

Submission to the public review on common data elements for ME/CFS: Problems with the Chalder Fatigue Questionnaire

Wilshire, C.E., McPhee, G., and the Science for ME CFQ working group

The Chalder Fatigue Questionnaire (CFQ; Chalder et al., 1993) is among the scales being proposed to provide common data elements (CDEs) on fatigue for future NIH- and CDC-funded studies of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS).

Because the CFQ was used in the PACE trial it has received close scrutiny from patients and researchers who have been critical of the trial (e.g., Wilshire et al., 2016). Some of those same individuals were involved in the drafting of the present submission.

The Chalder Fatigue Scale

Many of the problems with the scale are obvious upon inspection, and so it is important to examine the scale. The complete scale, in its final 11-item form, is reproduced below (bolding is ours).

*We would like to know more about any problems you have had with feeling tired, weak or lacking in energy in the last month. Please answer ALL the questions by ticking the answer which applies to you most closely. If you have been feeling tired for a long while, then compare yourself to how you felt **when you were last well.***

1. Do you have problems with tiredness?
2. Do you need to rest more?
3. Do you feel sleepy or drowsy?
4. Do you have problems starting things?
5. Do you lack energy?
6. Do you have less strength in your muscles?
7. Do you feel weak?
8. Do you have difficulties concentrating?
9. Do you make slips of the tongue when speaking?
10. Do you find it more difficult to find the right word?
11. How is your memory?

[Response options: less than usual; no more than usual; more than usual; much more than usual.]

The scale items can be scored ‘bimodally’ or with ‘Likert’ scoring, as shown below. The scores for each item are then summed to produce an overall score.

Response	Bimodal score	‘Likert’ score
Less than usual	0	0
No more than usual	0	1
More than usual	1	2
Much more than usual	1	3

Problems with the scale

1. Few items appear clearly related to fatigue

Only three of the eleven items on the scale (#1, #2 and #5) appear to be clearly related to fatigue. For the rest, the scale assumes that memory problems, speech errors, sleepiness/drowsiness, muscle weakness and so on are indicators of fatigue, and that the more such symptoms a patient reports, the greater their overall fatigue. These assumptions are untested and their basis is unclear.

The item on ‘problems starting things’ is particularly puzzling. It appears to be probing for lassitude, a common symptom in depression. Indeed, similar items appear in several depression scales, such the Montgomery-Asberg Depression Scale (Montgomery & Asberg, 1979). The relationship of lassitude to fatigue outside the context of depression is unknown.

Chalder et al. (1993) defined ‘caseness’ as a bimodal score of 4 or more on the CFQ, which means that a patient could be defined as a fatigue case if their only symptoms were difficulties in concentrating, making slips of the tongue, word-finding, and having memory problems. This appears to be entirely inappropriate, since it is unclear whether any of these symptoms are effective at discriminating between those with fatigue and other types of complaints (for example, mild cognitive impairment).

The lack of obvious or validated relevance to fatigue of the majority of items on the scale would, on its own, appear to make the CFQ unfit for purpose as a fatigue scale.

2. Focus on change in fatigue rather than intensity

The CFQ asks patients who have been feeling tired for a long while to rate their fatigue compared to when they were last well. It does not take ‘no fatigue’ as its baseline.

For ME/CFS patients – who, by definition, must have been ill for some time in order to achieve a diagnosis – this means remembering how it felt to be well. Patients may have been unwell for anything from several months to several decades and their recollection may well not be accurate.

An added source of confusion is that respondents are told to compare themselves to ‘when [they] were last well’, but the response options ask whether respondents are having problems ‘less/more than usual’. ‘Usual’ to a patient with a chronic illness such as ME/CFS is clearly not the same as ‘when [they] were last well’, and this conflicting wording is likely to lead to response errors.

The fact that respondents can mark each fatigue problem as occurring ‘less than usual’ is also problematic. It is unclear how anyone could feel less tired than when they were well, and therefore unclear what a respondent means when they select this option. Confusingly, a score of zero on the ‘Likert’ scoring of the CFQ is therefore not the base-point of the scale; a patient who scores 11/33 is no more fatigued than when they were last well, not one who scores 0/33. This makes interpretation of the scale difficult.

3. Arbitrary weighting of physical and mental components

Chalder et al. (1993) report a principal components analysis indicating that the scale has two major components – mental and physical fatigue. They combine these into a single score in the CFQ but the weighting of these components appears arbitrary, and is based simply on the number of questions of the two types in the questionnaire.

Even putting aside concerns about the validity (particularly of some of the ‘mental’ fatigue items), the consequence of combining mental and physical fatigue questions is that the scale is not necessarily monotonic, as an improvement in one form of fatigue could be accompanied by a worsening of the other type.

4. Incompatibility of scoring schemes

There are two alternative scoring methods. The ‘bimodal’ method assigns a 0 or 1 to each response, depending upon whether the complaint is present or absent (maximum score 11). The ‘Likert’ method rates each response from 0–3. The minimum score of 0 is given only for ‘less than usual’ (paradoxically *less* fatigue than before illness). A response of ‘no more than usual’ scores a higher 1, even though it indicates full recovery. Scores of 2 and 3 are given for ‘more than’ and ‘much more than’ respectively (maximum score 33).

The relationship between the two scoring schemes is far from transparent. One of them counts the number of symptoms, the other weights the intensity of the symptoms (and confusingly, gives extra credit for being *even better* than before the illness). Indeed, these two methods can generate contradictory findings: in the PACE trial, 23 showed improvement at the primary trial endpoint based on one scoring method but actually scored more poorly based on the other.¹

5. Failure to directly measure fatigue intensity

In the table on p.3 of the Fatigue Subgroup Materials section of the CDE Public Review document (NINDS/CDC, 2017), the CFQ is described as an index of ‘fatigue intensity’. As noted above, the bimodal scoring method simply yields a count of symptoms on a present/absent basis, while the ‘Likert’ version blends the number of symptoms with their intensity in a manner that is impossible to interpret from the total score.

6. Ceiling effect

Kindlon (2010) has pointed out that findings reported by Morriss et al. (1998) indicate that ceiling effects are likely when the CFQ is used. These investigators applied the questionnaire to 136 CFS patients in an outpatient clinic, and reported near-maximal scoring on six physical fatigue-scale items from the questionnaire, irrespective of which scoring method is used.

Clearly, it is important to know whether ME/CFS patients are experiencing worsening fatigue – or even harm – in response to an intervention. It is also important to know whether fatigue correlates with a potential biomarker. The CFQ’s ceiling effect is therefore a problem.

¹ Dataset available at https://sites.google.com/site/pacefoir/pace-ipd_foia-qmul-2014-f73.xlsx?attredirects=0, ‘readme’ file <https://sites.google.com/site/pacefoir/pace-ipd-readme.txt?attredirects=0>

Conclusions

We have here identified a number of serious problems with the CFQ, and note that the Fatigue Subgroup Draft Recommendations document also summarises some problems with it (p.33, our bolding):

Scoring:

This instrument can be scored in two ways: Bimodal and Likert scoring. It appears that the choice of scoring method may result in significant differences in interpretation of outcomes. (Rebecca Goldin. Sense About Science USA. March 21, 2016 <http://www.senseaboutscienceusa.org/pace-research-sparked-patientrebellion-challenged-medicine/>). **This will need to be further researched.**

Thresholds have been reported for both methods. (Bimodal: Case (≥ 4) vs. non-case (> 4) Mean score = 9.14 (SD 2.73) and 3.27 (SD 3.21) for Community sample. Mean “Likert” score 24.4 (SD 5.8) and 14.2 (SD 4.6)). However, the study referenced for these thresholds in the Chalder instrument required patients to meet either Oxford or Fukuda. As NIH’s ME/CFS Pathways to Prevention report noted, Oxford could have selected patients with other fatiguing conditions. Thus, it is difficult to know if these thresholds apply to ME/CFS cohorts. **Further research is needed.**

We are pleased to see these problems acknowledged, but concerned to see a call for further research on a questionnaire which appears unfit for purpose, and which is unlikely to become so with even major modification.

We would much prefer to see a questionnaire developed from the ground up: one that begins with researchers conducting a narrative interview, and then identifies items worth including on the basis of their ability to discriminate severely fatigued individuals from healthy ones. Perhaps one already exists and is being considered – we do not know the wider literature – but it is clearly not the CFQ.

We are pleased also to see (p.6 of the document) that the Fatigue Subgroup is aware that a challenge in assessing fatigue in ME/CFS is not only symptom variability, but also that symptoms are exertion-dependent. It is perfectly possible for a patient who is very severely disabled by ME/CFS to experience little fatigue most of the time because they are pacing themselves and restricting their activities to remain below their fatigue-triggering threshold.

We are grateful for the opportunity to contribute to the development of common data elements for our disease and will follow the work on this with great interest.

Wilshire, C.E., McPhee, G., and the Science for ME CFQ working group, January 22, 2018

References

- Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, Wallace EP. Development of a fatigue scale. *J Psychosom Res.* 1993;37(2):147-53.
- Kindlon T. Data on the level of maximal scoring (on the Chalder Fatigue Scale) would be useful. *BMJ* 2010;340:c1777. Available at: <http://www.bmj.com/rapid-response/2011/11/02/data-level-maximal-scoring-chalder-fatigue-scale-would-be-useful>
- Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry.* 1979;134:382–89.
- Morriss RK, Wearden AJ, Mullis R. Exploring the validity of the Chalder Fatigue scale in chronic fatigue syndrome. *J Psychosom Res.* 1998 Nov;45(5):411-7.
- NINDS/CDC. Public Review Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) Common Data Elements (CDE) Fatigue Subgroup Materials. 2017. Available at: https://www.commondataelements.ninds.nih.gov/Doc/MECF/03_Fatigue_Subgroup_CDE_Draft_Recommendations.pdf
- Wilshire C, Kindlon T, Matthees A & McGrath S. Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial. *Fatigue: Biomedicine, Health & Behavior* 2017;5.