Published in final edited form as:

Appl Nurs Res. 2012 February ; 25(1): 25–30. doi:10.1016/j.apnr.2010.06.001.

The validity of self-report weight and height as a surrogate method for direct measurement

Yolanda M. Powell-Young, PhD, PCNS-BC, CPN, CNE

University of Iowa, College of Nursing, 305 Nursing Building, Iowa City, IA 52242-1121; Dillard University, Division of Nursing, New Orleans, LA 70122

Abstract

Bland-Altman analysis used to determine the extent of bias, agreement, and precision between self-report and the “gold standard” of actual weight and height measurement revealed significant discrepancies between methods. Use of self-report data by health care providers and researchers should be made based on the clinical situation, patient safety and research goals.

Introduction

Among U. S. teens, 12 – 19 years of age, prevalence data show that overweight and obesity has more than tripled within the last 30 years. This is particularly significant for African American adolescent females who have a comparatively higher incidence of overweight and obesity than any ethnic-gender subgroup (Ogden, Carroll, Curtin, McDowell, Tabak, Flegal, 2006; Ogden, Carroll, Curtin, Lamb, & Flegal, 2010). There is also a greater prevalence of co-morbid illness previously unseen in this age group and overweight and obese youth are predisposed to remain overweight and obese as adults (Daniels, 2006). To better understand the epidemiology of overweight and obesity, the acquisition of accurate height and weight and calculated body mass index (BMI) is crucial for systematically establishing incidence and patterns; identifying at risk groups; monitoring outcomes; evaluating the effectiveness of preventive programs and interventions; and informing policy (Institute of Medicine, 2003).

Weight and height values, the basic elements for calculating BMI, are often collected to evaluate health, monitor growth, operationalize weight status, target treatment options for obesity-related disease processes, and calculate precise dosing for therapeutic medications. The ease and cost-effectiveness of obtaining self-report data translates to a methodology that is often used in clinical and epidemiology research. This is particularly relevant for large-scale or nationalized studies where in person data collection may not be practical (e.g., Behavioral Risk Factor Surveillance System [BRFSS], Youth Risk Behavior Surveillance System [YRBSS]). For clinical research, SR weight and height data may be used for the initial determination of study inclusion, as a co-variate in outcomes research, or as a predictor for the development of chronic disease using retrospective data from self-report surveys. Health care providers may also use self-report data as a matter of convenience or in

Yolanda M. Powell-Young, PhD, PCNS-BC, CPN, CNE, Post Doctoral Fellow, Clinical Genetics, The University of Iowa, College of Nursing, 305 Nursing Building, Iowa City, IA 52242-1121, Phone: 319-335-7062 (work); 504-289-5423 (cell), Fax: 319-353-5326, yolanda-young@uiowa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

the absence of functioning equipment. In situations where treatment is required but the adolescent is unable to respond, self-report weight and height information located on state issued identification cards, driving permits, or driver's licenses might also be used. However, the validity of using self-report weight and height data and consequently the calculated BMI values with pediatric populations in general, and African American youth specifically, has received little attention in the scientific literature (Himes, Hannan, Wall, & Neumark-Sztainer, 2004; Sherry, Jefferds, & Grummer-Strawn, 2007).

In addition to the question of dependability associated with the use of self-report data, incorrect analyses to comparatively evaluate the accuracy and clinical performance of different measurement methods against the "gold standard" have a long-standing history in biomedical research (Neiminen, Carpenter, Rucker, & Schumacher, 2006). A meta-analysis by Sherry et al (2007) concluded that comparative evaluation of results produced by studies investigating the accuracy between self-report and measured weight and height indices was made more challenging because of methodological variation, including but not limited to statistical analyses. Bland and Altman (1986, 1999, 2003) discussed the repeated use of inappropriate statistical testing, particularly correlation and regression analyses, for comparisons between measurement methods. Agreeability versus strength of relationship, unaltered correlations despite variations in scale measurement, the irrelevance of alpha, and a priori publication legitimacy were a few of the reasons cited for the improper use of these approaches.

