



The determinants of misreporting weight and height: The role of social norms

Joan Gil ^{a,*}, Toni Mora ^b

^a *Dep. of Economic Theory and CAEPS, University of Barcelona, Ave. Diagonal 690, 08034 Barcelona, Spain*

^b *Universitat Internacional de Catalunya and Barcelona Institute of Economics, Spain*

ARTICLE INFO

Article history:

Received 27 October 2009

Received in revised form 7 May 2010

Accepted 25 May 2010

JEL classification:

I10

I12

Keywords:

Self-reported weight

Height and BMI

Anthropometric biases

Social norms

Spain

obesity

ABSTRACT

Given the lack of availability of measured anthropometric data for the whole of Spain, this paper combines data from the 2006 Catalan Health and Health Examination Surveys to compute the size of weight and height self-reporting biases. The underlying determinants of these biases are then analyzed, placing special emphasis on the role played by social norms. Our findings show that social norms regarding “ideal” weight (proxied by the average weight of a reference group based on gender and age) tend to affect the self-reporting weight (relative) bias. This finding suggests that the more satisfied individuals feel with their own body image the less prone they are to under-report their weight, although this effect is contingent upon the definition of social norms and the correction of endogeneity. However, we found no evidence of a similar impact caused by the social norms governing height. The relationship found between the measured and self-reported anthropometric data was applied to the Spanish National Health Survey (NHS) so as to correct the self-reported information contained in it. After correcting for self-reporting errors, both the BMI and the prevalence of obesity were found to be significantly underestimated, with instances of misreporting being more prevalent among women.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The recent rapid increase in the prevalence of obesity is widely recognized as a major public health concern in economically developed countries. The condition is considered to be responsible for an increasing share of total health care expenditure as it is closely associated with a number of co-morbidities (e.g., heart disease, diabetes, certain cancers, arteriosclerosis, or sleep apnea), and for generating other socioeconomic costs (e.g., labor incapacity, absenteeism, wage and promotion discrimination, stigma or social exclusion).¹ Moreover, we cannot ignore the fact that

obesity is more prevalent among low-educated and poor income families and as such is unevenly distributed within society (Sobal and Stunkard, 1989; Zhang and Wang, 2004; Costa-Font and Gil, 2008). Despite its obvious significance, the severity of the epidemic has been tested to date mainly by means of self-reported anthropometric information.

In this respect, the International Obesity Task Force in its briefing on obesity in Europe complains that “few countries conduct systematic measured surveys to obtain reliable nationally representative data to assess the degree of overweight and obesity in their populations”. The Organization of Economic Cooperation and Development (OECD) even recognizes that the use of self-reported weight and height data is subject to significant biases which ultimately “underestimate the true prevalence of obesity in many OECD countries because of reporting biases” (IOTF, 2005). This means that self-reported surveys need to be corrected so as to obtain more accurate

* Corresponding author. Tel.: +34 93 4020109; fax: +34 93 4039082.

E-mail address: joangil@ub.edu (J. Gil).

¹ See, for instance, Averett and Korenman (1996), Cawley (2000, 2004) and Morris (2006).

measures of obesity.² There can be little doubt that access to reliable information is crucial for health policy makers dealing with this pandemic in order that they might avoid making inappropriate public health policy responses.

This paper also questions the validity of self-reported anthropometric data. Under the general assumption that self-reports on weight and height partly reflect, or are influenced by, people's desires regarding their ideal body image, the objective of this paper is threefold. First, we quantify the magnitude of weight, height and BMI self-reporting biases based on the most recently conducted Health & Health Examination Surveys for Catalonia (given the lack of availability of similar data for Spain as a whole). These biases, which are computed as deviations (in relative terms) between self-reported and actual or measured weight and height, can be interpreted as representing the individuals' degree of dissatisfaction with their own body shape.³ Secondly, the paper contributes to the literature by analyzing the impact of a set of determinants on these anthropometric biases, emphasizing the role played in this phenomenon by the so-called social norms regarding "ideal" or desired weight and height. Mirroring the literature on social norms or peer influences, we assume that social norms regarding what constitutes an ideal body weight also affect individuals' self-reporting decisions when answering anthropometric questions on health questionnaires (see *Etilé, 2007; Powdthavee, 2009*). However, whether this social environmental effect constitutes a general concern for researchers working with self-reported data is a question under empirical scrutiny and will depend on the size of this social norm effect (*Powdthavee, 2009*). Thirdly, following *Cawley and Burkhauser's (2008)* approach we transfer the relationship found in our sample between self-reported and measured data to the latest edition of the Spanish health survey in order to achieve a more accurate measurement of both BMI and obesity rates for the whole of Spain. This exercise allows us to assess whether the corrected Spanish self-reported survey based on our estimations differs greatly from the corrections performed when using, for instance, the conversion formulas provided by *Cawley and Burkhauser (2008)* based on dissimilar cultural and environmental patterns.

In line with previous studies, our findings corroborated the existence of significant biases in self-reported weight and height in the Catalan data sample. The biases followed the expected pattern for gender, age, education and health status, among other factors. In addition, by using a Seemingly Unrelated Regression (SURE) procedure to account for the presence of common unobservable effects, we found evidence confirming that social norms regarding ideal weight had an effect on weight bias. This effect was however contingent upon the definition of the social norm

covariate and the correction of endogeneity. The anthropometric biases were also explained by the incidence of other covariates (i.e., the BMI, waist-to-hip ratio, age, gender, nationality, education, health status and health-care variables). Finally, we found that both self-declared BMI and the obesity rate of the Spanish population were significantly underestimated (18.5% as opposed to 15%), with the degree of misreporting being higher among women.

The remainder of this paper is structured as follows. In Section 2 we measure and quantify the magnitude of bias in self-declared anthropometric data for the Catalan case. Section 3 outlines the econometric methodology, discusses the underlying determinants of the misreporting behavior, and presents our empirical evidence. Section 4 corrects the anthropometric data for Spain as a whole and the final section summarizes our main findings.

2. Bias in anthropometric data

2.1. Data description

Unfortunately, given the unavailability of a nationally representative examination survey for the whole of Spain (*Aromaa et al., 2003*),⁴ we assess and quantify the magnitude of bias in self-reported data on weight, height and BMI by using the 2006 editions of the Catalan Health Survey (CHS) and the Catalan Health Examination Survey (CHES), both produced by the Catalan government's Department of Health, providing valuable health information based on self-reports and clinical measurements. Although it is not testable, we claim the existence of a similar reporting pattern of health information between Catalan and Spanish societies on the grounds that the two populations share similar customs, cultural and historical traits, and economic and religious backgrounds.⁵ Indeed, our results regarding the size of the anthropometric biases are comparable to those found by *Quiles and Vioque (1996)* for the region of Valencia, another highly populated region in Spain.

The CHS-2006, which is representative of the resident and non-institutionalized Catalan population, gathers information on aspects such as self-perceived health status, healthcare utilization, consumption of medicines, chronic conditions, mental health and social support, preventive medicine, life styles, self-reported anthropometric data, and socioeconomic characteristics. The survey was carried out using computer-assisted personal interviews and was structured in three questionnaires following a complex multistage sampling procedure. The original sample comprised 18,126 individuals (15,926 adults aged 15+ and 2200 children) with information collected between December 2005 and July 2006. During 2006, a voluntary health examination survey (CHES-2006) was

² An example of this bias is provided by a self-reported adult obesity rate of 20% based on the BRFSS telephone survey in the US, which contrasted with estimates of 28% (men) and 34% (women) for the same year based on the NHANES data, using measurements provided by trained personnel.

³ See *Danubio et al. (2008)* for a survey of bias magnitudes for OECD countries.

⁴ The exceptions are a number of regional surveys covering their respective populations: the "Nutrition and Health Examination Survey 1994" of the Region of Valencia, and the 2002 and 2006 editions of the CHES.

⁵ Catalonia is the second most populated region in Spain with 16% of the country's total population and 18.8% of the Gross Domestic Product.

undertaken on a representative (by health regions, gender and age groups) subsample of interviewed individuals, aged 6 and over, obtained from the CHS-2006, to complement the information provided by the health survey and to improve the understanding of the distribution of cardiovascular disease (CVD) risk factors (diabetes, arterial hypertension or obesity) among the Catalan population. In common with other examination surveys, the CHES-2006 includes physical (height, weight, waist and hip circumference, blood pressure and heart rate) and biological (glucose and cholesterol tests) explorations and a short list of questions on food disorders.⁶ The measurements were obtained by trained personnel and carried out at individuals' homes during 2006. From a theoretical subsample of 4000 respondents extracted from the health survey, the CHES-2006 finally contained 2119 individuals, a response rate of 55.5%. Note that the two surveys were intentionally programmed to collect clinical or measured data (from the CHES) after reporting other health information required by the CHS so as to avoid any potential "framing effects".

