

ABSTRACT

Objectives

Consideration of health status in children and adolescents now includes broader concepts such as health-related quality-of-life (HRQoL). Globally there is a need for relevant preference-based HRQoL measures (PBMs) for use in children and adolescents, yet measurement of HRQoL in these groups presents particular challenges. This paper systematically reviews the available generic childhood PBMs, and their application and cross-cultural validation in sub-Saharan African (sSA).

Methods

A systematic review of published literature from 01/01/1990 to 02/08/2017 was conducted using MEDLINE (through OVID Sp), EMBASE (OVID Sp), EconLit (EBSCOhost), PsycINFO, Web of Science and PubMed.

Results

220 full-text articles were included in a qualitative synthesis. Ten generic childhood PBMs were identified, of which nine were adapted from adult versions and only one developed specifically for children. None of the measures were originally developed in sSA or other resource-constrained settings. The HUI3 and the EQ-5D-Y-3L were the only measures that had been applied in sSA settings. Further, the HUI3 and the EQ-5D-Y-3L were the only generic childhood PBM that attempted to establish cross-cultural validation in sSA. Five of the six of these validation studies were conducted using the EQ-5D-Y-3L in a single country, South Africa.

Conclusions

Review of childhood PB HRQoL measures in sSA

The findings show that application of generic childhood PBMs in sSA settings has hitherto been limited to the HUI3 and EQ-5D-Y-3L. Most adaptations of existing measures take an absolutist approach, which assumes that measures can be used across cultures. However, there is also need to ensure linguistic and conceptual equivalence and undertake validation across a range of sSA cultural contexts.

Highlights:

i. What is already known about the topic?

Since the 1990s, there has been growing interest in developing and using child and adolescent (childhood) preference-based health-related quality of life measures (PBMs). The increased prominence of childhood PBMs has been driven by their use in measuring health-related quality-of-life (HRQoL) as part of economic evaluations, thereby informing health care policies.

ii. What does the paper add to existing knowledge?

While this paper examines existing childhood PBMs and summarises their psychometric properties, it focuses on the validation and application of these measures in the sub-Saharan African (sSA) region. Development of existing childhood PBMs have been restricted to high-income countries, and none has been developed in low-income countries. Validation and application of childhood PBMs in the sSA region has been limited to the HUI3 and EQ-5D-Y-3L in two countries.

iii. What insights does the paper provide for informing health care-related decision making?

Findings from health technology assessments (HTA) and economic evaluations are increasingly used to inform health policy in sSA. These studies should incorporate local preferences for health outcomes. Preference-based values for health outcomes are commonly generated using tools that are adapted to local contexts. If the process of adapting these measures in sSA contexts is not robust, the findings from subsequent HTAs and economic evaluations may not be applicable.

Introduction:

In sub-Saharan Africa (sSA) i.e. the region below the Sahara desert, there has been a remarkable decline in mortality rates in infants and under 5-year olds since 2000,(1) partly driven by policy responses to the Millennium Development Goals (MDGs).(2) In 2015, the United Nations set 17 sustainable development goals as a replacement for MDGs for nations to achieve by 2030. Goal number three focuses on ‘good health and wellbeing’, specifically targeting child and maternal mortality and chronic disease reduction.(3) As childhood mortality continues to fall, and increasing numbers of children survive into adolescence and beyond, future interventions will increasingly focus on improving wider outcomes including children’s health-related quality of life (HRQoL)’.(4)

Preference-based HRQoL measures (PBMs) are designed to “assess patient preferences across broad areas including symptoms, physical functioning, work and social activities, and mental well-being”.(5) PBMs have increased in prominence over recent years because of their use in economic evaluations, which often inform policy.(6) This can only be meaningful if PBMs are relevant, valid and robust. However, most generic PBMs have been developed in the English language (7) and, as the development of new measures is time and resource intensive, use of PBMs in other contexts has mostly relied on translating and/or adapting existing measures, with an implicit assumption that adaptation produces an equivalent measure.(8) Adaptation of an existing measure requires an iterative process of translation and back translation with a qualitative review to establish both linguistic and conceptual equivalence.(7) In addition, the adapted measure needs to be re-validated using standard psychometric methods to re-establish validity and reliability for the new form of the measure in the new context.