The primary aims of this investigation were to determine the bias, level of agreement, and precision between subjective (self-report [SR]) and objective (measured [MD]) weight and height values and the BMI indices determined with these values, and to ascertain BMI misclassification frequency associated with using SR data. A secondary aim was to explore the effect of weight status on measurement differences and categorization among African American adolescent females.

Methods

Research Design

This project used data obtained as part of a larger descriptive, cross-sectional parent study that investigated the predictive capacity of body weight and self-perception on healthy lifestyle behaviors among African American female teens. As a part of a primary sub-analysis, we investigated the dependability of SR data as a surrogate for MD data. A secondary sub-analysis was conducted to compare measurement methods by weight category and describe these differences.

Sample

For the parent study, N = 310 essentially healthy African American adolescents, 14 to 18 years of age, were recruited from New Orleans, Louisiana using a combination probability/non-probability sampling strategy. Inclusion criteria included the ability to read English; enrollment in public high school; and the ability to provide assent and written consent from parent or legal guardian. Exclusion criteria included self-report history of acute or chronic conditions or use of medications that would affect weight, body composition or physical activity level (e.g., anorexia nervosa, asthma, Crohn's disease, steroids); pregnancy within the past 6 months or currently breastfeeding. Prior to the implementation of recruitment procedures, written approvals to conduct the parent study were obtained from a University Institutional Review Board (IRB) and school administrators. Informed consent was obtained from the participant's legal guardians. Assent was also obtained from the prospective participants prior to study inclusion.

Data obtained from $n = 264$ of the parent study participants who met the following criteria were eligible for inclusion in the sub-analyses for this project: complete subjective and objective height and weight values and complete demographic data. Based on a confidence interval (CI) of 95%, proportion (p) of .25, and CI_p of .45, a minimum sample size of 217 was required for statistical analyses (National Statistical Services, 2009).

Instruments & Measurements

Demographics—All participants completed a self-administered questionnaire that solicited standard demographic information including age, grade, weight to the nearest pound, and height in feet and inches. Enrollment in the National School Lunch Program was utilized as a proxy measurement for poverty. Subsequently, weight and height were measured in a private area with minimum clothing in stocking feet by a trained research assistant.

Weight, Height & Body Mass Index—Scale accuracy was assessed using a standardized weight at the beginning of each measurement day. Visual inspection of the scale for zero-balancing was also performed prior to initial use and before each participant was weighed. Weight was recorded to the nearest one-half pound using a professional dial floor scale (Health o meter®, Bedford, Heights, OH). Height was recorded to the nearest one-half inch using a portable height rod (Seca, Hanover, MD). Body Mass Index was calculated separately using SR and MD weight and height values substituted for mathematical variables within the formula, $[\text{weight (kg)}/\text{height (m)}^2]$ after converting weight in pounds to kilograms and height in inches to meters using the conversion factors of 2.2 (kg) and 0.025 (m), respectively. The appropriate CDC (2000) BMI-for-age growth chart for girls was used to plot and categorize weight as follows: underweight, $< 5^{\text{th}}$ percentile; healthy weight, $\geq 5^{\text{th}}$ percentile but $< 85^{\text{th}}$ percentile; overweight, $\geq 85^{\text{th}}$ percentile but $< 95^{\text{th}}$ percentile; and obese, $\geq 95^{\text{th}}$ percentile (CDC, 2009; Kuczmarski, et al., 2002).

Prior to the implementation of data collection procedures, the principle investigator trained research assistants on the appropriate techniques for obtaining, converting and recording weight and height according to study protocol. Written materials were provided that included illustrations on appropriate body/Frankfort horizontal plane positioning. Research assistants were also trained in correctly calculating, plotting and categorizing BMI using CDC growth charts. Practice sessions and return demonstration was used for evaluation and verification of technique and skills.

Statistical Analyses

Frequencies, central tendency measures, and measures of variability were used to evaluate normal distribution, generate descriptive summarization of the study variables, and calculate the incidence of BMI misclassification.

