Submitted to: Psychosomatic medicine

Psychometric properties and factor structure of a shortened version of the Cognitive Behavioural Responses Questionnaire (CBRQ)

Abbreviated running title: Psychometric properties of CBRQ

Elizabeth G. Ryan[1], PhD, Silia Vitoratou[1, 2, 3], PhD, Kimberley A. Goldsmith[1], PhD, and Trudie Chalder*[4,5], PhD

Dr Elizabeth Ryan, Dr Silia Vitoratou, and Dr Kimberley Goldsmith

[1]Biostatistics & Health Informatics Department, Institute of Psychiatry, Psychology and Neuroscience (IOPPN), King's College London, De Crespigny Park, Denmark Hill, London SE5 8AF, UK

[2] Centre of Implementation Science, Health Services and Population Research Department, Institute of Psychiatry, Psychology and Neuroscience (IOPPN), King's College London, De Crespigny Park, Denmark Hill, London SE5 8AF, UK

[3] Quality improvement team, South London and Maudsley NHS Foundation Trust, London, UK

[4]Academic Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience (IOPPN), King's College London, Weston Education Centre, 10 Cutcombe Rd London SE5 9RJ, UK

[5]South London and Maudsley NHS Foundation Trust, London, UK

*Professor Trudie Chalder (Corresponding author)

trudie.chalder@kcl.ac.uk

Phone: +44 20 7848 0406

Number of tables: 5

Number of figures: 0

Conflicts of Interest and Source of Funding:

All authors were in part supported by the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Psychology & Neuroscience, King's College London. TC has received royalties from Sheldon Press and Constable and Robinson. Funding for the PACE trial was provided by the Medical Research Council, Department for Health for England, The Scottish Chief Scientist Office, and the Department for Work and Pensions. We acknowledge the other investigators on the PACE trial, Peter D. White and Michael Sharpe, and the help of the PACE Trial Management Group. The views expressed in this article are those of the authors and not necessarily those of the NIHR or the NHS. ER, SV and KG do not have any conflicts of interest.

Abstract

Objective: Symptoms of chronic fatigue syndrome (CFS) can be perpetuated by cognitive and behavioural responses to the illness. We aimed to determine the factor structure, reliability and validity of the 40-item Cognitive Behavioural Responses Questionnaire (CBRQ) using data gathered from CFS patients. We also propose a short version CBRQ for greater clinical utility.

Methods: The psychometric analysis was performed on datasets drawn from two sources: a clinical service for CFS patients (N=576) and the PACE randomised controlled trial (RCT) of CFS treatments (N=640). An exploratory factor analysis (EFA) was conducted on the clinical dataset and a confirmatory factor analysis (CFA) was performed on the RCT dataset. Using these results, a short version of the CBRQ was proposed. Reliability, metric invariance across age and sex, and construct validity were assessed.

Results: The EFA (relative Chi-square 2.52; RMSEA 0.051; CFI 0.964; TLI 0.942) and CFA (relative Chi-square 4.029; RMSEA 0.069; CFI 0.901; TLI 0.892) revealed that eight factor models fitted the data well. Satisfactory Cronbach's alpha values were obtained for the final subscales (≥0.76). The shortened CBRQ was obtained by removing items that cross-loaded onto other factors and/or were the lowest loading items in each factor. The shortened CBRQ contained 18 items which had high factor loadings, good face-validity and reliability (Cronbach's alpha 0.67-0.88).

Conclusions: The CBRQ, long and short versions, are reliable and valid scales for measuring cognitive and behavioural responses of patients with CFS. Further research is needed to examine the utility of the CBRQ in other long-term conditions.

Keywords: Psychometrics, Reliability, Validity, chronic fatigue syndrome, CBRQ, Instrument development

Acronyms

AIC = Akaike information criterion

BIC = Bayesian information criterion

CBRQ = Cognitive and Behavioural Response Questionnaire

CBT = Cognitive behavioural therapy

CFA = Confirmatory factor analysis

CFI = Comparative fit index

CFQ = Chalder fatigue scale

CFS = Chronic fatigue syndrome

EFA = Exploratory factor analysis

HADS = Hospital Anxiety and Depression Scale

ME = Myalgic encephalomyelitis

MIMIC = Multiple indicator multiple cause

ML = Maximum likelihood

PACE = Pacing, graded activity, and cognitive behaviour therapy: a randomised evaluation

RMSEA = Root mean square error of approximation

RCT = Randomised controlled trial

SF36 = Medical Outcomes Study Short Form Health Survey

TLI = Taylor–Lewis Index

WLSMV = Weighted least squares means and variance adjusted estimator

WSAS = Work and social adjustment scale

INTRODUCTION

Chronic fatigue syndrome (CFS), or myalgic encephalomyelitis (ME), is a disorder that is characterised by severe and debilitating fatigue that is not alleviated by rest, persists for more than six months and has no identified medical cause (1). CFS patients often have difficulty working and recovery without treatment is uncommon (2, 3).

Most studies have found that CFS patients do not have significant differences in physiological test results compared to controls (e.g., 4, 5). Shortly after its recognition as an illness (6, 7), clinical researchers postulated a cognitive behavioural model of CFS that hypothesised that CFS symptoms are perpetuated by the patients' cognitive and behavioural responses to the illness (8, 9). Some of the unhelpful cognitive responses in patients with CFS include excessive focusing on fatigue symptoms (10), and catastrophic interpretations of symptoms which involves worrying about worst-case potential consequences (11). These cognitive responses are also associated with behavioural responses in CFS patients, which include reducing or avoiding activities to alleviate fatigue (12, 13), or alternating between high levels of activity when feeling well and resting excessively in response to symptoms. Activity avoidance in some CFS patients can be associated with the belief that activity can exacerbate their symptoms or it may also result from the embarrassment that they feel when experiencing fatigue symptoms in a social context.

