A Comparison of Field Measures of Adiposity Among Australian Adolescents from the Raine Study

Paola Chivers  
Helen Parker  
School of Health Sciences,  
The University of Notre Dame Australia, Fremantle

Beth Hands  
Institute for Health and Rehabilitation Research,  
The University of Notre Dame Australia, Fremantle  
Centre for Child Health Research,  
Telethon Institute for Child Health Research, Perth

Lawrence Beilin  
School of Medicine and Pharmacology,  
University of Western Australia, Perth

Garth Kendall  
Centre for Child Health Research,  
Telethon Institute for Child Health Research, Perth  
School of Nursing and Midwifery, Faculty of Health Sciences,  
Curtin University of Technology, Perth

Max Bulsara  
Institute for Health and Rehabilitation Research,  
The University of Notre Dame Australia, Fremantle  
School of Population Health, University of Western Australia, Perth.

Abstract

This study examined the possible inter-changeability of common proxy measures of adiposity in adolescent 14 year olds from a large cohort. A sample of 1,607 adolescents (n = 825 male and n = 782 female) was drawn from the Western Australian Pregnancy Cohort (Raine) Study. Pearson Product Moment Correlation, Bland-Altman method, and Bland regression were used to determine the level of agreement between common proxy adiposity measures. The Bland-Altman and Bland regression results supported the notion that the two indices (Waist-height-Ratio and Body-Mass-Index) are measuring
a similar construct \( (F = .974, p = .324) \). However, 95 percent limits of agreement differed between the methods (95% CI [21.04, 21.52] and [17.71, 24.90] respectively), with the Bland regression approach suggesting that the measures are not interchangeable. These results provide evidence that different adiposity measures are not comparable, particularly when tracking weight status over time. The validity of WHR as an adiposity measure is questioned, and the similarity in weight status groupings based on WHtR and BMI, indicate these measures are most comparable.

**Keywords**: Pearson Moment Correlation, Bland-Altman method, Bland Regression limits of agreement, adiposity, BMI, obesity, Raine Study, Waist-Height Ratio, Waist-Hip Ratio, waist girth.

**Introduction**

Obesity is considered an excess of body fat that predisposes an individual to adverse health consequences. When tracking and investigating the prevalence of obesity in large populations, valid and accurate field measures of body composition are required. However, it has been questioned how body fat can be accurately measured, and which proxy measure is a better indicator for adverse health consequences (Marshall, Hazlett, Spady, Conger & Quinney, 1991)? The selection of a measure in children and adolescents is constrained by time, cost, and reproducibility. Additional considerations must be made, especially for those with excess fat, as it is more difficult to measure consistently for obese individuals compared to their lean counterparts, across anthropometric measures (Heyward, 2001).

At the moment, a simple, direct field measure is not available. Instead there are many indirect measures of adiposity, each with their own limitations, assumptions, and criticisms (Goran, 1998; Marshall et al., 1991). Many factors including those related to gender, age, and puberty impact on the value of these commonly used adiposity measures. What constitutes classification of obesity itself is arbitrary, as each measure is more or less continuously related to the risk of adult disease. Not surprisingly, with no common percentile point for obesity across measures, there is the likelihood of differences in classification of continuous variables into categories of adiposity (Marshall et al., 1991).

Commonly used anthropometric measures consider body size and proportion such as height, weight, and circumference measures (Heyward, 2001; Hills, Lyell & Byrne, 2001). Anthropometric indices are a ratio between two different body measurements and include Body Mass Index (BMI), Waist-Hip Ratio (WHR), and Waist-Height Ratio (WHtR) (Hills et al., 2001). Anthropometric measures generally make for good field tests as they are relatively easy to administer, inexpensive, and require less technical skill compared to laboratory constrained tests such as the Dual energy X-ray absorptiometry (DXA). However, it is difficult to measure anthropometric measures consistently for obese individuals compared
A comparison of adiposity field measures
to their lean counterparts, with obese individuals more adverse to measurement of weight (Heyward, 2001).

The problem lies with which field measure to choose for studies using adiposity proxy measures. Is there a difference between these measures in weight status categorisation? For longitudinal studies, can we interchange or make comparisons between these proxy measures dependent upon which have been collected across time points?