Bland-Altman analysis was used to compare the degree of agreement between SR and MD methods. This method allows for the systematic evaluation of comparability, as a function of investigator defined use, between two measurement techniques (Bland & Altman, 1986). First, a scatter plot of weight and height measurements including a line of perfect agreement was created to provide a visual illustration of concordance and to facilitate the identification of any systematic variance between values obtained with both measurement methods. Pair-wise correlation coefficients (r) were then computed to test the null hypothesis of no linear relationship between measurements obtained with SR and MD methods. Finally, by plotting the between-method difference and average value for each subject against the mean of the differences (bias), the relationship between the methods is illustrated. Limits of agreement (LOA) reflect the likely distribution for the differences (precision). Provided this range

would not violate a priori investigator established criteria for satisfactory agreement, the viability of measurement substitution or replacement can be established. For the purpose of this study, LOA was defined as mean bias $\pm$ 1.96 SD. Ninety-five percent confidence intervals (CI) were calculated for the bias and LOA as a function of population generalizability.

Analysis of variance (ANOVA) was conducted to evaluate the effect of weight status (i.e., underweight, healthy weight, overweight, obese) on BMI categorization accuracy. Using a table for 3-group ANOVA with $\alpha = .05$, $\eta^2 = .20$, and power of .80, group numbers (i.e., 14) for each level were appropriate. All inferential calculations required an alpha of .05 or less for statistical significance.

Results

Demographic Characteristics

Mean participant age was 15.9 years (SD = 1.2) and mean grade was 10.3 (SD = 1.1). Averages for measured weight (kg) and height (cm) were (M = 64.4, SD = 12.2) and (M = 160.5, SD = 6.8), respectively. Detailed demographics are outlined in Table 1.

Agreement between Measurements & Methods

Scatterplots for weight and height values from both measurement methods with line of perfect agreement, indicating where all points would lay if values obtained from both methods were exactly the same, are illustrated in Figure 1. Correlation coefficients were: weight, $r(264) = .948$, $p < .001$ and height, $r(264) = .976$, $p < .001$. Subsequently, the correlation coefficient generated for BMI was also statistically significant; $r(264) = .931$, $p < .001$.

Methods comparison with this cohort indicates that SR weight, in general, is underestimated by approximately -1.8 kg. Self-reporting of weight may fluctuate around this difference, by between -1.3 kg and -2.3 kg. However, the extent of agreement between the two methods (i.e., ≈ 16 kg) was wide and ranged from -9.8 kg below MD weight to an overestimation of 6.2 kg above MD weight; suggesting considerable discrepancy estimates between the two methods. Self-report weight was within ± 2 SD of MD weight for 94% ($n = 250$) of the study population; ± 3 SD for 3% ($n = 7$); and $\geq \pm 4$ SD for 3% ($n = 7$).

In contrast to weight, SR height is generally overestimated by 0.4 cm with variability of 0.2 cm to 0.5 cm above this difference. Similar to SR weight, measurement differences for SR height can span 6 cm. African American adolescent females in this study may report values -2.8 cm below to 3.6 cm above MD height suggesting that the two measurement methods did not consistently provide similar measures. Self-report height was within ± 2 SD of MD height for 93% ($n = 246$) of the population, ± 3 SD for 5% ($n = 13$) and $\geq \pm 4$ SD for 2% ($n = 5$).

Self-report values obtained from this population to calculate BMI demonstrate an underestimation difference of -1.5 kg/m². SR calculated BMI was between 1.3 to -1.7 kg/m² of the MD values. Based on the results obtained from SR weight and height, it was not surprising that the LOM for BMI are demonstrative of significant variance (i.e., 1.7 kg/m² and -4.7 kg/m²). Values outside the 95% LOA were comparable to those for weight and height. Bland-Altman plot comparisons for weight, height and BMI are illustrated in Figures 2A through 2C. A comparative summary of the bias and limits of agreement for all anthropometric measures is presented in Table 2. Thirty-six (14%) of BMI categorization misclassifications resulted from the use of SR data.

