Center for Demography and Ecology

University of Wisconsin-Madison

The Accuracy of Self-Reported Anthropometry: Obesity among Older Mexicans

Alberto Palloni

Beth Soldo

Rebeca Wong

CDE Working Paper No. 2004-23

The Accuracy of Self-Reported Anthropometry: Obesity among Older Mexicans

Alberto Palloni (*)
Center for Demography and Ecology
University of Wisconsin

Beth Soldo (*)
Population Studies Center
University of Pennsylvania

Rebeca Wong (*)
Population Research Center
University of Maryland

Paper presented at the Population Association of America Meetings, Minneapolis: May 1-3, 2003.

(*) Authors' names are listed in strict alphabetical order

ABSTRACT

Recent surveys of older adults include batteries of questions or modules on self-reported chronic conditions as well as on limited self-reported anthropometry. Experience with such surveys in developed countries shows that some self-reported conditions possess reasonably high validity. There is much less information on the accuracy of self-reported anthropometry. In developing countries these problems are virtually unexplored. This is a problematic gap in our knowledge since no less than ten different surveys are currently in the field eliciting information on these characteristics.

In this paper we use a new data set to explore the accuracy of self-reported height and weight in a sample of older adults in Mexico. In this survey (MHAS), administered to a nationally representative sample of older adults fifty and over, actual measures of body weight and stature were collected for a sub-sample jointly with self-reported weight and height.

Our analyses probe the following four issues: (a) the degree of concordance between self-report and objective measures; (b) individual determinants of discordance (c) biases in estimates of determinants of obesity when assessed from self-reported height and weight, (d) biases in equations assessing the relation between obesity evaluated through self-reported height and weight and self-reported diabetes.

1. Introduction

Studies of elderly populations rely heavily on self-reported measures of health status and chronic conditions as well as on selected self-reported anthropometry, such as height and weight. Although self-reported health status has proven to be of considerable use, there are lingering doubts about how exactly we ought to interpret the quantities we obtain for a particular sample and, especially, how should we interpret contrasts across populations.

There is considerably more skepticism about the utility of self-reported chronic conditions and of anthropometry. This is quite problematic, as self-reports are often more practical and considerably less expensive to obtain than clinical assessments and physical anthropometry. Simple anthropometry (height and weight) is normally used to calculate Body Mass Index (weight/height^2), a useful indicator of population prevalence of obesity and as a predictor of diabetes. Survey participants' self-reported measures of selected chronic conditions are a keystone on which researchers base at least provisional descriptions of prevalence of chronic ailments, such as diabetes, hypertension, cardiovascular and circulatory diseases, cancers and the like. Thus, it is important to evaluate the precision (validity and reliability) of both self-reported chronic conditions and anthropometry. Such evaluations will give us an increasingly more accurate idea about the bounds of uncertainty one faces when relying on self-reported measures

The precision of anthropometry has been evaluated in a handful of studies (Stewart, 1982; Palta et al., 1982; Charney et al., 1976; Wing et al., 1979) whereas precious little research has been implemented on the degree of precision of assessments of self-reported chronic conditions, both in clinical and population-based settings (Vargas et al., 1997; Bush et al., 1989; Johansson et al., 1999; Kehoe et al., 1994; Torno et al., 2000; Martin et al., 2000; Harlow and Linet, 1989). In both cases, however, the assessment studies share one or more undesirable characteristics, including small sample sizes, potentially selected samples (e.g. reduced to volunteers or to individuals who are below or above some health threshold), and lack of overall representation. Most problematic of all, though is that, with a few exceptions (Goldman et al., 2002; Beckett et al., 2000), these studies take place in developed countries and with samples that do not always include any or an adequate representation of elderly people.

In this paper we provide a preliminary assessment of the reliability and validity of anthropometric reports (height and weight) using a nationally representative population of adults older than 50 who resided in Mexico during the year 2000-01. We also probe into uncharted territory and evaluate potential biases associated with the use of BMI from self-reported height and weight as predictor of obesity and self-reported diabetes.

2. Model and tests

In what follows we briefly describe a standard model to represent the relations between self-reports and true measures of the quantity of interest. We also describe a series of tests we apply to the information available to identify potential consequences of lack of precision of self-reports.

2.1. Basic model

We assume that a self-report, S(Y), is a measure of an underlying trait, y, that may be subject to a systematic error, **7**, as well as to random variation, **,**,

$$S(Y) = Y + 7 + ,$$

It is conventionally assumed that the random term has expectation 0 and is uncorrelated with both the bias and the true value of the trait. The expectation of the quantity **7** is called the bias of S(Y) and the difference (S(Y)-Y) is a measure of the degree of validity of S(Y). In turn, the variance of the error term is a measure of the reliability of S(Y).

In what follows we use the observed difference (S(Y)-Y) as the main object of analysis and we will refer to it as *****. The quantity S(Y) will be self-reported weight or height. In the case of physical anthropometry, the quantity Y is associated with physical measurements of height and weight.

It is, of course, possible that what we construe as the "true value of the trait", Y, may well be a measure which itself is subject to error:

$$Y = Y^{*} + >$$

so that

$$* = (S(Y)-Y) = ((S(y)-Y^{*}) = > + 7 + ,$$

Thus, > may be confounded with other error components. If > has a 0 expectation and is uncorrelated with the bias and the random component, it will not contaminate our measure of error, *. The only potential damage will be to increase the variance of the random component and to lead to underestimates of the reliability of the measure S(Y). If instead, > it has a non-zero expectation its final impact will depend on whether it helps to offset or to compound any bias that S(Y) may contain.

It is well known that, at least in linear models, measurement errors in a variable will cause no bias in estimated effect parameters (but will increase their variance) if the variable is used as a dependent variable. When the variable is used as a predictor it will induce attenuation of estimated coefficients. Except under very special conditions, measurement errors will almost always bias estimated effects in more general, non-linear, models. Thus, the existence of far-from perfect measurement precision (validity/reliability) not only leads to errors in measures of central tendency and prevalence but also, and more importantly, generates biases when such quantities are part of more general models.

