VALUE IN HEALTH 22 (2019) 348–354

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Patient-Reported Outcomes

Measuring Weight-Specific Quality of Life in Adolescents: An Examination of the Concurrent Validity and Test-Retest Reliability of the WAItE

Yemi Oluboyede, PhD, Tomos Robinson, PhD[*]

Health Economics Group, Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK

ABSTRACT

Objectives: To examine the concurrent validity of the Weight-specific Adolescent Instrument for Economic evaluation (WAItE) as compared with the generic, preference-based Child Health Utility 9D (CHU-9D) and the weight-specific Youth Quality of Life—Weight (YQOL-W) and also to examine the test-retest reliability of the WAItE. **Methods:** An online survey was used to administer the 3 instruments on a sample of adolescents (aged 11-18 years). Individual responses were converted into either utility scores (CHU-9D) or health-related quality-of-life scores (WAItE and YQOL-W). A 10% subsample of the respondents also completed the WAItE 1 week after completion to assess test-retest reliability. **Results:** One thousand adolescents completed the online survey. There was a strong correlation between the WAItE and both the CHU-9D (0.731; *P*<.001) and the YQOL-W (0.747; *P*<.001). All 3 instruments were able to discriminate according to different weight status categories and a measure of self-assessed health. Unlike the CHU-9D or the YQOL-W, the WAItE did not show a substantial ceiling effect. The WAItE also showed acceptable levels of test-retest reliability. **Conclusions:** The study results are encouraging, and illustrate that the WAItE can be used to reliably and accurately measure weight-specific outcomes in the younger population. The WAItE can also be used to assess outcomes in cost-effectiveness analysis of weight management interventions for young people, given the instrument is less likely to display ceiling effects and may thus be more sensitive in measuring change that results from interventions developed for this population.

Keywords: adolescence, CHU-9D, health-related quality of life, obesity, WAItE, YQOL-W

Introduction

Obesity has been described as an escalating global epidemic by the World Health Organization[1] and is estimated to cause about 3.4 million deaths per year worldwide.[2] As with the adult population, the prevalence of obesity in adolescence (approximately between the ages of 11 and 18 years) has been increasing in recent years.[3,4] This raises public health concerns, because it has been shown that pediatric obesity tracks into adulthood.[5] Obesity has been associated with negative consequences that have both immediate and long-term implications on health and health-related quality of life (HRQOL).[6]

Several obesity prevention initiatives have been established by public sector organizations, such as the National Child Measurement Programme, managed by Public Health England. Furthermore, in 2006 the National Institute for Health and Care Excellence developed the first national guidelines on the prevention, identification, assessment, and management of obesity in adults and adolescents in England and Wales.[7] These guidelines recommended several interventions for the prevention and management of obesity in young people, including interventions related to lifestyle, behavior, physical activity, and diet. Nevertheless, although such public health interventions may be effective in reducing levels of adolescent obesity, it is becoming increasingly important for efficient resource allocation to also assess the value for money of such interventions.[8]

Cost-utility analysis (CUA) is commonly used to inform whether new interventions should be made available within a publicly funded healthcare system. In CUA, benefits are commonly measured using quality-adjusted life-years (QALYs), which take into account both the length of life and the HRQOL of the patients. Preference-based measures (PBMs) can be used to calculate an individual utility score, where a value of 1 represents full health and a value of 0 represents death. PBMs differ from non-PBMs in the way the scoring algorithms have been derived. Specifically, these scoring algorithms are estimated from the

* *Address correspondence to*: Tomos Robinson, PhD, Health Economics Group, Institute of Health and Society, Newcastle University, Baddiley-Clark Bldg, Newcastle upon Tyne NE2 4AA, UK.

E-mail: tom.robinson@ncl.ac.uk

1098-3015/$36.00 - see front matter

https://doi.org/10.1016/j.jval.2018.10.001

values that patients place upon different aspects of health, instead of being a summative scoring procedure.[9] Although generic PBMs are in theory comparable across multiple clinical areas, condition-specific PBMs may be more sensitive to certain disease-specific improvements, and therefore may be a better way of valuing patient benefit.

