Looks that Kill: Predicting Adult Morbidity and Mortality from Adolescent Facial Characteristics in Yearbook Photographs

PAA Paper Submission
September 22, 2005

by

Eric N. Reither, Utah State University
Robert M. Hauser, University of Wisconsin-Madison
Sheri Meland, University of Wisconsin-Madison

4.1 Introduction

Analyses of NHIS data in the previous chapters revealed that birth cohort membership is an independent determinant of obesity. However, these analyses also clearly demonstrated that period effects are the predominant cause of the obesity epidemic in the United States. What social and economic changes in the U.S. have precipitated these strong period effects? This question has already received much attention from epidemiologists and public health officials, and it is beginning to draw considerable interest from economists and sociologists.

The Wisconsin Longitudinal Study (WLS) offers a promising source of information for social scientists who are interested in addressing this question. The WLS is a large cohort study of over 10,000 individuals who graduated from a Wisconsin high school in 1957 (Hauser et al. 1994). It has followed this cohort from 1957 into the new millennium (Sewell et al. 2004) and boasts a wide variety of socioeconomic, demographic and psychological indicators that social scientists could draw upon to evaluate competing theoretical explanations for the obesity epidemic. In addition, because all WLS subjects were born around 1939, neither age nor birth cohort can explain weight differences in this sample. Therefore, the WLS is ideally suited for the investigation of period effects. Taken together, these advantages suggest that the WLS is a resource rich with potential for improving our understanding of how secular changes have differentially affected patterns of weight gain and weight loss among Americans who are now entering retirement age.

Unfortunately, while the WLS collected data on height and weight in both the 1992-93 and 2003-05 waves, it did not collect data on either height or weight in the initial 1957 wave or in a subsequent wave in 1975. Of course, given the strengths of the WLS and the availability of two waves of data on height and weight that are chronologically situated within the obesity epidemic, it could be used to evaluate the etiology of weight change during the period 1993 to 2005. Nevertheless, the absence of information on the body mass of WLS respondents at baseline currently limits the utility of WLS data for assessing questions regarding the causes of the obesity epidemic.

In a recent study of facial attractiveness (Meland 2002), approximately 3,000 photographs were extracted from the senior high school yearbooks of WLS respondents. Extracted yearbook photographs provide visual information on the physical characteristics of the faces and necks of WLS respondents. Might these photographs have some utility in establishing the relative weight or "body mass" of WLS respondents at baseline? A review of the scientific literature on the associations between adiposity, BMI and physical characteristics of the face and neck suggests that the answer to this question is "Yes."[1]

Adiposity, Body Mass and Facial Characteristics

A number of clinical studies have investigated the relationships between adiposity, body mass and facial characteristics in samples of human subjects. These studies have demonstrated that deposits of adipose tissue in the cheeks and neck, neck circumference, and craniofacial morphology are all related to body mass and central adiposity. For instance, Levine, Ray and Jensen (1998) used computer tomography to measure cheek, visceral abdominal (i.e., intra-

[1] Unless context dictates otherwise, the term "facial characteristics" will refer to physical characteristics of the face and neck from this point forward.

abdominal) and abdominal subcutaneous (i.e., surface abdominal) fat in 25 patients who were being treated at the Mayo clinic for various conditions. Despite the small sample of subjects, Levine et al. demonstrated that the amount of cheek fat was significantly associated with deposits of visceral abdominal fat (r = 0.54, $p < 0.01$). Cheek fat was not significantly associated with either abdominal subcutaneous fat (r = 0.22, $p = 0.29$) or BMI (r = 0.39, $p = 0.07$), but this was likely the result of insufficient statistical power since these correlations were relatively strong. This was particularly true for the correlation between cheek fat and BMI, which approached statistical significance despite the small sample of subjects.

In another study, Laakso, Matilainen and Keinänen-Kiukaanniemi (2002) evaluated associations between neck circumference, abdominal obesity, general obesity, and several variables related to insulin resistance. Direct anthropometric measures—including height, weight, and circumferences of the neck, waist and hip—and indicators of insulin resistance were gathered from 541 participants in a population-based study of northern Fins born in 1935. Results of this study clearly demonstrated that neck circumference was strongly correlated with anthropometric measures of obesity. Among men, neck circumference was significantly associated with waist-to-hip ratio (r = 0.41, $p < 0.01$), waist circumference (r = 0.65, $p < 0.01$) and BMI (r = 0.69, $p < 0.01$). These correlations were nearly identical among women ($p < 0.01$ in each case). Importantly, these results suggest that neck circumference explains about half of the variation in BMI.

Like neck circumference, deposits of subcutaneous adipose tissue (SAT) in the neck are strongly related to central and upper body adiposity. To study the relationships between SAT deposits in 15 different body locations, Möller et al. (2000) utilized the LIPOMETER to assess

SAT-topography in 590 healthy adult subjects. Factor analyses of these data indicated that SAT deposits in the neck, upper back, front chest, lateral chest, upper abdomen, lower abdomen and hip were highly interrelated. Also, the neck was the *only* body site that did not exhibit significant gender differences. That is, women had significantly higher average SAT values than men at all body sites except the neck. This suggests that measures of neck adiposity may hold promise as gender neutral indicators of central obesity in the general population.

A common expression is to say that a person is heavy because he or she is “big boned.” Studies of the relationship between body mass and bone density support this notion. Bone mineral density is strongly (and positively) correlated with lean mass, weight and BMI (Barondess, Nelson and Schlaen 1997; Langendonck et al. 2002). For instance, Barondess et al. found that bone mineral density explained between 25 and 37 percent of the variation in BMI in ethnically diverse samples of males. Of particular importance for the purposes of this investigation, studies of craniofacial morphology have demonstrated that bone structures in the face are also associated with body mass in human subjects (Yu et al. 2003).

Sadeghianrizi (2003) and Örhn et al. (2002) compared cephalometric measurements of obese adolescents (similar in age to WLS respondents in 1957) to age and sex matched control subjects who were not obese. These studies revealed that the facial skeletal structures of obese adolescents are typically somewhat larger than their normal weight peers. Most notably, each study found that the mandibles of obese adolescents tend to be relatively long. For instance, Sadeghianrizi (2003) found that obese boys and girls had mandibles that averaged 10 and 8 millimeters longer, respectively, than their normal weight peers. Studies of adolescents have proven consistent with studies of adults, which have reported similar associations between facial bone structures and body mass (e.g., Liao et al. 2004; Paoli et al. 2001). The etiology of craniofacial differences between heavy and normal weight individuals may stem from hormonal factors, nutritional factors, or both (Örhn et al. 2002; Sadeghianrizi 2003).

Health Outcomes Associated with Adiposity of the Face and Neck

As noted in Chapter 1, elevated BMI is associated with a host of deleterious health outcomes, including hypertension, increased low-density lipoprotein (i.e., bad cholesterol), decreased high-density lipoprotein (i.e., good cholesterol), glucose intolerance, insulin sensitivity, ischemic stroke, osteoarthritis, coronary heart disease, type-II diabetes and some cancers (Manson, Skerrett and Willett 2002). In addition, research has indicated that the distribution of adipose tissue affects the probability of disease incidence. For instance, research has shown that central adiposity is an important determinant of conditions such as hypertension, diabetes and heart disease (Rexrode et al. 1998).

Although not widely recognized, scientific studies have also shown that facial characteristics are strongly associated with health complications such as type-II diabetes (Sierra-Johnson and Johnson 2004), sleep apnea (Mortimore et al. 1998) and weight cycling (Wallner et al. 2004). To illustrate, in the aforementioned epidemiological study of northern Fins, Laakso et al. (2002) found that neck circumference was an independent predictor of hypertension. Regardless of gender, Fins in the highest quintile of neck circumference were approximately three times more likely to have hypertension than Fins in the lowest quintile, even after controlling for BMI. This finding was corroborated by a study of 20 diabetic and 122 non-diabetic women that discovered that neck adiposity was better at discriminating type-II diabetics from non-diabetics than adiposity in 14 other body locations (e.g., upper abdomen), BMI, weight or body fat percentage (Tafeit et al. 2000).

A Promising New Approach to Measuring Body Mass

In a statement on the relationship between facial fat and insulin resistance, Sierra-Johnson and Johnson (2004) presented a Venn diagram depicting hypothetical interrelationships between facial fat accumulation, abdominal fat accumulation and obesity (see Figure 4.1). This review of the literature corroborates this diagram—facial adiposity is related to both central adiposity and BMI. Moreover, neck circumference, deposits of adipose tissue in the neck, and facial bone structures are all related to upper body obesity and BMI. Taken together, these scientific findings imply that the methodical examination of photographs may provide a novel and useful way to assess the relative body mass of human subjects. Furthermore, because facial adiposity is strongly associated with chronic conditions such as hypertension, insulin resistance and sleep apnea, the development of new methodologies capable of isolating facial characteristics is potentially important because it may yield new insight into the etiology and prevention of these conditions.

While this review of the literature has elucidated the scientific basis for expecting a strong relationship between facial characteristics and body mass, it has not addressed the ability of human beings to distinguish lean from heavy persons solely on the basis of facial photographs—and with good reason. An exhaustive search of the literature produced only one study (Rudin 1996) that required subjects to estimate weight from photographs.[2] Although social scientists have regularly studied ascribed traits such as attractiveness or stigma by contrasting photographs or drawings of obese figures to figures with other physical traits (e.g., Furnham and Radley 1989; Latner and Stunkard 2003; Richardson et al. 1961), they have not explored whether human subjects can reliably distinguish lean from heavy persons from visual stimuli provided exclusively by facial photographs.

Furnham and Radley (1989) did find, however, that a group of young subjects (ages 16 to 21) had the ability to rank order drawings depicting naked persons of varying adiposity along a continuum from very thin to obese. In this study, 130 subjects were presented with twelve drawings of adult males and twelve drawings of adult females and then asked to sort the drawings based on perceptions of relative weight. Impressively, all 130 subjects performed this task perfectly (i.e., *no* errors were committed). Obviously, there are many important differences between rank ordering drawings of entire naked bodies and attempting to perform a similar task given only photographs depicting the faces and necks of clothed individuals. Nonetheless, these results are encouraging since they imply that human beings have a refined capacity for differentiating the relative weight of other persons based solely on visual stimuli.

---

[2] Rudin (1996) required subjects to assign photographs of male and female adults—all taken from issues of issues of *Fortune* and *Working Woman* magazine—to one of three categories; "overweight," "thin," or "can't tell." It is unclear from the whether these photographs isolated facial characteristics, but the language in the article suggests that they did not. Moreover, because Rudin's study employed crude categories of relative weight and provided no formal assessment of reliability or criterion-related validity, it did not provide substantial evidence on the ability of human subjects to distinguish lean from heavy persons on the basis of photographs. However, Rudin did find a high level of agreement between male and females coders, which suggests that gender differences in body-image do not distort intersubjective agreement regarding the relative weight of anonymous persons in photographs.