2.2. Tests

In all cases we will estimate the mean and variance of * and implement appropriate statistical tests to determine if the assumption of a 0 mean is warranted in the total population and in selected subpopulations. We will also use the correlation coefficient between S(Y) and Y as an alternative metric of validity (Stewart, 1982).

To assess reliability we use the variance of *, Var(*). This measure is not a true measure of reliability since, as pointed out before, the value of Y may itself be subject to error. Furthermore, except in cases in which we can compare the measure of reliability in our sample with estimates of reliability retrieved from other samples and studies, there is no obvious baseline to judge the magnitude of the variance of the error term in any subgroup. For this reason, we compare Var(*) across selected subgroups to identify potential determinants of reliability. Comparisons across groups provide an opportunity for judging relative reliability and to identify its most relevant determinants.

We also implement a procedure to identify pitfalls in statistical inference that may be facilitated when using S(Y) (instead of Y) as a dependent or independent variable in selected models of interest. In order to do this we proceed with a three pronged strategy. We first estimate models to identify determinants (say Z) of biases and poor reliability; we then compare estimation of models using S(Y) and Y and, finally, we estimate simple models to produce improved estimates of Y from the observed values S(Y) and appropriate controls identified in the first step (Z).

3. The accuracy of self-reported weight and height

We begin by assessing validity and reliability of self-reported height and weight. While these two measures can be and are indeed used separately, they are more commonly combined to form the so-called Body Mass Index, computed as the ratio of the weight (in kilograms) to the square of the height (in meters). To evaluate precision of self-reported weight and height we use MHAS, a nationally representative study of the Mexican elderly (aged 50+) which includes both self-reports and direct measures of height and weight.

3.1. Population and data

The data are from the first wave of the new Mexican Health and Aging Study (MHAS). MHAS is a nationally representative, prospective panel study of Mexicans aged 50 and over as of 2000. Interviews were sought with spouse/partners of sampled persons regardless of their own age. Data were collected on multiple domains of health; demographic history, including the migration history of respondents, their parents and offspring; family support networks and transfers exchanged; aspects of work history; income by source; assets, including home ownership, business, real estate, capital assets and pensions; and aspects of the built environment. The core interview is about 80 minutes in length. Six Mexican states with high rates of out-migration to the U.S. were over sampled (with a ratio of 1.7 to 1).

The first wave of MHAS was successfully fielded during June-September 2001 in Mexico by the INEGI (*Instituto Nacional de Estadística, Geografía e Informática),* the equivalent of the U.S. Census Bureau. The sampling frame for MHAS was the household listing of about 136,000 dwelling units from the 4th Quarter of the ENE-2000 (National Employment Survey), also fielded by INEGI. Hence, the weighted first wave MHAS data are both a baseline for the MHAS panel and a representative cross-section of the approximately 13 million Mexicans aged 50 or older. Baseline interviews were completed in 9,845 households with about 15,000 respondents. By U.S. standards, the individual non-response rate of 10.5 percent for a population-based survey is very low. MHAS data have recently become public, and can be accessed at www.pop.upenn.edu/mhas.

The health measures included: self-reports of conditions, symptoms, functional status, life-style behaviors (e.g., smoking, drinking alcohol, exercising); the use, source, and expenditures in health care, depression, pain, reading and cognitive performances. In addition, individuals were requested to answer questions regarding their height and weight. The questions were of the following form: "Approximately what is your height without shoes?" and "Approximately how much do you weigh now?").

In a 20% randomly selected sub-sample (n=2332) investigators obtained direct measures of weight, height, waist and hip circumference, knee length, calf circumference, and timed one-leg stands. Height was evaluated by having the subject remove his/her shoes. Then, with the help of a standard stadiometer and while the individual stood as straight as possible against a wall, a tape measurement was taken. Weight was measured using standard, non-digital scales that were repeatedly calibrated to obtain maximum reliability. Height was measured in centimeters and weight in Kilograms rounded to the nearest 100th gram. Interviewers were specially trained to utilize both pieces of equipment by a crew of specialized technicians in a geriatric clinic in Mexico City. Although no study of validity and reliability of measures were carried out prior to the interview, application of the instruments during training sessions and experience with them in other settings suggests very high reliability and validity.

In what follows we take the measurements of weight and height retrieved as part of the anthropometry as our ***baseline*** measures, the "***true***" values of the underlying traits or the value of Y. The values of δ associated with weight and height will be referred to as δ-weight and δ-height respectively. The associated value for BMI will be referred to as δ-BMI.

3.2. Results

Table 1 shows basic statistics for the MHAS sub-sample we study. This sample consists of 2,332 individuals with characteristics displayed in the first panel of the table. Among the statistics displayed in this panel we include a few of the covariates we will use throughout our analysis. These are defined as follows: sex (males/females), age categories, 50-59, 60-69 and 70+; and education categories, 0 years of education, 1-6, and 7+. The mean baseline weight in the sample is 68.5 Kgs and the mean baseline height is 1.58 mts. These are remarkably close to the mean self-reported weight (69.8Kgs) and height (1.61mts). Unlike what normally occurs in younger populations, there is a slight overestimation of weight. Instead, as verified in other studies, height is slightly overestimated (see Table 1, panels a and b)

Table 1 about here

<u>Internal checks: threats to validity of our study.</u> Of the total original sample selected for physical measurements 107 individuals were unable or unwilling to submit to one or more of the protocols. Of these, 8 had missing weight only, 8 had missing height only, and 91 were missing both weight and height. Simple statistics (not shown) indicate that the small group of individuals for whom we have no baseline measures of either weight or height (n=107), are no different from the rest in terms of age, gender, education, or self-reported health status. Except when otherwise noted, our analyses use either the subset with valid height (or weight) or the subset with only valid measures on both baselines (e.g. n=2,225 individuals) (see Table 1, panel c).

Among those with valid measures on both baselines there are 487 (about 22 percent) with no self-reported height (but with available self-reported weight), 44 (about 2 percent) with no self-reported weight (but with available self-reported height), and 202 (about 9 percent) with neither self-report. The number of subjects in this target sub-sample with valid measures on both self-reported and baseline weight, e.g. where a *****-weight is computable, is 1,983. The number of subjects in the sub-sample with a computable *-height is 1,540. Thus, the attrition in the sample with available baseline measures on height and weight is about 11 percent due to missing self-reported weight and 31 percent due to missing self-reported height. The levels of attrition are only slightly different if, instead of using this highly restrictive sample, we used samples with an available baseline in the dimension of interest. Thus, there are 2,233 individuals with valid measures of weight (but not necessarily of height) and of these 250 (about 11.1 percent) are missing the corresponding self-reported weight. In turn, of the 2,233 individuals with valid measures of height (but not necessarily of weight) there are 693, or 31 percent, with no self-reported height.

Finally, these levels of attrition are almost identical to those in the larger sample. . Indeed, about 10 percent of the overall samples have missing self-reports on weight and close to 24 percent have a missing self-report on height. That is, there is no higher propensity to omit self-reports on height or weight in the sub-sample with physical or baseline measurements than there is in the total sample. Yet, no matter how one calculates them, these are non-trivial levels of attrition and deserve some scrutiny. Do they pose a threat to the validity of our analyses?

In a search for at least indications that they do, we first estimate a model for the probability of not providing a self-report in either weight or height. In all three possible cases (only weight missing, only height missing and both missing) the probability of a missing value is higher at younger ages, among those who are less educated, and among females. The effects of the corresponding variables are all significant and the fit of the models is good (results not shown). But this by itself does not tell us anything at all about whether or not individuals who must be discarded from the sub-samples because they lack self-reports are more or less likely to have been drawn from the population that tends to err more (or less) relative to baseline measures.

A second check is to study the association between combinations of missing variables. Table 1 (panel d) displays three measures to assess the association between missing values on self-reports and baseline measures. There is a strong association between not having baseline height and baseline weight (first row of the last panel of Table 1). This is expected, because the majority of individuals for whom interviewers could not retrieve anthropometric measures had physical disabilities and limitations.

There is also a strong association between not having a self-report on height and on weight (second line of last panel in Table 1). This is also an expected result for an important contributor to a missing self-report is lack of cooperation on the part of the respondent or a complete lack of knowledge about the target characteristic. The third through fifth lines of the table show that there is also a correlation between missing values in baseline measures and self-reports for both weight and height. That is, individuals who do not self-report weight (height) are also more likely to be among those for whom we do not have a baseline measure of weight. These last three sets of association measures, which assess cross-instrument correlations, are lower than those that apply to the same instrument (first two lines of the panel).

That the estimated associations are all considerable (and statistically significant at very conservative levels of significance in all cases) suggests that common conditions may have led to missing information on baseline, on the one hand, and self-reports, on the other. They also suggest that there must be conditions that increase the risk of missing information on both types of measures (and lack of cooperation of respondent may be one of them). But, as before, the evidence cannot be interpreted to suggest that the subpopulations with missing values in either physical measures or self-reports is more (less) likely to produce errors (delta) of higher magnitudes than all other individuals.

A more revealing check is to determine if those individuals who missed a self-response in **only one dimension** provide a self-report containing higher error in the other dimension. Suppose that it is true that individuals with missing self-reports are also more likely to be the ones who, if self-reports had been obtained, would have produced the largest errors or * values. It should then be the case that individuals who have at most one missing self-report must display errors in the non-missing self-report that are of higher magnitude than those in the general population. This is not so: the mean error in self-reported weight in the sub-sample is of about .42 kgs with a standard deviation of 4.90 kgs. Among those who do not have a self-report for height the mean error in weight is *lower*, .36 kgs, with a larger standard deviation, 5.45 kgs. The same pattern is found for height. The mean error in the total sample is 1.73 cms with a standard deviation of 4.85 cms. But among those who do not provide a self-report of their weight, the error in height is of smaller magnitude, 1.71 cms, with a larger standard deviation, 5.84 cms. This is **not** the pattern we would expect if those individuals who omit a self-report were more prone to errors than those who supply the information.

Our conclusion is that while attrition due to missing observations in the sub-sample is considerably higher than what would be desirable, especially for the analysis of height, it poses no obvious threat to internal validity though it could affect estimates of the variance of each of the * values we study.

Examination of estimates: the case of weight. Table 2 displays the estimated mean and standard errors of *-weight for the total sample and for selected sub-groups. In addition, the table also displays a 99% confidence interval and the R-squared between self-reports and baseline measures. In all cases, the sample sizes correspond to the number of subjects used in the calculations (e.g., with valid self-reports and baseline on the dimension of interest). The first feature worth noting is that the errors are all positive, namely, that individuals tend to self-report higher weight than what is recorded in the baseline, and that men do more so than women. This is the opposite of what has been found in younger populations in the US (Stewart, 1982; Palta et al., 1982), where invariably individuals underestimate their weight and women do more so than males. That there is a systematic bias in the overall sample and in some subgroups is indicated by the fact that the very conservative 99% confidence interval does not contain the value 0. The biases are more salient for males and for those in the oldest age group. They are less so for the other subgroups. Surprisingly, lower levels of education do not appear to be associated with the magnitude or even the existence of a bias.

An important feature in the table is that the correlation coefficient between self-reported and baseline weight is almost always above .93. This suggests a high level of validity but not as high as levels obtained with younger samples in developed countries, where the correlation coefficients invariably exceed .97.

The variance of *-weight is a good measure of the reliability of self-reports. The absolute levels of the measure are difficult to judge without comparing them to a standard but can be compared across groups. Tests of equality of variances suggest that females provide significantly more reliable self-reports than males, that younger people have much higher reliability than the ones in the oldest groups and that highly educated individuals supply more reliable information than do those with lower education.

One important source of errors affecting self-reports is a tendency to declare weights ending in 0 or 5 (Stewart, 1982). That this is the case is shown in Figures 1a and 1b for males and females respectively. In fact, estimated values for Whipple's index suggest heavy digit preference as the index values for the total population as well as all subgroups exceed 150 for all subgroups considered here. Values of this magnitude or higher are considered to reflect considerable heaping around the preferred digits. Note that in the total sample with valid self-reported weight, about 40 percent report a digit ending in 0 or 5, when one would expect no more than 20 percent would do so. The observed proportions self-reporting a weight ending in 0 or 5 are displayed in the last column of Table 2. The contrasts in digit preferences are sharper for males and females and, more significantly, across levels of education.

