Journal of
Obesity & Weight Loss Therapy

The BMI Bias: The Validity Problems of Using Self-Reported Data to Measure Body Mass Index

Karen J. Buhr*
Assistant Professor of Health Policy & Administration, Penn State Mont Alto, PA, USA

Abstract

Obesity has been increasing at an alarming rate over the past several years, affecting countries of every development status. The burden of obesity is largely carried by those of low income and this is a shift from previous periods in time. In order to adequately design effective health policies, there is a need for accurate measures of the health of the related populations. Having an accurate measure of the bodyweight status of entire populations is an integral part of this process, particularly so that the related health care providers can determine the most effective way for treating diseases related to body weight problems.

Keywords: Body mass index; BMI bias; Self-reported BMI

Introduction

Obesity has been increasing at an alarming rate over the past several years, affecting countries of every development status. The burden of obesity is largely carried by those of low income and this is a shift from previous periods in time. This major increase in obesity is hard to track, as the dietary and activity habits of people are changing at an accelerated speed [1]. In order to adequately design effective health policies, there is a need for accurate measures of the health of the related populations. Having an accurate measure of the bodyweight status of entire populations is an integral part of this process, particularly so that the related health care providers can determine the most effective way for treating diseases related to body weight problems.

Body mass index (BMI) uses individual's height and weight to categorize them as being underweight, of healthy weight, or overweight. This index is often used to measure large populations to determine if the given population is at a health risk, related strictly to their overweight and underweight status [2]. The BMI index is largely accepted as the universal measure for estimating adiposity in adults. Many authors would agree that use of the body mass index is a valid measure for the generalization of the adiposity of large adult populations [3]. However, many authors question whether self-reported data, the data most frequently used, accurately report the body weight status of adult and child populations due to potential for over and under reporting height or weight values [4-6]. While the validity of the body mass index in providing accurate data is generally considered to be accurate for adults, there are some studies that question how accurate the index is for measuring obesity in children [3].

The purpose of this study is to use econometric analysis to determine if self-reported data gives biased estimates of the body mass index. More specifically, the goal is to look at measured data from overweight and underweight males and females to determine if they are more likely to under or over report their weight based on their actual weight. The same analysis will also be done on height as well. The purpose of this analysis is to more accurately examine the expected bias in self-reported data so that future researchers and policy makers can correct for these errors in a revised models.

Literature Review

There is a wide range of literature that looks at self-reported height and weight bias among varied age groups and demographics. Elgar et al.[4], Brener et al. [7], Mendelson et al. [5], and Dietz and Bellizzi [3] all examine self-reporting among children and young adults, while

Rothman [8], Spencer et al. [9], Flood et al. [10], and Plankey et al. [6] all focused on self-reported data from adults. Gillum and Sempos [11] focus BMI bias from ethnic variation. The general conclusion gathered from the studies that focus on children, find that youth underreport body weight. However, none of these studies break down the children into their actual weight status and then determine if under reporting is universal across all weight statuses. Brener et al. [4] found that, on average, people over reported their height, while Mendelson et al. [5] cited self-esteem as a likely reason for such findings. This same study also found that females were more likely to underreport their weight than males.

Studies on adults came to similar conclusions. In general, it was also found that adults were more likely to underreport their weight. Spencer et al. [9] found that heavier men and women were more likely to underreport their smaller counter parts. This same study found that height was over estimated, just as was found with the children. Flood et al. [10] found the measured estimate of overweight or obesity to be almost doubles self-reported estimates.

Theoretical model

The main purpose of this study is to measure the deviation between self-reported BMI and actual BMI based on certain attributes such as gender, age, weight, and height. In addition to measuring this bias, we also wish to see if there is a correlation between BMI bias and self-reported error among certain attributes. To determine the BMI self-reported error (which will be referred to as BMI Error from here on) the following formula will be used:

BMI Error = Actual BMI - Self-Reported BMI (1)

$$BMI\ Error = \left(\frac{703 \times Actual\ Weight\ (lb)}{(Actual\ Height\ (in))^2}\right) - \left(\frac{703 \times Self\text{-}Reported\ Weight\ (lb)}{(Self\text{-}Reported\ Height\ (in))^2}\right)(2)$$

This formula will give us a measure of the BMI Error. This is the first step in comparing how much an individual's weight and height