This retrospective study reports on a large sub-sample of adolescents at average age 14 years ($M = 14.0$, $SD = 0.2$) recruited in utero as part of The Western Australian Pregnancy Cohort (Raine) Study. We compare height, weight, waist girth, hip girth, with the indices BMI, WHR, and WHtR. We test the similarities among these measures using the Bland-Altman method (Bland & Altman, 1986) with z-scores, an alternative regression approach (Bland, 2004; Bland & Altman, 2003), and the traditional Pearson Product Moment Correlation. The latter has been criticised as an inappropriate statistical analyses of similarity (Bland, 2004; Bland & Altman, 2003). We examine differences in weight status categorisation using the International Obesity Task Force (IOTF) BMI cut-offs (Cole, Bellizzi, Flegal & Dietz, 2000), WHtR cut-offs (Kahn, Imperatore, & Cheng, 2005; McCarthy & Ashwell, 2006), and waist girth cut-offs (Taylor, Jones, Williams & Goulding, 2000).

**Methods**

**Participants**

The sample for this study was drawn from The Western Australian Pregnancy Cohort (Raine) Study. The original study began in 1989 where investigators established a low risk, unselected cohort of pregnant women recruited at 16-18 weeks gestation with 2,868 children followed from birth (49.3% female). Details of data collection in this cohort have been previously described (Newnham, Evans, Michael, Stanley & Landau, 1993). The Raine cohort is well established and broadly representative of the West Australian population (Li, Kendall, Henderson, Downie, Landsborough & Oddy, 2008).

This retrospective study compares proxy adiposity measures collected on these adolescents in the 14th year survey wave, with their age at assessment ranging between 13-15 years, and a mean age of 14 years. By the 14th year survey wave there were 79% of the original cohort, with sample size varying for different measures, and ranging from 1,250-1,606 participants (refer to Table 1 for specific measure samples sizes). Across all measures 48% of the sample was female. The measures used in this study were only available for the 14th survey year.
Anthropometric Measures

Anthropometric measures were taken by trained staff of the Telethon Institute for Child Health Research. Where appropriate, adolescents were standing in the anatomical position, palms facing forward. Each area was measured at least twice in sequence with measures within one centimetre, or repeated until this criteria was met. All measures were taken at expiration.

Waist girth was measured over the belly button, with the flexible plastic tape horizontal and in the same plane. Hip girth was measured over light clothing with the measure taken over the widest part of the buttocks. Height was measured using a Holtain stadiometer with shoes off, and heels, bottom and head against the board. The chin was positioned to straighten the neck and the measure taken with a breath intake. Weight was measured to the nearest 100g using Wedderburn digital chair scales and the adolescent wearing light clothing (Burke et al., 2005).

Adiposity Indices

BMI was calculated from measured height and weight scores using the formula weight (kg) / height (m)². BMI cut-off points of 25 and 30, age and gender adjusted for children, were used to classify children as normal weight, overweight or obese as defined by the IOTF criteria (Cole et al., 2000), with the underweight category collapsed into the normal weight category. BMI categorisation for comparison with the two category normal weight and overweight WHtR (McCarthy & Ashwell, 2006) and waist girth (Taylor et al., 2000) was performed by collapsing underweight and normal weight into normal weight, and overweight

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (m)*</td>
<td>825 1.7 .09</td>
<td>781 1.6 .06</td>
<td>1606 1.6 .08</td>
</tr>
<tr>
<td>Weight (kg)*</td>
<td>825 58.7 14.2</td>
<td>782 56.6 11.9</td>
<td>1607 57.7 13.2</td>
</tr>
<tr>
<td>Hip girth (cm)*</td>
<td>650 89.0 9.3</td>
<td>600 92.1 8.5</td>
<td>1250 90.5 9.1</td>
</tr>
<tr>
<td>Waist girth (cm)*</td>
<td>815 76.4 11.5</td>
<td>767 74.5 10.0</td>
<td>1582 75.5 10.8</td>
</tr>
<tr>
<td>Waist-Height Ratio</td>
<td>815 .46 .07</td>
<td>766 .46 .06</td>
<td>1581 .46 .06</td>
</tr>
<tr>
<td>Body Mass Index**</td>
<td>825 21.1 4.2</td>
<td>781 21.5 4.1</td>
<td>1606 21.3 4.2</td>
</tr>
<tr>
<td>Waist Hip Ratio*</td>
<td>650 .85 .06</td>
<td>600 .81 .06</td>
<td>1250 .83 .06</td>
</tr>
</tbody>
</table>