Within the obesity literature, a number of weight-specific HRQOL instruments have been developed for adolescents, including the KINDL-obesity module,[10] the Impact of Weight on Quality of Life-Kids (IWQOL-Kids),[11] the Moorhead-Ardelt Quality of Life Questionnaire (M-A QoLQ),[12] Sizing Me Up,[13] Youth Quality of Life—Weight (YQOL-W),[14] and Oxford Pediatric Obesity Instrument (OPOI).[15] None of these tools are, however, preference-based, nor were they designed for this purpose. Furthermore, to date only the YQOL-W has been fully validated. In response to this, a weight-specific HRQOL instrument for adolescents, the Weight-specific Adolescent Instrument for Economic evaluation (WAItE), has been developed. The WAItE is a brief 7-item measure incorporating the views and experiences of adolescent girls and boys aged 11 to 18 years (for full details of the development of the WAItE, please see the study by Oluboyede et al[16]). The WAItE was developed to be feasible for a valuation study to be conducted to derive preference weight to operationalize the calculation of QALYs, and it is currently the only weight-specific tool for adolescents with this quality. To date, a valuation study is yet to be carried out. It is important to conduct a robust validation of the WAItE to provide evidence of its criterion validity and reliability to endorse future use.

The first aim of this study was to examine the concurrent validity of the WAItE, because it is imperative to evaluate the performance of the WAItE against previously validated HRQOL tools.[17] The second aim of this study was to evaluate the test-retest reliability of the WAItE to ensure that the WAItE is free of measurement error and gives consistent responses over 2 time points where no change in weight status has been observed.[18]

Methods

Sample

An online survey was developed with the survey company Survey Sampling International (SSI) for administration to a sample of adolescents aged 11 to 18 years residing in the United Kingdom. The survey contained 2 sections. The first section comprised a series of sociodemographic questions on age, sex, height, weight, self-assessed weight status, and self-assessed health. The second section comprised the WAItE, Child Health Utility 9D (CHU-9D), and YQOL-W instruments, randomized in order.

A sample of about 15 000 adult participants with children between the ages of 11 and 15 years was approached to complete the survey from SSI's large participant panel. The 11- to 15-year-old participants were then able to complete the survey, given the consent of their guardian. Furthermore, about 2500 16- to 18-year-olds were directly invited to complete the survey by SSI. The survey was left open until 1000 participants had completed the survey. Respondents to SSI surveys receive an average of £0.30 per 5-minute interview. The median time to complete the survey was 6 minutes.

To examine the test-retest reliability of the WAItE, a 10% subsample of the respondents was contacted again about 1 week later, and asked to complete a shortened version of the questionnaire for a second time. This 1-week time period was based on a similar study examining the test-retest reliability of the YQOL-W.[14] Most of those contacted again completed the shortened survey within 10 days, with the maximum time between the completion of the 2 surveys being 18 days (n = 2). The survey was completed between July 5 and September 25, 2017, and was approved by Newcastle University's Faculty of Medical Sciences Research Ethics Committee (project reference: 1262/12643).

The concurrent validity of the WAItE as compared with the CHU-9D and the YQOL-W was assessed in terms of its ability to differentiate between individuals of different weight statuses. The CHU-9D was chosen as a comparator because it is a generic, preference-based HRQOL tool commonly used in economic evaluation, whereas the YQOL-W was chosen as a comparator because it was the only pediatric weight-specific HRQOL tool available at the beginning of the study. A quota on weight status was stipulated initially, with an aim of having the 1000 respondents split equally between 3 weight groups: normal, overweight, and obese. The 3 weight groups were formed on the basis of the respondent's self-reported height and weight, and the age-specific cutoff points established by Cole et al.[19] Adolescents in the 85th percentile of the age- and sex-adjusted weight distribution were considered overweight, and those in the 95th percentile were seen as being obese.

HRQOL Scoring and Utility Measurement

Weight-specific Adolescent Instrument for Economic evaluation

The WAItE has 7 dimensions (related to tiredness, walking, participation in sports, concentration, embarrassment, unhappiness, and being treated differently), with a 5-level frequency response scale representing the increasing degrees of severity (ranging from "never" to "always"). The WAItE total score is calculated by simply summing the answers of the 7 dimensions, and is scored between 7 and 35.[16] In the analysis, this WAItE total score was reverse-coded so that a higher WAItE total score indicated a higher quality of life, in line with the CHU-9D.

Child Health Utility 9D

The CHU-9D has 9 dimensions (related to being worried, sad, in pain, tired, annoyed, schoolwork, sleep, daily routine, and ability to join in activities), each with a 5-level frequency response scale representing increasing degrees of severity (ranging from, eg, "I don't feel worried today" to "I feel very worried today").[20] The instrument has been validated for use in adolescent populations. We used the scoring algorithm on the basis of the preferences of the UK adult general population, using the syntax provided by the authors. These utility scores have a minimum value of 0.33 and a maximum value of 1.