The therapies that have been shown to be effective in CFS, such as cognitive behavioural therapy (CBT) (14) are based on identifying and targeting unhelpful cognitive and behavioural patterns in CFS patients (9). It follows that accurate measurement of these factors is critical to the assessment of whether treatments are working as expected and to the refinement of treatments, so that they have the maximum possible effect on changing beliefs and behaviours that perpetuate CFS symptoms.

The Cognitive and Behavioural Response Questionnaire (CBRQ) is a 40-item self-rated questionnaire that was designed to measure these cognitive and behavioural responses to patients' illness symptoms. In the development phase the scale was found to have five cognitive subscales and two behavioural subscales (15). It was subsequently used to assess symptom beliefs in patients with CFS (16, 17, 18). The cognitive subscales include fear avoidance, catastrophising, damaging beliefs, embarrassment avoidance, and symptom focusing. Four subscales assess the interpretation of symptoms, whilst symptom focusing assesses the attentional focus towards symptoms. The behavioural subscales consist of all-or-nothing behaviour and avoidance/resting behaviour. Table S1 in the Supplemental Digital Content describes the CBRQ items that make up the subscales. These subscales were based on a previously completed preliminary analysis, which was reported in conference proceedings (15). The aims of this paper were to see if the seven factor subscale structure (15) would be confirmed using new data sources, or whether a different subscale structure would be more appropriate, and to formally assess the validity of the CBRQ items. No formal psychometric analysis of the CBRQ has been published as yet. We also propose a shorter version of the CBRQ, based on the psychometric analysis results of the full questionnaire that retains cognitive and behavioural items and test the reliability and validity of this short version. Through the development and implementation of this short version of the CBRQ, we hope to reduce participant burden, reduce the risk of having random responses, and also obtain a questionnaire with as strong indicators as possible (by removing problematic items).

METHODS

Participants

CBRQ data were drawn from two sources: a routine clinical service and the pacing, graded activity, and cognitive behaviour therapy: a randomised evaluation (PACE) trial ((14); ISRCTN54285094). The CBRQ is a self-report questionnaire that the participants filled out on paper. Other measurements, such as the 36-item Medical Outcomes Study Short Form Health Survey (SF36) (19), Work and Social Adjustment Scale (WSAS) (20), Chalder Fatigue Scale (CFQ) (21), and anxiety and depression measured using the Hospital Anxiety and Depression Scale (HADS) (22) were also collected. For both datasets, only baseline measures were used in these analyses to avoid treatment effects.

Clinical dataset

Routine de-identified screening data were collected on 728 adult patients (≥18 years) that were assessed at the Chronic Fatigue Research and Treatment Unit in London, UK, between November 2007 and January 2014. These patients were selected in accordance with the Oxford criteria for diagnosing chronic fatigue syndrome (7). All participants were medically assessed by the specialist clinic doctors to exclude alternative diagnoses (23).

Patients were excluded from analysis if they: did not have CBRQ data (n=64); did not have a diagnosis of CFS (n=70); had bipolar (affective) disorder (n=4), an eating disorder (n=1), seizures (n=2), or cancer/were receiving chemotherapy (n=5) (1). Six entries were removed as they were duplicates of patients (original entries used). After inclusion/exclusion criteria were imposed, the sample was reduced to 576 patients.

This data was collected as part of a clinical audit of routinely collected outcomes; an Audit and Service Evaluation Project Proposal Form was submitted and approved by South

London and Maudsley's Psychological Medicine Clinical Academic Group Audit Committee and the clinical governance department, part of King's Health Partners.

RCT/PACE dataset

Data were collected on 641 patients that were recruited into the PACE study (14) between March 18, 2005, and November 28, 2008. These patients were aged 18 years or older and were recruited from consecutive new outpatients attending six specialist CFS clinics in the UK National Health Service. The patients fulfilled the Oxford criteria for CFS, and specialist clinic doctors assessed the participants to exclude alternative diagnoses. The PACE study was approved by the West Midlands Multicentre Research Ethics Committee (MREC 02/7/89) and was the largest UK trial to date of CFS treatments. PACE included three therapies and one medical treatment for CFS. The main results of the trial have been reported elsewhere (14). One of the PACE participants withdrew consent for use of their data, so the final dataset was n = 640.

Measures

The CBRQ was used to assess the patients' cognitive and behavioural responses to their symptoms. The development of the questionnaire is discussed elsewhere (Moss-Morris & Chalder, in preparation).

Each CBRQ item is measured on a five-point Likert scale, scored from 0 (strongly disagree) to 4 (strongly agree), where a higher subscale score indicates more unhelpful cognitions and behaviours. Two of the items, FA2 and FA9, are reverse-scored. To calculate the totals, items FA2 and FA9 must first be reverse-coded, and the items in each of the subscales (Table S1 Supplemental Digital Content) are added together to

create the subscale scores. Boxplots for the original subscale structures in each data set are displayed in Figure S1 in the Supplemental Digital Content.

The SF36 (19), WSAS (20), CFQ (21) and HADS (22) were also used in the analysis to determine construct validity for the CBRQ. The SF36 health survey is a 36-item self-report survey of patient health, where a low score indicates greater impairment. The WSAS is a five-item self-report scale that measures impairment in work, home management, social activities, private leisure activities and relationships due to an identified problem. The CFQ is an 11-item self-report scale that measures symptoms of physical and mental fatigue. From this scale, a total score or a bimodal score can be obtained. The HADS is a 14-item self-report instrument for detecting states of depression and anxiety in patients with medical illnesses. For the WSAS, CFQ and HADS high scores indicate greater impairment. These measures were collected in both datasets.