---

Corresponding author: Karen J. Buhr, Assistant Professor of Health Policy & Administration, Penn State Mont Alto, 1 Campus Drive Mont Alto, PA, USA, Tel: +1- 717-749-6047; E-mail: kjb44@psu.edu

actually deviate from the self-reported values and it will allow us to see if there are correlations between these deviations, and gender and age, and BMI Error. To determine that, the following theoretical model will be used for both genders:

$$BMI\ Error = \beta_1(Age) + \beta_2(Self\text{-}Reported\ Weight) + \beta_3(Self\text{-}Reported\ Height)$$
$$+ \beta_4(Consider\ Yourself\ Overweight)$$
$$+ \beta_5(Consider\ Yourself\ Underweight)$$
$$+ \beta_6(Consider\ Your\ Weight\ normal) + B_7(Want\ to\ Weight\ More)$$
$$+ B_8(Want\ to\ Weight\ Less) + B_9(Want\ to\ Weight\ Same) + B_{10}(Short)$$
$$+ B_{11}(Average) + B_{12}(Tall) + \beta_{13}(Underweight) + \beta_{14}(Normal\ weight)$$
$$+ \beta_{15}(Overweight) + \beta_{16}(Obese) + \beta_{17}Constant + e$$

Where:

Short=shorter than average height by more than one standard deviation; Averag =within one standard deviation of the average height; Tall=taller than average height by more than one standard deviation; Underweight is someone whose BMI is below 18.5; Normal is someone whose BMI is between 18.5 and 24.9; Overweight is someone whose BMI is between 25 and 29.9; Obese is someone whose BMI is above 30

For this particular study the respondents included in the analysis were between the ages of 20 and 79 years of age. A different method is used to calculate BMI for those under 20 and thus all those individuals were excluded from the analysis. The survey used grouped all those aged 80 and older as just 80 years old so we excluded these observations from our analysis as well. Our hypothesis about age is that younger people are more concerned about their weight and thus this might affect the BMI error in two ways. First, younger people might be more self-conscious about their weight and therefore they might be more likely to lie about their weight when self- reporting it so that their weight meets their perceived social norms. Secondly, younger people may be more conscious about their weight and in turn be more aware of their actual weight; therefore they may be more likely to report their weight accurately. For these reasons we include age as a descriptive variable.

The BMI error is calculated by determining the difference between actual BMI and self-reported BMI. By including self-reported weight and height as explanatory variables there may appear to be a problem of circular referencing and thus some clarification may be needed. BMI error is strictly how much self-reported numbers deviate from the actual numbers and thus is a measure of deviation. Our hypothesis people who report their weight higher than

Six dummy explanatory variables were used to describe how people felt about their weight. Our hypothesis is that those who feel that they either, weigh above the normal weight, or consider their weight to be heavy are more likely to underreport their weight, and vice versa for those who feel the opposite. If people feel their weight is not acceptable they may feel embarrassed or ashamed to share their actual weight and thus may report their weight closer to the acceptable level.

Seven other dummy explanatory variables were used to measure how BMI error varies across actual heights and weights. The first three dummy variables were set up so that if someone actual height was within one standard deviation from the gender mean height they were considered average height, if someone's actual height was more than one standard deviation from the gender mean height they were considered tall, and if someone's actual height was less than one standard deviation from the gender mean height they were considered short. With a normal distribution approximately 68% of the population will fall within on standard deviation from the mean, thus using this measure to gage if someone is tall or short seemed fitting as it accounts for the majority of the population to fall within the average height category. One of the goals of this analysis was to determine if people who were shorter or taller had any difference in BMI error. Along the same guidelines, people with an actual BMI that fell into the Centre for Disease Control's range for being underweight, normal, overweight or obese were put into one of the four dummy variables. This allows us to determine if someone's weight status impacts the way they reported their weight and height.

The fore mentioned model will allow for a thorough analysis of self-reported BMI bias. This model will provide answers to the following questions:

- Do taller people have a tendency to over or under report their BMI?

- Do overweight people tend to over or under report their BMI?

- Do underweight people tend to over or under report their BMI?

- Since men and women are split into separate equations we are able to determine significant differences between males and females in BMI self-reported error.

- On average, are men over reporting their height more frequently than women?

- On average, are women underreporting their weight more frequently than men?

In order to fill in some of the missing gaps in current research, and thus have a more accurate picture of the growing global obesity problem, a thorough analysis of the possible causes of BMI estimation error is required. Previous studies have not described BMI error to the extent that this study is does. By elucidating the problem in greater detail, one is able to better understand the problem and thus create better policies to fix the problem. Using the above estimation of BMI error, the following formula will be used to calculate BMI :

$$BMI = Self\text{-}reported\ BMI + BMI\ Error$$

Methods

Data was collected from the NHANES II survey. This is a cross-sectional survey of non-institutionalized civilian residences of the United States. The survey was conducted by the National Center for Health Statistics during the time period between 1976 and 1980. The survey used a stratified, multistage, clustered, probability sampling method. Our sample includes 51623 people between the ages of 6 months and 74 years old. Participants were given an interview portion conducted at their homes and later given a physical examination.

After all of the necessary data was extrapolated from the database, the 51,623 observations were sorted and any observations that lacked a response for one or more of the selected variables were removed. Then, the observations were sorted again and any observations that had the answer "Refused" or "Don't know" were removed. Once the non-response data were removed 21,060 observations were leftover. Of the 21,060 usable observations, 10,855 of the observations were female and 10,205 of the observations were male (Respectively, there were 26,263 and 25,360 female and male observations before "cleaning"). Given that we lost 59.76% of all of our female observations, and 58.67% of all our male observations, we can assume there was little to no non-response bias between genders, given a 1% difference[1]. Once the data had been cleaned, the total 21,060 observations were separated into two groups, males and females. The necessary missing variables were calculated using excel. Regression analysis was conducted on the data using ordinarily least squared (OLS).

---

1    $\dfrac{26263 - 10855}{26263} * 100 = 59.76, \dfrac{25360 - 10205}{25360} * 100 = 58.67$

Results

A standard OLS regression was used to model the data to explain the BMI error produced from self-reporting BMI. The results from the male data were enough different from the results from the female data to suggest that separate methods should be used when calculating BMI from self-reported data. At 95% confidence level, males' age had a statistical significance on BMI error, yet females' age had no statistical significance. Women's opinion on how they considered their weight and their desire to lose weight were all statistically significant at a 95% confidence level, yet, men's opinions and aspirations on weight and weight loss had no effect on BMI error at this same confidence level (Tables 1,2).

While males and females have some factors that differ from the opposite sex that will lead to biased BMI results when using self-reported data, there are also some similarities. In both the women and the men higher self-reported height leads to more BMI error, while higher self-reported weight leads to lower BMI error. Both variables are statistically significant. Compared to shorter people average height and taller people reduce BMI error. We are 99% confident that those who fall within the spectrum for being underweight and normal weight (based on actual BMI) are going to have a negative effect on BMI error. These people will accurately report their BMI better than those who are Obese. In both the men's and women's cases, those who are overweight (again based on actual BMI) have no statistical effect on BMI error. Those who are obese (based on actual BMI) have a positive effect on BMI error (Tables 1,2).

The demographics between the two sexes are fairly similar. Men have a slightly higher proportion of overweight individuals while females have a slighter higher percentage over obese individuals (Figure 1). Of those who are underweight based on BMI, 27% of females over reported their weight compared to only 8% of males (Figure 2). 52% of

the normal BMI females over reported their weight, while only 27% of men in the same group over reported their weight (Figure 3). There was not as large of a discrepancy in the overweight and obese groups (See Chart 7-10). 65% of overweight females over reported their weight while only 46% of overweight males over reported their weight (Figure 4). 70% of obese females over reported their weight, and 60% obese males over reported their weight (Figure 5) (See Chart 9 and 10). On average, at any particular height, women will have more error in their reported BMI than men; although, at large heights men and women have similar BMI Error (Figure 6). At any given weight, on average, women's BMI error will be about 1.5 index points higher than men at that some weight (Figure 7).

Looking at the correlation between actual BMI and BMI one can see that people who have lower BMIs have lower BMI error. On average, at a given BMI men under report their BMI more than women although this trend levels off at higher BMIs. On average, men and women who are underweight under report their BMI between -2 and 0 index points. Normal weight men and women on average report their BMI close to accurate, although men may slightly under report their BMI. Men and women who are overweight, on average, over report their BMI between 0 and 2 index points, with women over reporting their BMI more than men. Obese men and women over report their BMI, on average, approximately 2-4 index points(Figure 8).

Breaking down men and women into those who over estimated and under estimated their BMI we can easily analyse the general differences between genders. On average, females who over estimated their BMI or under estimated their BMI were only a few index points off from the correct value. On the other hand, on average, men who over estimated their BMI did so by a proximately 5 index points, a significantly larger number than women from the same group (1.5-2 index points). Interestingly, women, who over estimated their BMI, on average, overestimated their weight by almost 9 pounds, yet men from the same

Males					
Source	**SS**	**df**	**MS**	**Number of Obs**	**10855**
Model	19827.441	14	1416.246	Prob>F	0.000
Residual	42493.260	10840	3.920	R-Squared	0.318
				Adj R-Squared	0.317
Total	62320.701	10854	5.74173	Root MSE	1.980
BMI Error	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% Conf. Interval]**
Age***	-0.011	0.0012	-9.37	0.000	-0.013	-0.009
Self-Reported Height***	0.572	0.0102	55.94	0.000	0.551	0.592
Self-Reported Weight***	-0.029	0.0008	-34.58	0.000	-0.030	-0.027
Consider Self Overweight	-0.137	0.1598	-0.86	0.391	-0.450	0.176
Consider Self Underweight	(dropped)					
Consider Self About Right Weight	0.142	0.1498	0.95	0.345	-0.152	0.435
Want to Weight More*	-2.037	1.1517	-1.77	0.077	-4.295	0.220
Want to Weigh Less	-1.870	1.1451	-1.63	0.103	-4.114	0.375
Want to Weigh About the Same	-1.858	1.1443	-1.62	0.104	-4.101	0.385
Short***	(dropped)					
Average Height***	-0.820	0.0330	-24.88	0.000	-0.884	-0.755
Tall***	-1.074	0.0354	-30.36	0.000	-1.143	-1.005
Underweight (Actual BMI Based)***	-2.959	0.5020	-5.89	0.000	-3.943	-1.975
Normal (Actual BMI Based)***	-1.599	0.4820	-3.32	0.001	-2.544	-0.654
Overweight (Actual BMI Based)	-0.030	0.4820	-0.06	0.950	-0.975	0.915
Obese (Actual BMI Based)***	2.118	0.4840	4.38	0.000	1.169	3.067
Constant***	-27.074	1.3912	-19.46	0.000	-29.801	-24.347
***Statistically Significant at 1%, **Statistically Signifcant at 5%, *Statistically Signifcant at 10%						

Table 1: Regression results for the male sample.

Females					
Source	**SS**	**df**	**MS**	**Number of Obs**	**10205**
Model	10565.385	14	754.670	Prob>F	0.000
Residual	27203.046	10190	2.670	R-Squared	0.280
				Adj R-Squared	0.279
Total	37768.432	10204	3.70134	Root MSE	1.634
BMI Error	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% Conf. Interval]**
Age*	0.002	0.001	1.91	0.056	0.000	0.004
Self-Reported Height***	0.419	0.0084	50.12	0.000	0.403	0.436
Self-Reported Weight***	-0.019	0.0008	-24.7	0.000	-0.021	-0.018
Consider Self Overweight	-0.237	1.6429	-0.14	0.885	-3.458	2.983
Consider Self Underweight	-0.336	1.6408	-0.21	0.838	-3.553	2.880
Consider Self About Right Weight	-0.290	1.642	-0.18	0.860	-3.509	2.929
Want to Weight More	(dropped)					
Want to Weigh Less***	0.338	0.0905	3.73	0.000	0.160	0.51509
Want to Weigh About the Same***	0.268	0.0769	3.49	0.000	0.11772	0.41924
Short***	(dropped)					
Average Height***	-0.717	0.0292	-24.56	0.000	-0.774	-0.660
Tall***	-0.946	0.0315	-30.07	0.000	-1.008	-0.885
Underweight (Actual BMI Based)***	-2.291	0.3629	-6.31	0.000	-3.003	-1.580
Normal (Actual BMI Based)***	-0.970	0.3294	-2.94	0.003	-1.616	-0.324
Overweight (Actual BMI Based)	0.131	0.3279	0.4	0.690	-0.512	0.773
Obese (Actual BMI Based)***	1.629	0.3303	4.93	0.000	0.982	2.276
Constant***	-24.004	1.746	-13.75	0.000	-27.426	-20.581
***Statistically Significant at 1%, **Statistically Signifcant at 5%, *Statistically Signifcant at 10%						

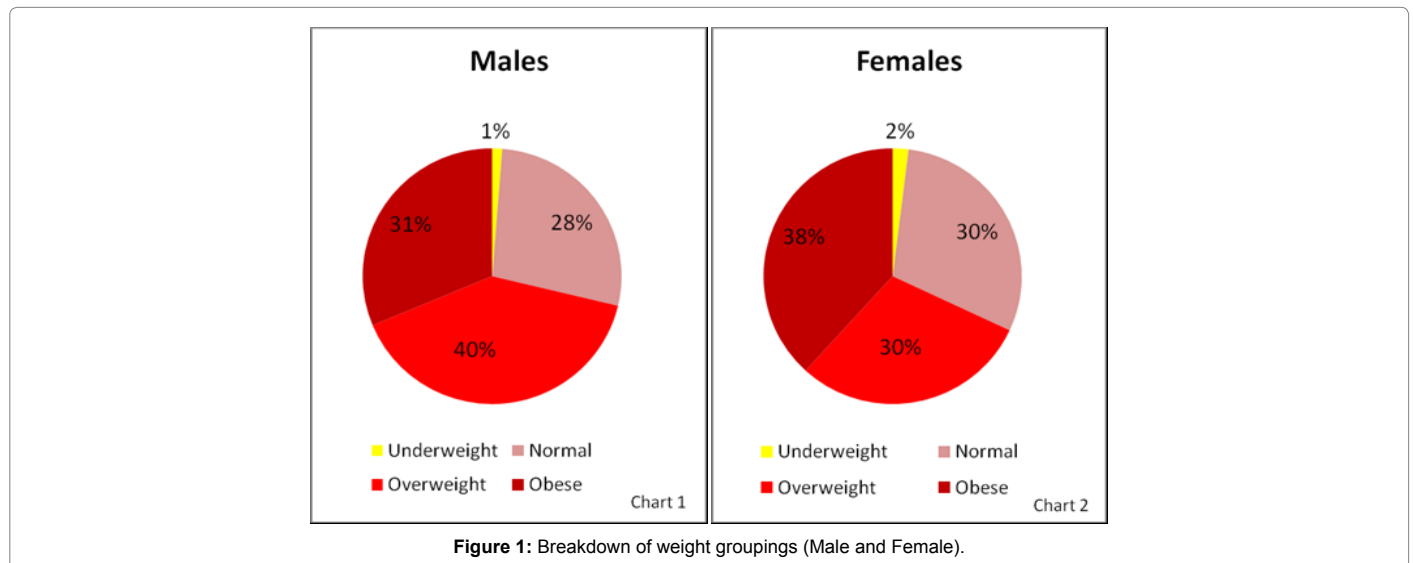**Table 2:** Regression results for the female sample.

Figure 1: Breakdown of weight groupings (Male and Female).

group under estimated their weight by about 6.5 pounds. Women, who under estimated their BMI, on average, also under reported their weight (by around 4 pounds), yet men from the same group did the opposite and over reported their weight by around 7 pounds (Figure 9).

Discussion

To begin our discussion, it only seems fitting to answer the original questions posed in the methods section and expand from there.

- Do taller people have a tendency to over or under report their BMI?

Our regression analysis and a linear estimation between height and BMI error provide somewhat contradictory results. The regression analysis suggests that as a person gets taller their BMI error is reduced. In the linear equation, shown in graph one, BMI error is fairly constant across all heights. Society does seem to project being tall as a desired trait. Thus it would be consistent that shorter people are more likely to lie and say they are taller than they actually are, at the same time if you are tall and thus have the desired trait you need not lie and thus your self-reported results will be similar to your measured results.

- Do overweight people tend to over or under report their BMI?

Our regression analysis, charts 7-10, and graph 3 all suggest that overweight people are more likely to underreport their BMI.

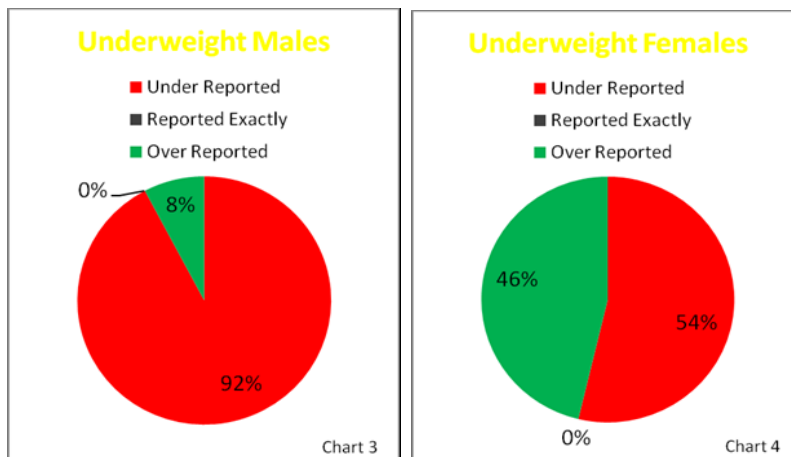- Do underweight people tend to over or under report their BMI?

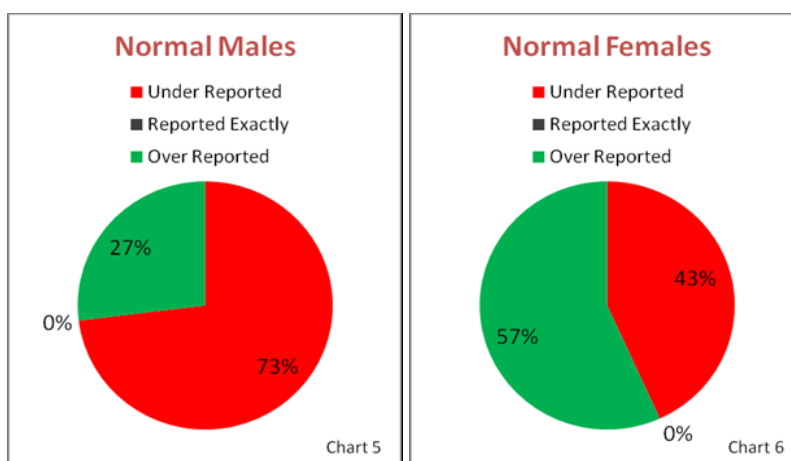**Figure 2:** Breakdown of under or over reporting of weight for underweight individuals (male and female).

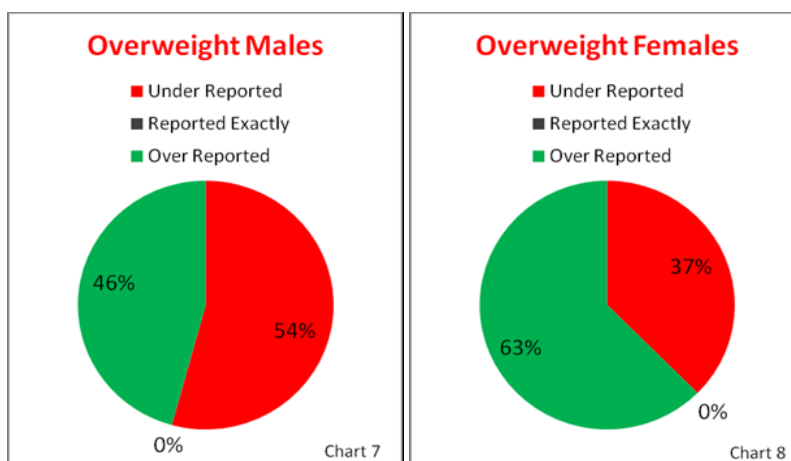**Figure 3:** Breakdown of under or over reporting of weight for normal weight individuals (male and female).

Figure 4: Breakdown of under or over reporting of weight for normal overweight individuals (male and female).
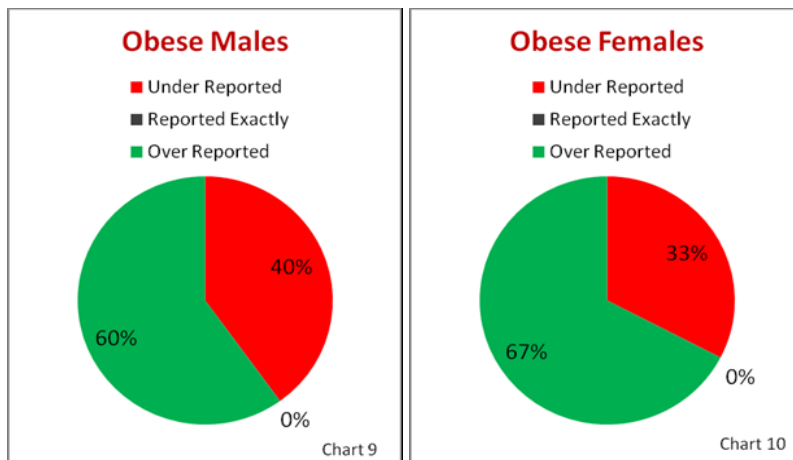
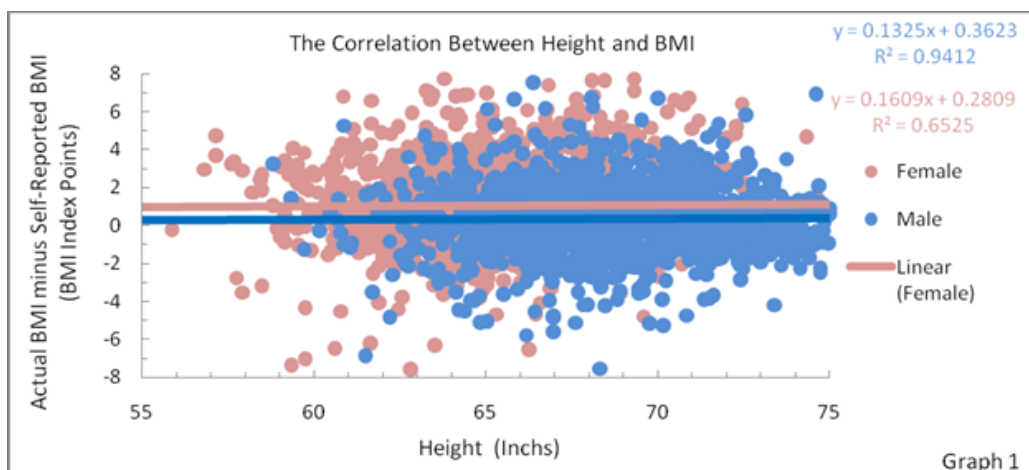**Figure 5:** Breakdown of under or over reporting of weight for obese individuals (male and female).
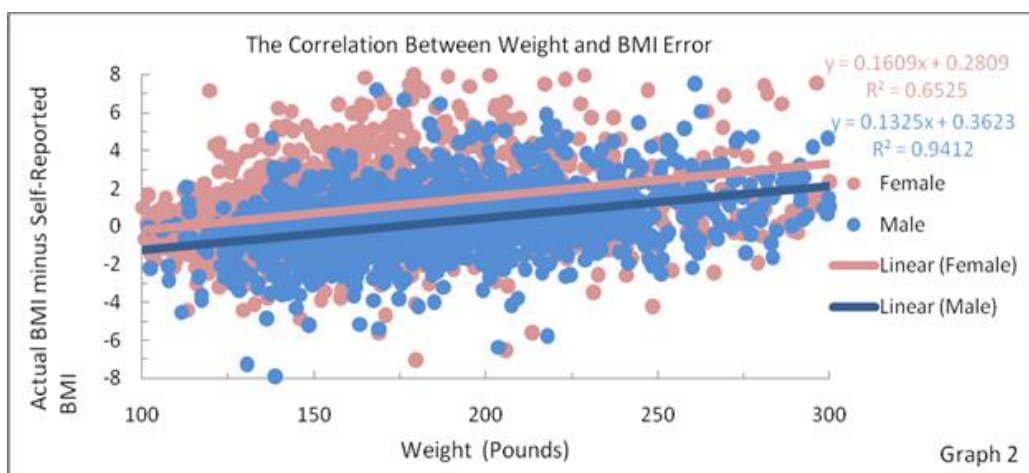

Figure 6: The Correlation between height and BMI.

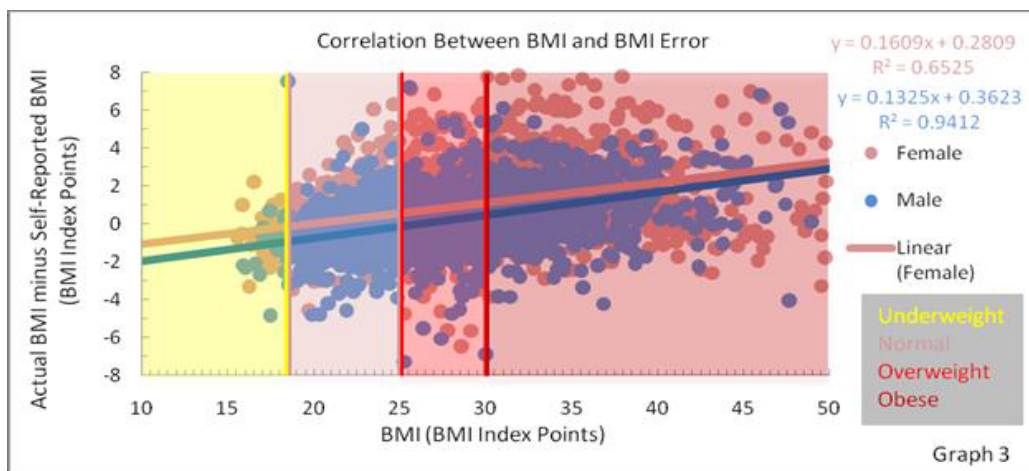**Figure 7:** The correlation between weight and BMI error.