Several reviews have assessed the development and application of childhood PBMs and non-PBMs (restricted to ages 6-17 years), (9-11) their cross-cultural validation, (12, 13) and availability of measures for young children (<8 years) in alternative settings.(14) However, to our knowledge, no review has specifically studied the use or validation of childhood PBMs across sSA.

We carried out a systematic review of generic childhood PBMs, particularly focussing on their use and cross-cultural validation in sSA settings. The objectives of the review were threefold: (1) to identify existing generic childhood PBMs and summarise their psychometric quality; (2) to explore which of these generic childhood PBMs have been applied in sSA; and (3) to assess the extent of cross-cultural validation of generic childhood PBMs for application in sSA.

Methods

We conducted the literature searches using a two-stage process compatible with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines.⁽¹⁵⁾ For the first stage, we searched the literature to identify existing childhood PBMs (objective 1). In the second stage (to address objectives two and three), we refined our search strategy to identify all published research studies that applied or cross-culturally validated any one of these PBMs in the sSA region. Cross-cultural validation in our search criteria was defined as i) linguistic equivalence (achieved through translation), ii) conceptual equivalence (achieved through focus group discussion or cognitive interviews) and iii) psychometric validation. The peer reviewed literature published between 1st January 1990 and 2nd August 2017 was examined for both stages of the review.

Search strategy

In stage one, we carried out systematic searches using an intersection of four sub-groups of search terms: 1) HRQoL terms; 2) childhood terms; 3) instrument terms; and 4) psychometric terms. The psychometric terms were included in order to identify validated measures only. A full list of search terms is provided in Appendix (Additional file 1). The search was performed in six databases via Ovid: PubMed, EMBASE, Web of Science, EconLit, PsycINFO and MEDLINE, as per COSMIN recommendations.⁽¹⁶⁾

In stage two, we searched through the same databases using search terms for the identified generic childhood PBMs and combined them with sSA terms to identify their application and/or cross-cultural validation in sSA settings. The full list of identified measures and sSA search terms used to address the second and third study objectives are provided in Additional file 2.

Inclusion and exclusion criteria

To address the first study objective, studies reporting development of measures were included if: (1) they reported the development of at least one childhood PBM (this was limited to the original work by the developers only); 2) the target age(s) for completion of the measure included children between 6-17 years; and (3) were published in the English language. Studies were excluded if they reported measures designed for adults (≥ 18 years) or young children (< 6 years) as the review focussed on measures with the potential for self-completion by children.⁽¹⁷⁾ Additionally, all studies that reported on psychometric properties of the identified childhood PBMs were restricted to work by the primary developers.

To address the second and third study objectives, studies were included if they had applied and/or cross-culturally validated any measure identified at stage one of the review in a sSA setting. Studies using generic childhood PBMs that were not carried out either in part or whole in a sSA setting were excluded.

Procedures

Three reviewers (LGN, KK and LS) independently assessed the titles, and subsequent abstracts, to select potential full-text articles for screening. Study selection was agreed upon at each stage (title, title and abstract, full article) by the three reviewers with disagreements resolved via consensus.

Data extraction and analysis

To meet **objective 1**, for each identified generic childhood PBM the following was extracted:

(i) General description of the measure

The descriptive system consisting of the classification system; where, how and who developed the classification system; target population; number of dimensions; response levels; and time taken to respond.

(ii) Underpinning utility theory

This included the number of health of states generated by the classification system, the valuation technique used, model used to extrapolate the utilities to all possible health states, range of utilities for each measure and when and how these were generated.

(iii) Psychometric validation

The quality of each identified childhood PBM was appraised against standard psychometric criteria (18-21) as outlined below.

Conceptual framework including a detailed description of the conceptual basis of the construct measured, is usually based on qualitative work with the target population.(20)

Cross-cultural language and conceptual equivalence/adaptation is the extent to which a scale shows linguistic and conceptual equivalence, usually assessed through qualitative methodology, including translation, cognitive interviews or focus group discussions with the target population.(8)

Reliability, the extent to which the instrument is free from random error, was assessed in terms of *internal consistency*, defined using Cronbach's $\alpha \geq 0.70$ for group level comparisons and ≥ 0.90 for individual comparisons;(19) and *test re-test* assessed by an Intraclass-Correlation Coefficient (ICC) ≥ 0.70 or Kappa $\kappa \geq 0.70$ for nominal scales.(18, 19, 21)

Validity, the extent to which a scale measures what it purports to measure, was assessed using: *Content validity* which assesses the scale's conceptual breadth, including all aspects of the construct being measured;(18, 20) *Criterion validity*, which assesses how the scale correlates with a gold standard measure of the same construct administered at the same time (*concurrent*) or at some point in the future (*predictive*), defined using correlation ≥ 0.70 ;(21) *Convergent validity*, the extent to which the scale is positively related to other scales that purport to measure the same thing; *Discriminant validity*, the extent to which the scale is not related to other scales that purport to measure different things, assessed using correlation coefficients; *Known groups validity*, which assesses the extent to which scores on the scale are different between two groups that are known to be different on another clinically defined variable;(21) and *Responsiveness*, the ability of the scale to detect clinically important change over time, which was assessed using effect size statistics, score changes or standardized response means.(18)

Interpretability of scores refers to the degree to which one can easily assign meanings to an instrument's quantitative scores, assessed in several ways including normative data.(18-21)

Practicality is the burden on those completing and/or those administering the measure and includes completion time, and completion and response rates.(21)

In our review, all psychometric properties described above were classified as (+) if they met the criteria; (-) if not; and (NR) if not reported. We assessed the psychometric properties as a whole rather than using a single statistic as evidence of psychometric quality.

To achieve **objective 2**, all studies included in this second stage of the review were assessed for their quality using the Critical Appraisal Skills Programme (CASP) toolkit for case-control

studies.(22) The CASP tool consists of ten questions that address issues such as study aim and design, data collection and analysis, and ethical issues raised by the published study.

In order to achieve **objective 3**, each identified generic childhood PBM was assessed according to whether there had been: i) cross-cultural translation into any native language spoken in sSA; ii) any conceptual equivalence through focus group discussion or cognitive interviews on word meanings and changes to concepts to fit what would be locally equivalent into a sSA setting and iii) psychometric validation. Cross-cultural validation was classified as (+) if it was appraised and performed; (-) if appraised but not performed; and (NR) if not reported.

This review was not part of any clinical trial and was therefore not registered as would have been required otherwise.

Results

Identifying childhood preference-based HRQoL measures (objective 1)

The first stage of our search identified 59,342 unique articles (Figure 1). 58,366 articles were excluded after screening the title and a further 734 articles excluded after screening the abstract. In addition to the remaining 242 articles, we identified 91 articles through snowball searching. We therefore reviewed the full texts of 333 articles for eligibility. 113 of these articles were excluded as they did not meet eligibility criteria; a large number (n=51) of these articles were excluded as they did not report use of a generic childhood PBM. In total, we identified 220 articles that reported development of at least one generic childhood PBM, from which we identified ten unique measures (Figure 1).

Identified childhood PBMs

i) General description

Our review identified ten existing childhood PBMs in Table 1: 16-Dimensional (16D), 17-Dimensional (17D), Adolescent Health Utility Measurement (AHUM), Assessment of Quality of Life-6Dimension (AQoL-6D), Child Health Utility 9Dimension (CHU-9D), EQ-5D-Youth (EQ-5D-Y), Health Utilities Index Mark 2 (HUI2), Health Utilities Index Mark 3 (HUI3), Quality of Well Being-Self Administered (QWB-SA) and, Child Health-6 Dimension (CH-6D). The CH-6D was a childhood PBM recently developed in South Korea for children aged 7-12 years.(23) All ten measures were developed in high income countries, adapted predominantly from existing adult versions with little or no involvement of children at the development stage (24) with the exception of the CHU-9D.(25)

ii) Underpinning utility theory of identified measures (Table 2):