Statistical Analyses

Exploratory and Confirmatory Factor Analyses

The factor structure of the CBRQ was assessed using factor analysis techniques for categorical data. Since no formal psychometric analysis of the CBRQ has been published, we began with an exploratory factor analysis (EFA) to determine possible factor structures. EFA was conducted with the clinical CFS dataset and confirmatory factor analysis (CFA) was performed with the RCT dataset. That is, the clinical dataset was used as a “learning” sample for EFA and the RCT dataset was used as the “testing” sample for CFA.

When performing the EFA, 6-8 factor structures were initially examined as these produced fairly parsimonious models with good model fit and interpretability, and reflected previous results (15). Eigenvalues for the sample correlation matrix, scree plots and

parallel analysis for categorical data (24) (using the package *random.polychor.pa* in R (25)) were also used to confirm this decision (results presented in Supplemental Digital Content, Figures S2 and S3). Initially all 40 items were included in the EFA. The EFA was used to determine the factor structures, and reliability analyses were performed on these factor structures.

Once the EFA was completed, CFA was performed with the RCT dataset to test these factor structures. A CFA using the 7-factor structure that was previously proposed (15); (Table S1, Supplemental Digital Content) was also performed. The EFA and CFA were performed in Mplus (Version 7.4; (26)) to allow for handling of the categorical Likert-scored data. For the EFA, GEOMIN rotation and the weighted least squares means and variance adjusted estimator (WLSMV) were used as recommended by (27). For the CFA, WLSMV was also used for estimation. We also employed the maximum likelihood (ML) estimator in CFA so that the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values could be obtained in Mplus for fit comparison of non-nested models.

Goodness-of-fit

Model fit was assessed and compared using the relative Chi-square value (χ^2/df, where values close to 2 indicate a good fit) (28), the root mean square error of approximation (RMSEA, where values less than 0.08 are required for an adequate fit) (29), the comparative fit index (CFI, values above 0.9 indicate a good fit) (30), the Taylor–Lewis Index (TLI, values above 0.9 indicate a good fit) (31), as well as the factor loadings and face-validity of the factors.

Item Reduction

To obtain the shortened version of the CBRQ, we removed items that had the lowest factor loadings (in each factor) and/or loaded saliently on more than one factor (cross-loading) according to EFA results. EFA was used to determine the factor structure of the

shortened scale (again, using the clinical dataset) and CFA was used to confirm the proposed shortened scale (using the RCT dataset). The final shortened version of the CBRQ was chosen based on the item loadings, goodness-of-fit statistics, reliability indices, and the face-validity of the items in the factors (according to CFS experts).

Reliability and Validity

Reliability analyses, based on Cronbach's alpha if item deleted and item-total correlation (within each factor), were conducted in SPSS (Version 22; SPSS, Inc, Chicago, IL, USA). Metric invariance across age (continuous) and sex, as well as dataset source (clinical or RCT dataset) was assessed using the multiple indicator multiple cause (MIMIC) structural equation model (32) in Mplus (Version 7.4). Construct validity was assessed by examining the correlations between the CBRQ total score and the SF36 total score, WSAS, CFQ, HAD total, HAD anxiety, and the HAD depression score. The results of the MIMIC analyses and construct validity are only presented for the shortened scale; the results for the full scale are available upon request.

To check for the consistency in the factor structure between the two datasets, we pooled the RCT and clinical CFS datasets into one dataset and randomly split them into two datasets, and performed EFA on one sample and CFA on the other.

All data available were used in these complete case analyses – no imputation of missing data was performed. In the clinical dataset, 0.7-1.7% of the values were missing in the CBRQ items. In the PACE dataset, 0.2-0.6% of the values were missing in the CBRQ items.

RESULTS

Summary of datasets

[Table 1 here]

The demographics of the RCT and clinical CFS datasets were quite similar in terms of the patients' age, sex, and marital status (Table 1). There were some differences between the patients in terms of work status ($P < 0.001$) – more patients were temporarily unable to work in the RCT dataset compared to the clinical dataset (14.4% vs 23%), whereas the clinical dataset had more patients that were permanently unable to work due to illness (16% vs 3%). There were also more unemployed individuals in the RCT dataset (8.7% vs 23.6%). There were also significant differences in ethnicity and highest educational qualifications between the two datasets. The duration of illness was generally longer in the clinical dataset compared to the RCT dataset (median 48 vs 31.5 months; $P < 0.001$). An inclusion criteria for the PACE study was that patients had a CFQ score ≥ 6; 473/576 (82.1%) patients in the clinical dataset met this criteria. The mean CFQ score (bimodal) was 9.3 (SD=2.8) in the clinical dataset, and the mean in the RCT dataset was 10.3 (SD=1.2) and a significant difference was found between these means ($P < 0.001$). PACE patients were also required to have an SF36 score ≤60, which was later increased to ≤65; 348/576 (60.4%) patients had SF36 ≤60 in the clinical dataset and 377/576 (65.5%) patients had SF36 ≤65. The mean SF36 value in the clinical dataset was 46.9 (SD=26.4); the mean SF36 in the RCT dataset was 38 (SD=15.7) ($P < 0.001$). Due to trial entry criteria, the RCT patients had worse fatigue and disability than the clinical CFS patients.

Regarding distribution of scores on the CBRQ subscales (scored using the original factor structure), the scores for the damage and embarrassment avoidance subscales were similar between the two datasets (Table 2); the fear avoidance, catastrophising, symptom focusing, all-or-nothing and avoidance/resting subscales had significant differences

between the two datasets in terms of their distribution. In particular, the median scores for the all-or-nothing and avoidance/resting subscales were higher in the RCT dataset (Table 2).