Youth Quality of Life—Weight

The YQOL-W was developed as a measure of weight-specific quality of life in youth, and was based on more than 50 interviews with adolescents living in the United States and Mexico.[14] The YQOL-W has 21 dimensions (including those related to depression, exercise, and social anxiety), each with 11 different levels of severity (ranging from "not at all" to "very much"). The YQOL-W can be reverse-coded and converted into a total score between 0 and 100, with 100 being the maximum score.

Statistical Analysis

To test the concurrent validity of the WAItE, we first compared its statistical properties with those of the generic, preference-based CHU-9D and the weight-specific YQOL-W. Descriptive analyses, including means, standard deviations, and medians, were initially estimated. The distribution of the CHU-9D utility score and the WAItE and YQOL-W HRQOL total scores were tested for normality using the Shapiro-Francia test,[21] and the nonparametric Spearman rank correlation coefficient was used to assess the level of agreement between the instruments.[22] In line with Cohen,[23] a coefficient higher than 0.8 was seen to indicate a high level of correlation, and a coefficient of 0.6 to 0.8 was seen to indicate a good level of correlation. In addition, Bland-Altman plots[24] were used to study the limits of agreement between the WAItE and both the CHU-9D and the YQOL-W. We hypothesized that the highest correlation would be seen between the WAItE and the YQOL-W, because both these instruments are weight-specific rather than generic.

The concurrent validity of the WAItE was further assessed by analyzing the performance of the WAItE compared with those of the CHU-9D and the YQOL-W in its ability to discriminate between sex- and age-adjusted body mass index categories, the 5 levels of self-reported general health, the 6 levels of self-reported weight status, and the presence or absence of a long-term illness or disability. It was expected that respondents with better health statuses (eg, those who reported themselves to have "excellent" or "good" health) would have a higher HRQOL than those who had worse health statuses. Because of the non-normal distribution of all 3 instruments, 2 nonparametric tests (the Kruskal-Wallis test[25] and the Mann-Whitney *U* test[26]) were adopted to compare the respective utilities and HRQOL scores between the various subgroups.

To examine the test-retest reliability of the WAItE, we used a 10% random subsample of the respondents who completed the WAItE again 1 week later, and compared the responses with those from the full estimation sample. The test-retest reliability of the WAItE was assessed using methods recommended by the Consensus-based Standards for the Selection of Health Measurement Instruments checklist manual.[27] To assess the test-retest reliability of the continuous WAItE total score, the intraclass correlations coefficient (ICC) was used.[28] This measure is seen to be superior to measures such as the Spearman rank correlation coefficient, because it takes into account the possibility of systematic error. Results from individual mixed-effects models are presented, as the "raters" (the survey participants) were fixed over time.[29] The guidelines of Cicchetti[30] were used to assess the

Table 1 – Participant characteristics.

Characteristic	n (%)	WAItE total score	CHU-9D utility score	YQOL-W quality-of-life score
Full sample	975 (100)			
Mean ± SD		25.39 ± 5.99	0.81 ± 0.14	69.31 ± 30.85
Median (IQR)		26 (21-30)	0.83 (0.70-0.92)	79.52 (44.29-98.57)
Sex				
Male	482 (49.4)	26.25 (5.72)	0.82 (0.14)	73.25 (31.10)
Female	493 (50.6)	24.55 (6.14)	0.79 (0.15)	64.89 (30.07)
P value	–	<.001	<.001	<.001
Age group(y)				
11-15	361 (37.0)	28.60 (5.43)	0.88 (0.12)	82.19 (25.79)
16-18	614 (63.0)	23.50 (5.48)	0.76 (0.14)	61.29 (30.97)
P value	–	<.001	<.001	<.001
Weight status				
Normal	353 (36.2)	27.10 (5.64)	0.84 (0.14)	82.18 (24.17)
Overweight	221 (22.7)	26.10 (5.67)	0.81 (0.14)	73.23 (26.85)
Obese	401 (41.1)	23.49 (5.93)	0.78 (0.14)	55.12 (32.41)
P value	–	<.001	<.001	<.001
Self-assessed weight				
Very overweight	97 (10.0)	19.66 (4.81)	0.68 (0.11)	27.07 (21.23)
Moderately overweight	141 (14.4)	22.15 (5.74)	0.75 (0.14)	45.83 (26.40)
Slightly overweight	251 (25.8)	24.23 (5.09)	0.78 (0.13)	61.83 (26.33)
About the right weight	418 (42.8)	28.26 (5.16)	0.87 (0.13)	88.21 (19.83)
Slightly underweight	59 (6.1)	27.22 (5.40)	0.84 (0.14)	86.34 (16.37)
Moderately underweight	7 (0.7)	25.86 (4.67)	0.74 (0.15)	85.85 (15.68)
Very underweight	2 (0.2)	22.50 (6.36)	0.69 (0.44)	52.38 (6.06)
P value	–	<.001	<.001	<.001
Self-assessed health				
Excellent	168 (17.2)	29.95 (4.72)	0.92 (0.10)	89.28 (22.66)
Very good	275 (28.2)	27.58 (5.03)	0.86 (0.11)	80.43 (24.55)
Good	313 (32.1)	23.72 (5.29)	0.77 (0.13)	59.54 (30.94)
Fair	157 (16.1)	21.96 (5.36)	0.72 (0.14)	55.13 (29.59)
Poor	62 (6.4)	20.48 (5.81)	0.68 (0.16)	46.62 (27.58)
P value	–	<.001	<.001	<.001
Illness or disability				
Yes	560 (57.4)	23.06 (5.68)	0.75 (0.14)	58.46 (31.20)
No	415 (42.6)	28.54 (4.85)	0.89 (0.11)	83.27 (23.89)
P value	–	<.001	<.001	<.001

