Reviewer's report

Title: The functional status and well being of people with ME/CFS and their carers

Version: 2 Date: 15 December 2010

Reviewer: Jin-Mann Lin

Reviewer's report:

1. Is the question posed by the authors well defined?
Yes, the question is well defined. The paper raised an important question, the quality of life in CFS/ME patients and their caregivers. Numerous studies have examined differences between caregivers and non-caregivers in perceived stress, depression, general subjective well-being, physical health, and self-efficacy; however, this paper adds to literature specifically on CFS/ME.

2. Are the methods appropriate and well described?
Yes, the methods are appropriate and well described. However, some revisions are required.
In the Methods section, the authors described that diagnosis confirmation was asserted by any of three diagnostic criteria: CDC-1994, Canada, and the Epidemiological Case Definition. It was not clear how these three case definition criteria were operationized. It is suggested that the authors add how they defined each definition in the context of including patients in for the study. This is especially important if the authors wish to make comparisons to patients/studies outside of the UK.

3. Are the data sound?
Overall, the data are sound; however, there are some revisions needed. From time to time, the authors described their cases as “well characterized”. However, there are some key information missing such as the duration of fatigue or CFS/ME illness and the age of illness onset. These factors could potentially explain the differences in standard mean scores with other studies. In the Methods section, the authors indicated that clinical variables were collected for confirmed cases. Since the cases were confirmed by any of three diagnostic criteria, it will be useful to describe the characteristics in each detailed criterion of the three criteria. In this way, the readers can have a better idea how the criteria was operationalized. In other words, if the case ascertainment was in vacuum how reliable conclusion could be drawn from the comparison pertaining to the different case ascertainment criteria. It is suggested that the authors included a table or add to Table 1 the breakdown of the CFS/ME cases by the three definitions.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
Yes, in general it adheres to the relevant standards for reporting; however, the authors should be more cautious of statistical significance rather than reporting any difference. Also, several studies have examined the minimal important differences in the SF-36 scales. This might be something to be included in the discussion section on the difference reported by the authors.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

Discussion and conclusion will have to fully address concerns in Methodology and Results sections to become well balanced.

6. Are limitations of the work clearly stated?

The authors have stated the strengths of the study and the limitation of self-reported SF-36 measures. The limitation of the small sample size for the “carers” group was stated in the discussion. Perhaps it will be easy to follow if organizing all the strengths and limitations of the study into a subsection.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

Yes, but some revisions are needed.

8. Do the title and abstract accurately convey what has been found?

In the Abstract, Methods need to be revised using statistical comparison rather than the layman term “on average” lower.

9. Is the writing acceptable?

Yes, it is acceptable. However, three page length of Discussions could be rewritten in grouping the similar theme and deliver more concise message. The discussions were conveyed effectively and often mixed with the results not previously presented in the Results section.

Major Compulsory Revisions

1. In the Methods section, some details are required to clarify the operation of case confirmation by three case definition criteria. For sensitivity and specificity, more appropriate methods are required. The authors also need to describe the caveat about the feasibility of Canadian CLINICAL working case definition and 1994 CDC REARCH case definition; especially the difference by the measure procedure in the samples from primary care setting or community dwellings. Regardless of insufficient data evidence, it is bothersome that the authors had discussed the specificity in several places in the Discussion section. Nonetheless, the investigation of specificity across the case definition criteria was not included as one of the study aims. If this is one of the study objectives, the authors should design a study which provides sound data for this specific objective and use statistical methods as appropriate to test the hypothesis.

2. In the “Data processing and analysis” paragraph, p. 7, the authors indicated that “We then compared the results with those for the US general population, gender-age and disease-specific norms [11].” The authors should be careful with scientific wording about “compare”as more statistically appropriate.
3. In Table 5, the SF-36 scale scores were compared between subjects meeting Canadian criteria and 1994 CDC criteria. However, the two subgroups were not exclusive. That is, some subjects could be confirmed cases by both Canadian and CDC criteria. In this case, the statistics reported in the table might not be robust. It would be more appropriate to divide subjects into the categories: 1). Meeting Canadian criteria only, 2). Meeting CDC-1994 criteria only, 3). Meeting both criteria.

Discretionary Revisions

1. Page 8. “Table 1 compares some characteristics of cases and their carers and shows the relationship between them.” The authors stated Table 1 for comparing some characteristics of cases and their carers while the table only summarize the frequency and percentage without p-value for chi-square or fisher exact test on categorical variables. Alternatively, replace “compares” with “describes”.

2. “Table 3 compares results by sex; in general men scored lower in the mental domain (particularly vitality), and women, in the physical domain (particularly physical functioning). “ Could this association be accounted by their co-morbid conditions? If the existing data allows, the authors should describe as part of case characteristics.

3. Please cite the reference for the source on SF-36 diseases-specific norms. It would be useful to list the sample size for each disease-specific norms in Table 4. Eg. Back pain/ sciatica (n=2635), Cancer (except skin) (n=253), etc. Another caveat is that the sample means from selected disease-specific studies are not quite comparable in the age inclusion to the current ME/CFS study.

4. Page 10. The authors stated that “unlike disease specific measures, it can be adequately used for comparisons between people with ME/CFS and healthy individuals and those with a range of other diseases.” Is this authors’ personal statement or it could be backed-up by literature? SF-36 has been used in various studies to compare the changes over time or treatment, across illness groups or case-control. It might not be disease- or illness-specific, but it can certainly pick up certain level difference in term of health-related quality of life (or functioning status).

5. Page 12. The authors indicated that “The ‘role physical’ scale was the most affected of all, suggesting this could be a suitable outcome measure in ME/CFS.” What is the data evidence leading to the statement? The norm-based scores in ME/CFS cases for Role Physical, Physical Functioning, General Health, Vitality, and Social functioning were 25.4 (SD=8.2), 27.7 (10.6), 28.3 (8.0), 28.4 (7.1), and 25.7 (9.8), respectively. There is no significant difference when comparing the mean scores.

Minor Essential Revisions

1. The 3rd paragraph in the Background section, please provide references to adjust the importance of examining caregivers’ health. Perhaps findings from studies for caregivers’ health in caring patients with chronic conditions.

Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in

2. Sample size/power calculation: Page 8. The authors stated “This sample size was adequate to detect SF-36 mean score differences of 0.5 standard deviation between sub-groups of similar size within the sample, with a power of 90% and a significance level of 0.05.” Please clarify what specific and # of sub-groups for mean score comparison. Eg. CFS/ME vs. carers. If more than two group comparison, what statistical test was considered for the sample size/power calculation for the multiple group comparison with p-value adjustment.

3. Throughout the Results section, the authors used the wording such as “on average”, “consistently”, “much”, “slightly”, “even”, and “actually” lower. Alternatively, the authors could determine the statistical significance using the sample mean, standard deviation, and sample size to calculate the t-statistics and p-value between the group mean from the current ME/CFS and that from selected disease-specific mean.

4. Page 6. Last sentence, Data collection paragraph, the authors stated that “The SF-36 has been used and validated in patients with ME/CFS in different settings [12-25].” Use of the SF-36 in ME/CFS is not equivalent to validation of the SF-36. The references that the authors listed were using the SF-36 instrument rather than validating the SF-36 in ME/CFS population.

5. Page 7. Please briefly describe what was measured for the 8 SF-36 scales in the line along with “eight health domain scales”. For example, Physical Functioning measures the functioning in running, caring groceries, climbing flight of stairs, bend, walking, bathing, etc; Role Physical includes measures on time reduction on work or activity, limitation to work or other activities, etc.

6. Page 9. The authors presented the correlation results on the SF-36 scales of cases and their caregivers in addition to unmatched analysis on the group mean difference. Will the finding differs from that using matched analysis on 44 matched pairs.

7. Page 10. From Table 7 and Figure 2, it was not clear if the authors presented the correlation or linear regression slope coefficient. If it’s linear regression, please include the adjusted R-square in Table 7 and Figure 2.

8. Page. 13. “The fact that the scores of cases meeting the Canadian criteria were consistently lower than those with the CDC-1994 criteria further suggests that diagnosis specificity is related to disease severity, and that specific diagnoses such as the Canadian may be more appropriate for research studies investigating risk factors and disease biomarkers.” As mentioned in the comment of Major Compulsory Revisions, the group comparison in Table 5 was not statistically sound due to the overlapping subjects in both groups. In addition to the questionable statistical significance presented in the table, the lower mean scores are necessarily indicating higher specificity or higher sensitivity. One should not try to draw this conclusion based on the group mean comparison on non-exclusive groups. It requires some statistical approaches such as Kappa agreement, concordant, disconcordant, specificity, sensitivity, etc, given there is no gold standard classification definition.
**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

None of all above