Effect of Weight Status on Measurement Differences

Analysis of Variance was conducted to evaluate the relationship between weight status expressed as healthy weight (HW), overweight (OW), and obese (OB) according to BMI using MD height and weight values and the mean differences between SR and MD weight and height reported as kg and cm. None of the study participants fell within the parameters for categorization as underweight; therefore, the independent variable (i.e., weight status) included three category levels: HW, OW, and OB. The dependent variable was the mean difference in weight (kg) and height (cm) when comparing measurement methods. Although ANOVA for SR height was not statistically significant, ANOVA was significant for SR weight, $F(2,261) = 6.02$, $p = .00$. The strength of the relationship between weight category and the magnitude of variation between SR and MD weight, as assessed by η^2 , was small, with the weight category factor accounting for 4.4% of the variance of the dependent variable.

Follow-up tests were conducted to using Tukey HSD to evaluate pair-wise differences among the means. Between the HW and OB groups, there was a significant difference in the number of kg that participants reported weighing versus what they actually weighed. There was no significant difference between the OW group with either the HW or OB group. Obese individuals showed a greater underestimation of SR weight in comparison to the HW group. This suggests a greater degree of agreement between methods with individuals who are HW or OW, but not for those who are OB. The 95% confidence intervals for the pairwise differences, as well as the means and standard deviations for the three weight category groups, are reported in Table 3.

Discussion

Healthcare providers and investigators may choose to solicit self-reports of weight and height when the “gold standard” of measured weight and height are not practically obtainable or conveniently accessible. These estimates may be used to calculate BMI for surveillance of overweight and obesity prevalence, determine drug dosage, evaluate pubertal growth, or as a determinant for study inclusion. Concerns associated with the collection and use of SR data include an established degree of agreement with MD data among pediatric populations, a lack of validity with ethnic subgroups, and the use of inappropriate statistical analyses to inform practice, scientific inquiry, and policy development (Himes, Hanna, Wall, Neumark-Sztainer, 2004; Sherry, Jefferds, Grummer-Strawn, 2007; Bland & Altman, 1986). The purpose of this study was to evaluate the extent of agreement and precision between SR and MD weight, height, and BMI indices and the impact of the agreement outcomes on the accuracy of weight categorization among African American adolescent females.

The results of this study indicate that African American adolescent females have a general tendency to underestimate weight by -1.8 kg and overestimate height by 0.4 cm. As a result, BMI calculated using estimates of SR data were underestimated by -1.5 kg/m². Obese adolescents in this study were more likely to underestimate their weight by an average of -4.1 kg than the non-obese participants.

There were both comparative similarities and differences with previous studies. Gender-specific analyses revealed a unanimous consensus among past research findings regarding the underestimation of weight by adolescent females (i.e., -0.4 kg to -3.8 kg) (Nieminen, Carpenter, Rucker, & Schumacher, 2006; Strauss, 1999; Morrissey, Whetstone, Cummings, & Owen, 2006). Consequently, calculated BMI using SR values were also underestimated (i.e., 0.46 kg/m² to 2.5 kg/m²). A literature review conducted by Sherry et al (2007) reported over- and underestimations associated with height ranging from -1.1 cm to 2.4 cm.

Generally, ethnic-specific analyses were sparse and did not differentiate by gender. Morrissey et al. (2006) reported that a sample of mixed-gender 10–16 year old African American adolescents from eastern North Carolina were more likely to overestimate height and underestimate weight and BMI. Similar to the findings reported for this study, those at the highest end of the weight spectrum had a greater underestimation of weight. Brenner et al. (2003) indicated that African American adolescents in 9th through 12th grades recruited from 20 U.S. states and the District of Columbia were, in general, more likely to overestimate height and underestimated weight. Strauss and Himes (2004) reported no significant differences by ethnic stratification.

Although SR weight and height appears to be relatively precise, the amount of potential error between the estimated and true values with this population revealed a lack of agreement that would make indiscriminant substitution between subjective and objective data impractical. Analyses variability made comparative examination, with respect to methods agreement and precision, with previous literature problematic. Sherry et al. (2007) and Bland Altman (2003) allude to this challenge. Despite the variety of statistical tests and the mixture of results reported in previous studies exploring the reliability of SR data, the general consensus was that the decision to use SR weight and height values in place of MD values should be determined based on the clinical situation and research goals.

There was also a significant variability when comparing weight classification using SR data to calculate BMI. Previous estimates for correct classification of BMI using SR data range between 81% and 98%. The use of SR weight and height resulted in the correct classification of overweight status for 86% of this population. Similar to the literature, the trend for those who are most heavy to have a greater degree of SR bias was also observed in this population.

The decision to use SR data in place of MD must be determined appropriately. The degree of agreement between methods, the quality of the data generated, and clinical feasibility are essential factors requiring both quantitative validation and qualitative judgment in the decision making process. Findings from the study present a general lack of agreement between SR and MD values for African American adolescent females in this study.

Because SR weight and height are easily obtainable, for a demographically similar cohort, findings suggest that subjective data could be appropriate for large scale weight surveillance (prevalence) studies. In behavioral assessment studies where strong associations are suspected, between risk factors obesity, the use of SR weight and height appears to be feasible. The use of SR data for determining dosing parameters, identifying authentic association, tracking longitudinal weight changes, measuring obesity related morbidity, or clinical trials surveillance is not recommended.

Limitations of this study include the omission of males, potential for limited generalizability to African American adolescent females beyond the Southern United States, and to a lesser extent; disproportional weight category. Healthy weight individuals comprised approximately 55% of the study sample. There was an under-representation of underweight or severely overweight individuals. In light of the aims of this study, the possibility of erroneous measurements, incorrect conversions and as such the reliability of objective data collected needs to be acknowledged. Because error in objective measurement is most often and significantly noted with serial and or longitudinal measurements of height and weight using multiple data collectors the impact of potential error to the conclusions generated for these analyses are felt to be negligible.

Future studies should endeavor to include a larger, more weight-diverse sample inclusive of males with stratification of the results by ethnicity and gender. Suggested areas of focus

include the impact of age and socioeconomic status on SR and MD weight, height and BMI categorization.

Conclusions

This study demonstrated that subjective estimates of height and weight by African American adolescent females may be highly inaccurate. This is a particular concern for those who are obese. Although measured height and weight are the “gold standard”, challenges may exist that limit the ability to obtain objective data. The use of self-report data by health care providers, clinicians and researchers should be made based on the clinical situation, patient safety and research goals.

Acknowledgments

Thanks to the participating youths and their parents.

Funding: This manuscript was supported by a grant from the NIH National Institute of Nursing Research (T32 NR007110, J.K. Williams, P.I.).

References

- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; i:307–310. [PubMed: 2868172]
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods Medical Residency*. 1999; 8:135–160.
- Bland JM, Altman DG. Applying the right statistics: Analysis of measurement studies. *Ultrasound of Obstetrics and Gynecology*. 2003; 22:85–93.
- Brener ND, McManus T, Galuska DA, Lowry R, Wechsler H. Reliability and validity of self-reported height and weight among high school students. *Journal of Adolescent Health*. 2003; 32:281–287. [PubMed: 12667732]
- Centers for Disease Control and Prevention. About BMI for Children and Teens. 2009. Retrieved on January 30, 2009, from http://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/about_childrens_bmi.html
- Daniels SR. The consequences of childhood overweight and obesity. *Future Child*. 2006; 16(1):47–67. [PubMed: 16532658]
- Himes JH, Hannan P, Wall M, Neumark-Sztainer D. Factors associated with errors in self-reports of stature, weight, and body mass index in Minnesota Adolescents. *Annals of Epidemiology*. 2004; 15(4):272–278. [PubMed: 15780774]
- Institute of Medicine. Unequal treatment: Confronting racial and ethnic disparities in health care. Washington, DC: National Academies Press; 2003.
- Kuczmarski, RJ.; Ogden, CL.; Guo, SS.; Grummer-Strawn, LM.; Flegal, KM.; Mei, Z., et al. 2000 CDC growth charts for the United States: Methods and development. National Center for Health Statistics. *Vital and Health Statistics*. 2002. Retrieved on January 30, 2009, from http://www.cdc.gov/nchs/data/series/sr_11/sr11_246.pdf
- Morrissey SL, Whetstone LM, Cummings DM, Owen LJ. Comparison of self-reported and measured height and weight in eighth-grade students. *Journal of School Health*. 2006; 76(10):512–515. [PubMed: 17096824]
- National Statistical Service. Sample size calculator. 2009. Retrieved on January 30, 2009, from <http://www.nss.gov.au/nss/home.NSF/pages/Sample+size+calculator?OpenDocument>
- Nieminen, P.; Carpenter, J.; Rucker, G.; Schumacher, M. The relationship between quality of research and citation frequency [Commentary]; *BMC Medical Research Methodology*. 2006. p. 1-8. Retrieved on January 30, 2009, from <http://www.biomedcentral.com/1471-2288/6/42>
- Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of Overweight and Obesity in the United States, 1999–2004. *JAMA*. 2006; 295:1549–1555. [PubMed: 16595758]

- Ogden CL, Carroll MD, Curtin LR, Lamb MM, Flegal KM. Prevalence of High Body Mass Index in US Children and Adolescents, 2007–2008. *JAMA*. 2010; 303(3):242–249. [PubMed: 20071470]
- Sherry B, Jefferds ME, Grummer-Strawn LM. Accuracy of Adolescent self-report of height and weight in assessing overweight status. *Archives of Pediatric and Adolescent Medicine*. 2007; 161(12):1154–1161.
- Strauss RS. Comparison of measured and self-reported weight and height in a cross-sectional sample of young Adolescents. *International Journal of Obesity*. 1999; 23:904–908. [PubMed: 10490794]

Figure 1. Scatter diagram with the line of equality for weight and height indices

XY scatterplot comparing self-report weight and height with actual weight and height measurements for all ($N = 264$) participants. Plot provides a visual demonstration of reproducibility between measurement methods. Identify line represents perfect agreement for data.

Figure 2. Bland-Altman plots for weight, height and BMI

Figures 2A–2C. Bland-Altman diagrams for comparison of self-report and actual weight, height and BMI values for all ($N = 264$) participants. Plotted is the difference between the two methods against their means (bias) with limits of agreement ± 1.96 standard deviations (precision).

Table 1

Population Demographics (N = 264)

Variable	Frequency (n)	Mean	Standard Deviation	Percent (%)
Age (Years)		15.9	1.2	
14	33			12.5
15	78			29.5
16	62			23.5
17	60			22.7
18	31			11.7
Grade		10.3	1.1	
9	77			29.2
10	65			24.6
11	74			28.0
12	48			18.2
Socioeconomic Status				
Poverty	251			95.1
Above Poverty	13			4.9
Weight (kg)		64.4	12.2	
Height (cm)		160.5	6.8	
BMI		26.1	5.1	
Normal weight	145			54.9
Overweight	67			25.4
Obese	52			19.7

Table 2

Comparison of weight, height and BMI measurements (N=264)

Measurement	Bias (SD)	95% CI _{bias}		Limits of Agreement		95% CI _{LOA}		% of values outside 95% LOA
		(Upper – Lower)		(Upper – Lower)		(Upper)	(Lower)	
Weight (kg)	-1.8 (4.0)	-1.3, -2.3		6.2, -9.8		7.0, 5.3	-9.0, -10.6	6%
Height (cm)	0.4 (1.6)	0.2, 0.5		3.6, -2.8		3.8, 3.2	-2.4, -3.1	7%
BMI (kg/m ²)	-1.5 (1.6)	1.3, -1.7		1.7, -4.7		2.0, 1.4	-4.4, -5.0	7%

Bias calculated as mean difference (i.e., self-report – actual measurements); LOA = limits of agreement; and CI = confidence intervals,

95% Confidence Intervals of Pair-wise Differences in Mean Changes in Accuracy of Self-Report Weight by Weight Status

Table 3

Weight category	Weight Difference (kg)				Healthy weight	Overweight
	N	M	SD			
Healthy weight	145	-1.4	2.3			
Overweight	67	-2.3	6.5		-.76 to 2.6	
Obese	52	-4.1	7.7		.82 to 4.7*	-.59 to 4.2

An asterisk indicates that the 95% confidence interval does not contain zero, and therefore the difference in mean is significant at the .05 significance using Tukey HSD