Note. As per Cole et al,[19] individuals in the 85th percentile of the weight distribution for their age and sex were classed as overweight, and those in the 95th percentile were classed as being obese.
CHU-9D indicates Child Health Utility 9D; WAItE, Weight-specific Adolescent Instrument for Economic evaluation; YQOL-W, Youth Quality of Life—Weight.

degree of agreement using the ICC. An ICC of less than 0.4 was seen to indicate poor agreement, an ICC of between 0.41 and 0.6 was seen to indicate fair agreement, an ICC of between 0.61 and 0.74 was seen to indicate good agreement, and an ICC of more than 0.75 was seen to indicate almost perfect agreement.

To assess the test-retest reliability of the ordinal, individual scales of the WAItE, weighted κ coefficients were used. This measure is seen to be superior to the standard κ coefficient or the proportion agreement, because it takes into account the possibility of chance agreement.[31] Results with both linear and quadratic weights are presented. In line with Canaway and Frew,[32] the guidelines of Landis and Koch[33] were used to assess the degree of agreement using κ coefficients. A κ coefficient less than 0.2 indicates poor agreement, a coefficient between 0.21 and 0.40 indicates fair agreement, a coefficient between 0.41 and 0.6 indicates moderate agreement, a coefficient between 0.61 and 0.8 indicates substantial agreement, and a coefficient more than 0.81 indicates almost perfect agreement. All statistical analyses were conducted using Stata version 14.1 (StataCorp, College Station, TX).[34]

Results

Comparison of the WAItE, CHU-9D, and YQOL-W

A total of 1000 participants completed the online survey. The survey did not allow the respondents to answer the next question in the survey until the current question had been fully completed, and therefore there were no missing data on any of the key variables. Nevertheless, there were 25 respondents excluded from the final sample because of unfeasible body mass index values. The final sample included more respondents classed as obese (n = 401 [41.13%]) and normal weight (n = 353 [36.21%]) than overweight (n = 221 [22.66%]), although there were still sufficient numbers in the 3 groups to conduct a robust empirical analysis. As presented in Table 1, the mean age of the respondents was 15.4 years, and 50.6% of the respondents were female; 22.5% of the adolescents reported their health as either "fair" or "poor," 24.4% reported being "moderately overweight" or "very overweight," and 57.4% reported as having some form of illness or disability.

The distributions of the WAItE, CHU-9D, and YQOL-W are given in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.10.001. The mean WAItE score was 25.39, whereas the median score was 26. The WAItE did not display a substantial ceiling effect, with only 3.1% of the responses reporting the maximum WAItE total score of 35. The mean utility of the CHU-9D was 0.81, whereas the median utility was 0.83. The distribution of the CHU-9D utility score was negatively skewed and showed evidence of having a ceiling effect, with 20.1% of the respondents having a reported utility of more than 0.95, and 9.6% of the respondents reporting a maximum utility value of 1. Previous studies using the CHU-9D have also reported this ceiling effect.[32,35,36] The mean YQOL-W score was 69.31, whereas the median score was 79.52. The YQOL-W score was extremely negatively skewed and showed evidence of a substantial ceiling effect, with 21.2% of the respondents reporting the maximum total score of 100.