Table 2 about here

So far, our assessment of validity measures has proceeded through examination of univariate tables. Patterns are quite different in a multivariate framework. Table 3 (panel a) reports results from a simple OLS model where *****-weight is defined as a function of gender, age, education and digit preference. These results suggest that, consistent with previous research, females tend to underestimate their weight more than males but differences are not statistically significant. As expected, older individuals overestimate their weight more than younger individuals do but, here again, differences are not significant. Differences by education appear to be unimportant. Much more important is self-reporting a weight that ends in 5 or 0 which is associated with heavy underestimation of baseline weight. This is a result that mimics others obtained with US samples (Stewart, 1982) . Finally, note that the effects of baseline weight are fairly large and statistically significant: the heavier an individual is the higher the magnitude of the underestimation in self-reports.

Table 3 about here

Can one predict the "true" value of weight from self-reports? If a robust relation could be estimated, it would be possible to adjust self-reports and, presumably, reduce the variance of errors. That this is possible is shown in Table 3 (panel b) where we display OLS estimates of regression equations predicting baseline weight as a function of self-reported weight while controlling for gender, age, and expressed digit preference (for 0 or 5). Both equations fit the data very well as the proportion of explained variance exceeds .88. The factor with the heaviest loading is self-reported weight but none of the other estimated regression coefficients is statistically significant from 0. This suggests that the univariate contrasts reviewed in Table 2 should not be taken at face value. The only significant effect is associated with preference for digit 0. The signs (positive in both cases) of the estimated effects indicate that those who round toward 0 and 5 tend to *underestimate* their weight (e.g., they are more likely to be drawn from among those whose weight is higher than the self-reported one). However, only the estimated effect associated with the dummy variable indicating a self-reported a weight ending in 0 leads to a statistically significant coefficient.

Examination of estimates: the case of height.

Table 4 displays basic statistics for δ-height. As occurred in the case of weight, there is a tendency to overestimate height relative to the baseline measures, the bias is noticeable in all groups though it is more powerful among females than among males, increases with age and decreases with level of education. These patterns are all consistent with those observed in other, younger populations in the US. That there is a systematic upward bias in self-reported height is revealed by the fact that none of the 99% confidence intervals include the value 0 (absence of bias). Note also that the validity coefficients are all below .79, and that they in all cases they are much lower than the validity coefficients associated with weight.

Preference for digits ending in 0 or 5 is pervasive and much more apparent than was the case for weight. Note that the proportion of individuals self-reporting heights ending in 0 or 5 exceed .40 when one would expect this figure to be around .20. Figures 2a and 2b display absolute frequencies of several heights and show characteristic spikes at ages ending in 0 and 5. Not surprisingly, the standard errors are fairly large, and reliability correspondingly smaller, particularly when the measure of dispersion is expressed as a ratio to the mean value of δ (coefficient of variation) There are important differences in reliability across all subgroups, but none is as marked as those that characterize age groups and educational categories.

In summary, recording of height through self-reports leads to higher biases and lower reliability than in the case of self-reported weight. This is consistent with other research that pertains to younger populations. Also, the fact that the upward bias increases and reliability decreases as age advances is in keeping with the idea that individuals tend to shrink as they age (Palta, et al., 1982).

As we did in the case of weight, we estimate two different regressions. The first includes δ-height as a function of a number of characteristics, including choice of a digit ending in 0 or 5. The results of this regression appear in Table 5 (panel a). The equation does not fit as well as the analogous one for weight but does provide an indication about the nature of the determinants of errors. First, the effects of being a female is to underestimate height (effect is significant at $p<.01$) while those of older age are in the opposite direction. Second, although at higher levels of education underestimation increases, the effects are not significant. Choice of a digit ending in 0 or 5 does not seem to be associated with a bias as the effect is negative but not significant. Finally, people who are taller tend to underestimate their height by a fair amount and the associated effects are highly significant.

Panel b displays estimates of a regression where the baseline value of height is predicted using self-reported height while controlling for a number of factors, including choice of a preferred digit. While the fit of this equation is poorer than that for weight (R-squared is about .80), it is still remarkably tight. This equation can be used to produce adjusted values of self-reports that are less subject to bias and that remove some of the observed irregularities, including digit preference.

4. The assessment of biases in estimated relations: BMI and self-reported diabetes

We pose now two different questions. The first regards the accuracy and reliability of a BMI, a derived measure that requires weight and height as inputs. The second regards the estimation of the effects of BMI on self-reported diabetes.

4.1. Is BMI affected by errors in self-reported height and weight?

To be consistent with the treatment of height and weight, Table 6 displays basic statistics associated with δ-BMI, the error measure associated with BMI ($weight/height^2$.)

First, the value of BMI derived from self-reports contains a negative bias. In fact, self-reported weight and height, contaminated by the inaccuracies described before, lead to a downward bias relative to baseline BMI, e.g., as calculated from physical measures. Second, the bias is lower for males than it is for females, is lowest for the very old and for the well educated. The reliability of BMI differs across subgroups but the most important contrast is one across education groups as reliability is poorest in the group with the lowest education. In all cases, the validity coefficient exceeds .84 but the values vary in a very narrow range, between .84 and .91.

Table 7 displays results for the OLS regression with δ-BMI and baseline BMI as dependent variables. In contrast to the univariate results, it appears that only gender exerts a marginally significant effect in the direction of underestimating BMI. The most important effect is that exerted by baseline BMI: individuals with a higher "true" BMI impart a downward bias on BMI calculated from self-reports. These effects are large and statistically significant. The second equation displays estimates for an equation where baseline BMI is predicted from self-reports while controlling for a number of conditions. As was the case before, the fit of this equation is acceptable and suggests that it is possible to use it to adjust reported BMI.

But, what are the actual consequences of using BMI from self-reports? An important goal of research in the area of health among adults is to assess the prevalence of obesity. Conventionally, an individual scoring 30 or higher in BMI is classified as obese and as morbid obese if the score exceeds 39. What is the sensitivity and specificity of a measure of obesity derived from self-reports? Table 8 (panel a) displays a cross-tabulation of the sub-sample according to obesity, self-reported and calculated from physical measurements. Both sensitivity and specificity are relatively high, .75 and .94 respectively. The second panel of the table shows what we could expect for the scores of sensitivity and obesity if instead of using observed BMI we employed the predicted values derived from the equation in Table 7. Using the "adjusted" BMI leads to increased sensitivity (from .75 to .80) while specificity remains stable. These are hardly significant advances.

If one were to use self-reports as a base for assessment, the proportion of the population that would fall in the category "obese" is about .24 when the "true" prevalence of obesity is .25. This suggests that, despite biases and imperfect reliability, classification of the population by obesity based on self-reports will not lead us astray.

4.1. The propagation of biases

The last statement will remain unsubstantiated unless we can prove that errors affecting BMI from self-reports do not translate into erroneous inferences. One relation that is consistently estimated in studies of elderly population is that between prevalence of diabetes and obesity. It is believed that obesity is a risk factor for diabetes and this has proven to be the case repeatedly in national samples in most of the developed world and in virtually all age segments.

But, is this relation estimated with (unspecified) biases when, instead of using actual BMI, we use self-reports? A negative answer to this question would enhance the value of self-reports and diminish the need for expensive implementation of physical measurements. Table 9 displays estimates from a logit regression model predicting diabetes while controlling for a number of confounding characteristics. Diabetes was self-reported and elicited from questions about medical diagnoses of diabetes in the past. The first set of estimates is obtained while using baseline BMI, whereas the second includes BMI from self-reports. The similarities are impressive indeed, and none of the estimated effects or standard errors is affected by more than trivial amounts.

The conclusion we can draw from this exercise is a strong one: while self-reported height and weight contain small biases and are not perfectly reliable, their combination in a BMI index does not lead to contamination that can hamper assessment of obesity nor the estimation of relation between obesity and a chronic conditions such as diabetes.

5. Conclusions

Self-reported weight and height in this sample of elderly Mexicans reflect biases and lack of reliability that closely follow patterns already discovered in other data sets. They are associated with gender, age and levels of education, but the relations are not tight and errors are not always of large magnitude. Both self-reported weight and height are heavily affected by heaping (on digits ending in 0 and 5) but this does not appear to translate into large biases.

In all cases, self-reported weight and height provide an accurate gauge to baseline measures and, when proper controls are introduced, it is feasible to derive equations to adjust self-reports and diminish the impact of biases and lack of reliability.

Neither estimates of prevalence of obesity nor prediction of individual obesity are influenced by errors in self-reported height and weight. Sensitivity and specificity of obesity from self-reports is quite high and hardly improved upon by adjustments to observed BMI. Finally, and perhaps more importantly, estimates involving self-reported height and weight on the right hand side of an equation predicting self-reported diabetes are completely unaffected by errors. The resulting estimates are almost indistinguishable from those that would be obtained if one used measures of baseline BMI.

We end the paper with a note of caution. All our results are based on a subsample with available baseline measures **and** self-reports. Although we attempted to establish that this was not a selected sample in terms of δ-errors, our demonstration was *ad-hoc* and far from being completely convincing. In this sense, the strong conclusions drawn before ought to be tempered until we find ways of providing a more convincing demonstration of absence of relevant selectivity.

References

Bush, T., S.R. Miller, A.L Golden et al.,1989. "Self-report and medical record report agreement of selected medical conditions in the elderly" *American Journal of Public Health* 79: 1554-1556.

Charney, E.G., H. Chanblee, M. McBride et al., 1976. "Childhood antecedents of adult obesity." *New England Journal of Medicine* 295: 6-9.

Goldman, Noreeen, I-Fen Lin, M. Weinstein and Yu-Hsung Lin, 2002. "Evaluating the quality of self-reports on hypetension and diabetes." Working Series paper No. 2002-3. Office of Population Research, Princeton University.

Harlow, S. and M. Linet, 1989. "Agreement between questionnaire data and medical records. The evidence for accuracy of recall." *American Journal of Epidemiology* 129: 233-248.

Johansson, J., M. Hellenius, S. Elofsson et al., 1999. "Self-report as a selection instrument in screening for cardiovascular disease risk" *American Journal of Preventive Medicine* 16: 322-324

Kehoe, R., S. Wu, C. Leske et al., 1994. "Comparing self-reported and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *Journal of Clinical Epidemiology* 49: 1407-1417.

Martin, L.M., M. Leff, N. Calonge et al., 2000. "Validation of self-reported chronic conditions and health services in a managed care population." *American Journal of Preventive Medicine* 18: 215-218.

Palta, Mari, R.J. Prineas, R. Berman and P. Hannan, 1982. “Comparison of self-reported and measured height and weight.” *American Journal of Epidemiology* 115 (2): 223-230.

Stewart, Anita L. 1982. “The reliability and validity of self-reported weight and height” *Journal of Chronic Diseases* 35: 295-309.

Torno, M., C. Navarro, M. Chirlaque et al., 2000. “Validation of self diagnosis of high blood pressure in a sample of the Spanish EPIC cohort: overall agreement and predictive values.” *Journal of Epidemiology and Community Health* 54: 221-226.

Vargas, C., V. Burt, R. Gillum et al., 1997. “Validity of self-reported hypertension in the National Health and Nutrition Examination Survey III, 1988-1991.” *Preventive Medicine* 26: 678-685.

Beckett, M., M. Weinstein, N. Goldman, and Lin Yu-Hsuan, 2000. “Do health interview surveys yield reliable data on chronic illnesses among older respondents?” *American Journal of Epidemiology* 151: 315-323.

Wing, R.R, L.H. Epstein, D.J. Ossip et al., 1979. “Reliability and validity of self-report and observer estimates of relative weights.” *Addictive Behavior* 4: 113-140.

Table 1: Basic statistics of sample used

Panel a: Main characteristics

Subgroup	n	p*
Total	2332	1.0
	Gender	
Males	1075	.46
Females	1257	.54
	Age	
50-59	1082	.46
60-69	744	.32
70+	506	.22
	Education	
0 years	594	.26
1-6 years	838	.36
7+ years	900	.38

* p is proportion in total sample

Panel b: Weight and height

	Baseline		Self-reports	
	Weight	Height	Weight	Height
Mean	68.5	1.58	69.8	1.61
Std. Deviation	14.2	10.1	14.2	10.1

Table 1 (cont.)

Panel c: Attrition

		Baseline			Self-Among those with valid baseline	
		Weight			Weight	
		valid	miss		valid	miss
Height	valid	2225	8	valid	1492	44
	miss	8	91	miss	487	202

*miss refers to a missing value and valid to a valid report or measure

Panel d: Measures of association (p-levels in parentheses)

Relations	chi-squared	gamma	kendall-tau
Missing baseline H and W	195.5(p<.000)	.99(p<.000)	.92(p<.001)
Missing self-reported H and W	365.6(p<.000)	.86(p<.000)	.40(p<.000)
Missing baseline and self-reported H	10.4(p<.01)	.32(p<.02)	.07(p<.05)
Missing baseline and self-reported W	51.6(p<.001)	.63(p<.01)	.15(p<.020)
Missing BOTH baseline and self-reported	28.8(P<.00)	.57(p<.00)	.11(p<.00)

Table 2: Summary information and test of validity and reliability of self-reported weight (δ-weight)

Sample	n	Mean	SD	99%CI	R-Squared*	Digit Preference**
Total	1974	.42	4.90	.14-.71	.88	.40
Gender:						
males	937	.61	5.03	.19-1.03	.87	.43
females	1037	.24	4.76	-.14-.62	.88	.38
Age groups:						
50-59	95	.09	4.70	-.30-.48	.89	38
60-69	635	.37	4.52	-.09-.83	.89	.39
70+	382	1.29	5.74	.53-2.05	.82	.39
Education:						
0 years	443	.29	5.77	-.42-.99	.85	.48
1-6 years	723	.49	4.92	.02-.96	.87	.37
7+ years	808	.41	4.31	.02-.80	.90	.34

* R-Squared: is the r^2 between baseline and self-report (the square of the coefficient of validity)

** Digit Preference: proportion of sample choosing a weight ending in 0 or 5

Table 3: Regressions (OLS) between δ-weight and baseline with covariates

Panel a: OLS regression of δ-weight

Variable	Estimate	SD	R-Squared=.072, n=1974
constant	7.10	.67	
females	-.95	.22	
age 60-69	.15	.24	
age 70+	.52	.30	
educ 1-6	-.50	.29	
educ 7+	-.29	.25	
digit pref*	-.67	.13	
baseline	-.09	.008	

Panel b: OLS regression of baseline weight

Variable	estimate	SD	R-Squared=.88, n=1974
constant	4.61	.69	
females	-.08	.23	
age 60-69	-.40	.25	
age 70+	-1.61	.30	
educ 1-6	-.04	.30	
educ 7+	-.05	.25	
digit pref*	-.93	.22	
self-report	.95	.008	

*digit pref is a 1/0 variable for cases who elicited a self-reported weight ending in 0 or 5

Table 4: Summary information and test of validity and reliability of self-reported height (δ-height)

Sample	n	Mean	SD	99%CI	R-Squared*	Digit Preference**
Total	1525	1.73	4.85	1.40-2.05	.76	.52
Gender:						
males	816	1.31	4.69	.89-1.73	.66	,56
females	709	2.23	4.98	1.74-2.71	.62	.48
Age:						
50-59	775	1.44	4.18	1.05-1.82	.81	.49
60-69	750	2.05	5.44	1.53-2.56	.73	.52
70+	255	2.30	5.63	1.39-3.21	.73	.61
Education						
0 years	270	1.98	5.98	1.04-2.91	.69	.67
1-6 years	509	1.88	5.51	1.24-2.51	.69	.56
7+	746	1.54	3.79	1.18-1.89	.84	.44

* R-Squared: is the r^2 between baseline and self-report (the square of the coefficient of validity)

** Digit Preference: proportion of sample choosing a weight ending in 0 or 5

Table 5: Regressions (OLS) between δ-height and baseline with covariates

Panel a: OLS regression of δ-height

Variable	Estimate	SD	R-Squared=.059, n=1525
constant	26.3	2.91	
females	-.96	.33	
age 60-69	.40	.28	
age 70+	.45	.15	
educ 1-6	-.42	.35	
educ 7+	-.15	,28	
digit pref*	-.28	.25	
baseline	-.15	.02	

Panel b: OLS regression of baseline weight

Variable	estimate	SD	R-Squared=.80, n=1525
constant	45.53	2.50	
females	-4.20	.28	
age 60-69	-.45	.25	
age 70+	-1.47	.32	
educ 1-6	-1.59	.32	
educ 7+	-.99	.26	
digit pref*	.06	.23	
self-report	.73	.014	

*digit pref is a dummy variable taking the value 1 for cases who elicited a self-reported height ending in 0 or 5 and 0 otherwise

Table 6: Summary information and test of validity and reliability of BMI from self-reported measures (δ-BMI)

Sample	n	Mean	SD	99%CI	R-Squared*
Total	1472	-.42	2.45	-.58;-.26	.77
Gender:					
males	788	-.21	2.26	-.42;-.003	.74
females	684	-.66	2.62	-.91;-.40	.78
Age:					
50-59	754	-.47	2.38	-.69;-.27	.79
60-69	475	-.48	2.50	-.77;-.19	.73
70+	243	-.19	2.54	-.61;.23	.70
Education:					
0 years	253	-.45	2.85	-.91; .01	.70
1-6 years	495	-.47	2.80	-.79;-.15	.73
7+	724	-.38	2.00	-.57;-.19	.82

* R-Squared: is the r^2 between baseline BMI and BMI from self-reports (the square of the coefficient of validity)