[Table 2 here]

EFA and CFA results

The eigenvalues for the full sample correlation matrix indicated that up to eight factors would be sufficient (see Figure S2, Supplemental Digital Content). Parallel analysis also indicated that up to eight factors should be considered (see Figure S3, Supplemental Digital Content – this is the point immediately before where the plots of the observed (Polychoric correlation Empirical FA line) and simulated eigenvalues cross one another (34)).

EFAs using routine clinical data revealed that the best fitting model that included all items was the eight factor model (relative Chi-square 2.52; RMSEA 0.051 (90% CI: 0.048, 0.055); CFI 0.964; TLI 0.942). The seven factor model also produced a good model fit (relative Chi-square 3.10; RMSEA 0.060 (90% CI: 0.057, 0.064); CFI 0.946; TLI 0.920).The factor loadings and reliability for the eight factor model can be found in Table 3. The factor loadings for the seven factor model may be found in the Supplemental Digital Content (Table S2). Since the fit indices are similar between the seven and eight factor models, one could choose the seven factor model on the grounds of parsimony. However, we argue that the eight factor model better explained the different subscales of the CBRQ since it separated the avoidance and the resting items into two factors and was more interpretable. Each factor in the eight factor model had satisfactory reliability (Cronbach's alpha >0.7; (35)) and the item-total correlations (within each factor) were between 0.44 and 0.78 (data not shown). Summary statistics for the eight subscales obtained from the EFA and CFA and the total CBRQ score are given in Supplemental

Digital Content, Table S3. Boxplots for each of the subscales in each data set are given in Figures S1 and S4 in the Supplemental Digital Content.

Apart from the factor on which the items had their largest loading, the items L3, L4 and L11 loaded also on the fear avoidance subscale; the items EA3 and EA4 loaded on both the embarrassment avoidance subscale and the avoidance subscale; items SF1, SF3 and L10 loaded also on the catastrophising subscale; items C1 and C2 loaded also on the damage subscale; items C4 and C6 loaded also on the symptom focusing subscale.

[Table 3 here]

CFA using the RCT dataset demonstrated that the seven factor model (using the original subscale structure) produced the best model fit (relative Chi-square 3.867; RMSEA 0.067 (90% CI: 0.064, 0.070); CFI 0.905; TLI 0.897; AIC 60282.085; BIC 61258.805). The eight factor structure found during the EFA also produced a good model fit (relative Chi-square 4.029; RMSEA 0.069 (90% CI: 0.066, 0.071); CFI 0.901; TLI 0.892; AIC 60408.311; BIC 61416.250).

Similar results were obtained when the two datasets were pooled and randomly split into two datasets with one dataset used for the EFA and the other used for CFA (results available upon request).

Shortened version of CBRQ

We chose to focus on the factor loadings in the eight factor model as it had a more distinct factor structure. Based on the results of the EFA and CFA, we have proposed a

shortened version of the CBRQ. We wanted to develop a shorter version of the CBRQ so that it would take a shorter amount of time for the patients to fill out, provided that the psychometric properties of the CBRQ would be improved by removing some items.

Model fit

Initially we removed all items with the lowest loadings within each factor and/or had cross-loading in the eight factor model loadings (Table 3), and performed an EFA. We were interested in having a questionnaire with as strong indicators as possible, and so we then omitted the (remaining) weakest items in each factor so that we had at least three items in each factor, whilst ensuring minimal impact on reliability. The items removed were not considered essential according to TC, in terms of content validity. The factor structure for the final version of the shortened CBRQ is presented in Table 4, along with the Cronbach's alpha value of each subscale. Summary statistics for each subscale and the total score for the shortened CBRQ are presented in Supplementary Digital Content Table S4.

[Table 4 here]

The EFA showed that a six factor model fit the data well: relative Chi-square 1.10; RMSEA 0.013 (90%: 0.00, 0.029); CFI 0.999; TLI 0.999. The CFA also confirmed that this six factor model provided a good fit: relative Chi-square 2.60; RMSEA 0.050 (90% CI: 0.043, 0.057); CFI 0.983; TLI 0.978. There were no cross-loadings in the shortened questionnaire. The Cronbach's alpha values for the factors within the shortened questionnaire indicated that the factors were reliable as the majority of the values were above 0.7. Comparing Tables 3 and 4, it can be seen that the reliability is almost unchanged in the shortened version of the CBRQ, even though there was a reduction in

the number of items. It is not surprising that the Cronbach's alpha values have slightly decreased since it is a function of the number of items. The shortened version of the CBRQ was thought to have good face-validity by CFS clinicians. The six factor model (with 18 items) explained 67% of variance in the data, whereas the eight factor model (with 40 items) explained 60% of the variance in the data. The short version CBRQ is presented in Supplementary Digital Content Table S5.

Metric invariance (MIMIC models)

We ran the MIMIC models using the shortened questionnaire structure presented in Table 4. Initially, we ran a MIMIC model to see if there was metric invariance in the items for the dataset source (RCT or clinical dataset). That is, we used MIMIC models to look at whether the loadings of the items to their corresponding factors differed according to the dataset source for people with the same underlying trait score. We then ran MIMIC models within each dataset to see if the items' factor loadings differed for the individuals' age (continuous) or sex (for the same underlying trait score). The items that had metric non-invariance for dataset source are summarised in Table S6, and the items that had metric non-invariance for age and/or sex are summarised in Table S7 in the Supplementary Digital Content.

