Reviewer's report

Title: How does weight affect utility-based quality of life? Evidence from the UK WAVES randomised controlled study

Version: 1 Date: 3 July 2015

Reviewer: Yemi Oluboyede

Reviewer's report:

General comments
This is potentially a very interesting paper and is of relevance to the journal as a similar area has been published in Pediatrics (Chen et al., 2014). The paper explores the relationship between weight status and generic utility-based HRQOL (using the CHU-9D instrument) in children aged 5-6 years. The paper also examines the construct validity of the CHU-9D instrument by reporting specifically on the discriminant and convergent validity (as compared with the PedsQL).

This said the paper needs more work. The background section needs to be expanded and provide justification for the use of the two instruments. More importantly, the groupings of weight status into Underweight & Healthy weight, in the analysis, is a major issue. Do they authors hypothesise that the quality of life of healthy weight underweight children are similar? Literature shows that this is not the case (for example see Grieken et al, 2013 (Impaired parent-reported health-related quality of life of underweight and obese children at elementary school entry)). The analysis for the healthy weight & underweight groups need to be re-run as separate groups.

Equally important, the paper does not adequately address or discuss the limitations of the study. A key issue that should be discussed is the distribution of the sample according to weight status. There is no reference to this in the results section or in the discussion section.

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

1. The title refers to how weight affects utility-based quality of life. I think there should be clarification that the paper is assessing generic utility-based HRQoL. Also some reference to weight status (as opposed to weight) might be beneficial. Perhaps the title should follow similar lines as from another publication from the WAVES study as this clearly states what instruments are being utilised in the paper: Measuring preference-based quality of life in children aged 6-7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y--the WAVES pilot study

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
1. I understand why the authors have not delved in detail into the literature around the different definitions and descriptions of validity however I believe it is necessary to provide a reader that may be unfamiliar with the literature to a few key references – particularly references relating to the two types of validity that were mentioned: discriminant & convergent validity [lines 74-79].

2. Clarification – [line 82] do the authors mean a utility scale anchored to dead and full health? On line 120 the authors refer to the CHU-9D scale ranging from 1 to 0.33 (so not 0-1).

3. Line 81-82 refers to the paper exploring the relationship between weight status category and utility-based HRQOL. It should be made clear that this is an assessment of a generic utility based instrument.....and also acknowledgement that the CHU-9D is being compared to a non-utility based instrument [line 84-85]...and the reasoning behind this (i.e. All this would be addressed if the authors add a paragraph in the background section explaining the reasoning behind their choice of instruments (the CHU-9D and PedsQL)).

4. Clarification – line 157 to 164. It is not clear what the 9 CHU-9D dimensions are....perhaps number them?

5. Clarification – line 206 check symbol after the word Spearman’s (this is repeated in Table 5)

6. Line 222-223 – the authors state that the paper contributes evidence on the use of the newly developed utility-based CHU9D instrument, within an ethnically and socioeconomically diverse UK population of young children. There is no mention of how the study population is reflected in terms of weight status (13.1% obese and 8.6% overweight). Perhaps the discussion could provide some comparative assessment to the latest National Child Measurement Programme (NCMP) report (2013/14) and explain any differences / similarities for the 4-5 years age group (though it is noted that the study population is from 5-6 years)?

Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

1. The background could be extended a bit. Justification for the use of the CHU-9D and PedsQL would be useful. Why was a condition specific measure of QoL not considered? Quite a few weight specific non-utility based QoL instruments exist for the younger population (i.e. KINDL-obesity module, Sizing me up, IWQOL – kids etc) explanation of why these were not utilised would have been good. An additional paragraph discussing the issues would be beneficial.

2. Justification for why the CHU-9D was not compared to another utility based instrument (instead of the PedsQL) such as the HUI would have been useful. If the HUI, or other potentially relevant utility instruments were not appropriate, this is fine, however this issue needs to be flagged. An additional paragraph discussing these issues needs to be included in the background section, perhaps between lines [79 and 81]….Reference to the WAVES pilot study comparing the EQ-5D-Y and CHU-9D would be good as part of the justification.

3. Clarification – line 129-131 refers to a visual response prompt. Where they the same for both instruments (presumably no?)? Was the visual prompt developed...
and utilised in the validation studies for the CHU-9D and PedsQL (if not were the prompt(s) tested before use in a pilot study – okay if not but some further explanation necessary)? More information on what the prompt looked like and how it was used would be useful as this has a direct effect on how the children would have answered the questions. A figure of what the prompt looked like would be useful.

4. Line 134-136 refers to there not being a gold standard to the measurement of utility based HRQoL – NICE recommends the use of the EQ-5D (NICE, Guide to the methods of technology appraisal, 2013) so a little clarity as to what the authors mean by there being no gold standard....are they referring to there being no gold standard in the younger population?

5. Clarification – Line 148 – 150. More explanation is needed regarding the rationale behind the hypothesis that children in the underweight category would report more problems in each dimension of the CHU-9D than the overweight and obese categories. Would being underweight not also be associated with its own issues regarding HRQoL? Should the underweight category be grouped with the healthy weight category (and if so why?)?

6. The discussion does not clearly state any study limitations. This would be useful. For example would the results have been any different if there were a higher proportion of overweight and obese children? Would it have been useful to include a condition specific QoL instrument (was this beyond the remit of the original study)? For the adult population, it has been suggested that generic instruments are not as sensitive to changes in BMI as weight specific instruments (Brazier et al., 2004). Might it be the case that generic utility scales may be missing important domains of quality of life (QoL) affected by weight status. Are there any other potential study limitations that could be flagged?

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests