

HOW **NOT** TO CONDUCT A RANDOMIZED CLINICAL TRIAL

Bruce Levin, Ph.D.

Department of Biostatistics
Mailman School of Public Health

Biostatistics in Action: Tips for Clinical Researchers
Fall 2018 Lecture Series

September 12th, 2018



ACKNOWLEDGEMENT

I am indebted to the following people for material for this talk:

- David Tuller, Dr.P.H., Univ of Calif Berkeley for courageous journalistic investigation
- Carolyn E. Wilshire, Victoria Univ of Wellington, NZ, for profound psychological insights
- Vincent Racaniello of Columbia for content on his *Virology Blog*
- ME/CFS patients, especially Alem Matthees and Julie Rehmeyer, for their perspectives

In February of 2011, U.K. researchers published a paper in *The Lancet* entitled, “Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial.”

P.D. White, M.C. Sharpe, T. Chalder, J.C. DeCesare, and R. Walwyn (2011).
The Lancet 377: 823–836.

The article reported that two rehabilitative treatments, cognitive behavioral therapy (CBT) and graded exercise therapy (GET), were safe and effective treatments for patients with chronic fatigue syndrome, a.k.a. myalgic encephalomyelitis or ME/CFS. The effects were statistically significant and seemed substantial.

The PACE trial received international attention and has had widespread impact on research, treatments prescribed for patients, and attitudes toward the illness of both the medical community and the public at large.

At a press conference promoting the *Lancet* paper, one of the lead investigators (Trudie Chalder) stated that “twice as many people on graded exercise therapy and cognitive behaviour therapy got back to normal” compared to those in the other two study arms, Adaptive Pacing Therapy (APT) and Specialized Medical Care (SMC).

Boseley S. 2011. Study finds therapy and exercise best for ME. *The Guardian*, 18 Feb. Available at:
<https://www.theguardian.com/society/2011/feb/18/study-exercise-therapy-me-treatment>

An accompanying Lancet commentary similarly claimed that these “back-to-normal” participants had met a “strict criterion for recovery.”

Bleijenberg G, Knoop H. 2011. Chronic fatigue syndrome: where to PACE from here? *The Lancet*, 377: 786-788

UNDERLYING HYPOTHESIS

The treatments investigated in the PACE trial were based on the hypothesis that ME/CFS patients harbor “unhelpful” convictions about having an ongoing organic disease and that the perpetuation of their devastating symptoms is the result of deconditioning. Therefore it is plausible that CBT or GET could help.

By contrast, a 2015 review from the IOM (now the U.S. National Academy of Medicine), reported that ME/CFS is a complex, multi-system illness characterized by neurological, immunological, autonomic, and energy metabolism dysfunctions.

The cardinal symptom noted in the review is a systemic intolerance to exertion; if patients exceed their available energy resources, they can suffer serious and prolonged relapses.

IOM (2015): Beyond myalgic encephalomyelitis/chronic fatigue syndrome: redefining an illness. The National Academies: Wash. DC.

**But PACE trial patients complained
that there were some troubling
PECULIARITIES**

PECULIARITIES

Participants' baseline scores for the two primary outcomes of physical function and fatigue could qualify them simultaneously as disabled enough for eligibility into the trial but *already* “recovered” on those indicators—*even before any treatment began.*

This happened because *the criteria for recovery were changed*, months after data collection was completed, sometime between when the Statistical Analysis Plan was finalized and the trial results were published. The investigators claimed this was done before “looking at the data” and with the approval of the ethics oversight committee, but...

- The trial was not masked regarding Tx group
- The revised endpoints were subjective, self-reported outcomes, not the original objective measures
- Not clear to what extent the oversight committee was independent of the investigators.
- The Statistical Analysis Plan was published only years later.

Definitions of **improvement** and **recovery** in the trial protocol versus those used in the final trial reports:

- **Improvement** was the primary outcome measure specified in the protocol. It was changed from one binary composite outcome to two continuous measures.
- Originally proposed measure: Percentage of patients who fulfilled pre-specified criteria for overall improvement 52 weeks after randomization.
- Replaced in May, 2010 by two continuous measures, fatigue (a score from the Chalder Fatigue Questionnaire) and physical function (a score from the SF-36 Physical Function subscale).

Definitions of **improvement** and **recovery** in the trial protocol versus those used in the final trial reports:

- **Recovery** was also a pre-specified secondary binary composite measure, based on the Chalder Fatigue Questionnaire (CFQ), the SF-36 Physical Function subscale, the Clinical Global Impression (CGI), and various definitions of “caseness” (see below).
- The criteria for the above components were substantially changed in the revised recovery binary outcome.

Overall Improvement

Specified in trial protocol:

- Minimum score of 75 on the 100-point SF-36 physical function scale *or* a score increase of 50% or more, and...

Used in published reports:

- At least an 8 point increase in the 100-point SF-36 physical function scale, and...

Overall Improvement

Specified in trial protocol:

- ...and of the 11 fatigue items on the Chalder Fatigue Questionnaire (CFQ), three or fewer rated as worse/much worse than prior to illness OR the total number of items rated worse/much worse dropped by at least 50%.

Used in published reports:

- ...and at least a 2 point decrease on the 33-point CFQ (Likert scoring method).

Recovery

Specified in trial protocol:

- Minimum score of 85 on the 100-point SF-36 physical function scale, and...

Used in published reports:

- Minimum score of 60 on the 100-point SF-36 physical function scale., and...

Recovery

Specified in trial protocol:

- ...and of the 11 items on the CFQ, three or fewer rated as worse/much worse than prior to illness, and...

Used in published reports:

- ...and maximum score of 18 on the 33-point CFQ, and...

Recovery

Specified in trial protocol:

- ...and overall health self-rated as “very much better” on the Clinical Global Impression (CGI) scale, and...

Used in published reports:

- ...and overall health self-rated as “much better” or “very much better” on the CGI scale, and...

Recovery

Specified in trial protocol:

- ...and the final “caseness” criterion was met if the patient no longer fulfilled: The Oxford case definition of CFS; the CDC criteria AND the London ME criteria. (As determined by a non-blinded assessor.)

Used in published reports:

- ...and the revised “caseness” criterion was met if ANY of the following applied: a) the patient did not meet the standard Oxford case definition; OR b) on the CFQ, they rated less than six of the 11 fatigue items as being worse than prior to illness; OR c) their SF-36 Physical Function score was greater than 65.

PECULIARITIES

- As a result of these changes, 13 percent of patients *were already “recovered” or “within the normal range” at the start of the study for self-reported physical function.*
- The investigators did not publish the findings for the original endpoints as specified in the protocol, even as sensitivity analyses.
- For years they rejected requests from patients for the release of the findings for the originally specified endpoints as “vexatious.”

PECULIARITIES

- As already noted, the PACE claims of successful improvement and recovery were based solely on subjective outcomes.
- All but one of the objective measures from the trial—a walking test, fitness (VO_2 max by step test), data on days lost from work 6 months after primary endpoint, and the receipt of financial benefits—failed to provide any evidence to support claims of efficacy.

PECULIARITIES

- The only objective measure to show improvement was the six-minute walk test, which found that GET participants walked reliably farther than Control participants at the primary, 52-week endpoint.

PECULIARITIES

- However, after an entire year, this group walked an average of just 67 m farther than baseline, and around 30 m farther than Controls. To put this in context, a sample of Class II chronic heart failure patients with similar baseline walking distances increased their distance by an average of 141 m after only three weeks of a gentle graded exercise program.

PECULIARITIES

- In the middle of the study, the PACE team published a newsletter for participants that included glowing testimonials from earlier trial subjects about how much the “therapy” and “treatment” helped them.

PECULIARITIES

- The newsletter also included an article informing participants that the two interventions pioneered by the investigators and being tested for efficacy in the trial, graded exercise therapy and cognitive behavior therapy, had been recommended as treatments by a U.K. government committee “based on the best available evidence.”

PECULIARITIES

- But the newsletter article did not mention that a key PACE investigator was also serving on the U.K. government committee that endorsed the PACE therapies.