Table 1 also presents WAItE, CHU-9D, and YQOL-W values according to key sociodemographic variables and health status. Across the WAItE, CHU-9D, and YQOL-W, statistically significant differences were observed for age and sex, with girls and older adolescents reporting a lower level of HRQOL than boys and younger adolescents, respectively. All 3 instruments were also able to discriminate according to varying levels of self-reported weight and health status, as well as those who reported themselves as living with a long-standing illness or disability. For example, those respondents who self-reported their weight as being "about right" had an average WAItE total score of 28.26, compared with an average WAItE total score of 19.66 for those who reported their weight as being "very overweight." In addition, those respondents who self-reported their health as being "excellent" had an average WAItE total score of 29.95, compared with an average WAItE score of 20.48 for those who reported their health as being "poor."

Figures 1 and 2 display scatterplot comparisons between both the WAItE and the CHU-9D and the WAItE and the YQOL-W, respectively, with clear evidence of positive correlation in both cases. The Spearman rank correlation coefficient between the WAItE total score and the CHU-9D utility score was 0.731, whereas the correlation between the WAItE total score and the YQOL-W total score was 0.747. The limit of agreement of the instruments was explored using a Bland-Altman scatterplot.[24] On average, only 4.9% of the respondents were outside the 95% limits of agreement when comparing the WAItE with the CHU-9D, and 5.7% of the respondents lay outside the 95% limits of agreement when comparing the WAItE with the YQOL-W, implying a high level of agreement in both cases.

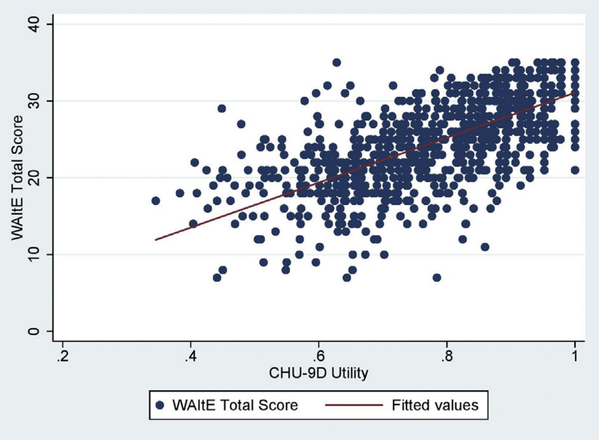

Fig. 1 – Scatterplot comparison of the WAItE total score and the CHU-9D utility score. CHU-9D indicates Child Health Utility 9D; WAItE, Weight-specific Adolescent Instrument for Economic evaluation.

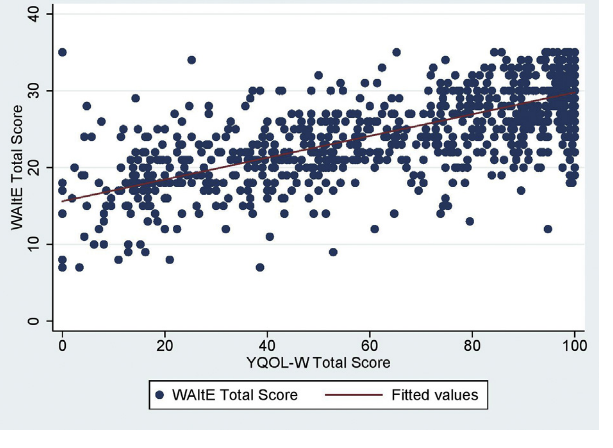

Fig. 2 – Scatterplot comparison of the WAItE total score and the YQOL-W total score. WAItE indicates Weight-specific Adolescent Instrument for Economic evaluation; YQOL-W, Youth Quality of Life—Weight.

Table 2 – ICCs for the WAItE total score (n = 97).

ICC model	ICC (95% CI)
One-way individual random-effects	0.795* (0.709-0.858)
Two-way individual random-effects	0.796* (0.708-0.859)
Two-way individual mixed-effects	0.801* (0.716-0.862)

CI indicates confidence interval; ICC, intraclass correlation coefficient; WAItE, Weight-specific Adolescent Instrument for Economic evaluation.
* P<.010.

The correlations between the individual dimensions of the WAItE and the CHU-9D and YQOL-W dimensions are provided in Supplemental Materials. Tiredness, a common dimension between the WAItE and the CHU-9D, had the highest correlation (r = 0.562) when comparing the instruments. Several other dimensions also had relatively large correlations, particularly those related to the "work" and "activity" CHU-9D dimensions. Nevertheless, several dimensions between the 2 instruments were only modestly correlated. For example, the "sports" WAItE dimension was relatively poorly correlated with several CHU-9D dimensions, including "pain" (r = 0.231) and sleep (r = 0.228), and the "tired" WAItE dimension was relatively poorly correlated with the "sleep" (r = 0.226) and "work" (r = 0.227) CHU-9D dimensions.