Based on the MIMIC analyses, there were a number of items in the dataset that appeared to have significantly different loadings across dataset source). The effect size from the MIMIC models were large for the AL3 and L7 items; moderate for the EA2, L2 items; and small for SF5 and AL1 (37, 38). Some items had significant metric non-invariance for sex or age (EA1 and L2 for sex in RCT dataset; FA10 and L7 for age in clinical dataset; L2 for age in RCT dataset). However in most cases (for instance, FA10) the effect size was small (37, 38). Future research is required to test the replicability of these effects.

Construct validity

Construct validity can be assessed by comparing the measure of interest to a similar (or different) measure. There are currently no other scales which specifically measure CFS patients' views/beliefs about their symptoms. Instead, we assessed evidence towards construct validity by examining the correlations between the shortened CBRQ total score and other measures of impairment: the SF36 total score, WSAS, CFQ, HAD total score, HAD anxiety score, and the HAD depression score (Table 5). The shortened version of the CBRQ had significant correlation with these impairment measures which were low to moderate in size.

[Table 5 here]

DISCUSSION

Summary of main findings

In this paper we investigated the subscale structure of the CBRQ to determine whether the scale originally developed by Moss-Morris and Chalder (in preparation) and used by (16, 17, 18) in CFS populations, was reliable and valid, or whether a different subscale structure was more appropriate. We also proposed a short version of the CBRQ which consisted of 18 items and was derived using the results from the EFA and CFA, as well as the expertise of CFS clinicians.

EFA was carried out on data from routine clinical practice (n=576) and then CFA was conducted using data collected as part of a large RCT (14) (n=640). Similar results were obtained from both the EFA and CFA where seven and eight factor models produced the best fits. The original subscale/factor structure proposed by Moss-Morris and Chalder (in

preparation) fit the data well. However, it appeared that the avoidance and resting items should be split into two factors since the resting items had high cross-loading in a seven factor model. This makes sense since at face value, these items appear to measure different constructs. Large Cronbach's alphas (34) were obtained for all the final subscales.

Although the catastrophising items have been used in previous papers (16, 17) and catastrophising was found to be one of the mediators of the effects of CBT and GET for CFS, we found these items to be problematic, with low factor loadings and evidence that they cross-loaded on to other factors. It may be that the items were not specific enough and/or overlapped with other constructs. In the proposed short version of the questionnaire we have removed these items since we wanted to obtain a questionnaire with as strong indicators as possible. The six factor model (with 18 items) explained 67% of variance in the data, whereas the eight factor model (with 40 items) explained 60% of the variance in the data. By removing problematic items, the latent structure became clearer.

We also removed the avoidance items from the short version of the questionnaire as they cross-loaded on the fear avoidance factor or the catastrophising factor. It is likely that these items are not required in the presence of the fear avoidance items. The subscales/factors in the short version had good reliability and the items had large factor loadings and no cross-loading.

MIMIC models indicated that there may be metric non-invariance in certain items in the short version, i.e., that the loadings of the items on their corresponding factors may differ according to age, sex, or dataset source for people with the same underlying latent trait score. The magnitude of the estimates can inform us as to the degree of metric non-invariance present. The effect size estimates for metric non-invariance for dataset source were quite large for some of the items (AL3 and L7), which means that the relationship between these items and the trait differs according to the dataset. This could be

explained by the fact that PACE patients had to meet certain criteria to be included in the clinical trial and these patients had more self-reported disability.

It is important to note that the effect size estimates for the items that had significant metric non-invariance across age and/or sex (Table S7, Supplemental Digital Content) are small (37), the largest odds ratio being 1.56. One could use the methods proposed by (38) to convert the log odds ratios to a Cohen's d value, and come to the same conclusion. This suggests that age and sex differences in the loadings might not be reproduced in smaller samples and should be explored in future research. It may be the case that females respond differently to men in their thinking and coping styles, and older patients may respond differently to younger patients due to experience. Future analyses could adjust for the items that had metric non-invariance for sex and/or age.

We used only baseline data in this study to focus on measurement avoiding the effects of treatment. Although we did not examine the post treatment measures of the CBRQ in this study, future studies will use longitudinal data analyses of the routine clinic dataset to investigate whether the item responses change over time (test-retest reliability). This has been investigated to some extent where mediation analyses were performed using the different subscales of the CBRQ as mediators on the primary outcome (measured by the CFQ and the physical function subscale of the SF36) using the RCT dataset (17) and a clinical CFS dataset (16).

Validity

When assessing construct validity, correlations with theoretically similar measures, should be high, whilst correlations with theoretically dissimilar measures should be low. The short version CBRQ total score had significant and moderate, positive correlations with the WSAS, as well as the HAD total, HAD anxiety and HAD depression scores, which makes

sense as increases in these measures correspond to more disability. The short version CBRQ had a small and positive correlation with the CFQ, which is not unexpected since CFQ measures fatigue, rather than patient beliefs. The short version CBRQ total score had a small, negative correlation with the SF36 total score, which makes sense since a lower SF36 score corresponds to more disability. The SF36 measures physical function, rather than beliefs about symptoms, and so it is not unexpected that the SF36 and (short version) CBRQ have small correlations.

The CBRQ scale has good face-validity and it also has good predictive validity since it changes over time with CBT (16). The scale items were based on a model of understanding symptom perception in which cognitive and behavioural responses are all important in determining outcomes in terms of symptom severity and disability (9). Two recent studies which examined the role of cognitive behavioural responses as mediators during the process of CBT support the use of this measure as a predictive tool (16, 17).

The clinic attenders and those who agreed to take part in the RCT may not be representative of the wider population of people with CFS. Those who agreed to take part in the RCT differed to those recruited in the clinic in that the routine clinic attenders had been ill for longer and were more ethnically diverse. Ingman et al. (18) found that black and minority ethnic individuals had more extreme baseline cognitive behavioural responses, but that these differences did not affect treatment outcome.

This paper was not designed to assess aetiology of CFS; its main aim is to assess the reliability and validity of the CBRQ in people with CFS. The CBRQ is designed to assess cognitive behavioural responses in relation to symptoms and one cannot make assumptions about causality. As fatigue is ubiquitous these responses could be important in the context of other diseases.

The short version of the CBRQ was developed using psychometric analyses performed on two datasets. This short version has not yet been piloted on patients, and so future analyses should re-validate this proposed short version of the CBRQ with an independent sample of patients that represent the same target population.

In summary, the long and short versions of the CBRQ, are reliable and valid scales for measuring cognitive and behavioural responses of patients with CFS. Further research examining the utility of the CBRQ in other conditions would be welcome.

REFERENCES

1. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: A comprehensive approach to its definition and study. Annals of Internal Medicine 1994; 121: 953–959
2. Wessely S, Hotopf M, & Sharpe M. Chronic Fatigue and its Syndromes. Oxford: Oxford University Press; 1998
3. Cairns R, Hotopf M. A systematic review describing the prognosis of chronic fatigue syndrome. Occupational Medicine 2005; 55: 20-31.
4. Wessely S. Chronic fatigue syndrome. J Neurol Neurosurg Psychiatry 1991; 54(8): 669-671.
5. Gibson H, Carroll N, Clague JE, Edwards RHT. Exercise Performance and Fatiguability in Patients with Chronic Fatigue Syndrome. Journal of Neurology Neurosurgery and Psychiatry 1993; 56(9): 993-998.
6. Holmes GP, Kaplan JE, Gantz NM, Komaroff AL, Schonberger LB, Straus SE, Jones JF, Dubois RE, Cunningham-Rundles C, Pahwa S, Tosato G , Zegans LS, Purtilo DT, Brown N, Schooley RT, Brus I. Chronic fatigue syndrome: a working case definition. Ann Intern Med 1998; 108(3): 387-389.

7. Sharpe M, Archard LC, Banatvala JE, Borysiewicz LK, Clare AW, David A, Edwards RH, Hawton KE, Lambert HP, Lane RJ. A report -- chronic fatigue syndrome: guidelines for research. J R Soc Med 1991; 84(2): 118-121.
8. Wessely S, David A, Butler S, Chalder T. Management of chronic (post-viral) fatigue syndrome. J R Coll Gen Pract 1989; 39(318): 26-29.
9. Deary V, Chalder T, Sharpe M. The cognitive behavioural model of medically unexplained symptoms: a theoretical and empirical review. Clinical Psychology Review 2007; 27: 781–797.
10. Hughes A, Hirsch C, Chalder T, Moss-Morris R. Attentional and interpretive bias towards illness-related information in chronic fatigue syndrome: a systematic review. Br J Health Psychol. 2016; doi.org/10.1111/bjhp.12207 [Epub ahead of print]
11. Lukkahatai N, Saligan LN. Association of catastrophizing and fatigue: a systematic review. J Psychosom Res 2013; 74(2): 100-109.
12. Chalder T, Power MJ, Wessely S. Chronic fatigue in the community: 'a question of attribution'. Psychological Medicine 1996; 26: 791–800.
13. Nijs J, De Meirleir K, Duquet W. Kinesiophobia in chronic fatigue syndrome: assessment and associations with disability. Archives of Physical Medicine and Rehabilitation 2004; 85: 1586–1592.
14. White PD, Goldsmith KA, Johnson AL, Potts L, Walwyn R, DeCesare JC, Baber HL, Burgess M, Clarke LV, Cox DL, Bavinton J, Angus BJ, Murphy G, Murphy M, O'Dowd H, Wilks D, McCrone P, Chalder T, Sharpe M. 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; 377: 823–36.
15. Moss-Morris R & Chalder T. Illness representations: where to from here? 16th Conference of the European Health Psychology Society, 2003; Kos, Greece.

16. Stahl D, Rimes KA, Chalder T. Mechanisms of change underlying the efficacy of cognitive behaviour therapy for chronic fatigue syndrome in a specialist clinic: a mediation analysis. Psychol Med 2013;44:1331-44.
17. Chalder T, Goldsmith KA, White PD, Sharpe M, Pickles AR. Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. Lancet Psychiatry 2015;2(2): 141-152
18. Ingman T, Ali S, Bhui K, Chalder T. Chronic fatigue syndrome: comparing outcomes in White British and Black and minority ethnic patients after cognitive–behavioural therapy. British Journal of Psych 2016; 209 (3): 251-256
19. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.
20. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. British Journal of Psychiatry 2002;180: 461–464.
21. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, Wallace EP. Development of a fatigue scale. Journal of Psychosomatic Research 1993; 37:147–153.
22. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica 1983; **67** (6): 361–370.
23. Reeves WC, Lloyd A, Vernon SD, Klimas N, Jason IA, Bleijenberg G, Evengard B, White PD, Nisembaum R, Unger ER. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. BMC Health Serv Res 2003; 3: 25.
24. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965; 30: 179-185
25. Presaghi F, Desimoni M. Random.polychor.pa: A Parallel Analysis with Polychoric Correlation Matrices. R package version 1.1.4-2. 2016.