Table 7: Regressions (OLS) between δ-BMI and baseline BMI with covariates

Panel a: OLS regression of δ-BMI

Variable	Estimate	SD	R-Squared=.04, n=1472
Constant	4.41	.36	
females	-.22	.12	
age 60-69	-.11	.14	
age 70+	-.15	.18	
educ 1-6	-.07	.17	
educ 7+	-.11	.13	
baseline	-.17	.01	

Panel b: OLS regression of baseline BMI

Variable	estimate	SD	R-Squared=.78, n=1472
Constant	2.78	.39	
females	.53	.13	
age 60-69	-.07	.14	
age 70+	-.55	.19	
educ 1-6	.11	.18	
educ 7+	.17	.14	
self-report	.91	.04	

Table 8: Relation between obesity evaluated from BMI from baselines and self-reports

Panel a: Specificity and sensitivity of BMI based on self-reports

		Baseline BMI	
		not obese	obese
BMI self-reports	not obese	1020	98
	obese	61	293

Specificity=.94; Sensitivity=.75

Panel b: Specificity and sensitivity of BMI based on adjusted self-reports

		Baseline BMI	
		not obese	obese
BMI adjusted self-reports	not obese	1008	88
	obese	73	303

Specificity=.94; Sensitivity=.79

*Adjusted self-reports refers to BMI calculated as predicted values from the equation in Table 7 (panel b)

Table 9: The relation between self-reported diabetes and BMI

	Logit Models using			
	Baseline BMI		BMI from self-reports	
Variable	Estimate	SD	Estimate	SD
Female	.03	.15	.04	.15
age 60-69	.21	.16	.21	.16
age 70+	.25	.21	.24	.21
educ 1-6	.29	.20	.29	.20
educ 7+	.10	.15	.09	.17
baseline	.35	.16	-	-
self-reports	-	-	.33	.16