The strength of correlation between the individual dimensions of the WAItE and the YQOL-W was in general larger than the strength of correlation between the individual dimensions of the WAItE and the CHU-9D. The highest correlations were between the "embarrassment" and "unhappy" WAItE dimensions and various measures of numerous YQOL-W dimensions, including a dimension relating to difficulties finding appropriate clothes (r = 0.632 and 0.612) and a dimension related to being ashamed (r = 0.634 and 0.595). Nevertheless, not all dimensions were as strongly correlated as these examples. This was particularly evident for the "tired" dimension of the WAItE, which had particularly low correlations with the YQOL-W dimensions related to inclusion (r = 0.165) and difficulties finding employment (r = 0.161).

Test-Retest Reliability

Tables 2 and 3 present the results for tests for the test-retest reliability of the WAItE. As presented in Table 2 and intuitively in the scatterplot of Figure 3, the WAItE total score showed a high level of test-retest reliability, with individual ICCs higher than the 0.75 threshold value outlined by Cicchetti[30] to indicate excellent reliability. As presented in Table 3, in general the WAItE dimension scores also showed acceptable levels of test-retest reliability, with percentage agreements ranging between 86.86% and 90.21%, and all weighted κ coefficients indicating either moderate or substantial agreement, depending on whether the linear or quadratic weighting strategy was used. We also conducted a "sense check" of the reliability of the self-reported weight status measure. Although there were some instances where adolescents reported movements between weight statuses between the 2 time points, the measure was generally stable (all weighted κ statistics indicated substantial agreement irrespective of which weighting strategy was used).

Discussion

This article has presented the methods and findings from an empirical study analyzing the concurrent validity and test-retest reliability of the WAItE. This is the first study to compare the WAItE with other validated HRQOL instruments, and the first study to formally assess the reliability of the WAItE. Ideally, the performance of the WAItE would have been tested against a well-established weight-specific preference-based tool that followed the criterion standard of instrument development.[37] The CHU-9D and YQOL-W were seen as being the best comparators to test the performance of the WAItE against, given that another instrument that follows these criterion standards does not yet exist.

First, the findings of this study demonstrate that the WAItE has good levels of concurrent validity, indicated by the high correlation of the WAItE with both the CHU-9D and the YQOL-W, and also the comparability of the WAItE with both instruments in its ability to accurately differentiate between both general and weight-specific health statuses. This indicates that the same latent construct of HRQOL is being measured in the WAItE and the previously validated measures. As expected, the correlation between the WAItE and the YQOL-W was higher than the correlation between the WAItE and the CHU-9D, almost certainly because of the fact that the WAItE and the YQOL-W were designed to specifically measure weight-specific HRQOL, whereas the CHU-9D is a generic measure. Second, the findings of this study demonstrate that the WAItE overall displays acceptable levels of test-retest reliability (ICC = 0.795), indicating that the WAItE is stable and reliable over time. As a point of comparison, the YQOL-W has previously been shown to have an ICC of 0.77.[14]

As well as showing good levels of concurrent validity in relation to the CHU-9D and the YQOL-W, it appears that the WAItE may have a significant advantage compared with both of these measures, because it does not exhibit a substantial ceiling effect. A substantial ceiling effect can indicate limited content validity

Table 3 – Test-retest correlations and percentage agreement for individual WAItE attributes (n = 97).

Measure and item	Test-retest correlation (Spearman rank)	Percentage agreement (weighted κ)	Weighted κ coefficient (linear weights)	Weighted κ coefficient (quadratic weights)
Tired	0.584*	88.66*	0.517*	0.606*
Walking	0.678*	87.11*	0.521*	0.663*
Sports	0.811*	90.21*	0.710*	0.798*
Concentration	0.684*	86.86*	0.563*	0.665*
Embarrassed	0.661*	87.11*	0.590*	0.674*
Unhappy	0.691*	86.60*	0.488*	0.628*
Treated different	0.608*	88.92*	0.522*	0.602*

WAItE indicates Weight-specific Adolescent Instrument for Economic evaluation.
* P<.010.