It is also encouraging to note that Kato and Higashiyama (1998) found that undergraduate subjects were generally able to provide accurate estimates of height based on full-length photographs of Japanese persons standing a standard distance in front of a homogenous white wall. For each photograph ($n = 50$), a mean rating score for estimated height was calculated for all thirty subjects. The correlation between actual height and mean values of estimated height was very strong ($r = 0.92$; $p < 0.001$). The estimation of height from standardized, full length photographs may not generalize to the estimation of relative body mass from yearbook photographs. However, this study provided additional evidence that human beings have a refined capacity for differentiation of other persons based on their physical characteristics.

In conjunction with considerable evidence on the association between facial characteristics and body mass, limited but compelling evidence on the human ability to estimate the height and weight of others from visual stimuli suggests that there is considerable promise in developing a new methodology to estimate relative body mass from yearbook photographs. This study represents an initial attempt to construct such a methodology. As a first step, 293 photographs were randomly selected from the yearbooks of WLS respondents for a pilot study designed to provide an initial assessment of the validity and reliability of the new method. Based on encouraging results from the pilot study, a total of 3,027 photographs from the WLS were subsequently assigned values for relative body mass by a team of six coders. Results of the complete study of 3,027 photographs indicate that the new measure of relative body mass is valid and reliable, which is exciting news for social scientists who are interested in using the WLS to study the obesity epidemic.

4.2 Methods

Study Population

The Wisconsin Longitudinal Study (WLS) is a random sample of 10,317 persons who graduated from a public, private or parochial high school in Wisconsin in 1957 (Sewell et al. 2004). In this initial wave, the WLS collected information on academic ability, socioeconomic background, attitudes toward higher education, educational and occupational aspirations, and a handful of contextual factors (Hauser 2005). Subsequent waves in 1964, 1975, 1992-93 and 2003-05 collected data from WLS respondents (or their parents) on a wide range of issues that are essential to studies of the life course, including educational and occupational histories, indicators of socioeconomic status, military service, marital status, family characteristics, social participation, psychological well-being, health behaviors and health outcomes (Hauser 2005; Sewell et al. 2004). Data on the height and weight of WLS respondents were also gathered in the 1992-93 and 2003-05 waves, which make it possible to assess BMI change during the heart of the obesity epidemic in the U.S. Although the WLS is not nationally representative, its respondents resemble over two-thirds of Americans who are now entering retirement age in terms of academic achievement and ethnic background (Hauser 2005).

The value of the WLS as a resource for studies of obesity will be enhanced by the addition of a measure of body mass at baseline. Consequently, this study measured the relative body mass of WLS respondents in 1957 by systematically examining photographs that were extracted from high school yearbooks. Through a cluster sampling design that selected schools based on probabilities proportional to size (PPS), a subsample of 93 schools representing 3,130 WLS respondents was randomly chosen for inclusion in this study. Because schools were selected on a PPS basis, each of the 10,317 WLS respondents had an equal probability of selection into this subsample of 3,130 respondents. Yearbooks from 1957 were collected from schools or libraries in the towns where selected schools were located. Photographs of WLS respondents (who were seniors at the time) were subsequently extracted through computerized scanning technologies. Due to a small number (103) of missing photographs, the final sample of photographs available for this study was 3,027. A single cluster of schools representing 293 photographs (171 girls and 122 boys) was randomly chosen for use in the pilot study.

Scale Development

Separate scales for boys and girls were developed to measure the relative body mass of WLS respondents in 1957. The term "relative body mass" (RBM) was chosen because it was hypothesized that the scale will result in a proxy for BMI. The rationale for this hypothesis was that extracted yearbook photographs generally do not provide direct visual evidence about the height of respondents. Therefore, as coders attempt to rate the photographs for relative weight, they should do so independently of the height of WLS respondents. In this way, the RBM scale is analogous to BMI which, of course, is also a measure of weight that controls for the height of subjects (BMI = weight(kg)/height(m)^2). Additionally, because the *actual* body mass of WLS respondents is unobserved in the 1957 yearbook photographs, it is not possible to measure this quantity through the RBM scale. Rather, the RBM scale is designed to assess the body mass of WLS respondents relative to one another.

Meland (2002) developed an eleven point scale to rate WLS photographs for physical attractiveness. Photographs of individuals of varying attractiveness were used to label points on the attractiveness scales and guide coders as they rated each yearbook photograph. Meland's attractiveness scales also used verbal descriptions as anchors—"not at all attractive" at one end of the scale and "extremely attractive" at the other. This methodology for coding facial attractiveness resulted in outstanding estimates of inter-rater and intra-rater reliability.

The literature also offers support for Meland's approach. In a text on scale development, DeVellis (1991) wrote:

> A desirable quality of a scale is variability. A measure cannot covary if it does not vary. [However] Another issue related to the number of response options is the *respondents' ability to discriminate meaningfully*. How fine a distinction can the typical subject make? (PP. 64-65)

Alwin (1992, 1997) has indirectly answered this query via research showing that subjects can reliably distinguish between as many as eleven points on attitudinal scales. Analyses of 99 measures from the 1973-74 General Social Survey and 1956-60 National Election Study indicated that reliability increased rapidly as the number of scale points increased from two to about seven (Alwin 1992). Reliability continued to increase monotonically from seven to eleven scale points, but at a slower pace. Although Alwin did not examine scales with more than eleven points, the functional form of the reliability distribution suggested that scale reliability would asymptote soon after eleven. Another advantage of adopting a scale with eleven points is that it has an odd number of response categories, which provides the option of a midpoint for respondents who hold indifferent or moderate attitudes toward a particular object or issue. This may enhance reliability relative to scales with an even number of response categories, which can inflate random error by encouraging respondents to report attitudes that are slightly more or less intense than their true attitudes (Alwin and Krosnick 1991).

Scales that attach verbal labels or other cues (e.g., photographs) to response options are preferable to those that use only numbers, since the latter "probably involve some inherent ambiguity of meaning" (Alwin and Krosnick 1991:152). Indeed, a number of studies (e.g., Finn 1972; Peters and McCormick 1966; Weng 2004) have found that reliability improves as the share of scale points that display labels increases. However, in a study of measurement error in survey data, Andrews (1984) found that data quality deteriorated when all scale points were labeled. In recognition of the need for clear labels but the potential pitfalls of over-labeling, Meland (2002) provided verbal labels or photographic cues for five of the eleven points on her attractiveness scale.

Given the consistency of Meland's methodology with the scientific literature on scale development and its success in reliably rating WLS photographs for physical attractiveness, I adopted her attractiveness scale with appropriate modifications for the measurement of RBM. Therefore, like Meland's attractiveness scale, the RBM scale has eleven points, two verbal anchors and five photographs (for males and females, separately) which are used as response cues for every other scale point. Photographs were chosen instead of verbal labels because they clearly and succinctly illustrate facial characteristics that are known to relate to BMI.

One simple modification was to change the scale anchors from "not at all attractive" and "extremely attractive" to "not at all heavy" and "extremely heavy," respectively. More difficult was the selection of male and female photographs for use as response cues on the RBM scales. Response cues were chosen by sorting through several hundred photographs of WLS respondents who were not chosen for inclusion in this study and either (1) discarding them or (2) placing them into separate folders for (a) very thin, (b) somewhat thin, (c) average body mass, (d) somewhat heavy and (e) very heavy. From these folders containing 5 to 10 photos each, I selected individual photographs that seemed representative of each folder and provided a continuum when placed onto the gender-specific RBM scales.

As a new methodology, no one has previous experience developing RBM scales. Therefore, it was impossible to seek guidance from expert reviewers with regard to my choice of photographs for response cues. Nevertheless, to evaluate my selection of photos, I requested critical feedback from two scholars in the Sociology department at the University of Wisconsin-Madison. One of these scholars (Karen Swallen) has completed several epidemiological and sociological studies of obesity and is currently pursuing formal medical training. The other scholar (Sheri Meland) has prior experience with scale construction and the difficulties presented by choosing photos to illustrate scale points. Although feedback from these scholars tended to corroborate my choices, some minor modifications did result from the process. To protect respondent confidentiality, these scales are not reproduced here. However, similar scales were developed for illustrative purposes from a pool of non-WLS respondents who also graduated from Wisconsin high schools in 1957 (see Figure 4.2 and Figure 4.3).

Measures

For every photograph of a WLS respondent, coders recorded a RBM scale score ranging from one to eleven. The RBM scores of individual coders were used for reliability and validity assessments and may also be used for latent variable modeling (see Chapter 5). Additionally, coder-specific scale scores were combined into two indexes of RBM. First, an unstandardized relative body mass index (URBMI) was constructed by calculating the mean of the RBM scores for individual coders. Second, a standardized relative body mass index (SRBMI) was constructed to eliminate differences in means and variances between coders that could cause certain coders to have disproportionate influence over URBMI. SRBMI was calculated separately for male and female photos by (1) generating coder-specific z-scores, (2) summing z-scores across coders and (3) dividing the sum of z-scores by the number of coders in the study. That is,

$$\text{SRBMI} = \frac{\sum_{i=1}^{n}[(x_{ij} - \bar{x}_{ij}) / \sigma_{ij}]}{n},$$

where i is an individual coder, n is the number of coders in the study, j is the gender of the WLS respondent in the photographs and x_{ij} is the series of RBM scale scores for coder i and gender j with mean $\bar{x}_{ij}$ and standard deviation σ_{ij}.

In some applications, SRBMI was treated as both a continuous and a categorical variable. This permitted the evaluation of possible non-linear associations between SRBMI and outcomes such as chronic disease and mortality. It also permitted SRBMI to be divided into standard BMI classifications for adolescents—underweight, normal weight, risky weight and overweight. Previous research (Ogden et al. 2002) has used BMI percentiles from CDC growth charts (Centers for Disease Control 2000) to define underweight at or below the 5th percentile, normal weight between the 5th and 85th percentiles, at risk for overweight between the 85th and 95th percentiles and overweight at or above the 95th percentile. To provide sufficient statistical power for each subgroup, I altered these percentile ranges slightly for body mass categories derived from SRBMI; underweight was defined at or below the 10th percentile of SRBMI, normal weight between the 10th and 80th percentiles, at risk for overweight between the 80th and 90th percentiles and overweight at or above the 90th percentile.

Several measures from the 1993 WLS were used to assess the discriminant and criterion-related validity of the RBM scale, including self-reported height, self-reported weight, BMI, self-rated health, and several symptoms and conditions indicative of health problems. The WLS measured height and weight in inches and pounds, respectively, which were then converted into a measure of BMI.[3] As discussed in the previous chapter, the range of BMI from measured values of height and weight in NHANES II, III and Continuous was approximately 12-70. Values of BMI in the WLS were truncated within this range, which necessitated the recoding of only one case with a reported BMI of 83. Indicators of obesity (BMI ≥ 30) and class II obesity (BMI ≥ 35) were also created to estimate the effect of overweight in 1957 on obesity in 1993.

Self-rated health was measured via a survey question that asked "How would you rate your health at the present time?" Response options to this question were very poor, poor, fair, good and excellent. Although the WLS asked other questions regarding self-rated health, indexes incorporating those items did not yield different results and resulted in some data loss. Therefore, results were only reported for this single measure of self-rated health.