Utility values were generated using time trade-off (TTO) for the AQoL-6D, AHUM and CH-6D; combining TTO with a visual analogue scale (VAS) for the EQ-5D-Y; a rating scale (RS) for 16D, 17D, and QWB-SA; and combining a RS and the standard gamble (SG) approach for the HUI2 and HUI3. The utilities were elicited using adult preferences - whether own or proxy – for the HUI2, HUI3, QWB-SA, 17D, CHU-9D, EQ-D-Y, AHUM and CH-6D. The AQoL-6D and 16D were the only two measures that elicited the preferences of adolescents. Through either statistical additive or multiplicative models, the utilities generated for a sample health states were then extrapolated for all health states of the measures. For the 16D, 17D, AHUM, QWB-SA, EQ-5D-Y, CHU-9D

this was through an additive model and through a multiplicative model for the HUI2, HUI3 and AQL-6D.

iii) Psychometric validation:

Table 3 presents the psychometric appraisal of the 10 identified measures. The psychometric evidence was mixed. Development of a conceptual framework, content validity, and convergent validity were most frequently evaluated. Only six measures (HUI2, HUI3, 16D, 17D, CHU-9D and EQ-5D-Y) had at least one piece of adequate evidence for both reliability and validity. A detailed description of the psychometric properties can be found in Additional file 3.

Application and cross-cultural validation of generic childhood PBMs in sub-Saharan Africa

The second stage of our search aimed to identify the application and cross-cultural validation of generic childhood PBMs in sSA. We identified 40 unique articles; 30 articles were excluded after screening the title leaving ten articles for full-text assessment. Four of these articles were excluded because they pertained to adults. Of the remaining six articles, two articles reported on the application (objective 2) of two childhood PBMs in sSA,(26, 27) but also provided evidence on aspects of validation (Figure 1). The included studies were of good quality, meeting most of the criteria as defined by the CASP guidelines (see Additional file 4).

Application of generic childhood PBMs in sSA (objective 2)

The two studies reporting the use of childhood PBMs in sSA were undertaken in Uganda and South Africa.(26, 27) The quality of both studies was good according to the criteria defined by the CASP

checklist with clearly focussed topics, methodology, analysis and interpretation of the results . These two studies used the HUI3 and the EQ-5D-Y, respectively (Table 4).

Sims-Williams and colleagues(26) used the HUI3 to evaluate the HRQoL of children with spina bifida and that of their caregiver adult proxies (self-reported from their perspective). There was little difference in HRQoL perception between children and their proxies (0.575 and 0.549, respectively, Pearson correlation=0.848). This study also found that male sex, urinary incontinence, large family size and hydrocephalus were factors that were associated with children's HRQoL.

Jelsma and Ramma(27) used the EQ-5D-Y to compare HRQoL amongst children (aged 7-12 years) attending open schools (healthy, n=567) and those attending special schools (with functional limitations, n=41); and to compare the responses to the measure provided by the children and their parents in South Africa. The parents (proxies), without consulting their children, completed the EQ-5D-Y proxy version whilst the children completed the self-report version of the EQ-5D-Y. The parents' responses to the EQ-5D-Y proxy version required them to rate their child's HRQoL from the perspective of the child. The study found no differences in perception of HRQoL between children in open and special schools in this population. However, differences were observed between HRQoL reported by children and their female parent proxies in open schools but not in special schools.

Cross-cultural validation of generic childhood PBMs in sub-Saharan Africa (objective 3)

The HUI3 and the EQ-5D-Y were the two generic childhood PBM measures that attempted cross-cultural validation in six studies in sSA settings (Table 5 summary and Additional file 5 detailed).(28) Five of the six studies were for the validation of the EQ-5D-Y and this was in one country alone, South Africa.

The HUI3

i) Linguistic equivalence

The HUI3 language translation involved forward translation only from the original English language into several local languages spoken in Uganda.(26) Members of the study team translated HUI3 into Luganda, Lugisu, Lugwere, Lunyole, Lusoga, Kiswahili, Ateso, Dhapadhadola, Samia and Kubsabiny as they went along interviewing families that they saw. However, the authors didn't specify whether the study members were qualified translators or that a backward translation was done prior to using the translated questionnaire.

ii) Conceptual equivalence

Cognitive interviews were not conducted prior to using the translated HUI3 among children and parents conversant in the translated language in Uganda. The study members translated the HUI3 into the local language as it was being administered to children with spina bifida and their proxies.

iii) psychometric validation

The validation involved comparison of the performance of the HUI3 tool and the VAS marked 0 to 10, and the scores of children and their proxies (completed from the child's perspective). There was moderate correlation (0.488) between the HUI3 and VAS scores of children in open and special schools, and even lower correlation (0.380) between those of children and their proxies.

The EQ-5D-Y

i) Linguistic equivalence

The EQ-5D-Y was not translated into any native South African language as the EQ-5D-Y English UK version was used.

ii) Conceptual equivalence

Cognitive interviews to establish conceptual equivalence were done for the EQ-5D-Y in four countries (Germany, Italy, Spain and Sweden), but not in South Africa during the development stage.(28) The adaptation involved revising wording of the adult EQ-5D dimensions into a child-friendly version followed by psychometric validation in five countries including South Africa.

iii) Psychometric validation

The psychometric validation study demonstrated that the EQ-5D-Y was feasible in South Africa since only 2% of the children had inappropriate or missing responses.(29) Further, there was convergent validity between the EQ-5D-Y and both the KIDSCREEN-10 and KIDSCREEN-27 on similar dimensions (correlation coefficient $r = -0.17$ to -0.52); the EQ-5D-Y and self-reported

general health ($r = 0.25-0.27$); and with the EQ-5D-Y VAS (Pearson, $r = -0.56$ to 0.55). The negative correlation between the EQ-5D-Y and KIDSCREEN is because these are scored in opposite directions (higher score on EQ-5D-Y indicates better HRQoL whereas for the KIDSCREEN higher scores mean worse HRQoL). Known-group validity was evident in that the EQ-5D-Y was able to differentiate reported problems between with ($p < 0.05$).⁽²⁹⁾ The p-value refers to the statistical significance between those reporting 'fair' or 'poor' self-reported health and more problems on EQ-5D-Y dimensions.

Scott et al.⁽³⁰⁾ found that the EQ-5D-Y demonstrated test-retest reliability across both the descriptive component and the VAS scores (kappa, $k = 0.365$ to 0.653 and $ICC = 0.77$); convergent validity with the PedsQL amongst the acutely ill children in the study ($p < 0.001$); and known-group validity amongst groups with different health states both by composite score (derived from a model in order to generate a single utility score for the descriptive component) and VAS scores (all differences $p < 0.001$), but not on composite scores.

In another South African psychometric validation study of the EQ-5D-Y, Jelsma and Ramma⁽²⁷⁾ demonstrated no inter-rater reliability between children and their adult proxies on all dimensions ($r < 0.33$), and no known-group validity between children with and without disabilities on 'doing usual activities', 'having pain or discomfort', and 'feeling worried, sad or unhappy' dimensions and on the VAS scores, mean 88.4 and 87.9, respectively. However, there was evidence of known-group validity on the 'mobility' and 'looking after myself' dimensions ($p < 0.001$). The last of the EQ-5D-Y validation studies, by Jelsma,⁽³¹⁾ demonstrated that the EQ-5D-Y and the EQ-5D

shouldn't be used interchangeably since the former performs better in children and adolescent populations in this setting. The EQ-5D-Y had statistically significant fewer missing responses on both dimensions ($\chi^2=9.404$; d.f.=1; $p=0.002$) and the VAS scale ($\chi^2=26.159$; d.f.=1; $p<0.001$), respectively, than the adult EQ-5D.

Summary assessment of results

Our review found that in previous studies both the HUI and the EQ-5D-Y did meet some but not all of the cross-cultural validation criteria. The HUI was translated into a native sub-Saharan language but the translation did not involve backward translation nor cognitive interviews to establish conceptual equivalence. Additionally, some but not all of the psychometric properties were evaluated for the HUI3.

On the other hand, previous studies of the EQ-5D-Y have established conceptual equivalence, but have not translated the measure into a native language spoken in sSA. Similar to the HUI3, we did not find any single study that had evaluated all the psychometric properties of the EQ-5D-Y. When all the findings from these individual EQ-5D-Y studies are considered together, it is evident that the psychometric properties have been comprehensively evaluated for the EQ-5D-Y except for internal consistency, which was not reported in any of the studies. In summary, we found that the majority of psychometric properties were not reported by individual studies. However, of those that did report, the majority did meet the psychometric criteria.

Discussion

This review aimed to identify the available childhood PBMs and to determine the extent to which they have been used and/or cross-culturally validated in the sSA region. Ten PBM measures were identified that were all originally developed in high-income countries and mostly adapted from adult versions of the measures. The HUI3 and the EQ-5D-Y were the only two measures used in child and adolescent populations in sSA. Our review also highlighted the lack of involvement of children and adolescents themselves in the process of adapting these measures for use in sSA.

Existing childhood PBMs

There is a need for greater involvement of children and adolescents themselves in the development of conceptual frameworks underlying childhood PBMs.(24) The CHU-9D is the only measure identified by our review whose dimensions were generated directly from children through qualitative interviews and analysis (although we have no data for the recently developed CH-6D).(24) Even though it has been demonstrated that adult domains/dimensions can relate to children and adolescents,(28) it is also well known that children have different perspectives and goals, implying that dimension structures should ideally be developed directly from children themselves within specific contexts.(32) This is consistent with previous authors(14) who have emphasised that interpretation of HRQoL may differ across cultures.(8, 33)

PBMs in sub-Saharan Africa

The use of generic childhood PBMs in sSA settings is sparse. In part, this may reflect a lack of policy relevance in sSA settings; although funders have set standards for childhood outcomes as evidenced by the Sustainable Development Goals (SDGs), these have primarily focussed on

reducing child mortality.(2) While this is important, there should be some effort to measure the HRQoL gained through the SDG interventions using the existing preference-based measures. It may also reflect the fact that the measurement of HRQoL in children has lagged behind that of their adult counterparts due to the challenges of eliciting robust self-reported information across the developmental stages of childhood. To be able to use PBMs effectively in sSA requires more research that develops methods of eliciting reliable information from children themselves via questionnaires and other modes of administration.

Our findings suggest that there is potential for the EQ-5D-Y to be used in sSA, but more work is needed to investigate its conceptual equivalence in this setting. There has been no cross-cultural validation in sSA of the ten identified childhood PBMs with the exception of the HUI3 and the EQ-5D-Y. The cross-cultural validation of the HUI3 involved translation and cognitive interviews with children and parents, and psychometric evaluation.(26) Three theoretical underpinnings for adapting preference-based HRQoL measures have previously been described, depending on the impact of culture: absolutist, universalist and relativism.(33) Some adaptations of existing measures take an absolutist approach, which assumes culture does not have an impact. This means that, without seeking to establish a conceptual equivalence, measures are exported, after crude translation, into the local cultural milieu.(7, 14) However, it is well known that local culture has an impact on health and health-related concepts.(8) It is important therefore that conceptual equivalence, which is part of the adaptation process, be established before using PBMs cross-culturally. Developers of the EQ-5D-Y themselves suggest that future studies should consider other aspects of HRQoL that are important to children and adolescents for inclusion in the measure.(34) Moreover, the measures have only been applied in a small selection of sSA countries,

raising questions about their applicability across sSA as a whole. Apart from Uganda, where the HUI3 was applied, the use of the childhood PBMs was limited to South Africa alone, which is not representative of most economies in sSA as it is ranked as an upper middle-income country according to the World Bank.(28, 29) The use of the EQ-5D-Y therefore needs to be explored in other countries outside South Africa in order to assess generalisability across sSA settings.

Strengths and limitations

This is the first study to our knowledge that assessed the application and cross-cultural adaption of childhood PBMs in sSA. Our search generated a high volume of data, and a large volume of articles were excluded during the title screening of our review, which could have led to some relevant articles being excluded. Nevertheless, the findings related to our first study objective are largely consistent with the most recent review of existing generic childhood PBMs.(10) Although we appraised the measures for their psychometric properties, this was limited to the original work by the developers. Despite the fact that psychometric evaluations for widely used measures, such as the HUI2 and HUI3, are ongoing, for practical reasons we only included the original psychometric evaluations of each measure.(35) Our searches were also limited to peer reviewed material, which could have overlooked important evidence in the grey literature. In relation to this, we **excluded** all studies other than those in the English language, which could have had the effect of excluding some local studies. Finally, the quality appraisal of the studies applied in sSA settings was performed using the CASP case-control checklist due to the lack of toolkits designed specifically for cross-sectional studies.

Conclusion

The findings of this review show that application of generic childhood PBMs in sSA settings has hitherto been limited to the HUI3 and EQ-5D-Y-3L. Most adaptations of existing measures take an absolutist approach, which assumes that measures can be used across cultures. However, there is a need to undertake cross-cultural linguistic and conceptual equivalence, testing and psychometric validation across a range of sSA cultural contexts. Unless cross-cultural validation of PBMs is robust when applied in the sSA region, the application and interpretation of economic evaluations reliant on preference-based outcome measures will fail to reflect local relevance in this part of the world.

Additional files:

Additional file 1: Search terms for identifying available generic childhood preference-based HRQoL measures

Additional file 2: Search terms for identifying applications and cross-cultural validations of generic childhood preference-based HRQoL measures

Additional file 3: Detailed psychometric properties as evaluated by developers of the identified generic childhood PBMs

Additional file 4: Quality appraisal of studies identified for application or adaptation of generic childhood PBMs in sSA (n=6)

Additional file 5: Detailed cross-cultural validation of generic childhood preference-based HRQoL measures in sSA (n=6)

REFERENCES

1. WHO G, Switzerland. UNESCO SCIENCE REPORT. last accessed on 11th October 2017.
2. WHO G, Switzerland. Millenium Development Goals and Beyond 2015.
3. WHO G, Switzerland. Sustainable development goals. <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>. last assessed 15th September 2017.
4. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*. 1998; 28: 551-8.
5. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. *Journal of Health Services Research and Policy*. 1999; 4: 174-84.
6. Wallander J, Schmitt M, Koot H. Quality of Life Measurement in Children and Adolescents: Issues, Instruments, and Applications. *Journal of Clinical Psychology*. 2001; 57: 571-85.
7. Bowden A, Fox-Rushby JA. A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Soc Sci Med*. 2003; 57: 1289-306.
8. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol*. 2015; 68: 435-41.
9. Janssens A, Thompson Coon J, Rogers M, et al. A systematic review of generic multidimensional patient-reported outcome measures for children, part I: descriptive characteristics. *Value Health*. 2015; 18: 315-33.
10. Chen G, Ratcliffe J. A Review of the Development and Application of Generic Multi-Attribute Utility Instruments for Paediatric Populations. *Pharmacoeconomics*. 2015; 33: 1013-28.
11. Thorrington D, Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. *PLoS One*. 2015; 10: e0135672.
12. Colbourn T, Masache G, Skordis-Worrall J. Development, reliability and validity of the Chichewa WHOQOL-BREF in adults in Lilongwe, Malawi. *BMC Res Notes*. 2012; 5: 346.
13. Westmoreland K, Reeve BB, Amuquandoh A, et al. Translation, psychometric validation, and baseline results of the Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric measures to assess health-related quality of life of patients with pediatric lymphoma in Malawi. *Pediatr Blood Cancer*. 2018: e27353.
14. Paltzer J, Barker E, Witt WP. Measuring the health-related quality of life (HRQoL) of young children in resource-limited settings: a review of existing measures. *Qual Life Res*. 2013; 22: 1177-87.
15. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009; 62: 1006-12.
16. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010; 19: 539-49.
17. Canaway AG, Frew EJ. 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. *Qual Life Res*. 2013; 22: 173-83.
18. (FDA) UDoHaHSFaDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Rockville, MD: FDA, 2009.
19. Trust SACotMO. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002; 11: 193-205.
20. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013; 22: 1889-905.