PECULIARITIES

- The PACE protocol contained an explicit commitment to tell prospective participants about conflicts of interest. The main investigators had longstanding financial and consulting ties with disability insurance companies, having advised them for years that cognitive behavior therapy and graded exercise therapy could get claimants off benefits and back to work.

PECULIARITIES

- Yet in seeking informed consent, prospective participants were NOT told about any insurance industry links and the information was not included on consent forms. The authors did include the information in the “conflicts of interest” sections of the published papers.

SUBSEQUENTLY

- In August of 2016, pursuant to a Freedom of Information request by Australian patient Alem Matthees, a U.K. tribunal ordered Queen Mary University of London to release a subset of raw trial data from the PACE study so that the pre-specified outcomes of the trial protocol could be analyzed.

SUBSEQUENTLY

- Analyses of these data have confirmed that the extensive outcome-switching allowed the investigators to report dramatically better findings than the null or minimal results obtained with the pre-specified measures. Specifically...

SUBSEQUENTLY

- On the original protocol-specified primary outcome measure—overall improvement—there was a significant effect of treatment group. However, the groups receiving CBT or GET did not significantly outperform the Control group after correcting for the number of comparisons as pre-specified in the Statistical Analysis Plan.

SUBSEQUENTLY

- On the original protocol-specified primary outcome measure—overall improvement—there was a significant effect of treatment group. However, the groups receiving CBT or GET did not significantly outperform the Control group after correcting for the number of comparisons as pre-specified in the Statistical Analysis Plan. [*AND WHAT ABOUT BIAS?*]

SUBSEQUENTLY

Overall improvement at 52 weeks (primary)

As appeared in the published report:

CBT

59% ($p=0.01^+$)

GET

61% ($p<0.01$)

SMC

45%

As specified in the trial protocol:

CBT

20% ($p=0.015$)

GET

21% ($p=0.010$)

SMC

10%

n.s. after protocol-specified Bonferroni correction

SUBSEQUENTLY

- Also, **rates of recovery** were consistently low and not significantly different across treatment groups.

SUBSEQUENTLY

- Also, **rates of recovery** were consistently low and not significantly different across treatment groups.

As appeared in the published report:

<u>CBT</u>	<u>GET</u>	<u>SMC</u>
22% ($p < 0.001$)	22% ($p < 0.001$)	7%

As specified in the trial protocol:

<u>CBT</u>	<u>GET</u>	<u>SMC</u>
7% ($p = 0.12$)	4% ($p = 0.76$)	3%

SUBSEQUENTLY

- Finally, on other secondary measures, significant effects were almost entirely confined to self-report measures. These effects did not endure beyond two years.

SUBSEQUENTLY

- **Conclusions:** These findings raise serious concerns about the robustness of the claims made about the efficacy of CBT and GET.

SUBSEQUENTLY

- **Conclusions:** These findings raise serious concerns about the robustness of the claims made about the efficacy of CBT and GET. **The modest treatment effects obtained on self-report measures in the PACE trial do not exceed what could be reasonably accounted for by participant reporting biases.**

Wilshire C et al. 2018. Rethinking the treatment of chronic fatigue syndrome—a reanalysis and evaluation of findings from a recent major trial of graded exercise and CBT. *BMC Psychology*; published online 22 March. Available at: <https://bmcp psychology.biomedcentral.com/articles/10.1186/s40359-018-0218-3>

BIAS ISSUES

- Heavy reliance on self-reports from participants who were aware of their treatment allocation.

BIAS ISSUES

- Heavy reliance on self-reports from participants who were aware of their treatment allocation.
- In behavioural intervention trials, full blinding is often not possible. Nevertheless, it is the researchers' responsibility to consider the possible effects of lack of blinding on outcomes, and to ensure such factors are insufficient to account for any apparent benefits.

BIAS ISSUES

- Heavy reliance on self-reports from participants who were aware of their treatment allocation.
- In behavioural intervention trials, full blinding is often not possible. Nevertheless, it is the researchers' responsibility to consider the possible effects of lack of blinding on outcomes, and to ensure such factors are insufficient to account for any apparent benefits.
- In a trial that is not blinded, self-reported outcomes can produce highly inflated estimates of treatment-related benefits.