* Significant gender difference p < .01.
** Significant gender difference p < .05.
and obese into overweight. WHtR was calculated using the formula waist girth (cm)/height (cm). WHtR cut-offs were selected to represent the cut-offs for BMI as proposed by Kahn and colleagues (2005), namely ≥ 0.539 obese (BMI ≥ 30), 0.490-0.539 at risk or overweight (BMI 25-30) and < 0.490 (normal) (BMI < 25) (Kahn et al., 2005), with the McCarthy and Ashwell (McCarthy & Ashwell, 2006) 0.5 cut-off used for the two category comparison normal weight and overweight. WHR was calculated using the formula waist girth (cm) / hip girth (cm). Age and gender adjusted waist girth cut-offs between normal and overweight was calculated according to those reported in Taylor and colleagues (Taylor et al., 2000).

**Statistical Analysis**

The statistical software SPSS for Windows, Rel.17.0.0. 2008 (SPSS Inc., Chicago, IL) was used for all statistical processes. Pearson Product Moment Correlation was calculated between each of the adiposity measures taken at average age 14 years. The categorisations of weight status using BMI, WHtR, and waist girth were compared using a Kappa Chi-square test. The Bland-Altman method (Bland & Altman, 1986) was used to assess the limits of agreement, with raw scores converted to z-scores prior to analysis and the resulting confidence intervals converted back into associated raw score units. The Bland regression method (Bland, 2004) determined 95 percent prediction limits at the mean.

**Results**

Descriptive statistics sample size, mean and standard deviations for adiposity measures are summarised in Table 1 for both total and gender separated samples. Apart from Waist-Height Ratio, all adiposity measures were significantly different across gender (p < .05). The results from the Pearson Product Moment Correlations between adiposity measures are presented in Table 2.

The Kappa Chi-square test for reliability showed an association between the BMI (Cole et al., 2000) and WHtR (Kahn et al., 2005) categories (n = 1,581, Kappa = 0.620, p < 0.001). For comparative purposes, BMI was then categorised into two groups by collapsing the IOTF overweight and obese into one group (group 1 = underweight and normal, group 2 = overweight and obese), and a WHtR cut-off of 0.5 was used (McCarthy & Ashwell, 2006). Associations were found between the two-category BMI and WHtR measures (n = 1,581, Kappa = 0.701, p < 0.001), and between waist girth categorisation with BMI (n = 1,581, Kappa = 0.671, p < 0.001) and WHtR (n = 1,581, Kappa = 0.615, p < 0.001).

Comparisons of BMI, WHR and WHtR measures are presented in Figure 1 using the Bland-Altman method (Bland & Altman, 1986). Plot A represents the
relationship (correlation) between the indices using raw scores, whilst plot B represents the level of agreement using z-scores. Regression analysis of the level of ‘agreement’ (z-scores) and confidence intervals (CI) (raw score units) between adiposity measures BMI, WHR, WHtR and waist girth are presented in Table 3. The regression of the Bland-Altman calculated bias and mean was not different from zero (that is, not significant), indicating no systematic relationship between bias and mean, hence a high level of agreement between measures. The small confidence intervals indicated small variation in differences between measures, with an acceptable degree of agreement.

The Bland regression analysis (Bland, 2004) reported a significant relationship between all comparisons ($p < .005$), suggesting that one variable could predict the other. However, calculation of the 95 percent prediction interval (CIs) at the mean (reported in Table 3) found that these prediction intervals (CIs) had a large variation, and prediction was not within acceptable limits for the clinical setting. For example comparison of BMI versus waist girth reported a confidence interval of between 17.7 kg/m$^2$ and 24.9 kg/m$^2$, which using the Cole cut-offs would classify a 14 year old male as normal weight or overweight.

**Discussion**

In this study our purpose was to evaluate the similarities between common adiposity measures in a large sample of adolescents using conventional Pearson Product Moment Correlation, the Bland-Altman method (Bland & Altman, 1986) and the Bland regression method (Bland, 2004). In addition an evaluation of different weight status categorisation methods (Cole et al., 2000; Kahn et al., 2005; McCarthy & Ashwell, 2006; Taylor et al., 2000) is reported. Results presented provide evidence that different adiposity measures are not comparable, particularly when tracking weight status over time.
Figure 1: Correlation Scatter Plots with Regression Line Fit and ‘Level of Agreement’ of Adiposity Measures Body Mass Index (BMI), Waist-Hip Ratio (WHR) and Waist-Height Ratio (WHtR) Using the Bland-Altman Method. Circle = Female, Square = Male, Plots A: Line is Fit at Total.
Initially the study development process saw the use of adiposity measure z-scores in the Bland-Altman method as an appropriate way to compare measures using different units. However, conversion to z-scores is not always considered appropriate for this method, as the limits of agreement are dependent on the variability of the sample. Alternatively, the Bland regression method was used as a more appropriate method for measures with different units (M Bland, personal communication, 2009, Bland & Altman, 2003). We compared results from both these methods; the 95 percent limits of agreement (Bland-Altman method) and 95 percent prediction interval (Bland regression method), and found large discrepancies in the confidence intervals identified by each method. Importantly the Bland regression method showed that prediction did not agree well with the actual measurement, the adiposity measures did not provide similar information, and they were not interchangeable.

Our findings support the work of Kahn and colleagues (2005) and Taylor and colleagues (2000) who reported the strongest significant correlations between waist girth and WHtR (0.938) and then with BMI (0.894). Waist girth and BMI

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Measure</th>
<th>Bland-Altman Method</th>
<th>Bland Regression Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Limits of agreement test&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Confidence Intervals</td>
<td>Confidence Intervals</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>BMI vs Waist girth (n = 1,581)</td>
<td>BMI</td>
<td>.974</td>
<td>.324</td>
</tr>
<tr>
<td></td>
<td>Waist girth</td>
<td>74.83 – 76.07</td>
<td>65.99 – 84.96</td>
</tr>
<tr>
<td>BMI vs WHtR (n = 1,581)</td>
<td>BMI</td>
<td>.906</td>
<td>.341</td>
</tr>
<tr>
<td></td>
<td>WHR</td>
<td>.456 - .463</td>
<td>.405 - .529</td>
</tr>
<tr>
<td>BMI vs WHR (n = 1,249)</td>
<td>BMI</td>
<td>.381</td>
<td>.537</td>
</tr>
<tr>
<td></td>
<td>WHR</td>
<td>.826 - .838</td>
<td>.721 - .936</td>
</tr>
<tr>
<td>Waist girth vs WHR (n = 1,250)</td>
<td>Waist girth</td>
<td>.373</td>
<td>.542</td>
</tr>
<tr>
<td></td>
<td>WHR</td>
<td>.826 - .836</td>
<td>.740 - .915</td>
</tr>
<tr>
<td>Waist girth vs WHtR (n = 1,581)</td>
<td>Waist girth</td>
<td>.001</td>
<td>.979</td>
</tr>
<tr>
<td></td>
<td>WHtR</td>
<td>.457 - .463</td>
<td>.424 - .424</td>
</tr>
<tr>
<td>WHR vs WHtR (n = 1,249)</td>
<td>WHR</td>
<td>.266</td>
<td>.606</td>
</tr>
<tr>
<td></td>
<td>WHtR</td>
<td>.456 - .466</td>
<td>.371 - .547</td>
</tr>
</tbody>
</table>

Note. BMI = body mass index; WHtR = waist-height ratio; WHR = waist-hip ratio.
1. Regression analysis of the level of agreement for each adiposity comparison was non-significant, indicating a high level of agreement.
A comparison of adiposity field measures

both have high reproducibility, with BMI slightly better (Freedman et al., 2007). Together these findings support the National Health and Medicine Research Council (NHMRC) stand that age adjusted BMI is a useful field measurement tool for adiposity, with waist girth an important clinical tool for determining health risk in children and adolescents (National Health and Medical Research Council, 2003).

The proxy adiposity ratios of WHtR and BMI were also strongly correlated (0.891), with the Bland-Altman method and Bland regression method results supporting the notion that there is a strong relationship. The sub-grouping of the sample by gender and three weight status BMI categories showed that the IOTF (Cole et al., 2000) and Kahn (Kahn et al., 2005) cut-offs provided similar groupings, which were highly correlated \( (p < .001) \). The BMI IOTF cut-offs (Cole et al., 2000) identified slightly fewer individuals as normal weight, more individuals as overweight, and less individuals as obese compared to the WHtR Kahn cut-offs (Kahn et al., 2005). Although the two category groupings (McCarthy & Ashwell, 2006) for BMI had a higher Kappa value with WHtR, there were more individuals categorised differently.

Based on the results from the Pearson Product Moment Correlation, Kappa Chi-Square, Bland-Altman method, and Bland regression method, WHtR should be considered as a valid adiposity measure for several reasons. Firstly, waist girth, which is recognised as a measure of central adiposity, is included in the derivation of the index and is strongly related to adverse health risk in children and adults (Garnett, Baur, Srinivasan, Lee & Cowell, 2007; Hsieh, Yoshinaga & Muto, 2003; Li, Ford, Mokdad & Cook, 2006; McCarthy, Jarrett, Emmett, Rogers & ALSPAC Study Team, 2005). Also, research suggests WHtR is better at predicting metabolic risk than other proxy measures, and is related to level of physical activity (Hsieh et al., 2003; McCarthy & Ashwell, 2006). Further, WHtR removes the psychological sensitivity associated with taking weight measurements. We also were able to show that the weight status classification of participants into the three categories normal weight, overweight and obese, was very similar for BMI and WHtR.

Our results lead us to question the usefulness of WHR as an adiposity measure in obesity studies and although statistically significant in this analysis, WHR had the weakest relationships with the other measures, possibly due to a greater potential for error in hip measurements. Hip girth correlated strongly with WHtR. There was no correlation between hip girth and its derivative WHR. An additional concern with WHR is that hip girth does not account for skeletal differences (Taylor et al., 2000). Its predictive ability of adverse health outcomes in children has been questioned (Fredriks, van Buuren, Fekkes, Verloove-Vanhorick & Wit, 2005), and the NHRMC does not consider WHR as a clinical measure (National Health and Medical Research Council, 2003). There are also no internationally agreed cut-offs for obesity using WHR among children, with those available relating only to adults (Australian Institute of Health and Welfare, 2005).
Most critically, we demonstrated the use of several statistical methods to examine similarity and inter-changeability of adiposity measures. The Bland-Altman method with z-scores reported strong agreement between some measures. However, the more appropriate Bland regression method, found large variation in limits that are unacceptable clinically. The adiposity measures examined in this study, although sharing strong relationships, appear to not provide similar information. These findings highlight the need to interpret most field-based measures of obesity with caution.

Conclusion

The findings from this large sample study have enabled an examination as to the most feasible and accurate adiposity measure for obesity studies in adolescents. Results were based on the statistical Bland regression method, whilst supplementing considerations of most appropriate statistical processes. The value of WHR as an adiposity measure is questioned due to its poor associations with other measures and based on limited evidence on its clinical value. The close relationship between WHtR and BMI indicates they are probably more similar measures of adiposity. Waist girth alone shows value as an adiposity measure and should be considered, particularly given its reported associations to health risk. As WHtR is similar to BMI, especially in weight status group categorisation, it may be a valid alternative, especially when individuals tend to be more sensitive to having their weight measured than their waist girth. The simplicity of waist girth and WHtR as adiposity measures would be enhanced if there were international standards for cut-off points to identify obesity among children through to adulthood. However the application of arbitrary cut off points to classify the weight status of individuals for variables that are frequently related to risk of ill health carries its own conceptual hazards, particularly at times when population levels of body fat are changing rapidly in children and adults.

In summary, the results provide evidence that different adiposity measures are not comparable, particularly when tracking weight status over time. The validity of WHR as an adiposity measure is questioned and the similarity in weight status groupings based on WHtR and BMI, indicate these measures are most comparable.

Acknowledgements

We are extremely grateful to all the families who took part in this study and the whole Raine Study team. The Raine Study is funded by The Telethon Institute for Child Health Research, the University of Western Australia, the Women’s and Infant’s Research Foundation at the King Edward Memorial Hospital, and
the Raine Medical Research Foundation at the University of Western Australia, along with previous grants from Healthway and NH&MRC.

**Disclosure**

The authors declared no conflict of interest.

**References**


Corresponding Author:
Paola Chivers
School of Health Sciences, Fremantle
The University of Notre Dame Australia
PO Box 1225
Fremantle Western Australia 6959
Ph: +61 8 9433 0200
Fax: +61 8 9433 0210
Email: pchivers1@nd.edu.au