21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007; 60: 34-42.
22. Programme CAS. CASP Systematic Review Checklist. 2018.
23. Kang E. Validity of Child Health-6 Dimension(CH-6D) for Adolescents. *Value in Health.* 2016; 19.
24. Stevens KJ. How Well Do the Generic Multi-attribute Utility Instruments Incorporate Patient and Public Views Into Their Descriptive Systems? *Patient.* 2016; 9: 5-13.
25. Stevens K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Qual Life Res.* 2009; 18: 1105-13.
26. Sims-Williams HJ, Sims-Williams HP, Mbabazi Kabachelor E, et al. Quality of life among children with spina bifida in Uganda. *Arch Dis Child.* 2017; 102: 1057-61.
27. Jelsma J, Ramma L. How do children at special schools and their parents perceive their HRQoL compared to children at open schools? *Health and Qual Life Outcomes.* 2010; 8: 72.
28. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res.* 2010; 19: 875-86.
29. Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res.* 2010; 19: 887-97.
30. Scott D, Ferguson GD, Jelsma J. The use of the EQ-5D-Y health related quality of life outcome measure in children in the Western Cape, South Africa: psychometric properties, feasibility and usefulness - a longitudinal, analytical study. *Health Qual Life Outcomes.* 2017; 15: 12.
31. Jelsma J. A comparison of the performance of the EQ-5D and the EQ-5D-Y health-related quality of life instruments in South African children. *Int J Rehabil Res.* 2010; 33: 172-7.
32. Ratcliffe J, Huynh E, Stevens K, et al. Nothing About Us Without Us? A Comparison of Adolescent and Adult Health-State Values for the Child Health Utility-9D Using Profile Case Best-Worst Scaling. *Health Econ.* 2016; 25: 486-96.
33. Herdman M, Fox-Rushby J, Badia X. 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires *Quality of Life Research.* 1997; 6: 237-47.
34. Connolly MA, Johnson JA. Measuring quality of life in paediatric patients. *Pharmacoeconomics.* 1999; 16: 605-25.
35. Feeny D. The Health Utilities Index: A tool for assessing health benefits. *PRO Newsletter.* 2005; 34.
36. Feeny D, Furlong W, Boyle M, et al. Multi-attribute health status classification systems. *Health Utilities Index. Pharmacoeconomics.* 1995; 7: 490-502.
37. Richardson J, McKie J, Bariola E. Multiattribute utility instruments and their use. In: Culyer AJ, ed., *Encyclopedia of health economics.* San Diego: Elsevier, 2014.
38. Seiber WJ, Groessl EJ, David KM, et al. Quality of Well Being Self-Administered (QWB-SA) Scale user's manual. . San Diego: Health Services Research Center, University of California, 2008.
39. Apajasalo M, Rautonen J, Holmberg C, et al. Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). *Qual Life Res.* 1996; 5: 532-8.
40. Apajasalo M, Sintonen H, Holmberg C, et al. Quality of life in early adolescence: A sixteen-dimensional health-related measure (16D). *Quality of Life Research.* 1996; 5: 205-11.
41. Richardson J, Day N, Peacock S, et al. Measurement of the Quality of Life for Economic Evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument. *The Australian Economic Review.* 2004; 37: 62-88.
42. Beusterien KM, Yeung JE, Pang F, et al. Development of the multi-attribute Adolescent Health Utility Measure (AHUM). *Health Qual Life Outcomes.* 2012; 10: 102.
43. Stevens K. Assessing the Performance of a New Generic Measure of Health-Related Quality of Life for Children and Refining it for Use in Health State Valuation. *Appl Health Econ Health Policy* 2011; 9: 157-69.

44. Kaplan R, Ganiats T, Sieber W, et al. The Quality of Well-Being Scale: critical similarities and differences with SF-36. *International Journal for Quality of Health Care*. 1998; 10: 509-20.
45. Kaplan R, Sieber W, Ganiats T. The quality of well-being scale: comparison of the interviewer-administered version with a self-administered questionnaire. *Psychol Health*. 1997; 12: 783-91.
46. Richardson JR, Peacock SJ, Hawthorne G, et al. Construction of the descriptive system for the Assessment of Quality of Life AQoL-6D utility instrument. *Health Qual Life Outcomes*. 2012; 10: 38.
47. Richardson J, Peacock S, Iezzi A, et al. Construction and Validation of the Assessment of Quality of Life (AQoL) Mark II Instrument. Research Paper 24. Melbourne: Monash University, 2007.