BIAS ISSUES

Given this discrepancy in the effects of blinding on subjective and objective measures, it appears unlikely that the improvement and recovery effects reflect genuine health benefits.

BIAS ISSUES

Given this discrepancy in the effects of blinding on subjective and objective measures, it appears unlikely that the improvement and recovery effects reflect genuine health benefits. **A more plausible explanation is that they are expectation-related artefacts** that reflect the operation of attentional biases that favour the reporting of events consistent with one's expectations, or recall/confirmation biases that enhance recollection for expectation-consistent events.

BIAS ISSUES

The PACE investigators have argued that expectancy effects alone cannot account for the positive self-reported improvements, because at the start of treatment, patients' expectations of improvement were not greater in the CBT and GET groups than in the other groups.

BIAS ISSUES

However, they fail to point out that CBT and GET participants were *primed during treatment* to expect improvement.

BIAS ISSUES

- The manual given to CBT participants at the start of treatment proclaimed CBT to be “a powerful and safe treatment which has been shown to be effective in... CFS/ME”.

BIAS ISSUES

- The manual given to CBT participants at the start of treatment proclaimed CBT to be “a powerful and safe treatment which has been shown to be effective in... CFS/ME”.
- The GET participants’ manual described GET as “one of the most effective therapy strategies currently known”.

BIAS ISSUES

- The manual given to CBT participants at the start of treatment proclaimed CBT to be “a powerful and safe treatment which has been shown to be effective in... CFS/ME”.
- The GET participants’ manual described GET as “one of the most effective therapy strategies currently known”.
- Both interventions emphasized that faithful adherence to the program could lead to a full recovery.

BIAS ISSUES

- Such messages — from an authoritative source — are likely to have substantially raised patients' expectations of improvement.

BIAS ISSUES

- Such messages — from an authoritative source — are likely to have substantially raised patients' expectations of improvement.
- **Importantly, no such statements were given to the other treatment groups.**

BIAS ISSUES

- When we add to this the fact that the CBT programme, and to a lesser extent GET, was designed to reduce “symptom focusing”, which may have further influenced self-report behaviour in the absence of genuine improvement, these findings start to look very worrying indeed.

CONSEQUENCES

- The U.S. Agency for Healthcare Research and Quality (AHRQ) downgraded its recommendations for CBT and GET.
- This downgrading occurred after the agency removed from its analysis the PACE trial and other studies using overly broad selection criteria that generated cohorts of patients with a grab-bag of fatiguing conditions.
- While the PACE trial claimed that GET is safe, AHRQ found that the therapy was associated with more adverse events.

CONSEQUENCES

- A year ago last summer, the U.S. Centers for Disease Control abandoned the recommendations that ME/CFS patients be treated with CBT and GET, having already removed references to the PACE trial.
- A couple of months later, the U.K. National Institute for Health and Care Excellence announced that it would pursue a full update of its 2007 guidance, citing concerns about the reliability and validity of the evidence base.

CONSEQUENCES

- Early this year, a report from the Dutch Health Council recommended that GET should not be used in the Netherlands as a treatment for the illness.

Health Council of the Netherlands. 2018. More scientific research on ME/CFS is needed to serve patients better. 19 March. Available at: <https://www.gezondheidsraad.nl/en/news/more-scientific-research-on-mecfs-is-needed-to-serve-patients-better>

CONSEQUENCES

- In March, a group of leading American clinicians who specialize in ME/CFS unanimously agreed that the two PACE treatments are inappropriate and possibly harmful for patients with the illness and should therefore not be prescribed.

Tucker M. 2018. Much can be done to ease 'chronic fatigue syndrome' symptoms. Medscape, 12 March. Available at:
<https://www.medscape.com/viewarticle/893766>

THANK YOU VERY MUCH.

Most of today's talk (and much more) can be found on Virology Blog here:

<http://www.virology.ws/2018/08/13/trial-by-error-open-letter-to-the-lancet-version-3-0/>