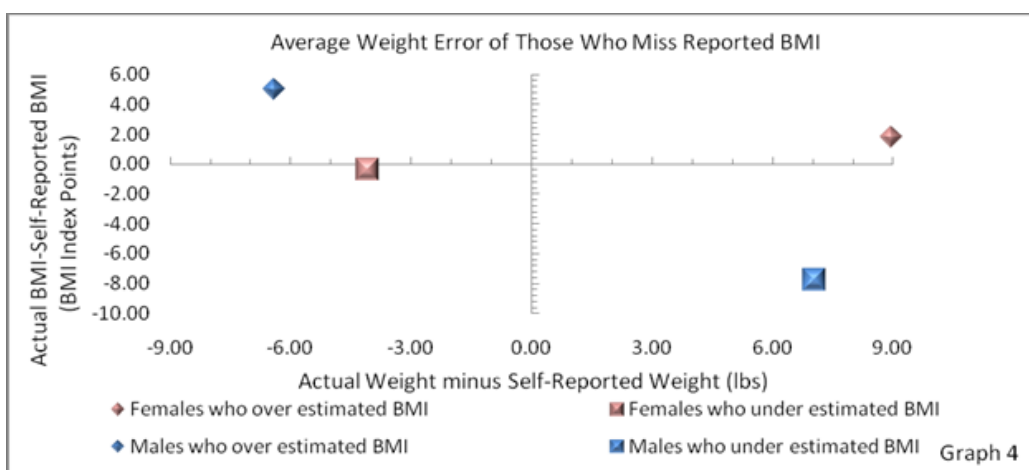**Figure 8:** The correlation between BMI and BMI error.

Figure 9: Average weight error of those who miss reported BMI.

Our regression analysis, charts 3 and 4, and graph 3 all suggest that underweight people underreport their BMI

- On average, are men over reporting their height more frequently than women?

94.4% of females under reported their weight while 5.5% over report their height. 29.2% of males under reported their weight while 70.6% of males over reported their height. Men are significantly more likely to report their height higher than women. This factor must be included when trying to estimate the BMI from self-reported data.

- On average, are women underreporting their weight more frequently than men?

Contrary to what some might expect, in each weight status group a higher percentage males underreported their weight than women (See Charts 3-10).

Prior research done by other academics concluded that people over report their weight (Brener et. al. 2003)[4]. Our findings found this to be true when peoples' BMI categorized them as obese, overweight, or normal. However, underweight people under reported their weight.

Even more importantly our regression analysis shows that in order to correct self-reported data different adjustments need to be made based on a person's weight status. In addition, to accurately adjust for self-reported data, a person's height must also be factored into adjustments. If they are taller or shorter than the average person this will factor into how much their self-reported data varies from the actual amount. Contrary to the study done by Mendelson et al. [5], it appears that peoples' feelings about their weight have no statistical significance on BMI error.

Conclusion

Using self-reported BMI data has advantages and disadvantages over using BMI measured data. It requires far few resources to send out a survey that asks people their weight and height than it does to send out people to measure each person's weight and height. However, self-reported data lacks the accuracy of measured data. By determining a model that estimates how much self-reported data deviates from the actual values, massive resources can be saved while retaining the accuracy. This study provides a model that deviates from some of the more common models and the assumptions those models make. To estimate BMI error, weight status, height, and age must be taken into consideration.

Refrerences

1. Popkin BM, Gordon-Larsen P (2004) The nutrition transition: Worldwide obesity dynamics and their determinats. Int J Obes 28: S2-S9.

2. Kuczmarski RJ, Carroll MD, Flegal KM, Troiano RP (1997) Varying body mass index cutoff points to describe overweight prevalence among U.S. adults: NHANES III (1988 to 1994). Obes Res 5: 542-548.

3. Dietz WH, Bellizzi MC (1999) Introduction: The use of the body mass index to assess obesity in children. Am J Clin Nutr 70S: 123S-125S.

4. Brener ND, Mcmanus T, Galuska DA, Lowry R, Wechsler H (2003) Reliablity and validity of self-reported height and weight among high school students. Adolesc Health 32: 281-287.

5. Mendelson MJ, Mendelson BK, Andrews J (2000) Self-esteem, body esteem, and body mass in late adolescence: is a competence X importance model needed? Journal of Applied Developmental Psychology 21: 249-266.

6. Plankey MW, Stevens J, Flegal KM, Rust PF (1997) Prediction equations do not eliminate systematic error in self-reported body mass index. Obes Res 1997;5:308-314.

7. Elgar FJ, Roberts C, Tudor-Smith C, Moore L (2005) Validity of self-reported height and weight and predictors of bias in adolescents. J Adolesc Health 37: 371-375.

8. Rothman KJ (2008) BMI-related errors in the measurement of obesity. Int J Obes 32: S56-S59.

9. Spencer EA, Appleby PN, Davey GK, Key TJ (2002) Validity of self-reported height and weight in 4808 EPIC-Oxford participants. Public Health Nutr 5: 561-565.

10. Flood VM, Webb K, Lazarus R, Pang G (2000) Use of self-reported to monitor overweight and obesity in populations: some issues for consideration. Aust N Z J Public Health 2000; 24: 96-99.

11. Gillum RF, Sempos CT (2005) Ethinc variation invalidity of classification of overweight and obesity using self-reported weight and height in American women and men: The Third National health and Nutrition Examination Survey. Nutr J 4:27.