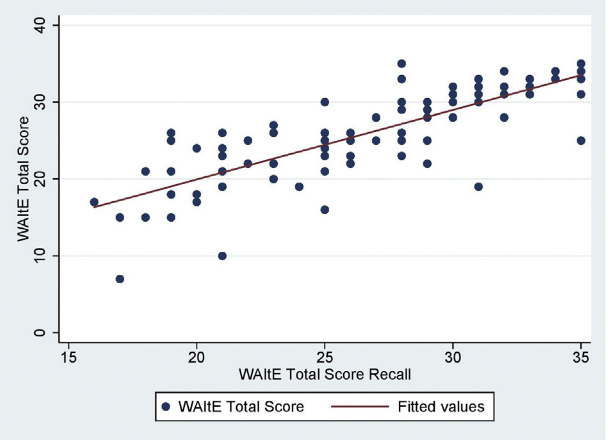

Fig. 3 – Scatterplot comparison of the WAItE total score and the WAItE total score recall. WAItE indicates Weight-specific Adolescent Instrument for Economic evaluation.

and a reduced reliability,[38] and can be considered a problem if 15% to 20% of respondents achieve the best possible score.[39] Given the lack of a ceiling effect, the WAItE may be more appropriate when measuring weight-specific adolescent HRQOL, because it is more likely to pick up meaningful changes in HRQOL, particularly at the higher end of the distribution.

This study has several limitations. First, the original aim was to have the sample equally split between normal weight, overweight, and obese adolescents to conduct a robust comparison between the 3 groups. Ultimately, the final estimation sample included 353 (36.2%) adolescents with a normal weight, 221 overweight adolescents (22.7%), and 401 obese adolescents (41.1%), and therefore cannot be seen to be truly representative of the UK population. Recent research indicates that about 35% of adolescents can be considered either overweight or obese in the United Kingdom[40] compared with 63.8% in our sample. Self-reported health was also lower than one would expect, with 22.3% of the adolescents reporting their health as either “fair” or “poor.” Relatively recent nationally representative figures from the UK Household Longitudinal Study have shown that this figure is more likely to be about 7%.[41] Furthermore, an unusually high number of respondents in our sample (57.2%) reported themselves as having some form of illness or disability compared with a national average of about 12%, as reported in the 2015 Labour Force Survey.[42] Although the overrepresentation of overweight and obese adolescents in the sample meant that a higher than average number of individuals reporting some form of illness or disability was expected, this was still a surprising finding. Nevertheless, the purpose of the study was not to be representative of objectively measured weight status according to the UK population. The oversampling of overweight and obese adolescents in the sample was seen to be necessary, considering that a principal aim of the study was to identify whether the HRQOL instruments were able to differentiate between individuals with different weight statuses, and therefore a certain number of individuals were needed in each subgroup to conduct a robust statistical analysis. In addition, our sample can be seen as more representative of the individuals who may benefit from a weight management intervention, and therefore can be considered the population of interest.

The limitations associated with the accuracy of self-reported measures of height and weight should also be noted[43]; it is, however, usual for measurements of weight status to rely on self-report in the current field of study. Furthermore, as with any study using online data collection methods, there is no guarantee that the respondents completed the survey themselves, and there is also the potential for both sample-selection bias and nonresponse bias resulting from the survey sampling methods. Finally, there is also a risk that the 1-week retest period may have been too short, therefore biasing the results through memory effects.

Planned future research includes the administration of the WAItE to different population groups and the estimation of a mapping algorithm to map between the weight-specific WAItE and the generic, preference-based CHU-9D. This mapping algorithm will enable researchers to generate a valid prediction of the CHU-9D utility score from the WAItE, and therefore facilitate an indirect estimate of QALYs using the WAItE. As argued by Brazier et al,[44] however, “a mapping exercise is always a second best exercise compared to either the direct use of [a generic PBM] or a valuation of the condition-specific instrument.” Therefore, further planned research also includes the generation of preference weights using a representative UK population to calculate QALYs, as per the National Institute for Health and Care Excellence recommendations.[6] Currently, the WAItE is a reliable and valid tool that would be appropriate for direct use in the measurement of health outcomes in public health or weight management interventions. The generation of a WAItE utility score would also enable assessment of CUA of weight management interventions and facilitate further comparisons between the WAItE and the CHU-9D.

Conclusions

Overall, the findings of this study are encouraging, and illustrate the potential for the WAItE to be used when measuring weight-specific HRQOL in adolescence. The WAItE therefore may be a key tool for decision makers now and in the future in the evaluation of weight management services and for organizations that run weight management services.

Acknowledgments

We thank 3 anonymous reviewers and Elena Olariu for constructive comments on earlier versions of the manuscript. All errors are solely those of the authors.

Source of financial support: This study was funded by the Health Foundation (project reference: C0150R4035).

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.10.001.