26. Muthén LK, Muthén BO. Mplus User's Guide. 6th edition. Los Angeles: Muthén & Muthén; 1998-2011
27. Muthén B, du Toit SHC, Spisic D. Robust inference using weighted least squares and quadratic estimating equations in latent variable modelling with categorical and continuous outcomes. Unpublished technical report. 1997
28. Hoelter JW. The analysis of covariance structures: goodness of fit indices. Sociol. Methods Res 1983; 11: 325–344
29. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS editors. Testing Structural Equation Models. Newbury Park: Sage; 1993.
30. Bentler, PM. Comparative fit indexes in structural models. Psychol. Bull. 1990; 107:238–246
31. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychol. Bull. 1980; 88: 588–606
32. Muthén, B. A method for studying the homogeneity of test items with respect to other relevant variables. J. Educ. Behav. Stat. 1985; 10: 121–132
33. Horn J. A rationale and test for the number of factors in factor analysis. Psychometrich 1965; 30: 179-185.
34. Nunnally JC. Psychometric theory. 2nd ed. New York: McGraw-Hill; 1978.
35. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the Use of Exploratory Factor Analysis in Psychological Research. Psychological Methods 1999; 4(3): 272-299.
36. de Castro MA , Baltar VT, de Castro Selem SS, Marchioni DML, Fisberg RM. Empirically derived dietary patterns: interpretability and construct validity according to different factor rotation methods. Cadernos de Saude Publica 2015; 31(2): 298-310.

37. Chen H, Cohen P, Chen, S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. Communications in Statistics – Simulation and Computation 2010; 39(4): 860-864
38. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. Psychological Bulletin 1995; 117: 167-78.

Table 1. Demographics of patients in clinical and RCT datasets

Variable	Clinical dataset (n=576)	RCT dataset (n=640)	Comparison
Sex	Male: 148 (25.7%) Female: 428 (74.3%)	Male: 145 (22.7%) Female: 495 (77.3%)	P = 0.23
Age	39 (18-80)	38 (18-77)	P = 0.40
Time with CFS (months)	48 (2.3-492)	31.5 (5-382)	P < 0.001
Ethnicity	White: 420 (72.9%) Other: 81 (14%) None given: 23 (4%) Missing: 52 (9%)	White: 595 (93%) Other: 40 (6.2%) None given: 5 (0.8%)	P < 0.001
Marital status	Single: 251 (43.6%) Married/living together: 261 (45.3%) Separated/divorced/widowed: 55 (9.5%) Missing: 9 (1.6%)	Single: 249 (38.9%) Married/living together: 332 (51.9%) Separated/divorced/widowed: 59 (9.2%)	P = 0.12
Highest educational qualifications	None: 16 (2.8%) Secondary school: 130 (22.6%) University: 290 (50.3%) Other: 129 (22.3%) Missing: 11 (1.9%)	None: 24 (3.8%) Secondary school: 237 (37%) University: 312 (48.8%) Other: 67 (10.5%)	P < 0.001
Work status	Working (fulltime, part time, casual): 243 (42.2%) Student: 33 (5.7%) Unable to work/study temporarily due to illness: 83 (14.4%) Unable to work permanently due	Working (fulltime, part time, casual): 240 (37.5%) Student: 20 (3.1%) Unable to work/study temporarily due to illness: 147 (23%) Unable to work permanently	P < 0.001

	to illness: 92 (16%)	due to illness: 19 (3%)
	Unemployed: 50 (8.7%)	Unemployed: 151 (23.6%)
	Retired (due to age): 19 (3.3%)	Retired (due to age): 3 (0.5%)
	Looking after the home: 29 (5%)	Look after the home: 20 (3.1%)
	Missing: 27 (4.7%)	Other: 40 (6.3%)

Data are presented as numbers (%) or medians (range).

Table 2. Medians and ranges of clinical measures in clinical dataset and RCT dataset

Subscale	Potential range	Clinical dataset (n=576)	RCT dataset (n=640)	Comparison Mann-Whitney U-test p-value
Fear avoidance	0-24	14 (1-24)	15 (2-24)	P < 0.001
Catastrophising	0-16	8 (0-16)	8 (0-16)	P = 0.015
Damage	0-20	11 (0-20)	11 (0-20)	P = 0.42
Embarrassment avoidance	0-24	12 (0-24)	12 (0-24)	P = 0.44
Symptom focusing	0-24	14 (0-24)	13 (0-24)	P = 0.011
All-or-nothing	0-20	9 (0-20)	14 (4-20)	P < 0.001
Avoidance/resting	0-32	13 (1-31)	19 (5-31)	P < 0.001

Table 3. Factor loadings (in descending order for each factor) for eight-factor model obtained from EFA (GEOMIN Rotation) performed on the clinical dataset, n=576

	Factor							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
FA1	**0.723**	0.001	0.086	-0.001	0.07	-0.002	0.03	-0.066
FA12	**0.697**	-0.054	-0.046	0.082	0.01	0.089	-0.036	0.057
FA2	**0.571**	0.010	-0.01	0.049	-0.018	-0.114	-0.125	0.030
FA17	**0.564**	0.150	-0.028	0.044	0.113	0.051	0.029	0.019
FA3	**0.547**	0.121	-0.046	0.088	0.036	-0.023	0.235	-0.027
FA14	**0.432**	0.068	0.085	-0.029	-0.114	0.095	0.003	-0.020
FA10	0.016	**0.818**	0.055	-0.002	-0.040	0.062	-0.064	0.021
FA4	0.052	**0.808**	0.047	-0.019	-0.050	-0.046	-0.045	0.008
FA9	-0.192	**0.709**	-0.005	0.091	0.033	-0.023	0.011	0.097
FA15	0.161	**0.615**	-0.116	-0.024	0.168	0.038	0.038	-0.001
FA16	0.159	**0.412**	0.079	-0.002	0.123	0.030	0.135	-0.089
SF5	-0.009	0.035	**0.858**	-0.021	-0.004	0.029	0.113	-0.022
SF12	-0.048	0.056	**0.842**	-0.027	-0.004	-0.040	0.131	-0.063
SF1	0.087	-0.095	**0.763**	***0.306***	0.028	-0.023	-0.076	0.058