Having merged the two samples using the identification number and having removed the data relating to subjects aged below 15 and above 65, we ended up with a sample comprising 1568 observations (49.7% men and 50.3% women) for whom we had both self-reported and measured anthropometric data together with socioeconomic and health information.⁷

2.2. Bias magnitudes

Table 1 shows the estimation of self-reported and measured weight, height and BMI (Body Mass Index) average values according to different population subgroups in the Catalan sample described above. The table also reports the magnitude of the bias or gap, computed simply as the difference between each self-reported anthropometric value and the corresponding measurement and the share of the bias in percentage terms. In line with standard practice, the BMI indicator was calculated as weight in kilograms divided by the square of height in meters (kg/m^2). The following categories adopted by the WHO (2000) are considered: (1) underweight: BMI <18.5; (2) normal weight: BMI 18.5–24.9; (3) overweight: BMI 25–29.9 and (4) obese: BMI \geq 30. The estimations excluded the influence of multiple outliers detected by applying the Hadi (1992) procedure for multivariate data.⁸

⁶ Weight and height measurements were taken with shoes off and with subjects clad only in a light robe or undergarments and rounded to the nearest 0.5 cm. Waist and hip circumferences were measured using a flexible, inelastic measuring tape with the subject in standing position. The tape was placed horizontally (parallel to the floor) around the abdomen at the level of the navel (waist circumference) and around the maximal circumference at the level of the trochanter (hip circumference). In both cases measurements are recorded to the nearest 0.5 cm.

⁷ We excluded the elderly population from the sample to reduce the incidence of their higher morbidity on the self-reporting of weight and height.

⁸ We excluded 20 (15) outliers for self-reported weight (height) data and just 13 outliers for self-reported BMI data.

As a criterion of validity, Pearson correlation coefficients between self-reported and measured weight and height were computed. All were statistically significant ($p < 0.001$) using Bonferroni-adjusted significance levels for all age groups within each gender. Specifically, correlation coefficients between self-reported and measured weight for all age groups and both genders ranged between 0.90 and 0.96, whereas correlations for height and the same population subgroups were somewhat lower, ranging between 0.80 and 0.91. Despite these strong correlations, our data confirmed that individuals tended to commit systematic errors when reporting anthropometric data.

Our estimations showed that the mean difference between self-reported and measured weight was -1.42 kg (95% CI: -1.63 to -1.21), confirming the widely reported underestimation of weight together with a higher degree of underestimation among women than among men (-1.45 kg vs. -1.38 kg). For instance, using data for Sweden, Boström and Diderichsen (1997) reported that adult women tended to underestimate their weight significantly more than did men (-1.64 kg vs. -0.74 kg). Similar results were found by Niedhammer et al. (2000) for a sample of middle-aged French subjects (-0.85 kg for women, -0.54 kg for men), by Bolton-Smith et al. (2000) based on a sample of Scottish adults (-0.95 kg for women, -0.63 kg for men) and by Quiles and Vioque (1996) using data from participants in the 'Survey of Nutrition and Health, 1994' conducted in the Autonomous Community of Valencia (-1.81 kg for women, -1.21 kg for men). However, other studies reported a more pronounced underestimation of weight among the male population (Gunnell et al., 2000; Spencer et al., 2002). In the case of height, the mean difference between self-reported and measured height was estimated to be 0.64 cm (95% CI: 0.47–0.81) clearly suggesting that height was overestimated. Interestingly, this gap or bias was substantially higher among adult Catalan women (0.92 cm) than it was among men (0.36 cm), suggesting perhaps a greater dissatisfaction with their height among these women. The findings in the literature on this point are again somewhat mixed: while some papers report a higher overestimation in self-reported height among females (Quiles and Vioque, 1996; Boström and Diderichsen, 1997; Paccaud et al., 2001), others present a larger overestimation among males (Gunnell et al., 2000; Kuczmarski et al., 2001; Spencer et al., 2002).

At this point, it is worth stressing the differentials found in the computed biases based on covariates. Reflecting previous results for another region of Spain (Quiles and Vioque, 1996), Table 1 shows that the height bias increased with age, in accordance with other studies (Boström and Diderichsen, 1997; Kuczmarski et al., 2001). Hence, the oldest adult group (aged 56–65) tended to over-report their height significantly more (1.42 cm) than any other age group.⁹ However, this finding could not be corroborated fully for weight since the greatest weight bias was found for the 25–35 age group. Next, educational

⁹ Reductions in stature and body weight in old age have been reported; this would ultimately explain why older adults are in general more unaware of their current height.

Table 1
Self-reported and measured mean height (cm), weight (kg) and BMI (CHS and CHES, 2006).

	Female	Male	Age 15–24	Age 25–35	Age 36–45	Age 46–55	Age 56–65	Low education	Second education	University education
<i>Weight:</i>										
Self-reported	62.56 (11.08)	78.14 (12.90)	63.95 (12.13)	68.64 (14.11)	72.00 (14.10)	72.73 (14.99)	73.57 (13.89)	72.48 (13.39)	69.04 (14.70)	71.17 (14.00)
Measured	64.00 (11.61)	79.53 (13.94)	65.10 (12.60)	70.35 (15.02)	73.27 (14.82)	74.02 (15.64)	75.05 (14.28)	73.70 (13.70)	70.54 (15.59)	72.59 (14.40)
Bias/gap	-1.45 (4.12)	-1.38 (4.34)	-1.15 (4.29)	-1.71 (4.57)	-1.27 (3.90)	-1.29 (4.03)	-1.48 (4.21)	-1.22 (4.44)	-1.50 (4.08)	-1.42 (4.34)
Bias (%)	2.3%	1.7%	1.8%	2.4%	1.7%	1.7%	2.0%	1.7%	2.1%	2.0%
<i>Height:</i>										
Self-reported	161.26 (6.57)	173.82 (7.51)	169.93 (9.44)	169.19 (9.21)	167.69 (9.16)	165.36 (9.74)	164.34 (8.60)	164.91 (8.83)	167.44 (9.32)	170.45 (9.59)
Measured	160.34 (6.59)	173.47 (7.57)	169.45 (9.21)	168.79 (9.22)	167.20 (9.40)	164.63 (9.90)	162.92 (9.30)	164.21 (9.91)	166.82 (9.24)	169.79 (9.91)
Bias/gap	0.92 (3.30)	0.36 (3.42)	0.48 (3.58)	0.40 (3.34)	0.48 (3.40)	0.73 (2.91)	1.42 (3.59)	0.69 (3.64)	0.61 (3.36)	0.66 (3.09)
Bias (%)	0.6%	0.2%	0.3%	0.2%	0.3%	0.4%	0.9%	0.4%	0.4%	0.4%
<i>BMI:</i>										
Self-reported	24.12 (4.40)	25.83 (3.87)	22.00 (3.00)	23.84 (3.72)	25.50 (3.89)	26.50 (4.51)	27.16 (4.08)	26.66 (4.53)	24.50 (4.16)	24.37 (3.54)
Measured	24.99 (4.77)	26.39 (4.24)	22.58 (3.40)	24.57 (4.16)	26.11 (4.21)	27.14 (4.61)	28.26 (4.53)	27.36 (4.77)	25.23 (4.54)	25.05 (3.94)
Bias/gap	-0.87 (1.90)	-0.56 (1.74)	-0.58 (1.65)	-0.73 (1.92)	-0.61 (1.68)	-0.64 (1.85)	-1.09 (1.96)	-0.70 (1.95)	-0.74 (1.78)	-0.68 (1.80)
Bias (%)	3.5%	2.1%	2.6%	3.0%	2.3%	2.4%	3.9%	2.6%	2.9%	2.7%
	SAH 1	SAH 2	SAH 3	Measured obese	Measured overweight	Measured normal weight	Measured underweight			
<i>Weight:</i>										
Self-reported	69.85 (13.98)	71.47 (14.90)	74.78 (17.81)	88.70 (13.40)	76.84 (9.33)	61.85 (9.07)	51.40 (8.42)			
Measured	71.32 (14.49)	72.66 (15.14)	76.13 (19.10)	92.24 (13.91)	78.29 (10.35)	62.21 (8.71)	49.80 (6.74)			
Bias/gap	-1.47 (4.16)	-1.18 (4.40)	-1.35 (5.62)	-3.54 (4.82)	-1.45 (4.31)	-0.36 (3.60)	1.60 (4.17)			
Bias (%)	2.1%	1.6%	1.8%	3.8%	1.9%	0.6%	-3.2%			
<i>Height:</i>										
Self-reported	168.17 (9.46)	165.11 (8.75)	163.74 (9.75)	167.65 (9.26)	168.24 (9.35)	167.29 (9.00)	167.46 (11.64)			
Measured	167.52 (9.68)	164.52 (8.93)	162.98 (10.25)	165.78 (10.16)	167.72 (9.78)	166.96 (9.09)	168.17 (10.89)			
Bias/gap	0.65 (3.29)	0.59 (4.04)	0.76 (4.68)	1.88 (3.59)	0.52 (3.62)	0.33 (3.11)	-0.72 (2.71)			
Bias (%)	0.4%	0.4%	0.5%	1.1%	0.3%	0.2%	-0.4%			
<i>BMI:</i>										
Self-reported	24.58 (3.89)	26.21 (4.87)	28.03 (5.65)	31.47 (3.74)	27.07 (1.40)	22.03 (2.05)	18.23 (1.45)			
Measured	25.32 (4.26)	26.80 (5.21)	28.79 (5.93)	33.44 (3.31)	27.76 (2.38)	22.22 (1.73)	17.53 (0.67)			
Bias/gap	-0.74 (1.75)	-0.59 (2.09)	-0.75 (2.17)	-1.96 (2.18)	-0.69 (1.94)	-0.20 (1.42)	0.70 (1.46)			
Bias (%)	2.9%	2.2%	2.6%	5.9%	2.5%	0.9%	-4.0%			