Participants in the 1993 WLS mail survey responded to a series of questions probing whether they had experienced particular health-related symptoms in the past six months. Of the 22 symptoms covered in the WLS, twelve were examined in this study. Ten of these symptoms (muscle aches, stiff or swollen joints, back pain or strain, chest pain, shortness of breath,

---

[3] The standard BMI formula uses metric measurements (i.e., BMI = weight(kg)/ height(m)^2). The conversion formula for English measurements is BMI = [weight(lbs)/height(inches)^2]*703 (U.S. Dept. of Health and Human Services 2004).

respiratory problems, dizziness or faintness, trouble sleeping, excess sweating and numbness) have known associations with obesity (Lean, Han and Seidell 1999; Stafford, Hemingway and Marmot 1998; Stunkard 1996) and were consequently used to assess the criterion-related validity of the RBM scale. Because obesity is not a known risk factor for the other two symptoms (ringing ears and nausea), they were included to assess discriminant validity.

WLS participants also responded to a series of questions about chronic conditions that have been diagnosed by a medical professional. Of the seventeen chronic conditions covered in the WLS, eleven were retained. Obesity is a known risk factor for nine of these conditions (arthritis, serious back trouble, high blood pressure, diabetes, heart trouble, circulation problems, cancer, asthma, and kidney or bladder problems), which were used to further evaluate the criterion-related validity of the RBM scale (Manson et al. 2002; Pi-Sunyer 2002; Stunkard 1996). In addition, a measure of comorbidity was constructed to determine whether a medical professional had diagnosed WLS respondents with two or more of these nine chronic conditions. Because obesity is not a known cause of either ulcers or allergies, these conditions were included to provide further assessment of discriminant validity. Both health symptoms and chronic conditions were coded as dummy variables in this study (i.e., 1 = symptom/condition present; 0 = symptom/condition not present).

Finally, measures of all-cause and cause-specific mortality were used to provide further assessment of the criterion-related validity of the RBM scale and determine whether the associations between the RBM scale and self-rated health, health symptoms and chronic conditions might be attenuated by selection bias. To provide an indicator of all-cause mortality, WLS personnel searched the Social Security Death Index (SSDI) for WLS participants in April of 2004. Although the SSDI does not provide information on the cause of death, it captures all deaths of U.S. citizens that were reported to the Social Security Administration after 1962 (Social Security Administration 2005).[4] Of the 3,027 WLS respondents in this study, 242 died between 1962 and 2004 according to the SSDI.

Determination of death and cause of death was also made by searching the National Death Index (NDI) for WLS respondents (National Center for Health Statistics 1999). WLS personnel most recently searched the NDI database in October of 2001. Of the 3,027 WLS subjects in this study, the NDI search indicated that 159 individuals died between 1979 and 1998. Seventeen subjects had already died by 1979 (according to SSDI data) and were eliminated from the NDI measures. Data from the NDI were matched with cause of death codes from the International Classification of Diseases (World Health Organization 1977) to construct measures of mortality resulting from (1) all causes, (2) all non-accidental causes, (3) all major diseases of the heart, and (4) all malignant cancers. Other major causes of mortality were also examined (e.g., diabetes), but these results were not reported due to insufficient statistical power arising from the low number of deaths from these causes.

Data Collection

Data collection was divided into two phases—the pilot study and a final data collection project. In the pilot study, five graduate students were recruited from the Sociology department at the University of Wisconsin-Madison to code 293 photographs that were chosen at random

[4] The SSDI was not automated until 1962, and does not systematically capture deaths that occurred prior to that date. Deaths occurring before 1962 are, generally speaking, only found in the SSDI if they were reported in 1962 or a later date.

from the extracted sample of 3,027 photographs. Two of the pilot coders were male and three were female. Coders were non-Hispanic White and varied in age from 26 to 33 years. Prior to coding the photographs, coders signed confidentiality agreements, read a basic set of written coding instructions, heard an oral summary of the project and were provided with several opportunities to ask questions (see Appendix D). All data collection for the pilot study and the final data collection project utilized personal computers located within the WLS main office in the Social Science Building at the University of Wisconsin-Madison.

The protocol for coding the photographs was straightforward. Coders were instructed to compare the photograph of the WLS respondent to the scale photographs located along the top of the computer monitor (see Figure 3.3). Note that the box at the bottom of the monitor is where the photograph of the WLS respondent appeared. Coders were given 10 seconds to make this initial comparison and, using a computer mouse, click the point that corresponded to their impression of where the photograph fit along the scale continuum. The chosen scale point immediately changed color from black to red and coders were presented with the question "Does the red symbol indicate your choice?" Before answering, coders were instructed to examine the facial features of the WLS respondent in more detail. If coders subsequently answered "Yes," they were presented with the next photograph. Coders answering "No" were provided with an opportunity to change their selection before moving on to the next photograph.

Given encouraging results from the pilot study, the RBM scale was not modified in any way for the final coding project of 3,027 photographs. All five coders from the pilot were retained for the complete project. One additional coder (a non-Hispanic White male) was recruited in the final study to balance out the gender distribution and maximize reliability estimates. The protocol for the complete project, which occurred over a three week period in January of 2005, was exactly the same as the pilot study. This research received clearance from the Social and Behavioral Science Institutional Review Board at the University of Wisconsin-Madison as part of general approval granted for research involving yearbook photographs from WLS participants.

After completing the final project, coders were instructed to reevaluate 100 male and 100 female photographs that were randomly chosen from the sample of 3,027. Reevaluation of these photographs was not conducted until at least one week after the end of the project to reduce the likelihood of recall bias. Data gathered from this phase of the project were used to assess the intra-rater reliability of individual coders.

Statistical Analyses

SAS 9.1 (2003) was used to manage data, generate descriptive statistics and analyze the reliability and validity of the RBM scale. SPSS 8.0 (1997) was used to evaluate the functional form of the relationship between URBMI and BMI. SPSS 8.0 was also used for certain graphics applications.

Coefficient alpha (Cronbach 1951) was used to evaluate inter-rater reliability, which may be expressed as follows:

$$\alpha = \frac{n}{(n-1)}\left[1 - \frac{\sum_{i=1}^{n} \sigma^2{}_{Yi}}{\sigma^2{}_{X}}\right],$$

where n is the number of coders in the study, $\sigma^2{}_{Y_i}$ is the variation in the RBM scale for coder *i* and $\sigma^2{}_X$ is the total variation in the RBM scale for all coders. Alpha is commonly used to assess the degree of internal consistency for measures of unobserved variables (Traub 1994). Coefficients of 0.70 or greater are generally thought to reflect an acceptable degree of internal consistency (Nunnally 1978).

Because reliability "refers to the correlational linkage between observed variables and some conception of a 'true' variable" (Alwin and Krosnick 1991:141), latent variable modeling may also be used to assess inter-rater reliability. An advantage of latent variable modeling is that it separates error from true score variation, thereby maximizing associations between latent variables. However, given the high reliability of the RBM scale, associations between RBM and BMI were strengthened only slightly when RBM was treated as a latent variable instead of an index (results not shown). For examples of RBM as a latent variable in covariance structure models of changes in body mass, please refer to Chapter 5.

Intra-rater reliability ((i.e., the agreement of each coder with herself (or himself) at two different points in time)) was evaluated via the test-retest coefficient of reliability (Traub 1994). As noted earlier, each coder reevaluated 200 randomly selected photographs one week after assigning RBM scores to all 3,027 photographs. Under the assumption of parallel tests, test-retest reliability was estimated as the correlation coefficient between RBM scores assigned at these two different times.

The criterion-related and discriminant validity of the RBM scale was assessed by generating correlation matrices of URBMI, SRBMI, BMI, height, weight and self-rated health.

Additionally, the validity of the RBM scale was evaluated by conducting logistic regression analyses to estimate the risk of obesity in 1993 as a function of weight classification in 1957. Logistic regression analysis was also used to evaluate the effect of RBM in 1957 on reports of illness symptoms and physician diagnosed chronic conditions in 1993. Finally, indicators of mortality from the SSDI and NDI were regressed on SRBMI (both in continuous and categorical forms) to evaluate whether RBM in 1957 predicted all-cause and cause-specific mortality in subsequent years. These analyses were performed for the entire sample of WLS respondents and also stratified by gender.

4.3 Results

The Pilot Study

Results of the pilot study were encouraging in several respects. First, the RBM scale exhibited good measurement properties. Mean and median scores for individual coders and URBMI clustered around 6.0, which is the midrange score of the RBM scale (see Table 4.1). The range of URBMI for all 293 photos was 2.2-10.6. In conjunction with coder scores ranging from one (Coder 1) to eleven (Coders 2-5), this demonstrated that the full range of the RBM scale was utilized in the pilot. Additionally, URBMI was normally distributed, as evidenced by slight skewness and kurtosis. Approximately 68 percent of the 293 photos scored between 4.5 and 7.7 on URBMI; 95 percent of the photos scored between 3.2 and 9.2.

Second, the reliability of the RBM scale was excellent. For all 293 photos, Cronbach's alpha was 0.91, indicating a high degree of agreement between coders (see Table 4.1). All five coders contributed to the reliability of the scale, as evidenced by (1) strong correlations between the scores of individual coders and the total inter-item variation and (2) lower alpha scores when individual coders were removed from the analysis. Inter-rater reliability for the 122 male photos ($\alpha = 0.93$) was somewhat better than the 171 female photos ($\alpha = 0.89$) but, regardless, the degree of agreement between coders was very high.

Third, the RBM scale was a valid indicator of body mass (see Table 4.2). Due to various forms of attrition, data on height and weight were not available for 28 percent of the respondents (22 percent of females and 36 percent of males) in the 1993 wave. Correlations for the remaining 212 respondents (134 females and 78 males) demonstrated that neither URBMI nor SRBMI was significantly related to height. This was indicative of discriminant validity and supported previous assertions that the RBM scale is independent of height. Significant correlations between the RBM indexes and BMI (e.g., $r_{BMI.SRBMI} = 0.29$; $p < 0.01$) provided evidence of criterion-related validity. Correlations between SRBMI and BMI were stronger among females ($r = 0.41$; $p < 0.01$) than males ($r = 0.18$; $p > 0.05$). The lack of a significant association between SRBMI and BMI among males was caused not only by the relatively weak association, but also by relatively low statistical power ($n = 78$). In any event, because the RBM indexes were not related to height and, on the whole, useful predictors of BMI, the notion that the RBM scale is a valid proxy for body mass was supported.

The preceding analyses were based on the assumption of linear relationships between the RBM indexes and outcome variables, especially BMI. The examination of scatterplots and various functional forms indicated that linearity was a reasonable assumption. To illustrate, the scatterplot in Figure 4.4 shows that as scores on URBMI increased, BMI tended to increase in a monotonic fashion. The proportion of variance explained by quadratic and cubic models was less than one percent greater than the linear model. Scatterplots and functional forms were also evaluated in gender specific models (results not shown). In each of these models, the relationship between the URBMI and BMI was essentially linear.