SF3	0.078	-0.036	**0.662**	***0.346***	-0.042	0.028	-0.007	0.011
SF9	-0.016	0.076	**0.660**	0.008	0.012	0.055	0.094	-0.012
SF2	-0.067	0.170	**0.573**	0.179	0.092	0.003	-0.030	0.018
C2	-0.030	***0.354***	0.054	**0.731**	-0.005	-0.013	-0.009	-0.033
C1	0.068	***0.296***	0.022	**0.657**	-0.065	-0.007	0.055	0.049
C6	-0.019	0.003	***0.253***	**0.431**	0.047	-0.029	-0.021	0.018
C4	0.041	0.132	***0.260***	**0.350**	0.097	0.069	0.098	0.058
EA1	0.033	-0.057	0.067	-0.027	**0.909**	-0.024	-0.058	0.039
EA2	0.021	-0.017	0.076	-0.052	**0.872**	0.068	-0.037	-0.001
EA5	-0.004	0.011	-0.030	0.035	**0.855**	-0.007	0.027	0.031
EA6	-0.06	0.083	0.016	0.034	**0.653**	-0.053	0.146	-0.020
EA3	0.010	0.065	-0.020	0.216	**0.513**	0.031	***0.316***	0.001
AL3	0.006	-0.022	0.048	-0.047	-0.033	**0.885**	-0.040	0.054
AL2	0.008	0.041	-0.055	0.179	0.015	**0.822**	-0.018	-0.028
AL1	0.113	0.010	-0.004	0.073	0.021	**0.786**	-0.011	-0.054
AL4	-0.020	-0.042	0.011	-0.047	-0.013	**0.742**	0.076	0.183
AL5	-0.053	0.084	0.034	0.015	0.103	**0.511**	0.170	0.050

L11	***0.343***	-0.018	0.038	-0.054	-0.025	-0.065	**0.658**	0.101
L10	-0.037	-0.081	-0.003	***0.286***	0.065	0.067	**0.653**	0.033
EA4	0.003	-0.014	-0.009	0.304	***0.281***	0.058	**0.593**	-0.141
L4	***0.348***	0.044	0.016	-0.045	-0.042	-0.048	**0.570**	0.151
L13	0.063	0.016	0.087	0.066	-0.019	-0.009	**0.556**	0.035
L9	0.001	-0.005	0.003	0.079	0.044	0.030	0.009	**0.853**
L7	-0.001	-0.001	-0.028	0.103	0.007	0.039	-0.037	**0.750**
L2	0.000	0.118	-0.012	0.004	0.047	-0.027	0.077	**0.628**
L3	***0.270***	0.015	0.064	-0.114	-0.04	0.033	0.214	**0.496**
Cronbach's alpha*:	0.76	0.80	0.89	0.77	0.88	0.87	0.80	0.80

The questions for each of the item abbreviations are written out in full in Supplemental Digital Content 1. The highest factor loading for each item is boldfaced. Items that have cross-loading (35, 36) on a factor are both boldfaced and italicised in their loading value. *The Cronbach's alpha values were obtained using the (boldfaced) items that make up each subscale.

Table 4. Factor loadings for proposed shortened version of the CBRQ obtained from EFA (GEOMIN Rotation) performed on the clinical dataset, n=576

	Factor					
	1	**2**	**3**	**4**	**5**	**6**
FA1	**0.720**	-0.005	0.182	0.039	0.000	-0.036
FA2	**0.574**	0.057	-0.021	-0.047	-0.110	0.022
FA12	**0.710**	-0.003	-0.020	0.007	0.095	0.087
FA4	0.064	**0.849**	-0.008	-0.012	-0.037	-0.039
FA9	-0.128	**0.639**	0.042	0.054	-0.002	0.107
FA10	0.016	**0.839**	0.004	-0.023	0.074	-0.003
SF5	0.007	-0.006	**0.929**	-0.014	0.017	0.023
SF9	0.022	0.055	**0.712**	0.025	0.063	0.014
SF12	-0.019	-0.001	**0.912**	0.001	-0.059	-0.012
EA1	0.005	-0.009	-0.001	**0.914**	-0.029	0.001
EA2	-0.011	-0.001	0.033	**0.869**	0.056	-0.013
EA5	0.017	0.015	-0.020	**0.849**	0.000	0.046
AL1	0.077	-0.007	0.037	0.022	**0.824**	-0.045
AL2	0.014	0.041	-0.033	0.027	**0.891**	0.021
AL3	-0.027	-0.016	0.007	-0.033	**0.819**	0.040
L2	-0.023	0.115	-0.004	0.040	-0.014	**0.624**
L7	0.016	-0.027	0.007	-0.043	0.054	**0.762**

L9	0.026	-0.005	0.012	0.014	-0.014	**0.920**
Cronbach's alpha*:	0.67	0.79	0.85	0.88	0.86	0.79

The questions for each of the item abbreviations are written out in full in Supplemental Digital Content 1. The highest factor loading for each item is boldfaced. *The Cronbach's alpha values were obtained using the three items that make up each subscale.

Table 5. Correlation (P-value) between total score from proposed shortened version of CBRQ and selected measures

Variable	**Correlation (P) in clinical dataset (n=576)**
SF36 total score	-0.25 (<0.001)
WSAS	0.35 (<0.001)
CFQ	0.20 (<0.001)
HAD anxiety	0.46 (<0.001)
HAD depression	0.45 (<0.001)
HAD total score	0.52 (<0.001)