REFERENCES

1. World Health Organization. Controlling the global obesity epidemic. http://www.who.int/nutrition/topics/obesity/en/. Accessed April 20, 2017.
2. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet.* 2014;384(9945):766–781.
3. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *J Am Med Assoc.* 2012;307(5):491–497.
4. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. *J Am Med Assoc.* 2014;311(8):806–814.

5. Singh AS, Mulder C, Twisk JW, Van Mechelen W, Chinapaw MJ. Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obes Rev*. 2008;9(5):474–488.
6. Søltoft F, Hammer M, Kragh N. The association of body mass index and health-related quality of life in the general population: data from the 2003 Health Survey of England. *Qual Life Res*. 2009;18(10):1293–1299.
7. National Institute for Health and Care Excellence. *Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children*; 2006. https://www.nice.org.uk/guidance/cg43/documents/obesity-consultation-nice-version2. Accessed November 3, 2017.
8. Drummond MF, Sculpher MJ, Claxton K, et al. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford, UK: Oxford University Press; 2015.
9. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. 2nd ed. Oxford, UK: Oxford University Press; 2017.
10. Ravens-Sieberer U, Redegeld M, Bullinger M. Quality of life after in-patient rehabilitation in children with obesity. *Int J Obes Relat Metab Disord*. 2001;25(S1):S63.
11. Kolotkin RL, Zeller M, Modi AC, et al. Assessing weight-related quality of life in adolescents. *Obesity*. 2006;14(3):448–457.
12. Moorehead MK, Ardelt-Gattinger E, Lechner H, Oria HE. The validation of the Moorehead-Ardelt quality of life questionnaire II. *Obes Surg*. 2003;13(5):684–692.
13. Zeller MH, Modi AC. Development and initial validation of an obesity-specific quality-of-life measure for children: sizing me up. *Obesity*. 2009;17(6):1171–1177.
14. Morales LS, Edwards TC, Flores Y, Barr L, Patrick DL. Measurement properties of a multicultural weight-specific quality-of-life instrument for children and adolescents. *Qual Life Res*. 2011;20(2):215–224.
15. Doyle S. Development of a new health related quality of life instrument for use in paediatric obesity. *Obes Rev*. 2011;12:77–78.
16. Oluboyede Y, Hulme C, Hill A. Development and refinement of the WAItE: a new obesity-specific quality of life measure for adolescents. *Qual Life Res*. 2017;26(8):2025–2039.
17. Murphy KR, Davidshofer CO. *Psychological Testing: Principles and Applications*. 4th ed. Upper Saddle River, NJ: Prentice-Hall; 1998.
18. van Agt HME, Essink-Bot M-L, Krabbe PFM, Bonsel GJ. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. *Soc Sci Med*. 1994;39(11):1537–1544.
19. Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. *Arch Dis Child*. 1995;73(1):25–29.
20. 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(8):1105–1113.
21. Shapiro SS, Francia RS. An approximate analysis of variance test for normality. *J Am Stat Assoc*. 1972;67(337):215–216.
22. Conover WJ. *Practical Nonparametric Statistics*. 3rd ed. New York, NY: John Wiley & Sons; 1999.
23. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. London: Routledge; 1988.
24. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307–310.
25. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc*. 1952;47(260):583–621.
26. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18:50–60.
27. Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol*. 2006;6:2.
28. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30–46.
29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420.
30. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284.
31. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213.
32. 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(1):173–183.
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
34. StataCorp LP. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp; 2015.
35. Chen G, Flynn T, Stevens K, et al. Assessing the health-related quality of life of Australian adolescents: an empirical comparison of the child health utility 9D and EQ-5D-Y instruments. *Value Health*. 2015;18(4):432–438.
36. Stevens K. Valuation of the child health utility 9D index. *Pharmacoeconomics*. 2012;30(8):729–747.
37. 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(8):1889–1905.
38. 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(1):34–42.
39. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293–307.
40. Niblett P, Lifestyle Statistics Team. *Statistics on Obesity, Physical Activity and Diet*. London: Health and Social Care Information Centre; 2015.
41. Siegler RBV. *Measuring national well-being—exploring the well-being of children in the UK*. 2014.
42. Office for National Statistics. Exploring the Well-being of Children in the UK Measuring National Well-being - Exploring the Well-being of Children in the UK; 2014. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/measuringnationalwellbeing/2014-10-08. Accessed November 3, 2017.
43. Rowland ML. Self-reported weight and height. *Am J Clin Nutr*. 1990;52(6):1125–1133.
44. Brazier JE, Kolotkin RL, Crosby RD, Williams GR. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. *Value Health*. 2004;7(4):490–498.