

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Health-related quality of life: Population epidemiology and concordance in 11-12 year old Australians and their parents
AUTHORS	Catchpool, Max; Gold, Lisa; Grobler, Anneke; Clifford, Susan; Wake, Melissa

VERSION 1 - REVIEW

REVIEWER	Julie Ratcliffe UNISA, Adelaide, Australia
REVIEW RETURNED	13-Mar-2018

GENERAL COMMENTS	<p>This is a well written paper describing a robustly conducted epidemiological study.</p> <p>I have a few minor comments only</p> <ol style="list-style-type: none">1. Introduction pg 5: preference based measures typically are based on weights generated from large representative samples of the general population but this is not always the case as in some instances patient samples have also been used hence please insert the word 'typically' here to account for this.2. Introduction page 5 consensus is now that 0 is equivalent to 'being dead' (seeing as death is a process) please amend and reference Brazier textbook.3. Introduction page 5. Not sure that we can claim that CHU9D is newly developed possibly relatively newly developed seeing as been in existence for over 8 years now.4. methods top page 7. We are informed that participants were included in the current analyses if they contributed any HrQoL data that could be scored. This statement needs further clarification - does this mean that complete data only were used or were missing data imputed? What was the amount of missingness and what were the reasons for this - missing at random or something else? How did the authors account for missing data and do the respective instrument scoring algorithms provide any guidance as to how missing data should be handled?5. Methods top page 8 change to 0=being dead6. Methods top page 8. Statement 'to assess concordance of HrQoL among child parent dyads using the CHU9D' - it needs to be made clearer here and indeed throughout the paper that what I think the authors are referring to here is concordance in relation to intergenerational HrQoL as opposed to concordance between child self assessment and parental proxy assessment of child HrQoL?
-------------------------	---

	<p>We are informed in second paragraph on page 8 that a proxy report version of PedsQoL is available - was proxy parental reporting used in this particular study and if so why are these data not reported? I think not used (or at least not reported in this paper) but at times it is confusing for the lay reader to be sure about what is meant by the term concordance exactly. In relation to this important point there could be a stronger rationale given at the outset as to why the authors (at least appear to) consider that child self-reporting of HrQoL in the 11-12 year old age group is more valid/reliable than parent proxy reporting to give a clearer justification for the methods and analytical approach adopted. 6. Discussion page 17. A clearer distinction needs to be given between differences in CHU9D values between adults and adolescents that are due to differences in measurement as opposed to valuation. Measurement refers to the way in which adults and adolescents respond to the descriptive system (and thereby classify themselves into CHU9D health states). Valuation refers to the way in which these classifications into unique CHU9D health states are scored e.g. via utilisation of adult vs adolescent preference based scoring algorithms. Whilst I do not disagree with the statement that 'adolescents do have different health preferences to adults and that it is important that adolescent values are used' - there are two elements here that need to be disentangled measurement (ie responses to the CHU9D descriptive system) and valuation (application of adult vs adolescent specific scoring algorithm). The point that I think needs to be made more clearly is that when conducting epidemiological investigations of young people/adolescent population health and/or economic evaluation of health services targeted at young people/adolescents their own unique responses to the CHU9D descriptive system and the adolescent specific valuation/scoring algorithm should be applied.</p>
--	--

REVIEWER	<p>Nilüfer Demirsoy Asst. Prof. Dr. (PhD). Eskisehir Osmangazi University Faculty of Medicine Dep of History of Medicine and Medical Ethics Eskisehir /Turkey</p>
REVIEW RETURNED	<p>17-Mar-2018</p>

GENERAL COMMENTS	<p>It is a very important concept to evaluate the quality of life of children and children by their parents' opinions. A large sample was used and it was a successful study. I congratulate the authors. Best regards</p>
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

Reviewer Comments	Author's Response	Page, line
Reviewer 1: Julie Ratcliffe, University of South Australia, Australia		
<p>1. Introduction pg 5: preference based measures typically are based on weights generated from large representative samples of the general population but this is not always the case as in some instances patient samples have also been used hence please insert the word 'typically' here to account for this.</p>	<p>The word “typically” has been inserted. The text now reads: “The weights are typically calculated based on the preferences of a representative sample of the general population.”</p>	Page 5
<p>2. Introduction page 5 consensus is now that 0 is equivalent to 'being dead' (seeing as death is a process) please amend and reference Brazier textbook.</p>	<p>The words “being dead” have replaced "death" and a reference to the Brazier textbook included. The text now reads: “Preference-based measures generally produce an overall score that ranges from a state of health equivalent to full or perfect health (1) to a state of health equivalent to being dead (0), and often to states considered worse than death (<0).”</p>	Page 5
<p>3. Introduction page 5. Not sure that we can claim that CHU9D is newly developed possibly relatively newly developed seeing as been in existence for over 8 years now.</p>	<p>We have inserted the word “relatively”. The text now reads: “Among these instruments, the Child Health Utility Nine Dimension (CHU9D) is a relatively newly developed paediatric generic preference-based measure of HRQL.”</p>	Page 5
<p>4. Methods top page 7. We are informed that participants were included in the current analyses if they contributed any HrQoL data that could be scored. This statement needs further clarification - does this mean that complete data only were used or were missing data imputed? How did the authors account for missing data and do the respective instrument scoring algorithms provide any guidance</p>	<p>Scoring algorithms, and their methods for dealing with missing data, vary by measure. The CHU9D scoring algorithm does not allow imputation and only complete cases are scored. The AQoL8D scoring algorithm permit a limited numbers of missing values be imputed within each of its dimensions. The PedsQL scores are a mean and incomplete cases were included without imputing, if the number of missing responses is below the maximum of 2 to 4 for each subscales. The Study Design and Participants section of the Methods the text now reads: “The scoring algorithm of each</p>	Page 7

Reviewer Comments	Author's Response	Page, line
as to how missing data should be handled?	<p>instrument specify how missing data are treated. The CHU9D required no missing data for a total score to be calculated. The AQoL8D scoring algorithm allows a limited number of missing values be imputed within each of its dimensions. The PedsQL allows two to four missing values in each subscale for a mean score to be calculated.”</p>	
<p>5. What was the amount of missingness and what were the reasons for this - missing at random or something else?</p>	<p>The number of participants with missing data are as follows: CHU9D – 47 children (2.5%) and 31 parents (1.7%); AQoL – 13 parents (0.7%); PedsQL – 27 children (1.4%).</p> <p>Data were missing for a combination of reasons. Of 47 child missing values, 9 children did not provide any questionnaire data and a further 2 children could not complete the questionnaire for medical reasons, 1 had all questionnaire data removed as unreliable, and 35 (15) did not complete sufficient items in the CHU9D (PedsQL) measure for a score to be calculated. Of 31 parent missing values, 4 did not provide any questionnaire data and 27 (9) did not complete sufficient items in the CHU9D (AQoL) measure for a score to be calculated.</p> <p>For both parents and children we cannot say with confidence that data were missing completely at random, although missingness was spread across all locations used in this study and all three assessment settings (“pop-up” assessment centre, smaller regional assessment clinic, home visit (as in Procedures))</p> <p>In the Study Design and Participants section of the Methods the text now reads: “Data completeness was high and participants were included in the current analyses if they contributed any HRQL data that could be scored (figure 1).”</p> <p>An additional reference (22) which provides more detail on data completeness/missingness for the CheckPoint study has been added at the final sentence of the Study Design and Participants section of the Methods.¹</p>	Page 7

Reviewer Comments	Author's Response	Page, line
6. Methods top page 8 change to 0=being dead	The words "being dead" have replaced "death". The text now reads: "Both techniques generate a score that is on the range of equivalence from perfect health (1) to being dead (0)."	Page 8

<p>7. Methods top page 8. Statement 'to assess concordance of HrQoL among child parent dyads using the CHU9D' - it needs to be made clearer here and indeed throughout the paper that what I think the authors are referring to here is concordance in relation to intergenerational HrQoL as opposed to concordance between child self assessment and parental proxy assessment of child HrQoL?</p>	<p>We recognise the potential confusion between parents self-rated HRQL (studied here) and their use as proxies in proxy-rated child HRQL (used elsewhere, not here). We have added some clarification throughout the text (Methods, Results and Discussion) to hopefully reduce this potential confusion (also see below).</p> <p>In the first paragraph of the Outcome Measure section of the Methods the text now reads: "Each respondent completed two self-rated HRQL measures as part of a larger questionnaire."</p> <p>At the end of the second paragraph of the Outcome Measure section of the Methods the text now reads: "To assess the distribution of HRQL data using the CHU9D, the adolescent scoring algorithm was applied to child data and the adult scoring algorithm on the adult data. However, to assess concordance of HRQL among child-parent dyads using the CHU9D, a common scoring algorithm needed to be applied. If this does not occur, any differences between HRQL of children and HRQL of their parents may be because of the different scoring methods used. Therefore, the adolescent algorithm was also applied to the parent CHU9D data, solely for the purposes of the intergenerational concordance analysis of this study."</p> <p>In the second paragraph of the Statistical Analysis section of the Methods the text now reads: "Concordance between HRQL of parents and HRQL of children for the CHU9D was assessed using 1) Pearson's correlation coefficients with 95% confidence intervals; and 2) linear regression with the child HRQL as the dependent variable and parent HRQL as the independent variable, adjusted for parent age, child age, disadvantage index, as well as parent and child sex in models including both sexes."</p> <p>At the start of the Intergenerational concordance of HRQL (table 4) section in</p>	<p>Page 7</p> <p>Page 7-8</p> <p>Page 9</p> <p>Page 15</p> <p>Page 17</p> <p>Page 17</p>
--	--	--

Reviewer Comments	Author's Response	Page, line
	<p>the Results the text now reads: "A weak positive correlation was present between child HRQL and parent HRQL as measured by the CHU9D (CC 0.13, 95% CI 0.09 to 0.18) (table 4)."</p> <p>In the Strengths and limitations section of the Discussion the text now reads: "This is the first study to provide HRQL concordance data between children's HRQL and the HRQL of their parents within a large cohort study."</p> <p>In the Findings in relation to other studies section of the Discussion the text now reads: "The mean score recorded for child HRQL was 0.8."</p>	
<p>8. We are informed in second paragraph on page 8 that a proxy report version of PedsQoL is available - was proxy parental reporting used in this particular study and if so why are these data not reported? I think not used (or at least not reported in this paper) but at times it is confusing for the lay reader to be sure about what is meant by the term concordance exactly. In relation to this important point there could be a stronger rationale given at the outset as to why the authors (at least appear to) consider that child self-reporting of HrQoL in the 11-12 year old age group is more valid/reliable than parent proxy reporting to give a clearer justification for the methods and analytical approach adopted.</p>	<p>Related to the above point, we have removed reference to the proxy-complete version of PedsQL.</p> <p>We agree that this paper is fully concerned with self-complete measures and mention of proxy-completion risks confusion.</p> <p>For similar reasons, we do not believe that this paper is the appropriate place to discuss validity/reliability of self-rating versus proxy-rating (because this is not an issue that is addressed in this study).</p> <p>At age 11-12, the child participants in this study are clearly within the age range seen as appropriate for self-rating by measure developers; these age ranges are presented in the Introduction or Methods section when describing the measures (from age 7 for CHU9D, page 5; from age 5 for PedsQL, page 8).</p> <p>In the Outcome Measure section of the Methods in the third paragraph when describing the PedsQL we have removed reference to parents being able to complete a proxy-report for the measure. The text now reads: "Children aged 5-18 years can complete the child self-report section of the questionnaire."</p>	<p>Page 7</p>

<p>9. Discussion page 17. A clearer distinction needs to be given between differences in CHU9D values between adults and adolescents that are due to differences in measurement as opposed to valuation. Measurement refers to the way in which adults and adolescents respond to the descriptive system (and thereby classify themselves into CHU9D health states). Valuation refers to the way in which these classifications into unique CHU9D health states are scored e.g. via utilisation of adult vs adolescent preference based scoring algorithms. Whilst I do not disagree with the statement that 'adolescents do have different health preferences to adults and that it is important that adolescent values are used' - there are two elements here that need to be dis-entangled measurement (ie responses to the CHU9D descriptive system) and valuation (application of adult vs adolescent specific scoring algorithm). The point that I think needs to be made more clearly is that when conducting epidemiological investigations of young people/adolescent population health and/or economic evaluation of health services targeted at young people/adolescents their own unique responses to the CHU9D</p>	<p>We agree. We have expanded this paragraph of the discussion and hope that this now more accurately conveys our meaning.</p> <p>The text now reads: "The difference in CHU9D values when applying adult and adolescent scoring algorithms to participants' self-rated CHU9D health states are expected and are in line with previous research.³⁴ This supports the fact that adolescents do have different health preferences to adults and that it is important that studies of adolescent health not only allow adolescents to provide their own responses to measures but also use adolescent values in any scoring applied to those measures."</p>	<p>Page 18</p>
--	---	----------------

Reviewer Comments	Author's Response	Page, line
descriptive system and the adolescent specific valuation/scoring algorithm should be applied.		
Reviewer 2: Nilüfer Demirsoy, Eskisehir Osmangazi University Faculty of Medicine Dep of History of Medicine and Medical Ethics Eskisehir, Turkey		
1. A large sample was used and it was a successful study. I congratulate the authors.	Thank you.	