre study of therapeutic interventions for chronic fatigue syndrome (CFS) in the United Kingdom (UK). The trial compared three therapies which were each added to specialist medical care (SMC): cognitive behavioural therapy (CBT), graded exercise therapy (GET), and adaptive pacing therapy (APT). [1] Henceforth SMC alone will be ‘SMC’, SMC plus CBT will be ‘CBT’, SMC plus GET will be ‘GET’, and SMC plus APT will be ‘APT’. Outcomes consisted of two self-report primary measures (fatigue and physical function), and a mixture of self-report and objective secondary measures. The trial’s co-principal investigators are longstanding practitioners and proponents of the CBT and GET approach, whereas APT was a highly formalised and modified version of an alternative energy management approach.

After making major changes to the protocol-specified “recovery” criteria, White et al. (2013) reported that when using “a comprehensive and conservative definition of recovery”, CBT and GET were associated with significantly increased recovery rates of 22% at 52-week follow-up, compared to only 8% for APT and 7% for SMC [2]. However, those figures were not derived using the published trial protocol (White et al., 2007 [3]), but instead using a substantially revised version that has been widely criticised for being overly lax and poorly justified (e.g. [4]). For example, the changes created an overlap between trial eligibility criteria for severe disabling fatigue, and the new

“normal range”. Trial participants could consequently be classified as recovered without clinically significant improvements to self-reported physical function or fatigue, and in some cases without any improvement whatsoever on these outcome measures. Approximately 13% of participants at baseline simultaneously met the trial eligibility criteria for ‘significant disability’ and the revised recovery criteria for normal self-reported physical function. The justification given for changing the physical function threshold of recovery was apparently based on a misinterpretation of basic summary statistics [5,6], and the authors also incorrectly described their revised threshold as more stringent than previous research [2]. These errors have not been corrected, despite the publishing journal’s policy that such errors should be amended, resulting in growing calls for a fully independent re-analysis of the PACE trial results [7,8].

More than six years after data collection was completed for the 52-week follow-up, the PACE trial investigators have still not published the recovery rates as defined in the trial protocol. Queen Mary University of London (QMUL), holder of the trial data and home of the chief principal investigator, have also not allowed access to the data for others to analyse these outcomes. Following a Freedom of Information Act (FOIA) request for a selection of trial data, an Information Tribunal upheld an earlier decision from the Information Commissioner ordering the release of that data (see case EA/2015/0269). On 9 September 2016, QMUL released the requested data [9]. Given the public nature of the data release, and the strong public interest in addressing the issue of “recovery” from CFS in the PACE trial, we are releasing a preliminary analysis using the main thresholds set in the published trial protocol. The underlying data is also being made available [10], while more detailed and complete analyses on the available outcome measures will be published at a later date.

Methods

Measures and criteria

Using the variables available in the FOIA dataset, ‘recovery’ from CFS in the PACE trial is analysed here based on the main outcome measures described by White et al. (2013) in the “cumulative criteria for trial recovery” [2]. These measures are: (i) the Chalder Fatigue Questionnaire (CFQ); (ii) the Short-Form-36 (SF-36) physical function subscale; (iii) the Clinical Global Impression (CGI) change scale; and (iv) the Oxford CFS criteria. However, instead of the weakened thresholds used in their analysis, we will use the thresholds specified in the published trial protocol by White et al. (2007) [3]. A comparison between the different thresholds for each outcome measure is presented in Table 1.

Outcome measure	Protocol-specified endpoint	Post-trial revised endpoint
Chalder Fatigue Questionnaire (CFQ)	≤ 3 points (range 0-11) when using the bimodal scoring system for the 11 items (0,0,1,1).	≤ 18 points (range 0-33) when using the Likert scoring system for the 11 items (0,1,2,3).
SF-36 physical function	≥ 85 points (range 0-100).	≥ 60 points (range 0-100).
Clinical Global Impression	A score of 1 i.e. “very much better” (range 1-7).	A score of 1 i.e. “very much better”, or 2 i.e. “much better” (range 1-7).
Oxford CFS criteria	Oxford CFS criteria not met.	Either Oxford CFS criteria not met, or a CFQ (bimodal) fatigue score of less than 6, or a SF-36 physical function score of more than 65.

Where follow-up data for self-rated CGI scores were missing we did not impute doctor-rated scores, in contrast to the approach of White et al. (2013), because the trial protocol stated that all primary and secondary outcomes are “either self-rated or objective in order to minimise observer bias” from non-blinded assessors. We discuss the minimal impact of this imputation below. Participants missing any recovery criteria data at 52-week follow-up were classified as non-recovered.

Statistical analysis

White et al. (2013) conducted an available-case analysis which excluded from the denominators of each group the participants who dropped out [2]. This is not the recommended practice in clinical trials, where intention-to-treat analysis (which includes all randomised participants) is commonly preferred. An available-case analysis may overestimate real-world treatment effects because it does not include participants who were lost to follow-up. Attrition from trials can occur for various reasons, including an inability to tolerate the prescribed treatment, a perceived lack of benefit, and adverse reactions. Thus, an available-case analysis only takes into account the patients who were willing and able to tolerate the prescribed treatments. Nonetheless, both types of analyses are presented here for comparison. We present a preliminary exploratory analysis of the frequency and percentage of participants meeting all the recovery criteria in each group, based on the intention-to-treat principle, as well as the available-case subgroup.

Neither the published trial protocol [3] nor the published statistical analysis plan [11] specified a method for determining the statistical significance of the differences in recovery rates between treatment groups. In their published paper on recovery, White et al. (2013) presented logistic regression analyses for trial arm pairwise comparisons, adjusting for the baseline stratification variables of treatment centre, meeting CDC CFS criteria, meeting London ME criteria, and having a depressive illness [2]. However, it has been shown that logistic regression may be an inappropriate method of analysis in the context of randomised trials [12]. While Fisher’s exact test would be preferable, a more rigorous approach would also take into account stratification variables, which unfortunately were not part of the available FOIA dataset. Nonetheless, there is reason to believe that the effect of including these stratification variables would be minimal on our analyses: the stratification variables were approximately evenly distributed between groups [1], and attempting to replicate the previously published [2] odds ratios and 95% confidence intervals using logistic

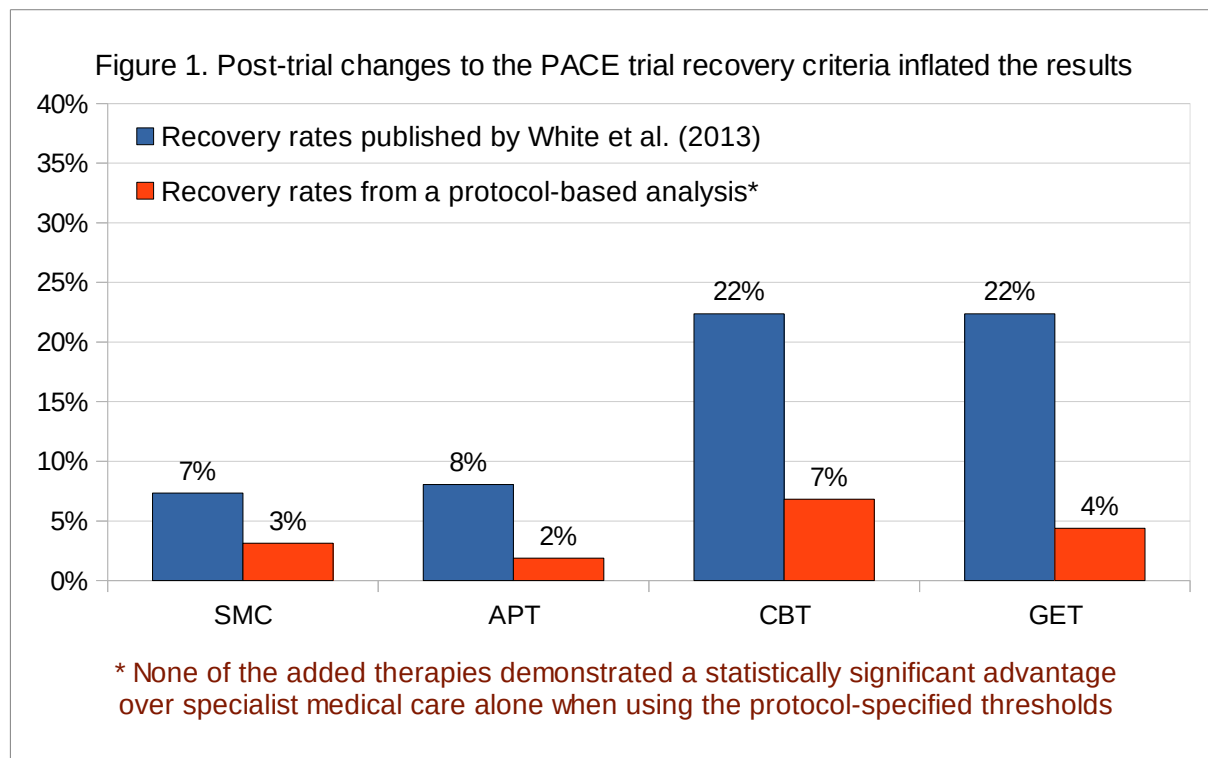
regression, but without stratification variables, yielded very similar results to the ones previously published (see Table 3).

We therefore present recovery rates for each group and compare the observed rates for each active treatment arm with those of the SMC arm using Fisher’s exact tests. The confidence intervals for recovery rates in each group and comparative odds ratios are exact 95% confidence intervals using the point probability method [13]. For sake of direct comparison with results published by White et al. (2013), we also present results of logistic regression analysis which included only the treatment arm as a predictor variable, with conventional approximate 95% confidence intervals.

Results

For our analysis of ‘recovery’ in the PACE trial, full data were available for 89% to 94% of participants, depending on the treatment group and outcome measure. Percentages are calculated for both intention-to-treat, and on an available-case basis. Imputing the missing self-rated CGI scores with doctor-rated CGI scores made no difference to the intention-to-treat analysis, as there were no participants with missing self-rated CGI scores with an assessor rating of 1, required for recovery; in the available-case analysis, the only effect this had was to decrease the CBT denominator by 1, and the assessor score for that participant was 3, “a little better”, therefore non-recovered. Table 2 provides the results and Figure 1 compares our recovery rates with those of White et al. (2013):

Intention-to-treat analysis	Available-case analysis
<u>Recovery rates (n/N, %, 95% CI):</u> SMC: 5/160, 3.1% (1.2-7.1) CBT: 11/161, 6.8% (3.6-12.0) GET: 7/160, 4.4% (2.1-8.7) APT: 3/159, 1.9% (0.5-5.5)	<u>Recovery rates (n/N, %, 95% CI):</u> SMC: 5/150, 3.3% (1.3-7.6) CBT: 11/142, 7.7% (4.1-13.3) GET: 7/143, 4.9% (2.3-9.7) APT: 3/149, 2.0% (0.6-5.9)
<u>Fisher’s exact test (two-tailed) with OR and exact 95% CI:</u> CBT vs SMC: p = 0.199, OR 2.27 (0.76-6.92) GET vs SMC: p = 0.770, OR 1.42 (0.42-4.67) APT vs SMC: p = 0.723, OR 0.60 (0.12-2.49)	<u>Fisher’s exact test (two-tailed) with OR and exact 95% CI:</u> CBT vs SMC: p = 0.124, OR 2.44 (0.81-7.43) GET vs SMC: p = 0.565, OR 1.49 (0.44-4.93) APT vs SMC: p = 0.723, OR 0.60 (0.12-2.49)
<u>Logistic regression (unadjusted):</u> CBT vs SMC: OR (95% CI) = 2.27 (0.77-6.70), p = 0.136 GET vs SMC: OR (95% CI) = 1.42 (0.44-4.57), p = 0.558 APT vs SMC: OR (95% CI) = 0.60 (0.14-2.54), p = 0.484	<u>Logistic regression (unadjusted):</u> CBT vs SMC: OR (95% CI) = 2.44 (0.82-7.19), p = 0.107 GET vs SMC: OR (95% CI) = 1.49 (0.46-4.82), p = 0.503 APT vs SMC: OR (95% CI) = 0.60 (0.14-2.54), p = 0.484



The CBT, GET, and APT groups did not demonstrate a statistically significant advantage over the SMC group in any of the above analyses, nor an empirical recovery rate that would generally be considered adequate (the highest observed rate was 7.7%). In the intention-to-treat analysis, the exact p value for the three degree of freedom chi-squared test for no overall differences amongst the four groups was 0.14. In the available-case analysis, the p value was 0.10. Given the number of comparisons, a correction for multiple testing might be appropriate, but as none of the uncorrected p values were significant at the $p < 0.05$ level, this more conservative approach would not alter the conclusion. Our findings therefore contradict the conclusion of White et al. (2013), that CBT and GET were significantly more likely than the SMC group to be associated with ‘recovery’ at 52 weeks [2]. However, the very low recovery rates substantially decrease the ability to detect statistically significant differences between groups (see the Limitations section). The multiple changes to the recovery criteria had inflated the estimates of recovery by approximately 2.3 to 5.1 -fold, depending on the group, with an average inflation of 3.8-fold.

Limitations

Lack of statistical power

When designing the PACE trial and determining the number of participants needed, the investigators’ power analyses were based not on recovery estimates but on the prediction of relatively high rates of clinical improvement in the additional therapy groups compared to SMC alone [3]. However, the very low recovery rates introduce a complication for tests of significance, due to insufficient statistical power to detect modest but clinically important differences between groups. For example, with the CBT vs. SMC comparison by intention-to-treat, a true odds ratio of 4.2 would have been required to give Fisher’s exact test 80% power to declare significance, given the observed margins. If we assume SMC has a probability of 3.1%, an odds ratio of 4.2 would have conferred a recovery probability of 11.8%, which was not achieved in the trial.

We believe that for our preliminary analysis it was important to follow the protocol-specified recovery criteria, which make more sense than the revised thresholds. For example, the former required level of physical function would suggest a ‘recovered’ individual could at least do most normal activities, but may have limitations with a few of the items on the SF-36 health survey, such as vigorous exercise, walking up flights of stairs, or bending down. The revised threshold that White et al. (2013) used meant that a ‘recovered’ individual could have remained limited on four to eight out of ten items depending on severity. We found that when using the revised recovery criteria, 8% (7/87) of the ‘recovered’ participants still met trial eligibility criteria for 'significant disability'.

Weakening the recovery thresholds increases statistical power to detect group differences because it makes the event (i.e. ‘recovery’) rates more frequent (i.e. less close to zero) but it also leads to the inclusion of patients who still, for example, have significant illness-related restrictions in physical capacity as per SF-36 physical function score. We argue that if significant differences between groups cannot be detected in sample sizes of approximately n=160 per group, then this may indicate that CBT and GET simply do not substantially increase recovery rates.

Lack of data on stratification variables

In order to increase the chance of being granted or enforced, the FOIA request asked for a ‘bare minimum’ set of variables, as asking for too many variables, or for variables that may be judged to significantly increase the risk of re-identification of participants, would have decreased the chance that the FOIA request would be granted. This was a reasonable compromise given that QMUL had previously blocked all requests for the protocol-specified recovery rates and the underlying data to calculate them. Some non-crucial variables are therefore missing from the dataset acquired under the FOIA but there is reason to believe that this would have little effect on the results.

Allocation of participants in the PACE trial was stratified [1]: *“The first three participants at each of the six clinics were allocated with straightforward randomisation. Thereafter allocation was stratified by centre, alternative criteria for chronic fatigue syndrome and myalgic encephalomyelitis, and depressive disorder (major or minor depressive episode or dysthymia), with computer-generated probabilistic minimisation.”*

This means that testing for statistical significance assuming simple randomisation results in p-values that are approximate and effect-size estimates that might be biased. The FOIA dataset does not contain the stratification variables. While the lack of these variables may somewhat alter the estimated treatment effects and the p-values or confidence levels, we expect the differences to be minor, a conclusion that is supported by Table 3 below. Table 1 of the publication of the main trial results (White et al., 2011) shows that the stratification variables were approximately evenly distributed between groups [1]. We have replicated the rates of “trial recovery” as previously published by White et al. (2013) [2]. We also attempted to replicate their previously reported logistic regression, without the stratification variables, and the results were essentially the same (see Table 3), suggesting that the adjustments would not have a significant impact on the outcome of our own analysis of recovery.

Table 3. Adjusted vs unadjusted logistic regression for recovery rates published by White et al. 2013 [2]	
PACE published adjusted ORs (95% CIs)	FOIA dataset unadjusted ORs (95% CIs)
CBT vs SMC, 3.69 (1.77–7.69) p<0.001	CBT vs SMC, 3.64 (1.76–7.55) p<0.001
GET vs SMC, 3.71 (1.78–7.74) p<0.001	GET vs SMC, 3.64 (1.76–7.55) p<0.001
APT vs SMC, 1.10 (0.47–2.58) p=0.83	APT vs SMC, 1.11 (0.47–2.59) p=0.82
CBT vs APT, 3.36 (1.64–6.88) p=0.001	CBT vs APT, 3.29 (1.62–6.69) p=0.001
GET vs APT, 3.38 (1.65–6.93) p=0.001	GET vs APT, 3.29 (1.62–6.69) p=0.001

If QMUL or the PACE trial investigators believe that further adjustment is necessary here to have confidence in the results, then we invite them to present analyses that include stratification variables or release the raw data for those variables without unnecessary restrictions.

Lack of data on alternative ME/CFS criteria

For the same reasons described in the previous subsection, the FOIA dataset does not contain the variables for meeting CDC CFS criteria or London ME (myalgic encephalomyelitis) criteria. These were part of the original definition of recovery, but we argue that these are superfluous because:

(a) While our definition of recovery is less stringent without the alternative ME/CFS criteria, these additional criteria had no significant effect on the results reported by White et al. (2013) [2]. (b) The alternative ME/CFS criteria used in the trial had some questionable modifications [14], that have not been used in any other trial, thus seriously limiting cross-trial comparability and validation of their results. (c) The Oxford CFS criteria are the most sensitive and least specific (most inclusive) criteria, so those who fulfil all other aspects of the recovery criteria would most likely also fail to meet alternative ME/CFS criteria. (d) All participants were first screened using the Oxford CFS criteria as this was the primary case definition, whereas the additional case criteria were not entry requirements [1].

Discussion

It is important that patients, health care professionals, and researchers have accurate information about the chances of recovery from CFS. In the absence of definitive outcome measures, recovery criteria should set reasonable standards that approach restoration of good health, in keeping with commonly understood conceptions of recovery from illness [15]. Accordingly, the changes made by the PACE trial investigators after the trial was well under way resulted in the recovery criteria becoming too lax to allow conclusions about the efficacy of CBT and GET as rehabilitative treatments for CFS. This analysis, based on the published trial protocol, demonstrates that the major changes to the thresholds for recovery had inflated the estimates of recovery by an average of approximately four-fold. QMUL recently posted the PACE trial primary ‘improvement’ outcomes as specified in the protocol [16] and that also showed a similar difference between the proportion of participants classified as improved compared to the post-hoc figures previously published in the Lancet in 2011 [1]. It is clear from these results that the changes made to the protocol were not minor or insignificant, as they have produced major differences that warrant further consideration.

The PACE trial protocol was published with the implication that changes would be unlikely [17], and while the trial investigators describe their analysis of recovery as pre-specified, there is no mention of changes to the recovery criteria in the statistical analysis plan that was finalised shortly

before the unblinding of trial data [11]. Confusion has predictably ensued regarding the timing and nature of the substantial changes made to the recovery criteria [18]. Changing study endpoints should be rare and is only rarely acceptable; moreover, trial investigators may not be appropriate decision makers for endpoint revisions [19,20]. Key aspects of pre-registered design and analyses are often ignored in subsequent publications, and positive results are often the product of overly flexible rules of design and data analysis [21,22].

As reported in a recent BMJ editorial by chief editor Fiona Godlee (3 March 2016), when there is enough doubt to warrant independent re-analysis [23]: “*Such independent reanalysis and public access to anonymised data should anyway be the rule, not the exception, whoever funds the trial.*” The PACE trial provides a good example of the problems that can occur when investigators are allowed to substantially deviate from the trial protocol without adequate justification or scrutiny. We therefore propose that a thorough, transparent, and independent re-analysis be conducted to provide greater clarity about the PACE trial results. Pending a comprehensive review or audit of trial data, it seems prudent that the published trial results should be treated as potentially unsound, as well as the medical texts, review articles, and public policies based on those results.

Acknowledgements

Writing this article in such a brief period of time would not have been possible without the diverse and invaluable contributions from patients, and others, who chose not to be named as authors.

Declarations

AM submitted a FOIA request and participated in legal proceedings to acquire the dataset.

TK is a committee member of the Irish ME/CFS Association (voluntary position).

References

1. White PD, Goldsmith KA, Johnson AL, Potts L, Walwyn R, DeCesare JC, Baber HL, Burgess M, Clark LV, Cox DL, Bavinton J, Angus BJ, Murphy G, Murphy M, O'Dowd H, Wilks D, McCrone P, Chalder T, Sharpe M; PACE trial management group. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *Lancet*. 2011 Mar 5;377(9768):823-36. doi: 10.1016/S0140-6736(11)60096-2. Epub 2011 Feb 18. PMID: 21334061. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3065633/>
2. White PD, Goldsmith K, Johnson AL, Chalder T, Sharpe M. Recovery from chronic fatigue syndrome after treatments given in the PACE trial. *Psychol Med*. 2013 Oct;43(10):2227-35. doi: 10.1017/S0033291713000020. PMID: 23363640. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3776285/>
3. White PD, Sharpe MC, Chalder T, DeCesare JC, Walwyn R; 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. *BMC Neurol*. 2007 Mar 8;7:6. PMID: 17397525. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2147058/>
4. A list of articles by David Tuller on ME/CFS and PACE at Virology Blog. <http://www.virology.ws/mecfs/>
5. Kindlon T, Baldwin A. Response to: reports of recovery in chronic fatigue syndrome may present less than meets the eye. *Evid Based Ment Health*. 2015 May;18(2):e5. doi: 10.1136/eb-2014-101961. Epub 2014 Sep 19. PMID: 25239244. <http://ebmh.bmj.com/content/18/2/e5.long>
6. Matthees A. Assessment of recovery status in chronic fatigue syndrome using normative data. *Qual Life Res*. 2015 Apr;24(4):905-7. doi: 10.1007/s11136-014-0819-0. Epub 2014 Oct 11. PMID: 25304959. <http://link.springer.com/article/10.1007%2Fs11136-014-0819-0>

7. Davis RW, Edwards JCW, Jason LA, et al. An open letter to The Lancet, again. *Virology Blog*. 10 February 2016. <http://www.virology.ws/2016/02/10/open-letter-lancet-again/>
8. #MEAction. Press release: 12,000 signature PACE petition delivered to the Lancet. <http://www.meaction.net/press-release-12000-signature-pace-petition-delivered-to-the-lancet/>
9. Queen Mary University of London. Statement: Disclosure of PACE trial data under the Freedom of Information Act. 9 September 2016 Statement: Release of individual patient data from the PACE trial. <http://www.qmul.ac.uk/media/news/items/smd/181216.html>
10. FOIA request to QMUL (2014/F73). Dataset file: 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>
11. Walwyn R, Potts L, McCrone P, Johnson AL, DeCesare JC, Baber H, Goldsmith K, Sharpe M, Chalder T, White PD. 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 Nov 13;14:386. doi: 10.1186/1745-6215-14-386. PMID: 24225069. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4226009/>
12. Freedman DA. Randomization Does Not Justify Logistic Regression. *Statistical Science*. 2008;23(2):237–249. doi:10.1214/08-STS262. <https://arxiv.org/pdf/0808.3914.pdf>
13. Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions*. 3rd ed. Hoboken, N.J: J. Wiley; 2003. 760 p. ISBN: 978-0-471-52629-2. (Wiley series in probability and statistics). <http://au.wiley.com/WileyCDA/WileyTitle/productCd-0471526290.html>
14. Matthees A. Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Ann Intern Med*. 2015 Dec 1;163(11):886-7. doi: 10.7326/L15-5173. PMID: 26618293.
15. Adamowicz JL, Caikauskaitė I, Friedberg F. Defining recovery in chronic fatigue syndrome: a critical review. *Qual Life Res*. 2014 Nov;23(9):2407-16. doi: 10.1007/s11136-014-0705-9. Epub 2014 May 3. PMID: 24791749. <http://link.springer.com/article/10.1007%2Fs11136-014-0705-9>
16. Goldsmith KA, White PD, Chalder T, Johnson AL, Sharpe M. The PACE trial: analysis of primary outcomes using composite measures of improvement. 8 September 2016. http://www.wolfson.qmul.ac.uk/images/pdfs/pace/PACE_published_protocol_based_analysis_final_8th_Sept_2016.pdf
17. BMC editor's comment on [Protocol for the PACE trial] (Version: 2. Date: 31 January 2007) http://www.biomedcentral.com/imedia/2095594212130588_comment.pdf
18. UK House of Lords. PACE Trial: Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. 6 February 2013. <http://www.publications.parliament.uk/pa/ld201213/ldhansrd/text/130206-gc0001.htm>
19. Evans S. When and how can endpoints be changed after initiation of a randomized clinical trial? *PLoS Clin Trials*. 2007 Apr 13;2(4):e18. PMID 17443237. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1852589/>
20. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010 Mar 23;340:c869. doi: 10.1136/bmj.c869. PMID: 20332511. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2844943>
21. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*. 2011 Nov;22(11):1359-66. doi: 10.1177/0956797611417632. Epub 2011 Oct 17. PMID: 22006061. <http://pss.sagepub.com/content/22/11/1359.long>
22. Wagenmakers EJ, Wetzels R, Borsboom D, van der Maas HL, Kievit RA. An Agenda for Purely Confirmatory Research. *Perspect Psychol Sci*. 2012 Nov;7(6):632-8. doi: 10.1177/1745691612463078. PMID: 26168122. <http://pps.sagepub.com/content/7/6/632.full>
23. Godlee F. Data transparency is the only way. *BMJ* 2016;352:i1261. (Published 03 March 2016) doi: <http://dx.doi.org/10.1136/bmj.i1261> <http://www.bmj.com/content/352/bmj.i1261>