The Complete Study

Because results from the pilot study indicated that the RBM scale was a valid and reliable measure of body mass, all 3,027 photos were subsequently coded without any changes in construction to the RBM scale. Results from the complete study of WLS photographs were generally similar to the pilot. As before, the RBM scale exhibited good measurement properties (see Table 4.3). In the full sample of 3,027 photos, URBMI was normally distributed (skewness = 0.17; kurtosis = -0.15) with median (6.33) and mean (6.48) values near 6.0, which is the midpoint of the RBM scale. Also as before, the full range of the RBM scale was utilized in the complete study, although coders were somewhat less apt to code photos as "not at all heavy" than to code them as "extremely heavy." Approximately 68 percent of the 3,027 photos scored between 5.0 and 7.7 on the RBM scale; 95 percent scored between 3.7 and 9.0. Descriptive statistics in Table 4.3 varied little by gender of the WLS respondents, or by the gender or age of the coders.

Results from the complete study confirmed that the RBM scale is a reliable measure of body mass (see Table 4.3). Overall, Cronbach's alpha was unchanged from the pilot study (α = 0.91), indicating a very high level of agreement between coders. Gender differences in inter-rater reliability in the pilot disappeared in the complete study, as shown by equivalent alpha scores for female photos (α = 0.91) and male photos (α = 0.91). Correlations between the scores of individual coders and the total inter-item variability were generally high, ranging from 0.64 to 0.81. Although Coder 6 had the lowest correlation with the total inter-item variation (r = 0.65; $p < 0.01$), he did not cause alpha to deteriorate in reliability analyses of either the full sample or gender-specific samples of photos. However, the removal of any other coder would have resulted in varying degrees of deterioration in alpha, indicating that they were important contributors to the reliability of the RBM scale.

Intra-rater reliability was also quite good, as evidenced by correlations ranging from 0.62 to 0.90 ($p < 0.01$ in all cases) for the sample of 200 photos that were coded at two different points in time (see Table 4.3). As with inter-rater reliability, Coder 6 exhibited the lowest degree of intra-rater reliability, indicating that he agreed least not only with other coders, but also with his own initial assessments of the photographs. This suggested that subsequent analyses might be improved if Coder 6 were dropped from the study. However, because eliminating Coder 6 did not improve inter-rater reliability but caused slight deterioration of the predictive validity of SRBMI (results not shown), he was retained in these analyses.

Generally speaking, female coders exhibited superior inter-rater and intra-rater reliability (see Table 4.3). Although intra-rater reliability was highest for Coder 1 (a male), it was clearly lower for Coder 2 and Coder 6 (both males) than the female coders. Inter-rater reliability was also low for Coder 2 and Coder 6 relative to the other coders. No discernable patterns in reliability were detected by age, which may be a function of the restricted age range of coders in the study. Given the small number of coders, it is impossible to determine whether gender differences in reliability were real or simply the result of sampling variability.

Results from the complete study confirmed that the RBM scale is a valid indicator of body mass. Correlations between URBMI, SRBMI, height, weight, BMI and self-rated health provided clear evidence of discriminant and criterion-related validity (see Table 4.4). Impressively, the correlation between SRBMI and height in the full sample of 1,999 photos was 0.00 ($p > 0.05$). This result varied little by gender or measure of RBM, demonstrating conclusively that the RBM scale was not confounded by height. Additional evidence of discriminant validity was found in somewhat stronger associations between RBM and BMI than RBM and weight in the full sample. For example, the correlation between SRBMI and BMI (r = 0.31; $p < 0.01$) was stronger than the correlation between SRBMI and weight (r = 0.25; $p < 0.01$). Robust associations between the RBM indexes and BMI were also evidence of criterion-related validity. Importantly, gender disparities in these associations that were detected in the pilot study disappeared when all WLS subjects were analyzed. In fact, the correlation between SRBMI and BMI was slightly stronger among males (r = 0.34; $p < 0.01$) than females (r = 0.30; $p < 0.01$) in the complete study. Finally, although correlations between the RBM indexes and self-rated health were weak, they were statistically significant in the full sample (r = -0.08; $p < 0.01$) and among females (r = -0.13; $p < 0.01$). Despite the passage of 36 years, persons with elevated RBM at baseline tended to report lower self-rated health in 1993.

Perhaps even more impressive was the effect of body mass classification in 1957 on the risk of obesity 1993. Relative to persons classified as having normal weight in 1957, underweight persons were 40 percent less likely to be obese in 1993 (Odds Ratio (OR) = 0.60; 95% Confidence Interval (CI) = 0.39-0.93). Obesity risks were also 2.7 times higher (CI = 2.01-3.73) among persons at risk for overweight and 3.4 times higher (CI = 2.48-4.56) among overweight persons, showing a graded effect of adolescent weight on the probability of obesity in middle age. The steepness of this gradient increased when class II obesity was considered. Compared to persons with normal weight in 1957, persons with risky weight had over threefold risk (OR = 3.33; CI = 1.99-5.57) and overweight persons had nearly *fivefold* risk (OR = 4.76; CI = 2.98-7.61) of exhibiting class II obesity in 1993. As with correlations between SRBMI and BMI, these associations did not vary substantially by gender.

Persons with elevated RBM at baseline also tended to report more illness symptoms and chronic conditions in 1993 than other WLS respondents. Results from a series of logistic regression analyses of illness symptoms on SRBMI (both in continuous and categorical form) were reproduced in Table 4.5. Highlights from this table include the following: In the full sample of 2,049 cases, SRBMI was a significant predictor of muscle aches (OR = 1.14; 95% CI = 1.02-1.26), stiff or swollen joints (OR = 1.21; CI = 1.07-1.36), dizziness or faintness (OR = 1.33; CI = 1.09-1.62), excess sweating (OR = 1.26; CI = 1.07-1.49) and numbness (OR = 1.21; CI = 1.02-1.44). Persons in the top decile of SRBMI (i.e., overweight persons) in the full sample were significantly more likely than persons between the first and eight deciles (i.e., normal weight persons) to report most of these same conditions. In addition, overweight persons were about two times more likely to report chest pain (OR = 2.16; CI = 1.36-3.44), shortness of breath (OR = 1.81; CI = 1.24-2.65) and respiratory problems (OR = 2.40; CI = 1.52-3.81) than normal weight persons.

Overweight females were roughly two times more likely to report back pain (OR = 1.61; CI = 1.08-2.41), chest pain (OR = 2.32; CI = 1.18-4.56) and dizziness or faintness (OR = 2.34; CI = 1.34-4.07) than normal weight females. Overweight males were about two times more likely to report chest pain (OR = 1.98; CI = 1.05-3.75), shortness of breath (OR = 2.40; CI = 1.43-4.03), excess sweating (OR = 2.50; CI = 1.27-4.90) and numbness (OR = 1.84; CI = 1.03-3.30) than normal weight males. Distressingly, overweight males were also over four times more likely to report respiratory problems (OR = 4.48; CI = 2.15-9.32). Because SRBMI (either in continuous or categorical form) was a significant predictor of most illness symptoms, further evidence was obtained that the RBM scale possesses criterion-related validity. Furthermore, significant associations were not detected between SRBMI and ringing ears or nausea in any of these analyses, providing additional evidence that the RBM scale possesses discriminant validity.

Results of logistic regression analyses of chronic conditions on SRBMI (in both continuous and categorical forms) were reproduced in Table 4.6. In the full sample, SRBMI was a significant predictor of arthritis (OR = 1.19; CI = 1.06-1.34), high blood pressure (OR = 1.25; CI = 1.10-1.42), diabetes (OR = 1.44; CI = 1.11-1.86), and two or more conditions (OR = 1.25; CI = 1.10-1.42). Overweight persons in the full sample were about two times more likely to report diabetes (OR = 2.38; CI = 1.33-4.28), heart trouble (OR = 1.79; CI = 1.08-2.95) and asthma (OR = 1.92; CI = 1.08-3.41) than normal weight persons.

Among females, SRBMI was a significant predictor of arthritis (OR = 1.31; CI = 1.13-1.53), high blood pressure (OR = 1.44; CI = 1.20-1.73), diabetes (OR = 1.70; CI = 1.17-2.47), ulcer (OR = 1.47; CI = 1.02-2.11) and two or more conditions (OR = 1.36; CI = 1.14-1.62). Because ulcer was included to assess discriminant validity, its significant association with SRBMI was unexpected. Overweight females were significantly more likely to report arthritis (OR = 1.58; 1.05-2.39) and diabetes (OR = 3.61; 1.66-7.85) than normal weight females. Despite associations generally in the expected direction, SRBMI was not a significant predictor of diabetes or any other condition among males. However, overweight males were two times more likely to report heart trouble (OR = 2.08; 1.14-3.80) and nearly three times more likely to report asthma (OR = 2.67; 1.08-6.61) than normal weight males. Underweight males were also three times more likely to report asthma (OR = 2.90; CI = 1.11-7.58).

According to data from the SSDI, SRBMI was a significant predictor of all-cause mortality among females (OR = 1.27; CI = 1.00-1.60) but not males (OR = 0.99; CI = 0.80-1.23) between 1962 and April of 2004 (see Table 4.7). This finding was corroborated by data from the NDI, which found significant associations between SRBMI and all-cause mortality between 1979 and 1998 for the entire sample (OR = 1.25; CI = 1.03-1.51) and females (OR = 1.44; CI = 1.10-1.88) but not males (OR = 1.08; CI = 0.82-1.42). The elimination of accidental causes of death strengthened these odds ratios for all groups, but did not result in a significant association between SRBMI and mortality for males (OR = 1.17; CI = 0.88-1.57). Relative to normal weight persons in the full sample, overweight persons had twofold increased risk of death from non-accidental causes (OR = 1.91; 1.20-3.04) and *fourfold* increased risk of death from heart disease (OR = 3.99; CI = 1.95-8.15). SRBMI was strongly predictive of heart disease mortality among females (OR = 2.11; CI = 1.08-4.11) but just failed to achieve the criterion of statistical significance among males (1.52; CI = 0.99-2.33). Moreover, despite a clear graded effect, neither risky weight (OR = 1.69; 0.56-5.08) nor overweight (OR = 1.85; 0.68-5.05) males were significantly more likely to die from heart disease than normal weight males. Neither continuous nor categorical SRBMI predictors were significantly associated with cancer mortality, although these associations were relatively strong and in the expected direction for women.

4.4 Summary and Conclusions

Despite many advantages, the WLS has been limited in etiological investigations of obesity by the lack of a baseline measure of body mass. Through the development of the RBM scale, this study has successfully removed that limitation. As hypothesized, the RBM scale was a valid and reliable proxy for body mass with good measurement properties. Indexes summarizing the RBM scores of individual coders (e.g., URBMI) were normally distributed, uncorrelated with height and more strongly correlated with BMI than weight. With a handful of exceptions, continuous and categorical measures of SRBMI also predicted the risk of obesity, illness symptoms, chronic conditions and mortality in later years. Given these encouraging results, scholars who intend to use the WLS to study the obesity epidemic and obesity issues in general may want to consider combining its rich set of socioeconomic, demographic, psychological and health measures with this new baseline measure of RBM.

However, rather than assuming that the RBM scale is essentially equivalent to anthropometric measures of adolescent body mass, it is important to compare its ability to predict BMI, morbidity and mortality in adulthood with those measures—particularly adolescent BMI. In recent analyses of data from the Bogalusa Heart Study, Freedman et al. (2005) found strong correlations between direct anthropometric measures of BMI for male (r = 0.74; $p < 0.01$) and female (r = 0.66; $p < 0.01$) adolescents (ages 15-17 at baseline) after fifteen years of follow up. Superficial comparison of these strong correlations to those observed between adolescent RBM and adult BMI in this study (see Table 4.4) suggests that the RBM scale may have underestimated the true association between adolescent and adult body mass. Of course, it is important to consider that the amount of time between measurements of adolescent and adult body mass was twenty years longer in the WLS than the Bogalusa Heart Study. Changes in weight and mortality selection that occurred during those twenty years may be responsible for attenuation in the association between adolescent and adult measures of body mass.

Indeed, there is compelling evidence to support this speculation. In a study of British cohort data, Wright et al. (2001) reported a correlation of 0.39 ($p < 0.01$) between BMI measured at the ages of 13 and 50. By comparison, Freedman et al. (2005) found much stronger correlations between BMI measured at the ages of 12-14 and again after ~16 years of follow up for both males (r = 0.71; $p < 0.01$) and females (r = 0.68; $p < 0.01$). The British cohort was followed for approximately the same length of time (38 years) as the WLS cohort between 1957 and 1993 (36 years). Also, both (1) the strength of association between measures of adolescent and adult body mass and (2) the degree of attenuation relative to the correlations reported by Freedman et al. were quite comparable between studies. Furthermore, just as Freedman et al. found slightly stronger associations between adolescent and adult BMI among males, so this study also found slightly stronger associations between adolescent RBM and adult BMI among males (see Table 4.4). Taken together, these findings strongly suggest that the RBM scale is comparable to adolescent BMI in terms of its ability to predict adult BMI.

Available evidence also suggests that the predictive validity of the RBM scale is comparable to adolescent BMI in relation to adult morbidity and mortality. Through analyses of 508 adolescents in the Harvard Growth Study, Must et al. (1992) found that overweight adolescents (defined as BMI $\geq 75^{th}$ percentile) were significantly more likely to report diabetes, coronary heart disease, arthrosclerosis and hip fracture than lean adolescents (defined as BMI between the 25^{th}-50^{th} percentiles) after 55 years of follow up. To illustrate, men who were overweight adolescents were 2.8 and 1.6 times more likely to report coronary heart disease and diabetes, respectively, than men who were lean adolescents; corresponding relative risks for women were 1.6 and 2.2. Similarly, Freedman et al. (2001) found that childhood BMI was a significant predictor of CHD risk factors such as cholesterol, triglycerides, insulin and blood pressure in adulthood. These results compare favorably to those reported in this study. Adolescent measures of RBM were significant predictors of several illness symptoms (muscle aches, stiff joints, back pain, chest pain, shortness of breath, respiratory problems, dizziness or faintness, excess sweating and numbness) and chronic conditions (arthritis, high blood pressure, diabetes, heart trouble and asthma) that are indicative of cardiovascular and musculoskeletal distress in adulthood.

Must et al. (1992) also found that overweight adolescents had elevated risks of mortality in adulthood. But in contrast to this study, Must et al. only found significant associations between adolescent overweight and adult mortality among men. Men who were overweight during adolescence were 1.8 times more likely to die from all causes and 2.3 times more likely to die from coronary heart disease than men who were lean. However, in a longitudinal study of 128,121 Norwegian adolescents, Engeland et al. (2004) found that overweight adolescents were at increased risk of adult mortality, regardless of gender. Relative to adolescents with medium BMI (defined as BMI between the 25^{th}-74^{th} percentiles), adolescents very high BMI (BMI $\geq 85^{th}$ percentile) were 40 percent more likely to die after 29 years of follow up.

Despite somewhat stronger correlations between RBM indexes and adult BMI among males (see Table 4.4), RBM was a much stronger predictor of mortality among females. The reason for this apparent contradiction is unclear, but one important possibility is that the RBM scale is more apt to measure adiposity among females than males. Wright et al. (2001) found no association between BMI at age 9 and body fat percentage after 41 years of follow up, which led the authors to argue that childhood BMI is strongly influenced by lean mass such as muscle tissue and bone density. By its very design, RBM was also influenced by lean mass. Feedback from coders in this study suggested that the contribution of lean mass to RBM was greater for males than females. On several occasions, coders raised questions about the proper coding of "big guys" with athletic characteristics (e.g., wide necks) and relatively lean builds. Because the RBM scale was developed explicitly as a proxy for BMI, coders were instructed to code persons who appeared to be heavy as heavy on the RBM scale, regardless of whether this was caused by lean mass or fat mass. Although lean but heavy women were also present in the WLS photographs, they were observed with less frequency than similarly built males. This gender disparity may account for the greater sensitivity of the RBM scale to indicators of mortality for females, which are influenced by adiposity rather than lean mass.

The inability of the RBM scale to differentiate lean mass from fat mass may be viewed as a limitation of the measure, but it is one shared with adolescent measures of BMI. Other limitations of this study include the following: First, there were relatively few coders of RBM, making it difficult to assess systematic differences in coding tendencies by gender, age, ethnicity or socioeconomic background. Of course, this was complicated by the similarity of the coders in terms of age, ethnicity and educational attainment. Second, although extracted photographs of WLS respondents were generally of similar type and quality, there were some inconsistencies that likely contributed to random error. For instance, coders remarked that certain poses (e.g., profile shots) and clothing options (e.g., turtlenecks) obfuscated facial features—particularly of the neck. Also, all WLS photographs were black and white, which could be disadvantageous relative to color photographs in the measurement of RBM. On a handful of occasions, coders expressed difficulty discerning whether an apparent facial feature (e.g., folds of adipose tissue in the neck) was real or the byproduct of shadows. If available, color photographs could help resolve those difficulties. Third, adult BMI, illness symptoms and chronic conditions were based on self-reports, which may have led to some attenuation in the associations between these outcomes and RBM. Fourth, although the sample size in this study was relatively large, it nevertheless accounted for less than one in three WLS respondents. This limited the statistical power of this investigation and will present some limitations to future investigations until a measure of RBM is available for all 10,317 participants in the WLS.

In future studies, it would be interesting to develop a measure of relative facial adiposity to complement the RBM scale. The development of a measure designed to isolate deposits of fat in the cheeks and neck from lean facial tissue would help resolve questions about gender disparities in the predictive validity of RBM, particularly in relation to mortality. Understanding the relative impact of lean, fat and total facial mass on the incidence of illness symptoms, chronic disease and mortality could help improve the understanding of disease etiology and lead to the development of new applications in research and clinical practice.

Future studies may also benefit from the inclusion of a larger and more diverse group of coders. Of course, the inclusion of additional coders would help maximize the reliability of the RBM, although it is worth reiterating that the degree of internal consistency achieved in this study was excellent. Additional coders would also provide more flexibility to retain only individuals who perform exceptionally well in terms of inter-rater reliability, intra-rater reliability and criterion-related validity. Clearly, human beings must possess varying levels of ability with regard to the visual assessment of RBM. Future studies could take advantage of this variation by identifying persons with the highest levels of ability. Furthermore, the inclusion of a more diverse set of coders with respect to age, ethnicity and socioeconomic background could help determine whether important differences exist across these groups. Limited existing research suggests that gender differences in body-image do not distort intersubjective agreement regarding the relative weight of persons in photographs (Rudin 1996), but more research is needed to answer questions regarding the effects of gender and other personal characteristics on coding tendencies.

Despite these limitations and intriguing areas for future study, the development of the RBM scale represents an important methodological development that should pave the way to an expanded set of research questions and projects in the WLS. Other longitudinal studies of adults without measures of body mass in childhood or adolescence may also benefit from development of similar RBM indicators. In such applications, the RBM scale would likely represent an exceptionally valid and reliable form of retrospective questioning, as it has in the WLS. Furthermore, emerging prospective studies of physical and psychosocial health could contribute to knowledge in this area by incorporating RBM measures. Importantly, this would provide opportunities to evaluate RBM measures contemporaneously with more traditional measures of body mass, providing more definitive evidence regarding their similarities and differences. Finally, it is encouraging to consider that while the RBM scale performed admirably in this study, it is a new construct. With additional research, refinements to RBM measures promise to unlock their full potential as refined and legitimate complements to traditional measures of body mass in human populations.

Table 4.1. Descriptive Statistics and Reliability Coefficients in the Pilot Study of WLS Yearbook Photographs

	Median Score	Mean Score	Standard Deviation	Range	Skewness	Kurtosis	Cronbach's Alpha	Correlation with Total Variation	Alpha if Coder is Removed
All Photos (n=293)							0.91		
Coder 1 (♂,33)†	6	6.02	1.69	1-10	-0.12	-0.44		0.80	0.88
Coder 2 (♂,26)	6	6.04	1.58	2-11	0.26	-0.33		0.70	0.90
Coder 3 (♀,28)	7	6.53	1.82	2-11	0.06	-0.12		0.82	0.87
Coder 4 (♀,28)	6	6.28	1.93	2-11	0.10	-0.46		0.75	0.89
Coder 5 (♀,25)	6	6.30	1.72	2-11	-0.10	-0.16		0.77	0.89
URBMI‡	6.20	6.24	1.50	2.20-10.60	0.28	-0.13			
Female Photos (n=171)							0.89		
Coder 1 (♂,33)	6	5.93	1.61	1-10	-0.32	-0.17		0.72	0.87
Coder 2 (♂,26)	6	6.02	1.61	2-11	0.03	-0.32		0.67	0.88
Coder 3 (♀,28)	7	6.43	1.63	2-10	-0.54	0.22		0.80	0.84
Coder 4 (♀,28)	6	6.36	1.75	3-11	0.16	-0.42		0.68	0.87
Coder 5 (♀,25)	6	6.05	1.68	2-11	-0.15	0.18		0.77	0.85
URBMI	6.00	6.16	1.37	2.20-10.60	0.16	0.44			
Male Photos (n=122)							0.93		
Coder 1 (♂,33)	6	6.15	1.80	3-10	0.03	-0.81		0.88	0.90
Coder 2 (♂,26)	6	6.08	1.52	4-10	0.66	-0.39		0.76	0.93
Coder 3 (♀,28)	6	6.68	2.05	3-11	0.39	-0.72		0.83	0.91
Coder 4 (♀,28)	6	6.16	2.17	2-11	0.12	-0.64		0.84	0.91
Coder 5 (♀,25)	7	6.66	1.72	3-11	-0.07	-0.65		0.80	0.92
URBMI	6.40	6.35	1.65	3.40-10.40	0.31	-0.71			

† Gender and age of coder in parentheses

‡ Unstandardized Relative Body Mass Index

Table 4.2. Criterion-Related Validity of Indexes in the Pilot Study of WLS Photographs

		Correlation Matrices						
		(1)	(2)	(3)	(4)	(5)	Mean	Standard Deviation
All Photos (n=212)								
URBMI†	(1)	1.00					6.26	1.47
SRBMI‡	(2)	1.00 **	1.00				0.02	0.84
Height	(3)	0.07	0.03	1.00			66.84	3.76
Weight	(4)	0.27 **	0.24 **	0.60 **	1.00		170.02	39.45
BMI	(5)	0.29 **	0.29 **	0.14 *	0.87 **	1.00	26.56	5.05
Female Photos (n=134)								
URBMI	(1)	1.00					6.21	1.36
SRBMI	(2)	1.00 **	1.00				0.03	0.82
Height	(3)	0.17	0.17	1.00			64.65	2.31
Weight	(4)	0.41 **	0.41 **	0.31 **	1.00		153.63	32.15
BMI	(5)	0.36 **	0.36 **	-0.03	0.94 **	1.00	25.78	5.16
Male Photos (n=78)								
URBMI	(1)	1.00					6.35	1.63
SRBMI	(2)	1.00 **	1.00				0.00	0.88
Height	(3)	-0.07	-0.07	1.00			70.58	2.64
Weight	(4)	0.14	0.14	0.40 **	1.00		198.19	34.74
BMI	(5)	0.18	0.18	-0.01	0.91 **	1.00	27.90	4.60

† Unstandardized Relative Body Mass Index

‡ Standardized Relative Body Mass Index

* $p < 0.05$, ** $p < 0.01$

Table 4.3. Descriptive Statistics and Reliability Coefficients in the Complete Study of WLS Yearbook Photographs

	Median Score	Mean Score	Standard Deviation	Range	Skewness	Kurtosis	Cronbach's Alpha	Correlation with Total Variation	Alpha if Coder is Removed	Intra-Rater Reliability§
All Photos (n=3,027)							0.91			
Coder 1 (♂,33)†	6	6.05	1.90	1-11	-0.09	-0.53		0.79	0.89	0.88
Coder 2 (♂,26)	7	6.61	1.49	2-11	-0.10	-0.41		0.70	0.90	0.71
Coder 3 (♀,28)	7	6.99	1.52	2-11	-0.43	0.31		0.80	0.89	0.80
Coder 4 (♀,28)	6	6.16	1.75	2-11	0.30	-0.56		0.77	0.89	0.85
Coder 5 (♀,25)	6	6.06	1.78	1-11	-0.09	-0.28		0.80	0.88	0.82
Coder 6 (♂,30)	7	6.99	1.37	2-11	-0.47	0.63		0.65	0.91	0.66
URBMI‡	6.33	6.48	1.37	2.33-11.00	0.17	-0.15				
Female Photos (n=1,622)							0.91			
Coder 1 (♂,33)	6	6.01	1.94	1-11	-0.22	-0.50		0.79	0.89	0.87
Coder 2 (♂,26)	7	6.77	1.45	2-11	-0.33	0.01		0.71	0.90	0.67
Coder 3 (♀,28)	7	6.91	1.46	2-11	-0.72	0.88		0.80	0.88	0.78
Coder 4 (♀,28)	6	6.28	1.74	2-11	0.12	-0.69		0.76	0.89	0.85
Coder 5 (♀,25)	6	6.02	1.89	1-11	-0.28	-0.48		0.81	0.88	0.84
Coder 6 (♂,30)	7	7.06	1.37	2-11	-0.68	1.03		0.64	0.91	0.62
URBMI	6.50	6.51	1.37	2.33-11.00	0.02	-0.14				
Male Photos (n=1,405)							0.91			
Coder 1 (♂,33)	6	6.10	1.86	1-11	0.09	-0.62		0.80	0.89	0.90
Coder 2 (♂,26)	6	6.42	1.51	2-10	0.16	-0.63		0.70	0.90	0.73
Coder 3 (♀,28)	7	7.08	1.59	2-11	-0.20	-0.29		0.81	0.89	0.82
Coder 4 (♀,28)	6	6.02	1.76	2-11	0.50	-0.30		0.78	0.89	0.86
Coder 5 (♀,25)	6	6.11	1.65	2-11	0.26	-0.11		0.80	0.89	0.81
Coder 6 (♂,30)	7	6.92	1.36	2-11	-0.24	0.26		0.66	0.91	0.70
URBMI	6.33	6.44	1.36	2.83-10.67	0.35	-0.11				

† Gender and age of coder in parentheses

‡ Unstandardized Relative Body Mass Index

§ Intra-rater reliability coefficients based on 200 photos (100 males, 100 females) randomly selected from pool of 3,027 photos

Table 4.4. Criterion-Related Validity of Indexes in the Complete Study of WLS Yearbook Photographs

		Correlation Matrices							
		(1)	(2)	(3)	(4)	(5)	(6)	Mean	Standard Deviation
All Photos (n=1,999)									
URBMI[†]	(1)	1.00						6.48	1.38
SRBMI[‡]	(2)	1.00 **	1.00					0.00	0.84
Height	(3)	-0.01	0.00	1.00				67.25	3.89
Weight	(4)	0.24 **	0.25 **	0.61 **	1.00			172.89	36.80
BMI	(5)	0.30 **	0.31 **	0.05 *	0.82 **	1.00		26.73	4.71
Self-Rated Health	(6)	-0.08 **	-0.08 **	0.02	-0.16 **	-0.21 **	1.00	4.16	0.68
Female Photos (n=1,104)									
URBMI	(1)	1.00						6.49	1.38
SRBMI	(2)	1.00 **	1.00					-0.01	0.83
Height	(3)	0.00	0.00	1.00				64.64	2.59
Weight	(4)	0.29 **	0.29 **	0.25 **	1.00			154.44	30.67
BMI	(5)	0.30 **	0.30 **	-0.18 **	0.89 **	1.00		25.99	5.21
Self-Rated Health	(6)	-0.13 **	-0.13 **	0.02	-0.21 **	-0.22 **	1.00	4.17	0.69
Male Photos (n=895)									
URBMI	(1)	1.00						6.47	1.37
SRBMI	(2)	1.00 **	1.00					0.02	0.84
Height	(3)	-0.03	-0.03	1.00				70.47	2.59
Weight	(4)	0.29 **	0.29 **	0.47 **	1.00			195.64	30.44
BMI	(5)	0.34 **	0.34 **	-0.01	0.88 **	1.00		27.65	3.82
Self-Rated Health	(6)	-0.02	-0.02	0.07 *	-0.14 **	-0.19 **	1.00	4.15	0.67

† Unstandardized Relative Body Mass Index

‡ Standardized Relative Body Mass Index

* $p < 0.05$, ** $p < 0.01$

Table 4.5. Odds Ratios for Associations Between Relative Body Mass in 1957 and Health Symptoms in 1993†

	Health Symptoms Experienced in Prior Six Months					
	Muscle Aches	Stiff or Swollen Joints	Back Pain or Strain	Chest Pain	Shortness of Breath	Respiratory Problems
All Cases (n=2,049)						
Continuous Predictor						
SRBMI‡	**1.14** (1.02-1.26)	**1.21** (1.07-1.36)	1.07 (0.96-1.19)	1.21 (0.99-1.49)	1.10 (0.94-1.29)	1.18 (0.96-1.45)
Categorical Predictors						
Underweight	0.96 (0.71-1.30)	0.74 (0.52-1.07)	0.88 (0.65-1.20)	0.99 (0.53-1.85)	1.02 (0.64-1.62)	1.25 (0.69-2.24)
Normal Weight§	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	1.07 (0.80-1.44)	0.92 (0.66-1.30)	1.06 (0.79-1.42)	0.85 (0.45-1.63)	**0.56** (0.32-0.98)	0.91 (0.48-1.74)
Overweight	**1.38** (1.03-1.84)	1.37 (1.00-1.87)	1.19 (0.89-1.60)	**2.16** (1.36-3.44)	**1.81** (1.24-2.65)	**2.40** (1.52-3.81)
Females (n=1,143)						
Continuous Predictor						
SRBMI	**1.19** (1.03-1.37)	**1.19** (1.02-1.39)	1.15 (1.00-1.33)	1.29 (0.95-1.75)	1.09 (0.87-1.35)	1.10 (0.85-1.42)
Categorical Predictors						
Underweight	0.85 (0.57-1.27)	0.85 (0.55-1.32)	1.02 (0.68-1.52)	0.84 (0.32-2.16)	0.88 (0.47-1.66)	0.99 (0.48-2.04)
Normal Weight	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	1.06 (0.72-1.56)	1.03 (0.67-1.56)	0.92 (0.62-1.38)	0.99 (0.41-2.39)	0.63 (0.31-1.28)	0.74 (0.33-1.65)
Overweight	1.45 (0.97-2.17)	1.32 (0.87-2.01)	**1.61** (1.08-2.41)	**2.32** (1.18-4.56)	1.32 (0.75-2.35)	1.75 (0.94-3.24)
Males (n=906)						
Continuous Predictor						
SRBMI	1.08 (0.92-1.27)	**1.28** (1.05-1.55)	0.98 (0.84-1.15)	1.14 (0.86-1.51)	1.12 (0.89-1.41)	1.39 (0.97-1.98)
Categorical Predictors						
Underweight	1.12 (0.70-1.79)	0.50 (0.25-1.03)	0.71 (0.43-1.16)	1.18 (0.51-2.71)	1.22 (0.62-2.42)	2.02 (0.74-5.53)
Normal Weight	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	1.08 (0.69-1.70)	0.74 (0.41-1.35)	1.26 (0.81-1.96)	0.73 (0.28-1.90)	0.46 (0.18-1.17)	1.43 (0.48-4.27)
Overweight	1.34 (0.88-2.04)	1.52 (0.95-2.45)	0.86 (0.55-1.32)	**1.98** (1.05-3.75)	**2.40** (1.43-4.03)	**4.48** (2.15-9.32)

† 95% confidence intervals (CIs) for odds ratios in parentheses; statistically significant coefficients ($p < 0.05$) are set in bold text

‡ Standardized Relative Body Mass Index

§ Normal Weight is referent category (i.e., fixed at 1.0)

Table 4.5 (continued). Odds Ratios for Associations Between RBM in 1957 and Health Symptoms in 1993†

	Health Symptoms Experienced in Prior Six Months					
	Dizziness or Faintness	Trouble Sleeping	Excess Sweating	Numbness	Ringing Ears	Nausea
All Cases (n=2,049)						
Continuous Predictor						
SRBMI‡	**1.33** (1.09-1.62)	0.97 (0.88-1.08)	**1.26** (1.07-1.49)	**1.21** (1.02-1.44)	0.96 (0.83-1.11)	0.99 (0.80-1.23)
Categorical Predictors						
Underweight	0.54 (0.26-1.12)	0.99 (0.73-1.33)	0.85 (0.51-1.42)	0.65 (0.36-1.17)	1.26 (0.83-1.89)	1.18 (0.66-2.13)
Normal Weight§	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	0.78 (0.42-1.43)	**0.69** (0.51-0.93)	0.66 (0.38-1.13)	0.66 (0.37-1.17)	0.92 (0.60-1.43)	0.78 (0.40-1.53)
Overweight	**1.71** (1.08-2.72)	1.06 (0.79-1.41)	**1.75** (1.17-2.59)	**1.71** (1.13-2.58)	1.34 (0.91-1.99)	1.36 (0.79-2.34)
Females (n=1,143)						
Continuous Predictor						
SRBMI	**1.41** (1.10-1.81)	0.99 (0.86-1.14)	1.21 (0.99-1.48)	1.15 (0.91-1.46)	0.99 (0.80-1.23)	1.00 (0.78-1.28)
Categorical Predictors						
Underweight	0.60 (0.25-1.41)	0.90 (0.61-1.32)	0.94 (0.53-1.64)	0.77 (0.38-1.58)	1.32 (0.76-2.31)	1.24 (0.65-2.38)
Normal Weight	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	0.80 (0.37-1.70)	0.61 (0.41-0.91)	0.66 (0.35-1.24)	0.49 (0.21-1.15)	0.79 (0.41-1.53)	0.78 (0.37-1.68)
Overweight	**2.34** (1.34-4.07)	0.92 (0.62-1.38)	1.58 (0.96-2.61)	1.59 (0.89-2.84)	1.44 (0.82-2.52)	1.35 (0.70-2.58)
Males (n=906)						
Continuous Predictor						
SRBMI	1.24 (0.89-1.72)	0.96 (0.82-1.14)	**1.50** (1.09-2.06)	1.28 (0.99-1.64)	0.92 (0.75-1.14)	1.03 (0.66-1.61)
Categorical Predictors						
Underweight	0.40 (0.09-1.68)	1.09 (0.67-1.77)	0.43 (0.10-1.83)	0.48 (0.17-1.35)	1.21 (0.67-2.21)	0.82 (0.19-3.58)
Normal Weight	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	0.73 (0.25-2.09)	0.79 (0.49-1.29)	0.59 (0.18-1.96)	0.90 (0.42-1.94)	1.07 (0.59-1.93)	0.73 (0.17-3.21)
Overweight	0.98 (0.40-2.38)	1.33 (0.86-2.04)	**2.50** (1.27-4.90)	**1.84** (1.03-3.30)	1.22 (0.71-2.10)	1.67 (0.61-4.57)

† 95% confidence intervals (CIs) for odds ratios in parentheses; statistically significant coefficients ($p < 0.05$) are set in bold text
‡ Standardized Relative Body Mass Index
§ Normal Weight is referent category (i.e., fixed at 1.0)

Table 4.6. Odds Ratios for Associations Between Relative Body Mass in 1957 and Health Conditions in 1993†

	Health Conditions Diagnosed by a Medical Professional					
	Arthritis	Serious Back Trouble	High Blood Pressure	Diabetes	Heart Trouble	Circulation Problems
All Cases (n=2,049)						
Continuous Predictor						
SRBMI‡	**1.19** (1.06-1.34)	1.09 (0.91-1.30)	**1.25** (1.10-1.42)	**1.44** (1.11-1.86)	1.21 (0.97-1.49)	1.07 (0.85-1.35)
Categorical Predictors						
Underweight	0.82 (0.57-1.17)	0.78 (0.45-1.36)	0.64 (0.43-0.98)	1.06 (0.48-2.39)	1.04 (0.56-1.95)	0.51 (0.22-1.19)
Normal Weight§	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	0.95 (0.68-1.33)	0.95 (0.58-1.58)	1.03 (0.72-1.47)	**1.77** (0.92-3.39)	0.64 (0.31-1.35)	0.48 (0.21-1.12)
Overweight	1.17 (0.85-1.62)	0.94 (0.57-1.56)	1.23 (0.88-1.72)	**2.38** (1.33-4.28)	**1.79** (1.08-2.95)	0.99 (0.53-1.84)
Females (n=1,143)						
Continuous Predictor						
SRBMI	**1.31** (1.13-1.53)	1.15 (0.90-1.47)	**1.44** (1.20-1.73)	**1.70** (1.17-2.47)	1.15 (0.79-1.66)	0.99 (0.73-1.34)
Categorical Predictors						
Underweight	0.77 (0.49-1.20)	0.99 (0.49-1.97)	**0.47** (0.24-0.89)	1.27 (0.43-3.74)	0.64 (0.19-2.12)	0.65 (0.25-1.66)
Normal Weight	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	1.09 (0.72-1.65)	0.97 (0.48-1.93)	1.08 (0.66-1.76)	1.90 (0.75-4.78)	0.63 (0.19-2.08)	0.25 (0.06-1.03)
Overweight	**1.58** (1.05-2.39)	0.95 (0.46-1.96)	1.48 (0.93-2.38)	**3.61** (1.66-7.85)	1.17 (0.45-3.06)	0.70 (0.27-1.79)
Males (n=906)						
Continuous Predictor						
SRBMI	1.04 (0.85-1.27)	1.03 (0.79-1.32)	1.09 (0.91-1.31)	1.22 (0.85-1.76)	1.22 (0.94-1.59)	1.20 (0.84-1.72)
Categorical Predictors						
Underweight	0.88 (0.48-1.61)	0.56 (0.22-1.42)	0.87 (0.50-1.52)	0.90 (0.27-3.04)	1.40 (0.66-2.95)	0.24 (0.03-1.80)
Normal Weight	1.00	1.00	1.00	1.00	1.00	1.00
Risky Weight	0.71 (0.38-1.32)	0.94 (0.45-1.96)	0.99 (0.59-1.65)	1.67 (0.67-4.17)	0.66 (0.26-1.69)	0.90 (0.31-2.62)
Overweight	0.78 (0.44-1.38)	0.92 (0.46-1.85)	0.99 (0.61-1.61)	1.46 (0.59-3.64)	**2.08** (1.14-3.80)	1.43 (0.61-3.34)

† 95% confidence intervals (CIs) for odds ratios in parentheses; statistically significant coefficients ($p < 0.05$) are set in bold text

‡ Standardized Relative Body Mass Index

§ Normal Weight is referent category (i.e., fixed at 1.0)

Table 4.6 (continued). Odds Ratios for Associations Between RBM in 1957 and Health Conditions in 1993†

	Conditions Diagnosed by a Medical Professional											
	Cancer		Asthma		Kidney or Bladder Problems		Ulcer		Allergies		Two or More Conditions	
All Cases (n=2,049)												
Continuous Predictor												
SRBMI‡	1.16	(0.83-1.62)	1.13	(0.88-1.44)	1.06	(0.84-1.34)	1.10	(0.82-1.46)	0.98	(0.84-1.13)	**1.25**	(1.10-1.42)
Categorical Predictors												
Underweight	1.38	(0.57-3.34)	1.51	(0.80-2.87)	0.78	(0.37-1.64)	0.54	(0.19-1.51)	1.07	(0.71-1.6)	0.69	(0.45-1.04)
Normal Weight§	1.00		1.00		1.00		1.00		1.00		1.00	
Risky Weight	1.08	(0.42-2.80)	0.81	(0.37-1.80)	1.03	(0.54-1.97)	0.77	(0.33-1.82)	0.79	(0.51-1.22)	0.89	(0.61-1.29)
Overweight	1.29	(0.53-3.12)	**1.92**	(1.08-3.41)	0.82	(0.41-1.67)	0.77	(0.33-1.81)	0.92	(0.61-1.39)	1.28	(0.91-1.80)
Females (n=1,143)												
Continuous Predictor												
SRBMI	1.31	(0.86-2.00)	1.20	(0.88-1.64)	1.16	(0.85-1.57)	**1.47**	(1.02-2.11)	1.02	(0.85-1.21)	**1.36**	(1.14-1.62)
Categorical Predictors												
Underweight	0.94	(0.28-3.19)	0.99	(0.41-2.37)	0.57	(0.2-1.62)	0.19	(0.03-1.4)	0.79	(0.47-1.32)	0.52	(0.29-0.93)
Normal Weight	1.00		1.00		1.00		1.00		1.00		1.00	
Risky Weight	0.92	(0.27-3.14)	0.63	(0.22-1.8)	1.33	(0.63-2.78)	0.97	(0.37-2.52)	0.64	(0.37-1.1)	0.77	(0.46-1.28)
Overweight	1.72	(0.64-4.64)	1.64	(0.78-3.48)	0.62	(0.22-1.75)	0.84	(0.29-2.42)	0.92	(0.55-1.52)	1.38	(0.88-2.19)
Males (n=906)												
Continuous Predictor												
SRBMI	0.92	(0.51-1.66)	1.02	(0.68-1.55)	0.93	(0.63-1.37)	0.64	(0.38-1.08)	0.91	(0.7-1.2)	1.13	(0.93-1.37)
Categorical Predictors												
Underweight	2.40	(0.65-8.89)	**2.90**	(1.11-7.58)	1.17	(0.40-3.42)	1.31	(0.38-4.56)	1.90	(0.97-3.74)	0.97	(0.54-1.76)
Normal Weight	1.00		1.00		1.00		1.00		1.00		1.00	
Risky Weight	1.42	(0.31-6.57)	1.25	(0.36-4.35)	0.51	(0.12-2.18)	0.38	(0.05-2.91)	1.22	(0.58-2.56)	1.07	(0.62-1.86)
Overweight	0.62	(0.08-4.87)	**2.67**	(1.08-6.61)	1.16	(0.44-3.08)	0.68	(0.16-2.99)	1.06	(0.51-2.22)	1.18	(0.71-1.96)

† 95% confidence intervals (CIs) for odds ratios in parentheses; statistically significant coefficients ($p < 0.05$) are set in bold text

‡ Standardized Relative Body Mass Index

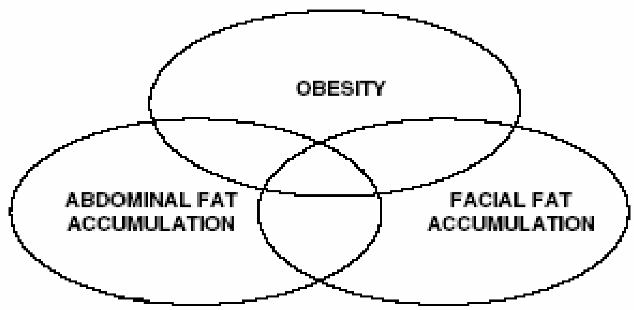
§ Normal Weight is referent category (i.e., fixed at 1.0)

Table 4.7. Odds Ratios for Associations Between Relative Body Mass in 1957 and Mortality Indicators from the Social Security Death Index (SSDI) and the National Death Index (NDI)†

	Mortality Indicator				
	SSDI All Cause	NDI All Cause	NDI Non-Accidental	NDI Heart Disease	NDI Cancer
All Cases (n=3,027)					
Continuous Predictor					
SRBMI‡	1.11 (0.95-1.30)	**1.25** (1.03-1.51)	**1.32** (1.08-1.61)	**1.69** (1.17-2.42)	1.16 (0.88-1.54)
Categorical Predictors					
Underweight	0.97 (0.61-1.54)	0.90 (0.50-1.62)	0.86 (0.45-1.63)	0.98 (0.29-3.29)	0.88 (0.37-2.07)
Normal Weight§	1.00	1.00	1.00	1.00	1.00
Risky Weight	1.31 (0.87-1.98)	1.22 (0.73-2.05)	1.38 (0.82-2.32)	1.28 (0.44-3.75)	1.01 (0.45-2.25)
Overweight	1.18 (0.77-1.81)	**1.69** (1.06-2.68)	**1.91** (1.20-3.04)	**3.99** (1.95-8.15)	1.49 (0.75-2.98)
Females (n=1,622)					
Continuous Predictor					
SRBMI	**1.27** (1.00-1.60)	**1.44** (1.10-1.88)	**1.47** (1.12-1.94)	**2.11** (1.08-4.11)	1.31 (0.93-1.84)
Categorical Predictors					
Underweight	0.84 (0.41-1.72)	0.96 (0.43-2.15)	0.86 (0.36-2.04)	•	0.59 (0.18-1.95)
Normal Weight	1.00	1.00	1.00	1.00	1.00
Risky Weight	1.42 (0.80-2.55)	1.22 (0.59-2.53)	1.27 (0.61-2.64)	•	1.17 (0.48-2.83)
Overweight	**1.99** (1.15-3.44)	**2.78** (1.56-4.97)	**2.90** (1.62-5.19)	•	1.89 (0.86-4.16)
Males (n=1,405)					
Continuous Predictor					
SRBMI	0.99 (0.80-1.23)	1.08 (0.82-1.42)	1.17 (0.88-1.57)	1.52 (0.99-2.33)	0.89 (0.53-1.51)
Categorical Predictors					
Underweight	1.12 (0.61-2.07)	0.85 (0.36-2.02)	0.87 (0.34-2.23)	0.43 (0.06-3.27)	1.56 (0.45-5.47)
Normal Weight	1.00	1.00	1.00	1.00	1.00
Risky Weight	1.25 (0.70-2.23)	1.25 (0.60-2.59)	1.52 (0.73-3.19)	1.69 (0.56-5.08)	0.51 (0.07-3.88)
Overweight	0.63 (0.31-1.28)	0.85 (0.38-1.89)	1.03 (0.46-2.34)	1.85 (0.68-5.05)	0.89 (0.20-3.92)

† 95% confidence intervals (CIs) in parentheses; statistically significant coefficients (p <0.05) set in bold text
‡ Standardized Relative Body Mass Index. § Normal Weight is referent category (i.e., fixed at 1.0)
• Maximum likelihood estimate may not exist. Model coefficients not reported.

Figure 4.1. Venn Diagram of the Hypothetical Relationships Between Obesity and Accumulations of Fat in the Face and Abdomen†



† Figure reproduced with the permission of Medical Hypotheses and Bruce Johnson.

Figure 4.2. Model of Relative Body Mass Scale Used to Code Male Photographs from WLS Yearbooks

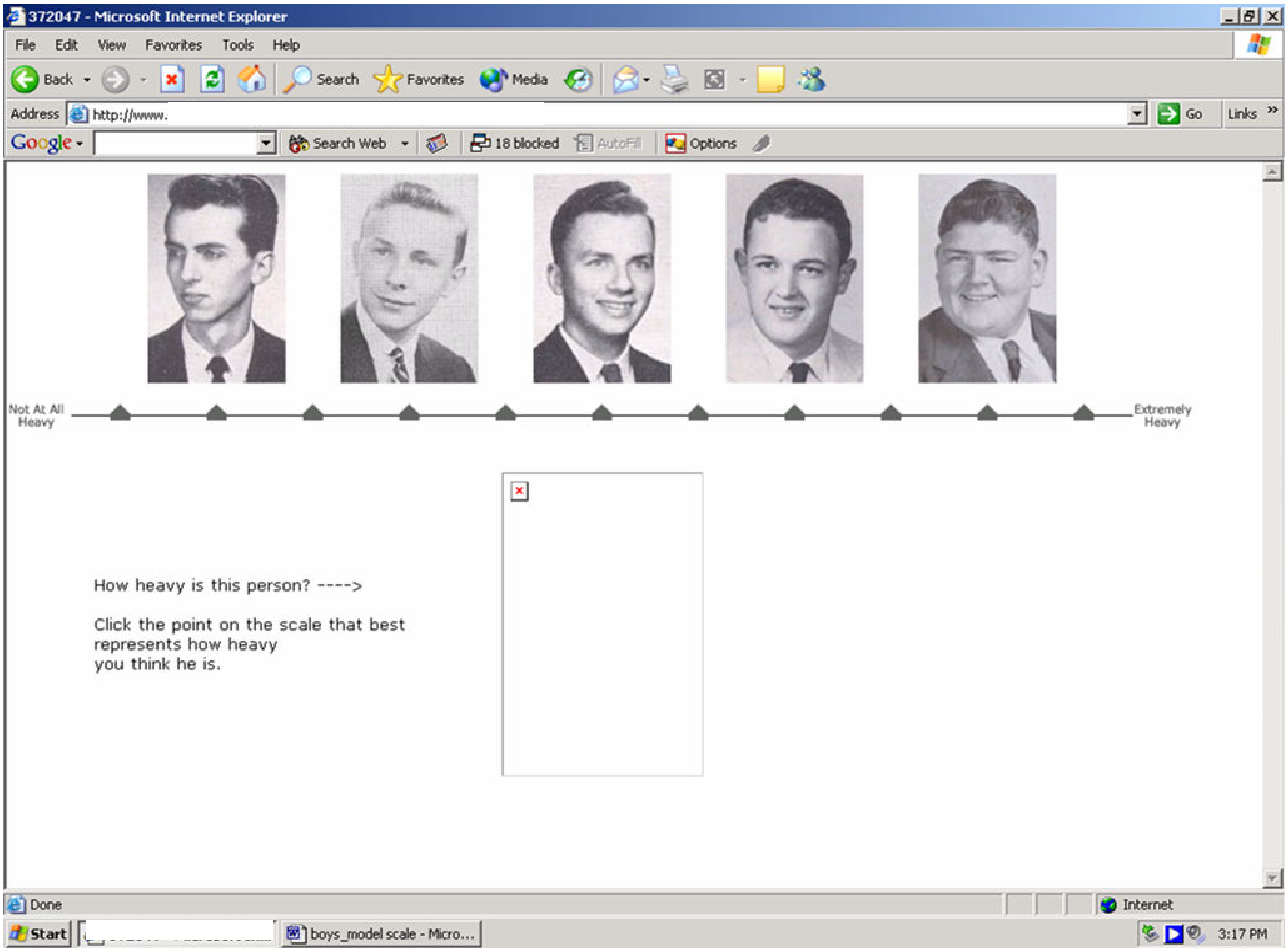


Figure 4.3. Model of Relative Body Mass Scale Used to Code Female Photographs from WLS Yearbooks

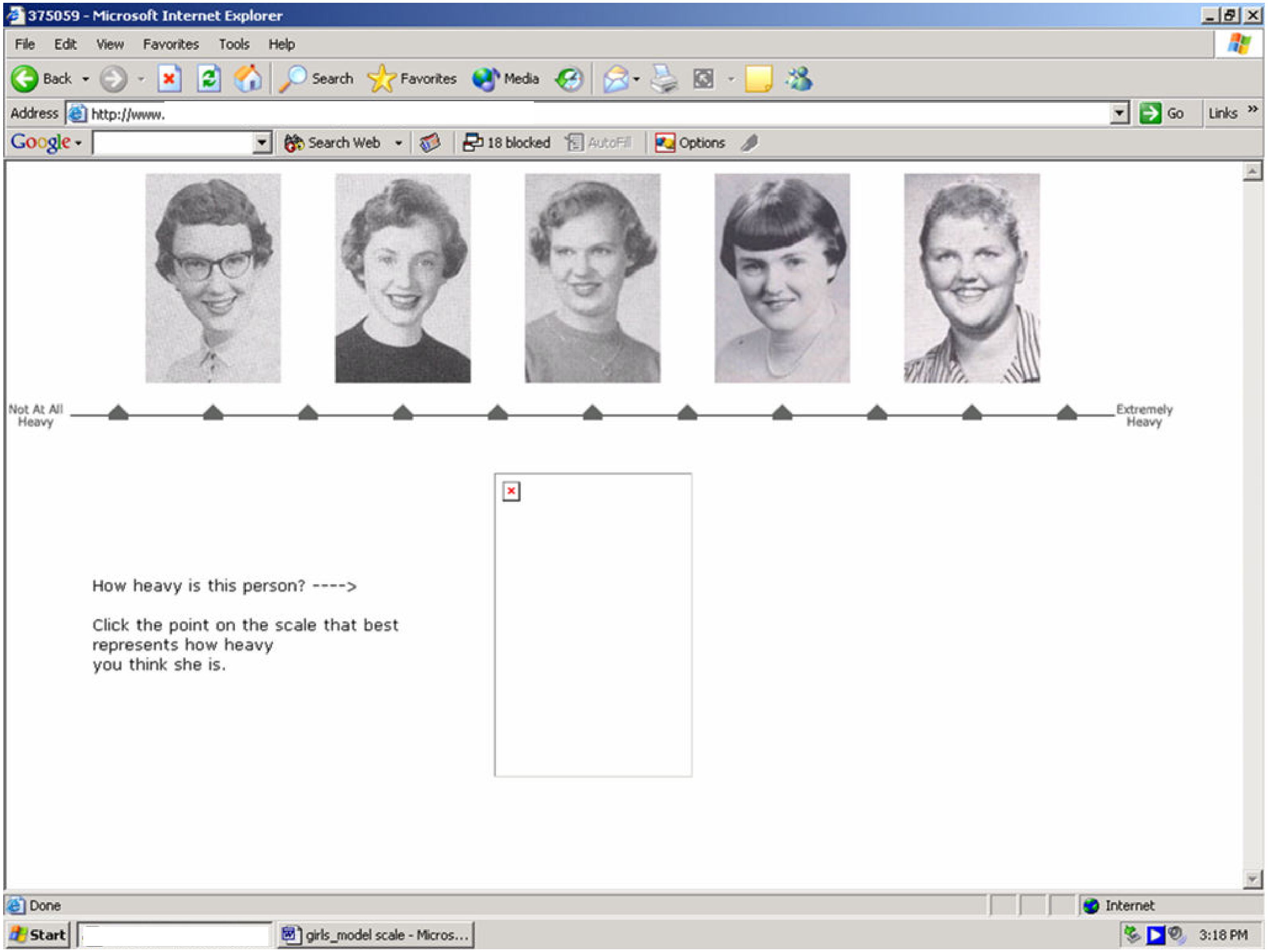


Figure 4.4. Scatterplot of the Unstandardized Relative Body Mass Index (URBMI) from the Pilot Study and Body Mass Index (BMI) in 1992