Note: Estimates performed using sampling weights and excluding outliers following Hadi (1992). Standard deviations are reported in brackets. SAH 1: good/very good and excellent; SAH 2: fair; SAH 3: poor self-assessed health status. Biases statistically significant at 5% are in bold typeface. Source: Catalan Health Survey (CHS) 2006 and Catalan Health Examination Survey (CHES) 2006.

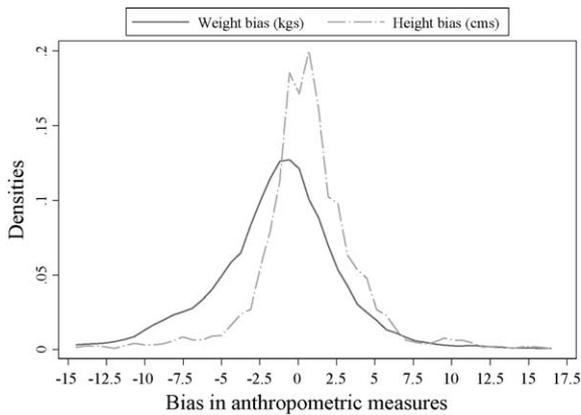


Fig. 1. Distributions of weight and height bias.

attainment played a notable role in the generation of biases. Interestingly, the lower the individual educational attainment, the greater was the bias in height, but the lower the bias in weight. Thus, our data indicate that these individuals seemed to be more worried about reporting their real height than their weight.¹⁰ Self-assessed health status also revealed certain trends in the misreporting. Both the weight and height biases were larger among those with better self-perceived health status, which could be interpreted as a consequence of greater levels of dissatisfaction with their own body size. Finally, as expected, our estimations showed that obese individuals under-reported their weight (-3.54 kg) and over-reported their height (1.88 cm) more significantly than their non-obese counterparts, while this degree of misreporting was also higher in relative terms.

On the understanding that these biases between self-declared and measured weight (height) may differ across the whole distribution, Fig. 1 displays kernel density functions to illustrate misreporting patterns across the anthropometric distributions. Two findings deserve particular mention. First, the height bias distribution only displays significant dissimilarities for heights 2.5 cm above the average. Second, the weight bias distribution was more skewed. In other words, greater dissimilarities in the shape of the distribution were seen for the weight bias, indicating that individuals misreported their weight more than they did their height. Therefore, bearing in mind these observations, the determinants of misreporting height and weight cannot be estimated jointly through the use of a single measure, i.e., the BMI indicator.

Finally, according to our data the difference between the average BMI calculated from self-reported and measured data was -0.72 kg/m² (95% CI: -0.81 to -0.63), revealing a systematic underestimation error when computing the BMI indicator using self-reports (Table 1). Indeed, the degree of underestimation of BMI

was much greater among women (-0.87 kg/m²) than it was among men (-0.57 kg/m²), as well as among those in the oldest age group (-1.09 kg/m² among those aged 56–65). Interestingly, based on this evidence, the negative BMI bias was roughly the same among those individuals with the lowest education attainment (-0.70 kg/m²) and those with a university degree (-0.68 kg/m²). A similar pattern was found among those who reported healthy vs. bad self-assessed health status. As expected, the category measuring obese individuals underestimated BMI (-1.96 kg/m²) much more significantly than was the case with any other body weight subgroup.

3. Determinants of misreporting behavior

3.1. Background and misreporting determinants

There exists an extensive literature in health economics questioning the validity of the self-reported health status of individuals, collected in socioeconomic surveys, as a measure of true health. As a subjective measure of health, self-reported health status may be prone to measurement error (see, among others, Groot, 2000; Baron-Epel and Kaplan, 2001; Lindeboom and van Doorslaer, 2004; Hernández-Quevedo et al., 2005; Lindeboom, 2006). In keeping with the explanations advanced by these reports for the existence of such self-reporting biases, we were able to anticipate similar reasons for the occurrence of this phenomenon – namely, evidence of weight and height biases when individuals self-declare such information in health questionnaires.

On the one hand, such biases can simply be understood as the consequence of a lack of information regarding one's own body weight and/or height. Thus, better informed individuals or individuals who have to be measured more frequently because of their health status will tend to declare their weight and height more accurately.¹¹ On the other hand, weight and height biases or errors could well be the consequence of a cognitive distortion affecting the individuals' perception of their own body shape. This would explain why many older individuals show a greater unawareness of their actual weight or height. In addition, since we assume that an individual's weight and height self-reports are partly driven by his or her own ideal or desired body image, our conceptual framework considers that such self-reports ultimately depend on the reporting behavior of a close reference group (social norms) as to what constitutes an ideal body weight (see, Etilé, 2007; Powdthavee, 2009).

In line with the discussion above and the literature on the determinants of weight and height self-reporting biases, our econometric specification contains the following set of covariates (Table 2):

- (i) Following Boström and Diderichsen (1997), Bolton-Smith et al. (2000), Gillum and Sempos (2005) and

¹⁰ Similarly, according to Boström and Diderichsen (1997) manual workers overestimated (underestimated) their height (weight) more (less) than non-manual workers. See also Niedhammer et al. (2000).

¹¹ Note that crucial to our analysis is the fact that at the time of the home interview individuals report their weight and height without knowing that they might be invited to participate in an examination survey during which some measurements will be taken.

Table 2
Variable definition and descriptive statistics.

Variable	Definition	Mean	SD
<i>Demographics and civil status</i>			
Male	=1 if male; 0 female	0.498	0.50
Age	Age of the individual	39.40	13.01
Single	=1 if single	0.343	0.47
Widow	=1 if widow	0.011	0.10
Divorced-separated	=1 if separated/divorced	0.048	0.21
<i>Nationality</i>			
African	=1 if African nationality	0.018	0.13
Latin-American	=1 if Latin America or Caribbean nationality	0.071	0.26
Asian	=1 if Asian nationality	0.004	0.06
European	=1 if European or other developed countries	0.019	0.14
<i>Education and wealth proxy</i>			
Low education	=1 if low education	0.230	0.42
Secondary education	=1 if secondary education	0.553	0.50
Non deprivation	=Index of individual deprivation (*)	0.769	0.18
<i>Health variables</i>			
Fair health	=1 if fair self-assessed health status	0.164	0.37
Bad health	=1 if bad self-assessed health status	0.036	0.19
Chronic diseases	Number of chronic conditions suffered by the individual	1.59	1.73
Checking cholesterol	=1 if cholesterol is controlled habitually	0.478	0.50
Medical visit	=1 if medical visit during last 15 days	0.257	0.44
Weight loss	=1 if the individual has lost more than 6 kg in the last 3 months	0.062	0.24
Fat perception	=1 if perceived as fat although others say the subject is too thin	0.129	0.34
<i>BMI and fatness</i>			
Measured BMI	Measured body mass index (Kg/m ²)	25.65	4.59
WHR	Waist-to-hip ratio (waist circumference divided by hip circumference)	0.895	0.15
Waist circumference	Measured waist circumference	88.70	14.41
<i>Controls</i>			
Interview duration	Number of minutes of interview duration	43.47	12.31
Time lag	No. of days between self-reported and measured anthropometric data	136.64	74.90
Height rounding	=1 if rounding one's height to the nearest '0' and '5' value	0.356	0.48
Weight rounding	=1 if rounding one's weight to the nearest '0' and '5' value	0.385	0.49
<i>Social norms</i>			
Social norms regarding height	Average height for a reference group (**)	166.186	6.83
Social Norms regarding weight	Average weight for a reference group (**)	71.364	8.01

Note: (*) Defined by assigning one if the home is equipped with bath and/or bathtub, hot water, heating, washing machine, dish-washer and air-conditioning. (**) The reference group considered is defined in terms of gender, age group and educational attainment level. Means computed using sampling weights and excluding outliers for weight and height following Hadi (1992).

Danubio et al. (2008) we include an individual's characteristics such as gender, age and age squared, marital status and ethnicity/nationality.

- (ii) To control for differences attributable to knowledge or education, we enter dichotomous variables for different levels of educational attainment. In principle, a lower degree of accuracy is expected among those with a lower level of educational attainment when self-declaring weight and/or height. However, more highly educated individuals can be expected to be more aware or better informed of their real weight and height and should deviate less in their self-reporting of this information.
- (iii) We also control for the existence of a socioeconomic gradient underlying this misreporting behavior. On the one hand, one can argue that poorer people might be expected to overestimate their height and under report their weight to a greater degree than is the case among their richer counterparts. In fact, some previous studies have shown that manual workers, in general, tend to overestimate their height more than non-manual workers do, while non-manual

workers tend to underestimate their weight (Bostrom and Diderichsen, 1997; Niedhammer et al., 2000). On the other hand, the opposite argument would also be plausible if, for instance, richer individuals are more influenced by social norms regarding an "ideal" body weight. Unfortunately, given the existence of many missing values in our dataset regarding the income variable, as an alternative we constructed an index of individual deprivation based on housing characteristics as a proxy of material wellbeing or wealth.¹²

- (iv) We include several health-related covariates which we assume affect the self-reports of weight and height. In particular we consider self-assessed health status, number of chronic conditions and proxies of health care utilization (e.g., control of cholesterol levels and medical visits).

¹² This index ranges from 0 to 1 and is calculated averaging whether the individual's home is equipped with bath and/or bathtub, hot water, heating, washing-machine, dish-washer and air-conditioning.

- (v) Two additional weight-related features are accounted for to measure the potential impact of distorted information regarding an individual's own body mass. On the one hand, a dichotomous variable seeks to identify whether interviewees had experienced a rapid and considerable weight loss (i.e., more than 6 kg in the previous 3 months). On the other hand, another dichotomous variable is included to capture whether individuals perceived themselves as being fat although others saw them as being too thin.
- (vi) Given that the reporting bias appears to be contingent on the particular weight and height level achieved by each individual (Boström and Diderichsen, 1997), we included the measured BMI (including a non-linear effect), the waist-to-hip-ratio (WHR) and the waist circumference as additional regressors. While BMI measures overall obesity, both WHR and waist circumference are proxy indicators of abdominal-visceral fat (central adiposity)¹³ and are estimated to be better predictors of the risk of several chronic diseases and death.¹⁴
- (vii) Finally, as additional controls, our econometric specification incorporates the CHS interview length, the number of days between the first CHS interview and the CHES examination survey¹⁵ and the rounding effect or tendency to round one's weight and height to the nearest '0' and '5' value (see Table 2).

3.2. The role of social norms

Given that the act of reporting anthropometric measures depend on either individual-specific or group-specific factors (Blume and Durlauf, 2006); this paper is particularly aimed at examining the influence of social norms regarding an "ideal" body weight on individuals' self-reports of weight and height. By social norms, we refer here to what Akerlof and Kranton (2000) call prescriptions, namely ideal characteristics that are specific to the individual's assigned social category. And since, in their model, prescriptions affect identity, such social norms enter into the individual's utility function. Other papers consider individuals' utility functions as being influenced by the group's commonly held social norm and people rationally select their weight (height) having in mind the body weight of their peers (Brock and Durlauf, 2001; Burke and Heiland, 2007; Etilé, 2007; Oswald and Powdthavee, 2007; Blanchflower et al., 2009). Based on this framework, the nature of the social norm effect that we consider involves each individual comparing his/her own weight and height to the group's commonly held norm or desired body weight, and this comparison is assumed to affect reporting behavior. That is, being shorter or fatter than average within a reference group

is assumed to affect self-reported data, since the ideal body image conditions self-declared weight and height responses. Hence we hypothesize that self-reports of one's weight and height are conditioned by the "ideal" body weight attributed to a reference social group.

Following standard practice, we constructed the social norms regarding the "ideal" weight (height) covariate as the average reported weight (height) of the individual's assigned reference group, once their own weight and height had been deduced. Assuming that an individual's self-reports are equivalent to his or her desires, we thus obtained a measure of "ideal" weight and height.¹⁶ According to Manski (1993), the identification of the social norm effect requires that the reference group to which the individual is assigned be adequately defined. As for the reference group, it should be noted that there is a long list of possible candidates including parents, friends, classmates or schoolmates, work colleagues or neighbors, who may participate in enforcing the social norms of body shape. However, given the limitations of our dataset, we assumed that the (baseline) reference group was defined on the basis of two individual characteristics: (i) gender and (ii) age (Etilé, 2007; Blanchflower et al., 2009). Alternative grouping criteria (including education but excluding age and including both age and education besides gender) were used for robustness reasons. Note that we preferred to employ education instead of occupation since the latter might suffer mismatch consequences. This was thought to be particularly important in the Spanish case since, in recent decades, the share of highly educated workers has increased notably. Nonetheless, this increase has not been matched by an equal increase in the supply of skilled jobs, thus resulting in over-education. Thus, we hypothesize a negative (positive) impact of the social norm on the weight (height) bias equation. In other words, the higher (lower) the average weight or desired weight (height) stated by the reference group, the lower (higher) the reporting weight (height) bias or the less inclined individuals are to under-report their weight.

3.3. Methods

To examine the influence of the above set of covariates on the calculated self-reporting biases for weight and height, we defined the latter in relative terms, on the understanding that the difference between what is reported and what is measured may not be linear.¹⁷ Nevertheless, for the ease of interpretation we hereafter define these biases or gaps in terms of positive values. Thus, our dependent variables are,

$$\text{Bias}_w = \frac{[w_m - w_d]}{w_m} \quad (1)$$

$$\text{Bias}_h = \frac{[h_d - h_m]}{h_m}$$

¹³ Since there is still some debate as to which indicator is the most appropriate adiposity marker, both measures of abdominal fat were included, albeit that waist circumference is easier to compute (NIH, 1998; Pischon et al., 2008).

¹⁴ See Pischon et al. (2008) for an analysis of the association between central and general adiposity with risk of death in Europe.

¹⁵ Note that during this period (approximately 4 months) the weight of these individuals might fluctuate.

¹⁶ Against this assumption, Etilé (2007) presents evidence using French data that self-declared and ideal BMI coincide for just 40% of the sample.

¹⁷ For instance, a gap of 2 cm between self-reported and measured height may not be the same for someone who is 1.8 m tall and someone who is 1.55 m tall.

where $w_m(h_m)$ stands for measured weight (height) and $w_d(h_d)$ represents declared or self-reported weight (height).

Before analyzing the empirical determinants of the misreporting behavior, we initially wondered whether the voluntary nature of participation in the CHES-2006 might have caused a sample selection problem, since self-reported and measured weight and height information were collected in different surveys (CHS and CHES). It is plausible that individuals might refuse to participate in the examination survey because they feel uncomfortable or, simply, as a consequence of their weight. Thus, gap measurements would be biased and, consequently, would under-represent a fraction of the overweight and obese population. Likewise, individuals might refuse to participate in a second interview due to work or family commitments.

We undertook two alternative specifications to examine the presence of this selection problem. Both models aim to explain why individuals are more likely to participate in the examination survey. First, we estimated the determinants of the weight and height (relative) biases through the use of a bivariate sample selection model.¹⁸ In both cases our estimations showed that the error terms of the participation and outcome equation were uncorrelated [Wald test = 0.05 ($p = 0.82$) for the relative weight bias and a Wald test = 0.23 ($p = 0.63$) for the relative height bias]. Similarly, the estimated coefficients of the inverse Mills ratio were not statistically significant. Second, we adjusted a two-part regression model specification or “hurdle model” assuming two separate processes underlying the censoring mechanism and the analyzed outcome. Again, our data confirmed that the coefficient of the participation regressor was not statistically significant. We therefore concluded that sample selection was not a problem in our data and hence OLS regression led to consistent parameter estimates of the relative bias equations.

Although we estimated the two self-reported relative biases or gaps separately by means of an OLS procedure, at this point we wondered whether common unobserved factors might affect both relative biases in weight and height. We therefore re-calculated our equations through the use of a SURE estimation method since all regressors could be considered exogenous. Note that this approach is useful when both residual terms are correlated, although the gain in efficiency is only considerable when this correlation is high. Then the estimated relative biases are,

$$\begin{aligned} \text{Bias}_{w,i} &= x_i'\beta + z_i'\gamma + \lambda_1 S_w + \varepsilon_{1,i} \\ \text{Bias}_{h,i} &= x_i'\delta + z_i'\gamma + \lambda_2 S_h + \varepsilon_{2,i} \end{aligned} \tag{2}$$

where x_i is the k -vector of explanatory variables; z_i denotes a vector of anthropometric measures; β , γ , δ and λ_1 and λ_2 are the k -vectors of unknown parameters; $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$

represent the random error terms which are assumed to be correlated and $S_w(S_h)$ represents the social norm regarding the weight (height) covariate proxied by the average self-reported weight (height) of the assigned reference group. Note we also enter the square of S_w and S_h in an attempt to test empirically the curvature of the social norm effect on the misreporting biases. Oswald (2008) reports evidence indicating that when individuals report their height they tend to use an approximately linear reporting function. A χ^2 test reporting the statistical significance of this correlation ($\text{Cov}(\varepsilon_{1,i}, \varepsilon_{2,i}) \neq 0$) confirms the validity of the SURE estimation although the gain in efficiency is small as the correlation is not very high. For this reason, we present only SURE estimations when reporting our empirical evidence.

Importantly, following Etilé (2007) note that the identification of the social norm effect hinges upon two key assumptions: (i) the non-linearity of the relationship between the social norm covariate and the variables defining group membership, which are also used as controls, although non-multicollinearity is needed to identify the model; and (ii) the exogeneity of the social norm measures.

3.4. Empirical evidence

Table 3 displays the SURE results for the determinants of the weight (first column) and height (second column) self-reporting (relative) biases.¹⁹ The social norm effect is defined by accounting for gender and age groups (baseline scenario). The estimated correlation coefficient between the error terms of the weight bias and the height bias is low (-0.12) but statistically significant. This correlation coefficient shows that there may be unobservable factors which are negatively related to misreporting weight and height. Note that weight and height biases were defined in opposite senses. The goodness of fit was reasonable (R^2 ranged from 0.11 to 0.18) given the microeconomic nature of our data.

Besides the presence of other expected covariate effects, the most interesting result is the finding of a social norms effect on individuals' reporting behavior regarding anthropometric measures. Interestingly, we find that social norms concerning an “ideal” weight exert a non-linear negative statistically significant effect on the self-reporting weight relative bias equation. That is, the greater the average “ideal” weight shared by the reference group, the lower the weight bias or the less inclined sample individuals are to under-report their weight. This result was statistically significant after controlling for the influence of a set of covariates. The estimated size of this weight norm effect is not negligible. Certainly, it is estimated that a 1% increase in the average weight of the reference group causes approximately a 22.3% reduction in the relative weight bias. Overall, this evidence confirms that social norms regarding “ideal” weight have an influence when individuals report their attributes (Etilé,

¹⁸ We introduced further covariates in the participation equation for identification reasons. Specifically, we included the year and month of the examination, health region, labor status, civil servant, type of labor contract, pregnancy, interviewer perceptions regarding the ease of obtaining responses and response sincerity, individuals reporting being the head of the family, and household size. Results are available from the authors upon request.

¹⁹ To account for the complex random design of the data we run the equations adjusting for sampling weights and the clustering of the data.

Table 3
SURE estimation of the weight and height self-reporting biases.

	Weight bias	Height bias
Social norms regarding weight	-0.0092 (0.00)**	-
Squared social norms regarding weight	0.0001 (0.00)**	-
Social norms regarding height	-	0.0005 (0.00)*
Measured BMI	0.0154 (0.00)***	0.0030 (0.00)***
Squared measured BMI	-0.0002 (0.00)***	0.0000 (0.00)***
Waist-to-hip ratio	0.0182 (0.01)**	-0.0202 (0.01)***
Waist circumference	0.0000 (0.00)	-0.0001 (0.00)
Male	0.0018 (0.01)	-0.0090 (0.00)**
Age	-0.0027 (0.00)***	-0.0005 (0.00)
Squared age	0.0028 (0.00)***	0.0007 (0.00)*
Single	-0.0023 (0.00)	-0.0017 (0.00)
Widow	-0.0017 (0.01)	0.0158 (0.01)**
Divorced-separated	0.0126 (0.01)	-0.0032 (0.00)
African	-0.0253 (0.02)	0.0067 (0.01)
Latin-American	-0.0032 (0.01)	-0.0013 (0.00)
Asian	0.0145 (0.01)	0.0234 (0.01)***
European-HIC	-0.0100 (0.02)	-0.0128 (0.00)***
Low education	-0.0040 (0.00)	-0.0033 (0.00)
Secondary education	-0.0008 (0.00)	-0.0005 (0.00)
Non deprivation	-0.0092 (0.02)	0.0004 (0.00)
Fair health	-0.0023 (0.00)	-0.0033 (0.00)**
Bad health	-0.0017 (0.01)	-0.0030 (0.00)
Chronic diseases	-0.0035 (0.00)***	0.0009 (0.00)**
Checking cholesterol	0.0068 (0.00)**	-0.0022 (0.00)
Medical visit	0.0077 (0.00)**	-0.0017 (0.00)*
Weight loss	-0.0310 (0.01)***	0.0000 (0.00)
Fat perception	-0.0098 (0.00)**	-0.0022 (0.00)
Interview duration	0.0001 (0.00)	0.0000 (0.00)
Time lag	0.0001 (0.00)**	0.0000 (0.00)
Weight (height) rounding	0.0081 (0.00)***	0.0016 (0.00)
N	1368	1368
F-test	7.92 (0.00)	4.43 (0.00)
R ²	0.1808	0.1072
Root MSE	0.054	0.018
$\rho(e_{1,i}, e_{2,i})$		-0.1163
χ^2 Breusch–Pagan test		18.49 (0.00)

Note: Weight bias is defined as the difference between measured and self-reported weight in relative terms whilst height bias is defined as the difference between self-reported and measured weight also in relative terms. Reference group accounts for gender and age groups. Estimations performed using sampling weights and adjusting for within-cluster correlation. Outliers for weight and height were excluded following Hadi (1992). These estimations included dichotomous variables for healthcare regions. Standard errors are displayed in brackets.

Source: CHS and CHES, 2006.

- * ($p < 0.1$).
- ** ($p < 0.05$).
- *** ($p < 0.01$).

2007; Powdthavee, 2009) and researchers should account for this effect when working with self-reported data. However, Table 3 indicates a negligible and hardly significant (at 10%) impact of social norms regarding an “ideal” height on the height bias.²⁰ Hereafter social norms on weight coefficients are the focus of the rest of the paper.

As long as we rely on a parametric identification of the social norm effect, we checked multicollinearity by means

of the Variance Inflation Factor for those variables defining group membership. Our results confirm that multicollinearity is not an issue and, hence, non-linearity is sufficient to identify the model.²¹ In addition, as a robustness check, alternative definitions of the reference group were used. Table 4 shows these results. The first two columns present the coefficient and the elasticity SURE estimates of the weight bias equation. Note we reproduced the baseline scenario in the first part of Table 4. In the second part we built reference group based on gender and educational attainment levels, i.e., excluding age groups. The presence of a negative and quadratic significant effect of the weight social norm on the relative weight bias was confirmed, and the size of this effect was stronger. Based on this new comparison group we found an elasticity figure of around -33.3%.²² At the bottom of the table we considered jointly the three features (gender, age groups and education). In this case we again found a negative and statistically significant but lower elasticity of -9.5%. Therefore, this would suggest that the narrower the definition of reference group the lower the impact of social norms on the (relative) weight bias. However, some caution is needed since the latter reference group split our database into excessively smaller groups which represents a lack of sample size representativeness. Summarizing, the evidence presented in the first two columns of Table 4 would ultimately reveal that our results are contingent on the definition of the social norm effect.

Note that one could even argue that education is endogenous casting some doubts or questioning the exogeneity of the social norm covariate. This may be the case if, for instance, the error term represents ability, self-control, personality or genetic traits that can influence both schooling (and, therefore, group membership) and obesity. Likewise, education could be endogenous as a consequence of heterogeneity in time preferences within educational groups (e.g., single individuals or married people without children do not display the same leisure preferences) or simply due to measurement errors. For this reason, we estimated through instrumental variable (IV) the self-reporting weight (relative) bias equation. The last two columns of Table 4 present the results. For the baseline scenario the social norms covariate is instrumented by means of the average self-reported weight of the adjacent reference group (Etilé, 2007), that is, those of the same gender with a differential age of no more than 5 years. As a result, we obtained two instruments for the endogenous covariate (i.e., the average weight of those above and below the individual age). The instruments were valid (Cragg–Donald Wald F statistic was considerably high:

²¹ For the weight bias equation the VIF values were: male (9.14); age (1.64), whereas the mean VIF value was 6.85. Multicollinearity was also absent in the subsequent definitions of group membership.

²² As an additional check we dropped educational attainment from the econometric specification (second part of Table 4) to avoid (potential) collinearity effects between our social norms effect and education as an exogenous covariate. Results were almost the same. Hence, the inclusion of education was not a problem in our econometric specification.

²⁰ Actually, this negligible impact of social norms on height is corroborated under alternative definitions of social norms.

Table 4
Estimation of the weight self-reporting bias: alternative social norms definitions.

	Weight bias: SURE estimation		Weight bias: IV estimation	
	Coeff.	Elasticity	Coeff.	Elasticity
<i>Reference group: age and gender (baseline)</i>				
Social norms regarding weight	−0.0092 (0.00)**	−22.30%	−0.0013 (0.00)***	−5.61%
Squared social norms regarding weight	0.0001 (0.00)**		–	
N	1368		1317	
R ²	0.1808		0.1588	
F-test (first stage) (p-value)	–		219.90 (0.00)	
Hansen statistic (p-value)	–		2.61 (0.11)	
<i>Reference group: gender and education</i>				
Social norms regarding weight	−0.0159 (0.00)***	−33.34%	−0.0013 (0.00)***	−5.49%
Squared social norms regarding weight	0.0001 (0.00)***		–	
N	1369		1365	
R ²	0.1792		0.1497	
F-test (first stage) (p-value)	–		7.36 (0.00)	
<i>Reference group: age and gender, and education</i>				
Social norms regarding weight	−0.0045 (0.00)**	−9.48%		
Squared social norms regarding weight	0.0000 (0.00)*			
N	1,358			
R ²	0.1755			

Note: Weight bias is defined as the difference between measured and self-reported weight in relative terms. Estimations performed using sampling weights and adjusting for within-cluster correlation. Outliers for weight and height were excluded following Hadi (1992). These estimations included dichotomous variables for healthcare regions. Standard errors are displayed in brackets.

Source: CHS and CHES, 2006.

* ($p < 0.1$).

** ($p < 0.05$).

*** ($p < 0.01$).

219.90, p -value = 0.00).²³ Our results show that the instrumented social norm weight effect is statistically significant and with the expected sign, although showing a lower impact since elasticity slumps to −5.6%.²⁴ Likewise, when gender and education are used to build the reference group and its corresponding social norm effect is instrumented using these characteristics of the adjacent reference group (i.e., one instrument is derived); we obtain a quite similar result (−5.5%).²⁵ Therefore, we can conclude that our previous weight social norm estimated effect, where the endogeneity of education was neglected, is somehow overestimated. Notwithstanding, this estimated effect remains considerably high compared to the remaining set of covariates, except the individual's BMI (12.51%).

The impact of several anthropometric measures on the self-reporting bias equations provided other interesting results (see Table 3). As expected, we find evidence of a positive and non-linear BMI effect on both the weight and the height relative bias. Thus, the higher the BMI the more likely individuals were to under-report their weight and

over-report their height. In the same vein, obese individuals with high levels of abdominal-visceral fat, proxied by the waist-to-hip ratio (WHR), were found to considerably under-report their weight. Likewise, the WHR was found to exert a negative influence on height reporting bias, possibly showing that those with more central adiposity were tall individuals who otherwise tended to report a lower relative height bias.

Regarding the effects of other covariates, Table 3 shows the existence of a clear-cut misreporting pattern according to age. Our findings revealed a negative and statistically significant non-linear age effect on the weight and height reporting bias. Consistently with some previous studies, the results document that men tend to under report their weight less than women. As expected, nationality played a notable role in the self-reporting behavior: Asian (Europeans and high income country people) individuals tended to over-report (under-report) significantly their height compared to Spaniards. Furthermore, health covariates (reporting a fair health status, number of chronic diseases, cholesterol checks and having visited a physician) seemed to play a significant role when explaining the bias magnitudes. Finally, the estimates show that those who have experienced a rapid weight loss or who self-perceive as obese people tended to under report less their weight.

4. Assessing the correction of self-reported data

Although several studies have documented the existence of substantial errors in self-reports of weight, height and BMI in developed countries, and there is an urgent

²³ At this stage, we computed an exogeneity test (Durbin-Wu-Hausman = 0.0399; p -value = 0.8416) indicating we cannot reject the null hypothesis of no endogeneity. The latter explains why OLS and SURE results do not differ. However, although the test indicates exogeneity, we preferred to instrumentalize the social norm impact as a consequence of the presence of convincing reasons to defend heterogeneous effects within educational levels.

²⁴ The estimated elasticity is roughly the same (−5.2%) when the definition of the reference adjacent group is extended to a 10 years range.

²⁵ Given the high number of restrictions in building the adjacent reference group when age, gender and education are used we decided to exclude the IV estimation of the social norm effect in Table 4.

need to ensure an adequate assessment of overweight and obesity across Europe (IOTF, 2005), to date the issue has not been investigated, as far as we know, in Spain's main health survey (the Spanish National Health Survey). Datasets of this kind usually lack measurements of weight and height or more accurate measures of obesity (e.g., waist circumference, hip circumference or WHR), and typically only include self-reported values. This section, therefore, has two primary aims: first, to quantify the magnitude of this reporting error or bias and to assess the degree of accuracy of self-reported anthropometric data in the latest SNHS conducted in 2006; and, second, to provide conversion formulas based on our merged dataset to enable social science researchers to use these estimations to calculate more accurate measures of weight and height when using their habitual datasets. The method we adopt for correcting reporting bias has been applied by many economists studying obesity (Cawley, 2000, 2004; Lakdawalla and Philipson, 2002; Chou et al., 2004 and Cawley and Burkhauser, 2008). Nonetheless, according to Plankey et al. (1997) correcting the reporting errors does not completely eliminate the errors obtained when using linear regression models. Consequently, measured variables are preferable when available.

In the first stage, the three panels of Table 5 report the results of the regressions of measured weight, height and

waist circumference by gender against self-reported anthropometric data, age and their corresponding squared terms (i.e., data commonly available in social science datasets) using our merged Catalan sample of adults from the CHS-CHES, 2006 (Section 2.1).²⁶ The R^2 scores of measured weight and height equations were relatively high (ranging from 0.73 to 0.88) indicating that the quadratic specification would fit the data quite well. Less valid, however, was the adjustment of the waist circumference (R^2 of 0.6 for males, 0.72 for females). The addition of more covariates, including education and civil status, to these econometric specifications barely improved the goodness of fit of the weight and height measurements.

In the second stage, the coefficients from these regressions were transferred to the sample of adults taken from the SNHS 2006 and, after multiplying the coefficients by the self-report values, we obtained measures of weight and height corrected for the reporting error.²⁷ Of course, this approach assumes data 'transferability', i.e., that the relationship between true and declared values is the same in both datasets (Cawley and Burkhauser, 2008). This is not an overly strong assumption to make as long as Catalan and Spanish societies can be considered to share similar customs, cultural and historical traits and economic and religious backgrounds. Table 6 shows the impact of this correcting procedure on the BMI. The second column displays the average BMI based on actual self-reporting information on weight and height, and the third column shows the estimated average BMI based on weight and height corrected for self-reporting. Interestingly, our results suggest that self-reported BMI values calculated using the latest available health survey (SNHS-2006) were underestimated for both genders, although the underestimation was much more pronounced for women. The differences between self-reported and corrected mean BMI were statistically significant at the 1% level.

For purely comparative purposes, the fourth column shows the corrected BMI values when applying the Cawley and Burkhauser (2008) estimated conversion coefficients for weight and height corresponding to the white male and female populations.²⁸ The figures suggested by these authors produced lower BMI values than our own method based on a Catalan sample (third column), which would suggest that the direct transfer of their estimations to the SNHS-2006 underestimates the BMI indicator.²⁹ Unfortunately, the SNHS 2006 did not include measurements to assess the validity of our BMI values corrected for self-reporting errors. While the sizes of the BMI differentials were relatively small, Table 6 showed that the correction of self-reports could have a significant impact on estimated

Table 5
Prediction of measured anthropometric data.

	MALE	FEMALE
<i>(A) Regression of measured weight on self-reported weight</i>		
Self-reported weight (kg)	1.8156***	0.9559***
Self-reported weight (kg) squared	-0.0050	0.0004
Age (years)	0.0482	0.0508
Age (years) squared	-0.0005	-0.0007*
Constant	-31.728	2.0064
No. of observations	947	967
R^2	0.82	0.88
<i>(B) Regression of measured height on self-reported height</i>		
Self-reported height (cm)	1.1036	-1.4495
Self-reported height (cm) squared	-0.0006	0.0070***
Age (years)	0.0852*	0.0913**
Age (years) squared	-0.0012**	-0.0019***
Constant	-0.1311	211.77**
No. of observations	936	957
R^2	0.77	0.73
<i>(C) Regression of measured waist circumference on self-reported weight and height</i>		
Self-reported weight (cm)	1.4529***	1.4089***
Self-reported weight (kg) squared	-0.0041*	-0.0036**
Self-reported height (cm)	-2.3858*	0.0389
Self-reported height (cm) squared	0.0059	-0.0009
Age (years)	0.3159***	-0.0041
Age (years) squared	-0.0011	0.0022***
Constant	232.26**	23.09
No. of observations	941	961
R^2	0.60	0.72

Note: The sample is formed by individuals aged 16 and over. Statistical inference based on cluster standard errors.

Source: CHS and CHES, 2006.

* ($p < 0.1$).

** ($p < 0.05$).

*** ($p < 0.01$).

²⁶ We deliberately excluded individuals aged 15 and under from the combined dataset since the adult sample of the SNHS 2006 started at the age of 16.

²⁷ Further details of the SNHS 2006, in which data are also collected in face-to-face interviews, can be accessed at <http://www.ine.es>.

²⁸ Cf. Tables 6 and 7 of Cawley and Burkhauser (2008). Their values in pounds and inches were converted into kilograms and centimeters (1 lb = 0.4536 kg; 1 in = 2.54 cm).

²⁹ Again, the differences between the mean BMI values in columns three and four were statistically significant at the 1% level.

Table 6
Actual and corrected mean BMI and obesity rate in the Spanish survey.

	<i>N</i>	Actual BMI based on self-reports	Corrected BMI for self-reporting	Corrected BMI Cawley and Burkhauser (2008)	Actual obesity based on self-reports	Corrected obesity for self-reporting	Corrected obesity Cawley and Burkhauser (2008)
Male	10,894	26.28 (3.97)	26.59 (3.86)	26.38 (4.05)	15.19% (0.36)	17.48% (0.38)	15.97% (0.37)
Female	15,310	25.08 (4.71)	26.06 (4.88)	25.85 (5.04)	14.85% (0.36)	19.60% (0.37)	19.36% (0.40)
Total	26,204	25.69 (4.39)	26.32 (4.40)	26.12 (4.57)	15.02% (0.36)	18.53% (0.40)	17.65% (0.38)

Note: Estimated means computed using sampling weights. Standard deviations are reported in brackets. Source: SNHS (Spanish National Health Survey) 2006.

Table 7
Estimated mean waist circumference (SNHS-2006).

	Male	Female
Waist circumference (cm)	96.29 (10.40)	86.74 (12.15)
“High Risk” [WC ≥ 102 cm (88 cm) in males (females)]	14.37%	21.21%
No. of Observations	10,894	15,310

Note: Estimated means computed using sampling weights. Standard deviations are reported in brackets.

obesity prevalence rates (columns 5–7). That is, correcting for the reporting error would result in an increase in the Spanish obesity prevalence of 3.5 percentage points (from 15% to 18.5%), while with the Cawley and Burkhauser (2008) approach the prevalence would not rise quite as much (to 17.6%). As expected, our findings indicated that the degree of underestimation in the obesity prevalence rate for women would be much greater if only self-reports on weight and height were used.

Finally, given that the fit of the measured waist circumference equation in Table 5 (panel C) was relatively weak, Table 7 tentatively shows the estimation of this more complete measure of fatness (central adiposity) for the Spanish adult population (aged 16 and over). Interestingly, average waist circumference was higher in males than in females (96.29 vs. 86.74). Using the US National Institutes of Health’s (NIH) threshold levels of 102 cm (40 in.) for men or 88 cm (35 in.) for women, it was estimated that 14.4% of the Spanish adult male population and 21.2% of the adult female population presented a relatively “high risk” of suffering type II diabetes, hypertension, and cardiovascular diseases.

5. Summarizing remarks

The use of social science databases containing declared or reported information on, among other items, individuals’ weight and height suffers from a serious problem: the existence of bias which may raise doubts about the calculations of the BMI and, hence, obesity assessment. This is one of the reasons why the medical profession has generally preferred alternative and more accurate proxies (e.g., the waist and hip circumference or the waist-to-hip ratio) to calculate the prevalence of obesity. Provided these alternative indicators are obtained through physical explorations, they are not conditioned by the same kind of biases as those using

self-declared weight and height information. However, the high cost of gathering a more complete set of information for large population samples explains the use of self-reported data by social scientists.

This paper has first of all examined the size of the self-reported weight and height biases with reference to a combined health and examination survey of the Catalan population (Catalan Health Survey 2006 and Catalan Health Examination Survey, 2006), given the lack of availability of a similar dataset for the whole Spanish population. Interestingly, we found that women, university graduates, subjects with a healthy self-assessed health status and the obese tended to underestimate their weight much more significantly than their counterparts, whereas height was overestimated by women, older age groups, lower educated individuals, those with healthy self-assessed health status and obese individuals. BMI was to a large extent underestimated by women, the older age groups and the obese. Ultimately, these results suggest that the self-reporting anthropometric data pattern found in our data (which could be extended to the whole of Spain) resembles that found in the literature for other developed countries.

More importantly, the paper has sought to examine the underlying determinants of the biases in this weight and height reporting behavior, and, among other covariates, the role played by social norms regarding “ideal” weight and height (Etilé, 2007; Oswald and Powdthavee, 2007; and Powdthavee, 2009). Interestingly, after taking into account the existence of sample selection problems, SURE estimations showed that social norms regarding “ideal” weight (proxied by the average weight of a reference group based on gender and age) exerted a non-linear, negative and statistically significant effect on the self-reporting weight (relative) bias equation, even after controlling for a set of covariates. This finding is, however, less intense after accounting for the plausible endogeneity nature of the social norm covariate. Therefore, the greater the average weight of the reference group recorded, the lower was the weight bias or the less inclined individuals were to under-report their weight. This finding seems to suggest that when individuals feel more satisfied with their own body image or are closer to the body weight of their peers they are less prone to misreport their weight. However, we found no evidence of a similar impact caused by the height social norm. Unfortunately, the magnitude and economic relevance of this effect is difficult to evaluate, since our results are in someway contingent on the definition of the

social interaction effect. As expected, we also found that other covariates (i.e., the BMI, waist-to-hip ratio, age, gender, nationality, education and health and health-care variables) significantly affected self-reported weight and height biases.

Finally, the paper has provided evidence to indicate that both the self-declared BMI and the obesity rate of the Spanish population were significantly underestimated, with this degree of misreporting being more prevalent among women. Therefore, the correction for the weight and height self-reporting errors would imply an increase in the Spanish obesity rate of 3.5 percentage points (from 15% to 18.5%). Interestingly, our correction procedure rendered a higher underestimation of the obesity rate compared to the more widely reported Cawley and Burkhauser (2008) conversion approach.

Acknowledgements

We thank H. Benítez-Silva, J. Costa-Font, P. García-Gómez, S. Jiménez-Martín, A. López-Nicolás, J. Pinilla, A. Sanz-de-Galdeano, A. Stoyanova, and participants at the XXVIII Jornadas de Economía de la Salud, the II-Workshop on Health Economics FEDEA-La Caixa, and the 7th European Conference on Health Economics for their helpful comments to a previous draft. We also thank the Department of Health of the Catalan Government for having provided us with the 2006 editions of the Catalan Health Survey & the Catalan Health Examination Survey. Institutional support from the Ministry of Science & Technology (Project ECO2008-04997/ECON) and the Catalan Government Science Network (Project 2009-SGR-359 and Project 2009-SGR-102) are also gratefully acknowledged. Finally, we gratefully acknowledge the interesting comments of the four anonymous referees. The usual disclaimers apply.

References

- Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. *Quarterly Journal of Economics* 115, 715–754.
- Aromaa, A., Koponen, P., Tafforeau, J., Vermeire, C., 2003. Evaluation of health interview surveys and health examination surveys in the EU. *European Journal of Public Health* 13 (Suppl. 3), 67–72.
- Averett, S., Korenman, S., 1996. The economic reality of the beauty myth. *Journal of Human Resources* 31, 304–330.
- Baron-Epel, O., Kaplan, G., 2001. General subjective health status or age-related subjective health status: does it make a difference? *Social Science & Medicine* 53, 1373–1381.
- Blanchflower, D.G., Oswald, A.J., van Landeghem, B., 2009. Imitative obesity and relative utility. IZA Discussion Paper Series No. 4010.
- Blume, L., Durlauf, S.N., 2006. Identifying social interactions: a review. In: Oakes, J., Kaufman, J.S. (Eds.), *Methods in Social Epidemiology*. Jossey-Bass, San Francisco, pp. 287–315.
- Bolton-Smith, C., Woodward, M., Tunstall-Perdoe, H., Morrison, C., 2000. Accuracy of the estimated prevalence of obesity from self-reported height and weight in an adult Scottish population. *Journal of Epidemiology Community Health* 54, 143–148.
- Boström, G., Diderichsen, F., 1997. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *International Journal of Epidemiology* 26 (4), 860–866.
- Brock, W.A., Durlauf, S.N., 2001. Discrete choice with social interactions. *The Review of Economic Studies* 68 (2), 235–260.
- Burke, M.A., Heiland, F., 2007. Social dynamics of obesity. *Economic Inquiry* 45 (3), 571–591.
- Cawley, J., 2000. An instrumental variables approach to measuring the effect of body weight on employment disability. *Health Services Research* 35 (5), 1159–1179.
- Cawley, J., 2004. The impact of obesity on wages. *Journal of Human Resources* 39 (2), 451–474.
- Cawley, J., Burkhauser, R.V., 2008. Beyond BMI: the value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics* 27 (2), 519–529.
- Chou, S.-Y., Grossman, M., Saffer, H., 2004. An economic analysis of adult obesity: results from the Behavioural Risk Factor Surveillance System. *Journal of Health Economics* 23, 565–587.
- Costa-Font, J., Gil, J., 2008. What lies behind socio-economic inequalities in obesity in Spain? A decomposition approach. *Food Policy* 33, 61–73.
- Danubio, M.E., Miranda, G., Vinciguerra, M.G., Vecchi, E., Rufo, F., 2008. Comparison of self-reported and measured height and weight: implications for obesity research among young adults. *Economics and Human Biology* 6 (1), 181–190.
- Etilé, F., 2007. Social norms, ideal body weight and food attitudes. *Health Economics* 16 (9), 945–966.
- Gillum, R.F., Sempos, C.T., 2005. Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the NHANES III. *Nutrition Journal* 4 (27), 1–8.
- Groot, W., 2000. Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics* 19, 403–420.
- Gunnell, D., Berney, L., Holland, P., Maynard, M., Blane, D., Frankel, S., Davey, G., 2000. How accurately are height, weight and length reported by the elderly, and how closely are they related to measurements recorded in childhood? *International Journal of Epidemiology* 29 (3), 456–464.
- Hadi, A.S., 1992. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B* 54 (3), 761–771.
- Hernández-Quevedo, C., Jones, A.M., Rice, N., 2005. Reporting bias and heterogeneity in self-assessed health. Evidence from the BHPS. The University of York, HEDG Working Paper 05/04.
- International Obesity Task Force, 2005. Obesity in Europe. EU Platform Briefing Paper. IOTF, Brussels, March.
- Kuczmarski, M.F., Kuczmarski, R.J., Najjar, M., 2001. Effects of age on validity of self-reported height, weight and body mass index: findings from the Third NHANES 1988–1994. *Journal of the American Dietetic Association* 101 (1), 28–34.
- Lakdawalla, D., Philipson, T., 2002. The growth of obesity and technological change: a theoretical and empirical investigation. NBER WP No. 8965.
- Lindeboom, M., van Doorslaer, E., 2004. Cut-point shift and index shift in self-reported health. *Journal of Health Economics* 23, 1083–1099.
- Lindeboom, M., 2006. Health and work of older workers. In: Jones, A.M. (Ed.), *Elgar Companion to Health Economics*. Edward Elgar, Cheltenham.
- Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies* 60, 531–542.
- Morris, S., 2006. Body mass index and occupational attainment. *Journal of Health Economics* 25, 347–364.
- NIH, 1998. Clinical Guidelines on the Identification, Evaluation and Treatment of Overweight and Obesity in Adults: The Evidence Report. National Institutes of Health Publication No. 98-4083 (NIH: Washington, D.C.).
- Niedhammer, I., Bugel, I., Bonenfant, S., Goldberg, M., Leclerc, A., 2000. Validity of self-reported weight and height in the French GAZEL cohort. *International Journal of Obesity* 24 (9), 1111–1118.
- Oswald, A.J., Powdthavee, N., 2007. Obesity, unhappiness and the challenge of affluence: theory and evidence. *Economic Journal* 117, F441–454.
- Oswald, A.J., 2008. On the curvature of the reporting function from objective reality to subjective feelings. *Economics Letters* 100, 369–372.
- Paccaud, F., Wietlisbach, V., Rickenbach, M., 2001. Body mass index: comparing mean values and prevalence rates from telephone and examination surveys. *Revue Epidémiologique et de Santé Publique* 49 (1), 33–40.
- Pischon, T., et al., 2008. General and abdominal adiposity and risk of death in Europe. *New England Journal of Medicine* 359 (20), 2105–2120.
- Plankey, M.W., Stevens, J., Flegal, K.M., Rust, P.F., 1997. Prediction equations do not eliminate systematic reporting error in self-reported body mass index. *Obesity Research* 5 (4), 308–314.
- Powdthavee, N., 2009. Ill-health as a household norm: evidence from other people's health problems. *Social Science & Medicine* 68, 251–259.

- Quiles, J., Vioque, J., 1996. Validez de los datos antropométricos declarados para la determinación de la prevalencia de la obesidad. *Medicina Clínica* 106, 725–729.
- Sobal, J., Stunkard, A.J., 1989. Socioeconomic status and obesity: a review of the literature. *Psychological Bulletin* 105, 260–275.
- Spencer, E.A., Appleby, P.N., Davey, G.K., Key, T.J., 2002. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutrition* 5 (4), 561–565.
- World Health Organization-WHO, 2000. Obesity: Preventing and Managing the Global Epidemic. Report of a WHO consultation. Technical Report Series No. 894.
- Zhang, Q., Wang, Y., 2004. Socioeconomic inequalities of obesity in the United States: do gender, age and ethnicity matter? *Social Science & Medicine* 58, 1171–1180.