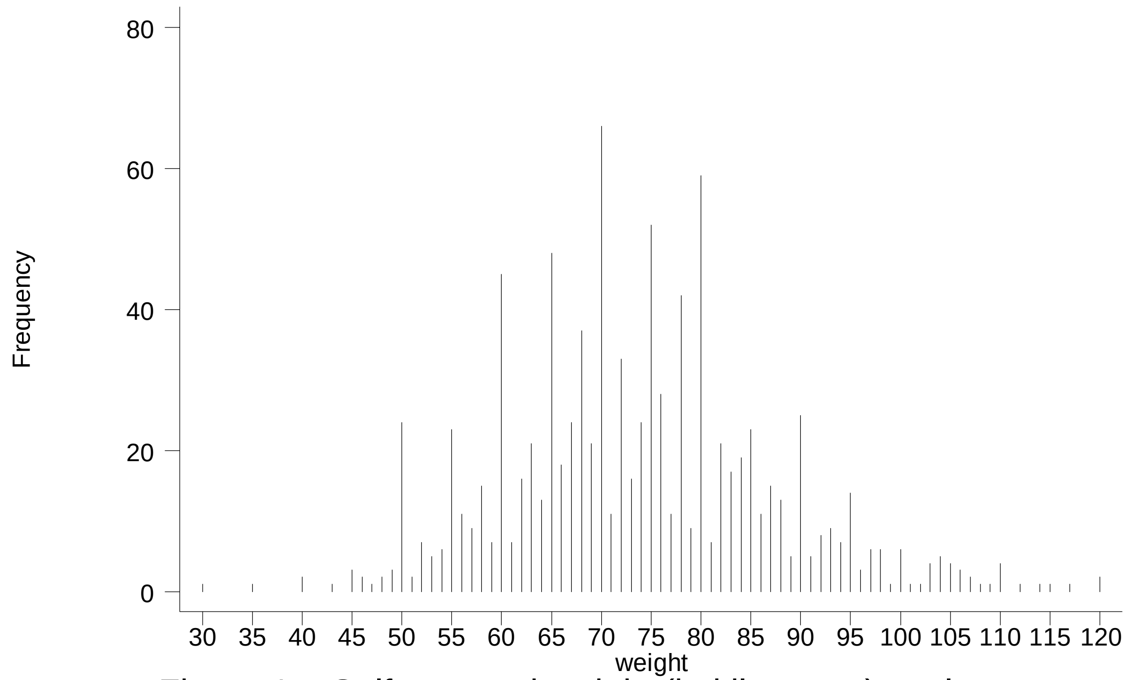

Figure 1a: Self-reported weight (in kilograms), males

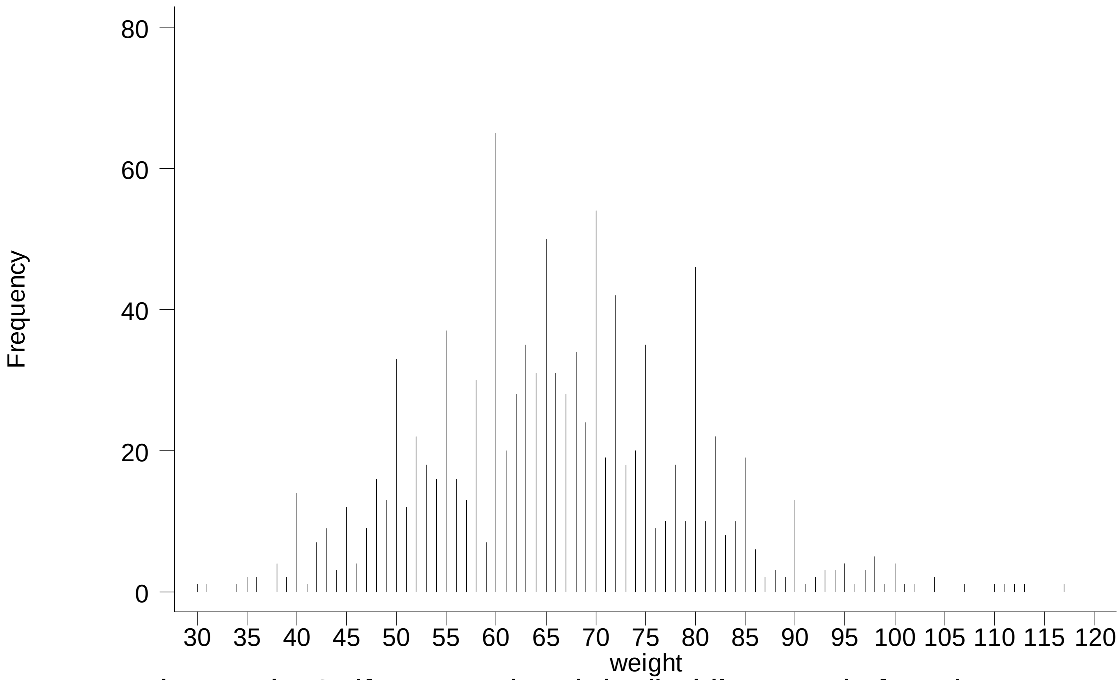

Figure 1b: Self-reported weight (in kilograms), females

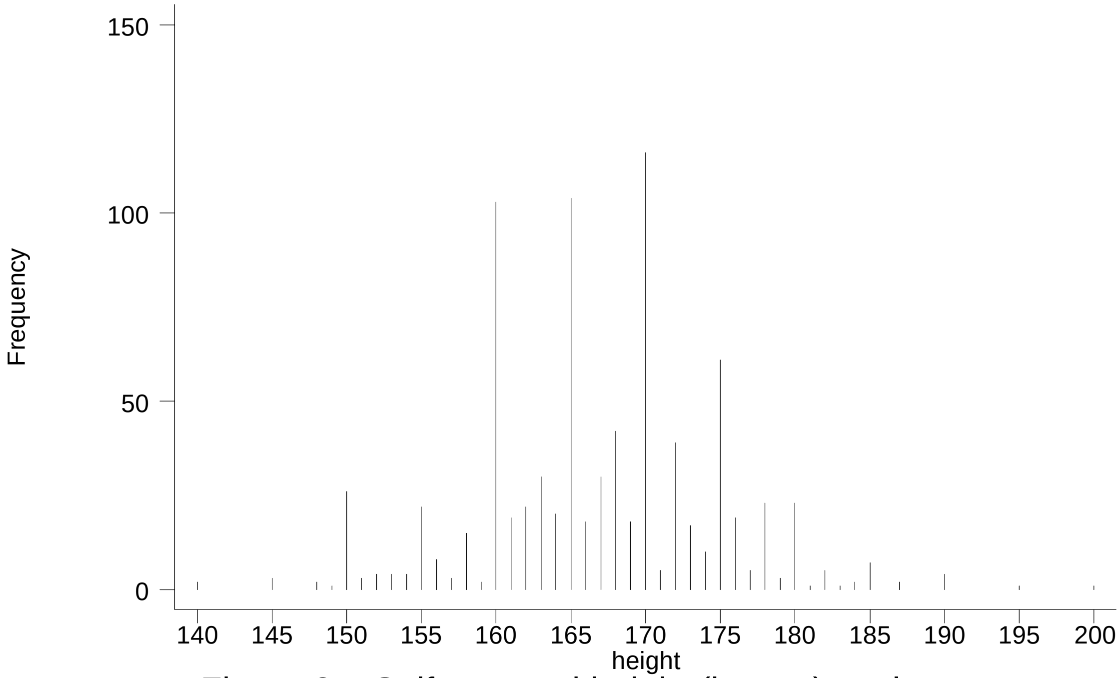

Figure 2a: Self-reported height (in cms), males

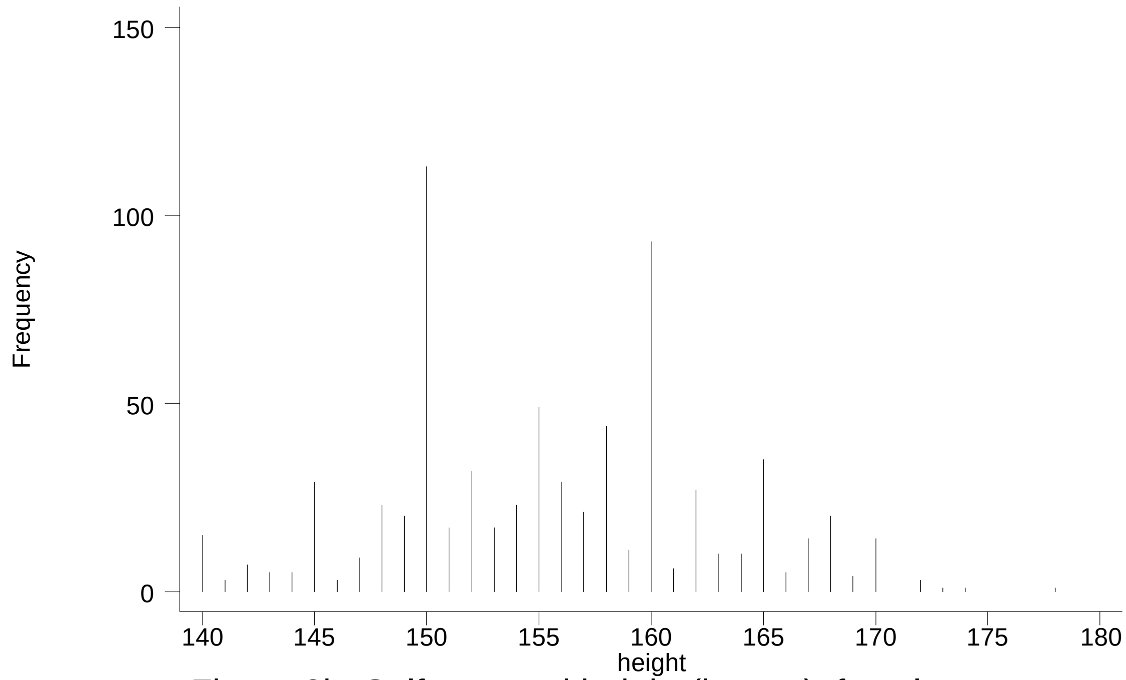

Figure 2b: Self-reported height (in cms), females

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: palloni@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu