# A model of errors in BMI based on self-reported and measured anthropometrics with evidence from Brazilian data

**Apostolos Davillas†**
University of Macedonia;
IZA, Bonn.


**Victor Hugo de Oliveira**
Instituto de Pesquisa e Estratégia Econômica do Ceará, Brazil


**Andrew M Jones**
Department of Economics and Related Studies,
University of York

### Abstract

The economics of obesity literature implicitly assumes that measured anthropometrics are error-free and they are often treated as a gold standard when compared to self-reported data. We use factor mixture models to analyse measurement error in both self-reported and measured anthropometrics with nationally representative data from the 2013 National Health Survey in Brazil. A small but statistically significant fraction of measured anthropometrics are attributed to recording errors. While, as they are imprecisely recorded and due to reporting behaviour, only between 10% and 23% of our self-reported anthropometrics are free from any measurement error. Post-estimation analysis allows us to calculate hybrid anthropometric predictions that best approximate the true body weight and height distribution. BMI distributions based on the hybrid measures do not differ between our factor mixture models, with and without covariates, and are generally close to those based on measured data. While BMI based on self-reported data under-estimates the true BMI distribution. "Corrected self-reported BMI" measures, based on common methods to mitigate reporting error in self-reports using predictions from corrective equations, do not seem to be a good alternative to our "hybrid" BMI measures. Analysis of regression models for the association between BMI and health care utilization shows only small differences, concentrated at the far-right tails of the BMI distribution, when they are based on our hybrid measure as opposed to measured BMI. However more pronounced differences are observed, at the lower and higher tails of BMI, when these are compared to self-reported or "corrected self-reported" BMI.

† Corresponding author. University of Macedonia, 156 Egnatia Street, GR-54636 Thessaloniki, Greece; a.davillas@uom.edu.gr

# 1. Introduction

Obesity is a strong predictor of overall mortality (Li et al. 2021; Prospective Studies Collaboration et al. 2009) and an important risk factor for several noncommunicable diseases such as cardiovascular diseases, diabetes, musculoskeletal disorder, and some cancers (Lin et al. 2020). A large literature has explored the economic and social ramifications of obesity, such as poorer labour market outcomes, increased health care utilization and associated public health costs (e.g., Cawley 2004; Cawley 2015; Rooth 2009). Moreover, studies have investigated and measured socioeconomic inequalities in obesity (e.g., Bilger et al. 2017; Davillas and Benzeval 2016; Zhang and Wang 2004).

Despite an influential report on the importance of physically measured health indicators for understanding how the social and economic environment may get under the skin, several multi-purpose social science datasets continue to collect only self-reported weight and height data (Cawley 2015). Some existing studies do use datasets that collect measured anthropometrics, often in addition to self-reported anthropometric data (e.g., Cawley 2015; Cawley et al. 2015; Davillas and Jones 2021; Gil and Mora 2011). Studies that analyse measurement error in anthropometric data typically compare self-reports and measured anthropometric data; this research explicitly assumes that measured anthropometric data are error-free "gold-standard" measures. Specifically, Cawley et al. (2015), using the US National Health and Nutrition Examination Survey (NHANES) for the period between 2003–2010, compare self-reports with measured weight and height data. They find that reporting error in self-reported data is non-classical, with those who are underweight, based on measured anthropometrics, tending to over-report and overweight and obese respondents tending to under-report their weight. Gil and Mora (2011) use data from the 2006 Catalan Health and Health Examination Surveys and compare self-reported with measured anthropometrics. They find that social norms regarding ideal weight may affect reporting bias in self-reported anthropometrics. Those respondents who are more satisfied with their own body image are less prone to under-report their weight, although these results are subject to the definition of social norms on body image.

Using UK data, Davillas and Jones (2021) conducted an experiment to explore the extent of measurement error in body mass index (BMI), when self-reported body weight and height data are compared to measured anthropometrics. This study shows non-classical reporting error in height and weight, taller people seem to report their height more accurately, and a sharp increase in reporting errors in self-reported body weight for

1

those of greater measured weight. Further analysis shows that heterogeneity in self-reported anthropometrics is associated with within-household measured BMI data. A study employing Swedish data (Ljungvall et al. 2015) finds the presence of reporting error in self-reports of BMI (called misreporting in the study), when compared to measured anthropometric data, and that there is systematic social patterning in misreporting which matters for estimation of the education and income gradients in BMI when based on self-reports. O'Neill and Sweetman (2013), using a selective sample of mothers from the Irish Cohort Study, find that self-reported BMI, as compared to measured anthropometrics, is subject to substantial measurement error, which also causes an overestimation of the relationship between BMI and income. Finally, the role of interviewers is examined by Olbrich et al. (2022). Using various datasets from the USA, UK and Germany this study shows that interviewers play an important role in differences between reported and measured body height data as well as on the changes in reported height over survey waves.

Over and above the fact that these existing studies assume that measured anthropometrics are error-free, they mostly compare self-reports and measured anthropometric data that were collected with a considerable time difference or where respondents were informed about the subsequent physical measurements (Cawley et al. 2015; Gil and Mora 2011). The related medical literature is often based on selected age groups, non-representative samples and neither aims to characterise the measurement error nor to quantify the implications of the measurement error for economic modelling (e.g., Engstrom et al. 2003; Gorber et al. 2007; Keith et al. 2011).

Despite the fact that measured anthropometrics are assumed to be error-free in much of the existing literature, the accuracy of measured anthropometrics may indeed be affected by several factors. For instance, recent evidence has documented the influence of interviewers on reliability of measured and self-reported body height data in different surveys (e.g., Finn and Ranchhod 2017; Olbrich et al. 2022). Potential sources of measurement error include both unintentional (such as accidental recording errors from measurement equipment to survey materials) and intentional (i.e., fabricating parts of the measurement or even conducting physical measurements on the wrong respondent) recording errors (e.g., Finn and Ranchhod 2017; Groves 2005; Olbrich et al. 2022). These may not be easy to detect if the interviewers visited the household to conduct the interview (Olbrich et al. 2022). Moreover, in some datasets, interviewers may visit the household more than once to complete a socioeconomic questionnaire and collect physical measurements, including anthropometrics; this increases the likelihood that mis-

identification takes place. More broadly, the literature has discussed the presence of measurement error in more objectively measured nurse-collected and blood-based health data (Davillas and Pudney 2020a,b). These studies use latent variable models to account for measurement error, but they do not aim to explicitly model measurement error or to explore its potential implications for economic models. Overall, there is limited research that has access to both self-reported and measured anthropometric data collected within the same survey wave.

Our paper contributes to the literature in various ways. We model potential measurement error in both self-reported and measured anthropometrics (i.e., body weight and body height). We use data from the 2013 National Health Survey (Pesquisa Nacional de Saúde; PNS-2013) of Brazil, which is a nationally representative dataset that allows for measured and self-reported data on body weight and height to be collected from the same individuals within the span of a household interview. In Brazil, obesity has systematically increased since the 2010s, with one in every five adults experiencing obesity (Triaca et al. 2020). Projections of the obesity-related costs in Brazil show that the annual health care costs may double from 2010 ($5.8 billion) to 2050 ($10.1 billion) — a total health care cost of $330 billion over 40 years (Rtveladze et al. 2013). As such, obesity is an important public health concern for Brazil.

To analyse measurement error in the Brazilian data we use a factor mixture model, initially proposed by Kapteyn and Ypma (2007). This Kapteyn and Ypma (KY) factor mixture model is applied and extended by Jenkins and Rios-Avila (2020) and Jenkins and Rios-Avila (2021, 2023a) to analyse measurement error in self-reported and administrative income data. To the best of our knowledge, the KY factor mixture model has not been used to analyse measurement error in self-reported and measured anthropometric data. Unlike the existing literature, that assumes no measurement error in measured body weight and height data, our analysis allows us to model different types of errors in both self-reported and measured anthropometrics. Specifically, we test the hypothesis that measured anthropometrics encompass data-recording errors. Moreover, the self-reported anthropometric data are assumed to be subject to a wider set of measurement errors. These include the precision of the scale for the self-reported data, which are only recorded as whole numbers (in cm or Kg), non-classical mean-reverting errors, and other types of remaining errors. As permitted by our data, we also estimate factor mixture models that account for individual-level covariates to explore the extent to which true latent anthropometrics as well as reporting error in self-reported anthropometrics, and their dispersion, may vary across population groups. Absence of

interviewer-level data, however, prevents us from exploring heterogeneity in measured anthropometrics due to interviewer characteristics.

Our analysis also allows us to estimate the probability of each type of measurement error in both self-reported and measured data. Of particular interest, given that measured anthropometric data are often considered error-free (e.g., Cawley 2015; Davillas and Jones 2021; Gil and Mora 2011), our results suggest that a small but systematic fraction of measured anthropometrics contain data recording errors. Turning to self-reported weight and height, the estimated probabilities that the self-reported anthropometrics equal the true body weight and height are relatively low, at 10% and 23% respectively.

Post-estimation analysis allows us to generate a set of predictions of the distribution of the true latent weight and height data that combine information from both self-reported and measured anthropometrics. Based on reliability measures and mean squared errors, estimated using simulated out of the sample predictions, we select the best performing predictions of latent weight and height distributions. After choosing our preferred prediction, for our factor mixture models with and without covariates, our sample data are used to compute body weight and height measures that approximate the true values; these are then used to calculate our proxies of the true BMI distribution.

Finally, we compare the distributions of BMI using self-reported, measured and our proxies of true BMI; the latter are very close to the distribution of BMI based on measured anthropometrics, while the BMI based on self-reported data under-estimates the true BMI distribution. We also employ the "corrected self-reported BMI" as an additional measure — a conventional measure used in the existing economics of obesity literature to correct self-reported data for reporting error (Cawley 2015). Our results show that these "corrected self-reported BMI" measures are not a good alternative to our "hybrid" BMI measures. In addition, we provide evidence to explore the potential implications of the measurement error in both self-reported and measured anthropometrics. As an illustration, we compare results when each of the self-reported, "corrected self-reported", measured and hybrid BMI measures are used as explanatory variables in linear regression models for the frequency of hospital admissions in the past 12 months. We find only moderate differences in the results between the hybrid BMI measure and the one based on measured anthropometrics and these are concentrated in the far-right tails of the BMI distribution. More pronounced disparities are observed, at the lower and higher BMI tails, when our hybrid measures are compared with BMI measures based on self-reported or "corrected self-reported" data.

Understanding and characterizing measurement error in both self-reported and measured anthropometrics has important public health implications. Self-reported and/or physical measurements of anthropometrics are collected in several nationally representative surveys. For example, the Survey of Health, Ageing and Retirement in Europe (SHARE), the European Community Household Panel (ECHP), the German Socio-Economic Panel (GSOEP) as well as the National Longitudinal Survey of Youth (NLSY), the Medical Expenditure Panel Survey (MEPS) and the Behavioural Risk Factor Surveillance System (BRFSS) are datasets that are frequently used for obesity research but are limited to self-reports of body weight and height data. Recent advances to survey measurement allow for measured anthropometrics to be collected as part of multi-purpose social science surveys to improve data reliability on anthropometric measurement (Cawley et al. 2015). Data from nationally representative surveys are used to estimate obesity prevalence at the national level as well as for international comparisons (Ng et al. 2014). Measurement errors that may contaminate both self-reported and measured anthropometrics may affect within and between country and region comparisons of obesity prevalence and estimates of the population at increased health risks. Depending on the size of the measurement error, this may mislead potential public policies to mitigate regional or cross-country differences in excess body weight. Moreover, studies that quantify the (public) health care costs associated with obesity and related diseases often rely on survey data (Cawley and Meyerhoefer 2012); this research is influential and is used to justify government programmes to prevent obesity on the grounds of external costs (USDHHS 2010). The extent to which these estimates may be biased due to measurement error in measured and/or self-reported anthropometrics collected in surveys is of relevance from a public health point of view given the cost savings from reducing obesity prevalence.

The rest of the paper is organized as follows. Section 2 present the methods used to analyse measurement error in both self-reported and measured anthropometric data. Our data source and descriptive statistics are presented in Section 3. The results of our analysis, post-estimation predictions and a preliminary analysis of the potential implications on measurement error in both self-reported and measured anthropometrics for economic research are presented in Section 4. Section 5 concludes and provides a summary of our findings.

## 2. Methods

We adapt the factor mixture model, proposed by Kapteyn and Ypma (2007), to model the relationship between measured and self-reported anthropometrics. This model has been applied and extended by Jenkins and Rios-Avila (2020) and Jenkins and Rios-Avila (2021, 2023a) to analyse measurement error in income data. In this study, we apply the KY model to measurement error in both self-reported and measured anthropometric data, on weight and height, using the 2013 National Health Survey of Brazil.

We assume that the true values of each anthropometric measure (weight or height) for an individual $i$ ($\xi_i$) are unobserved, but we can observe both measured ($r_i$) and self-reported ($s_i$) anthropometrics. Table 1 provides a description of the types of errors in measured and self-reported anthropometric data that can be captured by our factor mixture model.

Measured anthropometrics are collected at the end of the individual questionnaire in our dataset. According to the survey protocol, it is possible for the individual interview to be completed in more than one visit, so it may be the case that physical measurements (which are time consuming as they include anthropometrics and blood pressure measurements) may not take place on the same day. Also, the measured anthropometrics are recorded by the interviewer by hand in the survey materials. Thus, measured anthropometrics may suffer from (unintentional or intentional) recording error related to entering values from the measurement equipment to the survey materials, fabrication of the measurement of anthropometrics by the interviewer[1] or even physical measurements taken from the wrong household member (especially if the main interview and physical measurements are not collected on the same day). These measurement errors, although they may occur with low frequency, could have a non-negligible impact on data reliability.

Thus, in the case of measured anthropometrics, we assume that the distribution of each anthropometric measure is a mixture of two types of observation:

$$r_i = \begin{cases} \xi_i & \text{with probability } \pi_r \\ \zeta_i & \text{with probability } (1 - \pi_r) \end{cases} \tag{1}$$

---

[1] These fabrication errors (if they exist) are unlikely to result in mean reversion/mean divergence but may be fairly random errors. Existing studies have shown evidence of misperception of body size (Zelenytė et al. 2021), suggesting that interviewers may not be able to accurately predict participants' body weight/height (if not measured) and, thus, not be able to make guesses that may lead to mean reversion/mean divergence (i.e. guesswork that is strongly correlated with true body weight and height).

where, measured anthropometrics ($r_i$) equals the true value with probability $\pi_r$ (case R1). However, measured anthropometrics may be not equal to the true value for certain respondents with probability $1 - \pi_r$ (case R2); thus, an error-ridden measure ($\zeta_i$) is observed in this case. In the spirit of the KY factor mixture model, this erroneous anthropometric measure, which is incorrectly attributed to individual $i$, is denoted by $\zeta_i$.[2] The true values and those with recording errors are both assumed to be independently and identically normally distributed: $\xi_i \sim N(\mu_\xi, \sigma_\xi^2)$, $\zeta_i \sim N(\mu_\zeta, \sigma_\zeta^2)$; this implies that the marginal distribution of $r_i$ is a mixture of two normals. Given the type of errors that are captured by $\zeta_i$, as described above, we assume that there is no correlation between $\xi_i$ and $\zeta_i$[3]. The assumption that the erroneous measurements are uncorrelated with the true values contributes to the identification of the full model as it implies that these measurements are also uncorrelated with the self-reported anthropometrics.

Each of our self-reported anthropometrics (i.e., weight or height) are assumed to be a mixture of three types of observation:

$$
s_i = \begin{cases}
\xi_i & \text{with probability } \pi_s \\
\xi_i + \eta_i + \rho(\xi_i - \mu_\xi) & \text{with probability } (1 - \pi_s)(1 - \pi_\omega) \\
\xi_i + \eta_i + \rho(\xi_i - \mu_\xi) + \omega_i & \text{with probability } (1 - \pi_s)\pi_\omega
\end{cases}
\tag{2}
$$

Table 1 describes all sources of measurement errors in self-reported anthropometrics that are captured in equation 2. Specifically, we assume that the self-reported anthropometrics ($s_i$) equals the true latent value ($\xi_i$) with probability $\pi_s$ (case S1). The self-reported values are recorded as integers so this case only applies when the true value

---

[2] The factor mixture measurement error model proposed by Kapteyn and Ypma (2007) assumes that observed administrative income data are a mixture of correct matches and mis-matches (with survey data). However they argue that, over and above potential mismatches in the linkage between administrative and survey data, it is also likely that administrative and survey data may capture conceptually different things. As such, they argue that there is no loss of generality to assume that measurement error in administrative data may reflect different sources. Analogously, in our analysis measurement error in measured anthropometrics may reflect different sources (as described above); in particular interviewers' errors related to entering values from the measurement equipment to the survey materials, fabrication of the measurement of anthropometrics by the interviewer or even physical measurements for the wrong household member.

[3] Even in the case of fabricated interviews or when anthropometric measurement is not conducted for the intended respondent, this may be a strong assumption if quality control takes place. However, there is no such quality control undertaken in the dataset used in our analysis (as well as in many other multipurpose social science datasets that collect anthropometrics).

is a whole number[4]. Otherwise (cases S2 and S3), there must be some imprecision in $s_i$ due the scale of measurement. This imprecision, reflecting the different ways in which respondents may round their responses to whole numbers along with random noise in the self-reports, is captured by the error term $\eta_i$. This error is independent of the true value ($\xi_i$). In addition, we allow for the possibility of non-classical mean-reverting (or mean-diverging) error (survey measurement error, which is captured by term $\rho(\xi_i - \mu_\xi)$.[5] Existing studies comparing measured with self-reported data have shown the presence of mean-reverting errors in self-reported body weight (Cawley et al. 2015). The second case (S2), which allows for both sources of error, occurs with probability $(1 - \pi_s)(1 - \pi_\omega)$. The third case (S3), which occurs with probability $(1 - \pi_s)\pi_w$, adds a third source of measurement error ($\omega_i$) to allow for additional random noise that may occur in some observations who make additional errors in their self-assessments of height or weight (see Table 1). The measurement errors are both assumed to be independently and identically normally distributed: $\eta_i \sim N(\mu_\eta, \sigma_\eta^2)$, and $\omega_i \sim N(\mu_\omega, \sigma_\omega^2)$.

Note that the survey team undertook significant effort to minimize the risk of equipment failure for physical anthropometric measurements; our dataset employs international measurement protocols and validated equipment, which is calibrated daily to ensure reliability of the measurements. The procedures for taking anthropometric measures are defined to prevent biologically inaccurate measures and were done in partnership with the Laboratory for Nutritional Evaluation of Populations (LANPOP), part of the Public Health School in the University of São Paulo (Damacena et al. 2015; Szwarcwald et al. 2014). Also, the availability of two repeated physical measurements of body weight and height (we took the second measure for our main estimation and, for sensitivity analysis, the average of these measures) further reduces the likelihood of errors related to equipment failure. Thus, we do not capture this potential source of error in our factor mixture models.[6]

---

[4] Self-reported anthropometrics are collected as integer values (cm for height and Kg for weight), while the corresponding measured values are measured to one decimal point. In those cases where the respondent provided a non-integer value of their self-reported body weight and/or height (for example 61.5Kg), the interviewer recorded an integer value (such as 61Kg or 62Kg).

[5] Mean reversion ($\rho < 0$) means that respondents with high (low) values of true anthropometric measures, relative to the true mean, tend to under-report (over-report) their body weight and height in self-reports; the opposite is the case for mean divergence ($\rho > 0$).

[6] Moreover, one may argue that survey mode may influence measurement error in self-reported anthropometrics. For example, social desirability bias is much lower in the case of self-completion as opposed to the open interview (Bowling, 2005); thus, assuming that being taller and not of excess weight is more socially desirable, shorter people and those with excess weight may have distinct reporting patterns across collection modes. However, existing studies do not confirm the presence of such influences in reporting errors. Davillas and Jones (2021) find that measurement errors in

Table 1. Types of measurement error and their sources used in the factor mixture models

| Type of error | Sources |
|---|---|
| **Self-reported anthropometrics** | |
| Imprecision error | The survey questionnaire on self-reported anthropometric data requests the participants to respond to body weight and height questions with a precision of integer values (cm for height and Kg for weight). The precision error reflects the different ways in which respondents may round their responses to whole numbers (integers) along with random noise in the self-reports. Moreover, in those cases where the respondent provided a non-integer value of their self-reported body weight and/or height (for example 61.5Kg), the interviewer recorded an integer value (such as 61Kg or 62Kg). |
| Mean-reverting (or mean-diverging) error | Mean reversion means that respondents with high (low) values of true anthropometric measures, relative to the true mean, tend to under-report (over-report) their anthropometrics in self-reports; the opposite is the case for mean divergence. Existing studies have shown evidence of mean reversion in reporting errors in weight and height data (e.g., Cawley et al. 2015). |
| Additional random error | Captures any potential additional random noise (over and above imprecision and mean-reverting errors) that may occur for those who make additional errors in their self-assessments of height/weight. These errors may capture respondent's lack of awareness of their true anthropometrics. |
| **Measured anthropometrics** | |
| Unintentional or intentional recording errors | Recording errors in measured anthropometrics may capture errors related to entering values from the measurement equipment to the survey materials, fabrication of the measurement of anthropometrics by the interviewer or even physical measurements taken from the wrong household members (especially if the main interview and physical measurements are not collected the same day). |

The full KY model defines a mixture of six latent classes that correspond to the combination of cases R1 or R2 with S1, S2 or S3. Table 2 describes all the potential latent classes. For instance, the class 1 (R1, S1) consists of error-free self-reported (S1) and measured (R1) data and occurs with probability $\pi_r\pi_s$. The full model is a mixture of the six bivariate normal distributions for the observed outcome pairs ($r_i$, $s_i$), each with

---

anthropometrics do not differ according to the mode of interview, with similar patterns observed when self-reported anthropometrics are collected using randomly assigned open interview and self-completion modes. Along similar lines, Cawley et al. (2015) who also discuss mean reversion in reporting error in weight, highlight that interviewers do not amend/correct the self-reported anthropometrics based on measured data in their datasets and, thus, no additional interviewer effects are expected.

different means and covariance matrices (see Jenkins and Rios-Avila (2020, 2021) and Kapteyn and Ypma (2007) for full details).

The parameter estimates are obtained by maximizing the model log-likelihood (see Kapteyn and Ypma 2007, Appendix B), with identification relying on the existence of the "completely labelled" group that contains observations with error-free anthropometrics (class 1: R1-S1). Parameters $\mu_\xi$ and $\sigma_\xi^2$ are identified from these "completely labelled" observations and this contributes to identification of the other unknown parameters from the mixture of normals implied by the model specification (see Kapteyn and Ypma (2007) for further details on identification). Kapteyn and Ypma (2007) provide the expressions for the probability density functions and the associated log-likelihood function. Employing Jenkins and Rios-Avila's (2023b) user-written Stata command, we fit the full Kapteyn and Ypma (2007) model by maximum likelihood, assuming that the sample likelihood function is a finite mixture of latent class distributions. Our analysis is done separately for each of our anthropometric measures, i.e., for weight and height.

**Table 2: Groups (latent classes) in mixture model of self-reported and measured anthropometrics.**

| Groups ($i$) | Types | Probability ($\pi_j$) |
|---|---|---|
| 1 | R1,S1 | $\pi_r \pi_s$ |
| 2 | R1,S2 | $\pi_r(1-\pi_s)(1-\pi_w)$ |
| 3 | R1,S3 | $\pi_r(1-\pi_s)\pi_\omega$ |
| 4 | R2,S1 | $(1-\pi_r)\pi_s$ |
| 5 | R2,S2 | $(1-\pi_r)(1-\pi_s)(1-\pi_\omega)$ |
| 6 | R2,S3 | $(1-\pi_r)(1-\pi_s)\pi_\omega$ |

*Accounting for covariates*

Following Jenkins and Rios-Avila (2020; 2021), the factor mixture model is based on unconditional distributions. However, allowing the measurement error distributions to vary across observed characteristics has the advantage of increased flexibility and can be used to assess whether the distributions of measurement errors differ across population sub-groups (Jenkins and Rios-Avila 2023b). Goodness of fit tests based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used to compare our factor mixture models with and without covariates.

Jenkins and Rios-Avila (2023a), extend the Kapteyn and Ypma (2007) model, to allow transformations of relevant parameters to be specified as linear indices of characteristics ($X_i$):

$$G(\gamma) = \alpha_\gamma + \beta'_\gamma X_i. \tag{3}$$

where, for each factor mixture model parameter of interest ($\gamma$), $\alpha_\gamma$ is a constant, and $\beta_\gamma$ are the slopes associated with individual-level characteristics ($X_i$). The function $G(\cdot)$ is the specific transformation function of the parameter of interest. These are the identity function for means ($\mu$), the logarithmic function for SDs ($\sigma$), and Fisher's z transformation for correlations ($\rho$).

In practice, and for parsimony in the estimation of our factor mixture models with covariates, we parameterise errors in self-reported data using age, gender and region of residence. Existing studies argue that measurement error in self-reported body weight and height data depends on respondents' characteristics (e.g., Cawley et al. 2015; Davillas and Jones 2021). Specifically, for the self-reported anthropometrics, we model the $\mu$ and the $\sigma$ of the imprecision error ($\eta_i$) and of the additional random error ($\omega_i$) as a function of individual characteristics (age groups, gender and region of residence); we also condition the non-classical mean-reverting error on these respondent-level characteristics.

Moreover, for the latent true body weight and height, we assume that the mean ($\mu_\xi$) varies by respondents' age, gender and region of residence; the same covariates are also used for the SD equation ($\sigma_\xi$). Earlier research has considered these demographics as basic correlates of obesity (e.g., Baum and Ruhm, 2009, Davillas and Jones, 2020). Finally, we model the distribution of measured anthropometrics without covariates; given that the protocols on physical measurements of anthropometrics collected in surveys are the same for all respondents (irrespective of gender, age and region of residence). Interviewer characteristics, for those who are responsible for the physical measurements, might be more relevant sources of measurement error in measured anthropometrics (Olbrich et al. 2022).[7] However, these are not available in our dataset.

For the factor mixture models with covariates, we calculate the estimated parameters in their natural metrics, computing the Average Predicted Margins (APMs); for each measurement model parameter of interest ($\gamma$), we predict $\gamma$ for every individual in our sample using the fitted model and assuming all other covariates are at their observed values and, then, calculate the sample average of $\gamma$ (and its associated standard error). For presentation purposes, we report how each measurement error parameter ($\gamma$)

---

[7] Typically, failures of measurement equipment may be also relevant for measurement error in physical measurements of anthropometrics. However, we believe that the risk of equipment failure is less relevant in our dataset given the prevention mechanisms/protocols we describe above.

varies across covariates using the APMs (Jenkins and Rios-Avila 2023a,b). For example, for a gender dummy, we calculate the APMs, for males by setting all sample values of gender to male and then taking the average over the whole sample; APMs for females are calculated analogously. This allows us to test whether there are systematic gender differences in APM for each particular parameter of interest ($\gamma$). For comparison purposes, in addition to APMs across population groups (by gender, age groups and region of residence) we also report the corresponding APMs for all observations in the sample.

*Post estimation predictions*

As a post-estimation exercise, we generate predictions of the distribution of the true latent weight and height (e.g., Meijer et al. 2012). In line with Jenkins and Rios-Avila (2023a), we employ the most reliable prediction among all the potential hybrid measures of weight and height and then calculate BMI as weight (in Kg) over the square of height (in metres). We compare the distributions of hybrid, self-reported, and measured BMI. We take the estimated parameters of our mixture models, separately for the case of models with and without covariates, to create "hybrid" anthropometric predictions that combine information from both self-reported and measured anthropometrics.[8]

Specifically, in line with Meijer et al. (2012), both with and without covariates, we compare a number of approaches that combine measured and self-reported data to obtain the best prediction of the "true" anthropometrics of interest. Meijer et al. (2012) begin by deriving two predictors for the case of a single latent class (as described in Table 2 for our analysis): one that minimizes the mean squared error (MSE) and one that minimizes the MSE conditional on unbiasedness. Because class membership is unobserved, Meijer et al. (2012) proposed three ways to proceed: (1) compute the within-class predictors for each class and combine them in a weighted average using the (un)conditional class probabilities for weighting; (2) predict class membership and then use the within-class predictor for the predicted class; and (3) derive predictors that minimize the total mean squared prediction error. Because either the predictor based on MSE or on MSE conditional on unbiasedness could be the within-predictor for each of the three approaches listed above, there are six potential predictors in total. Finally, a system-wide predictor minimizes MSE under the assumption of linearity and imposing the condition of unbiasedness.

---

[8] The user written Stata command "ky_fit" predicts the seven "hybrid" measures proposed by Meijer et al. (2012). Table 6 in Jenkins and Rios-Avila (2023b) provides the descriptions of the predictors ("hybrid" outcomes), with the corresponding derivation of the formulae presented in their appendix.

As described above, following Meijer et al. (2012), seven "hybrid" measures to approximate the true body weight and height are generated in our study: 1) Weighted (unconditional), 2) Weighted (unconditional) unbiased, 3) Weighted (conditional), 4) Weighted (conditional) unbiased, 5) Two-stage, 6) Two-stage, unbiased, and 7) System-wide linear. Predictions 1 to 6 use two within-class predictors for $\xi$. The first set $\hat{\xi}_i^j$, used for predictors 1, 3, and 5, minimize the mean square error (MSE), $E\left[\left(\xi_i - \xi_i^j\right)^2 | \xi_i, i \in J\right]$. The second of set predictors, $\hat{\xi}_i^{Uj}$, used for predictors 2, 4 and 6, minimize the MSE conditional on $E\left(\xi_i - \xi_i^{Uj} | i \in J\right) = 0$. Predictors 1 and 2 provide weighted predictions using the unconditional within-class probabilities $\pi_j$. Predictors 3 and 4 provide weighted predictions using conditional or posterior within-class probabilities $\pi_j(r_i, s_i)$. Predictors 5 and 6 use a two-step Bayesian classification; i.e., the predicted class membership is obtained first and, then, the class-specific predictor of the predicted class is used. Finally, the seventh predictor ($\xi_{7i}$) is the system-wide predictor that minimizes MSE under the assumption of linearity and imposing the condition of unbiasedness.

To assess the precision of those predictions, we estimate reliability statistics and the MSE.[9] These are computed with respect to the seven "hybrid" measures that come from the sample simulations for body weight and body height based on estimated parameters for the factor mixture models both with and without covariates. Simulation analysis is done using the user-written Stata command "ky_sim" (Jenkins and Rios-Avila, 2023b).

We provide some further analysis to explore the implications of the measurement error in both self-reported and measured anthropometrics for empirical research on the association between obesity and health care utilization. Specifically, we compare results when each of the self-reported, measured and hybrid BMI measures are used as explanatory variables. If measurement error is non-classical, i.e., systematically associated with the measured values, it may cause bias in regression models that use anthropometrics as a regressor, even in the case where instrumental variable analysis is employed to deal with endogeneity or errors-in-variables (Cawley et al. 2015; O'Neill and Sweetman 2013).

---

[9] The mean square error is computed as $E(predictor - \xi)^2 = Bias^2 + Variance$. Reliability measures are computed as follows: $Rel1(r) = cov(\xi, r)/var(r)$, $Rel1(s) = cov(\xi, s)/var(s)$, $Rel2(r) = cov(\xi, r)^2/[var(\xi) \cdot var(r)]$ and $Rel2(s) = cov(\xi, s)^2/[var(\xi) \cdot var(s)]$. Further details can be found in Jenkins and Rios-Avila (2023a).

# 3. Data

Data on self-reported and measured anthropometrics are extracted from the 2013 National Health Survey of Brazil (Pesquisa Nacional de Saúde –PNS-2013).[10] This is a cross-sectional, nationally representative dataset for all Brazilian states and geographic regions. The survey focuses on use of health care services, population health conditions, and surveillance of chronic non-communicable diseases and their associated risk factors. The PNS-2013 collects demographics and socioeconomic characteristics of all household members. For each household, a randomly selected household member aged 18 or older is chosen for their body weight and height to be measured along with self-reports of the same anthropometrics.[11] This results in a working sample of 37,335 respondents, men and non-pregnant women aged 20 or older, with valid self-reported and measured weight and height data. We focus on adults (aged 20+) to avoid any puberty-related changes in body-size.

## 3.1 Self-reported and measured body weight and height data

Self-reported body weight and height data are collected as part of the survey questionnaire. Measured weight and height are collected twice by a trained survey team member at the end of the questionnaire. Weight is measured by a portable digital scale, following standard measurement protocols which require that the respondents remove their shoes, heavy clothes, accessories, and objects from their pockets (PNS 2013). Following common practice in the literature, when measured health data are used, we take the second measurement for weight and height for our base case analysis to reduce

---

[10] The 2013 National Health Survey of Brazil is publicly available online: https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=microdados.

[11] In PNS-2019, that collected data in 2019, body weight and height were measured for a much smaller sub-sample of respondents, due to the difficulties in physical anthropometric measurements for the full survey sample selected for individual interviews, (Reis et al. 2022). On the other hand, in PNS-2013, the anthropometric measurements were carried out on all residents selected for the individual interview, except pregnant women (Damacena et al. 2015). Collection of both self-reported and measured anthropometrics at the same wave is necessary for our research question and the estimation requirements of our factor mixture models. Given that measured anthropometrics are only available for a small fraction of the total survey sample in PNS-2019 and because time sensitivity is not a constraint for the scope and the nature of our research question for this study we have used the PNS-2013 data for our analysis.

any potential errors in measured anthropometrics (e.g., Johnston et al. 2009; Davillas and Pudney 2017). A sensitivity analysis is done using the average of the two measures.[12]

For height, a portable stadiometer is used to measure stature (PNS, 2013). Measurement protocols for body height require that the respondent must remove their shoes and other accessories, if possible, and keep at least three points of the body on the posterior surface of the stadiometer (PNS 2013). International measurement protocols together with validated and daily calibrated equipment are employed for anthropometric physical measurements. These procedures are settled in partnership with the Laboratory for Nutritional Evaluation of Populations (LANPOP), part of the Public Health School in the University of São Paulo, to prevent biologically inaccurate anthropometric measurements (Szwarcwald et al. 2014).

Our analysis allows for modelling all hypothesised errors in the measured and self-reported anthropometrics as relevant to our dataset and described in detail in Table 1. Along with the unconditional factor mixture measurement error models, we also estimate models that account for a parsimonious set of covariates to explore potential differential patterns in measurement errors across population groups. Specifically, in these models, we account for the respondent's gender, while respondents age is captured by a 6-category age group variable (20-29, 30-39, 40-49, 50-59, 60-69, and 70 or more). Region of residence is captured by a categorial variable for the five geographical regions (often called macro regions) of Brazil as defined by the Brazilian Institute of Geography and Statistics: North, Northeast, Central-West, Southeast and South.

### 3.2 Descriptive statistics

Figure 1 shows the histograms of the raw difference between measured and self-reported body weight and height data, as well as for BMI created from the measured and self-reported anthropometrics; a normal distribution is overlaid on each histogram. The horizontal axis is the number of units of raw reporting error; negative numbers indicate that self-reports are higher than measured values, and vice versa. The histograms would have been a single bar, with all the sample having zero reporting error, if every respondent reported the same measured and self-reported anthropometrics.

---

[12] Figure A1 (Appendix) plots the absolute differences between the 1st and 2nd body weight and height physical measurement. The graph shows that the mass of the absolute difference is concentrated at zero, and there are a few observations with absolute differences between the 1st and 2nd measurement that exceeds 1.5kg (for body weight) or 1.5cm (for body height).

**Figure 1: Histograms of the raw difference between measured and self-reported (measured − self-reported) anthropometrics**
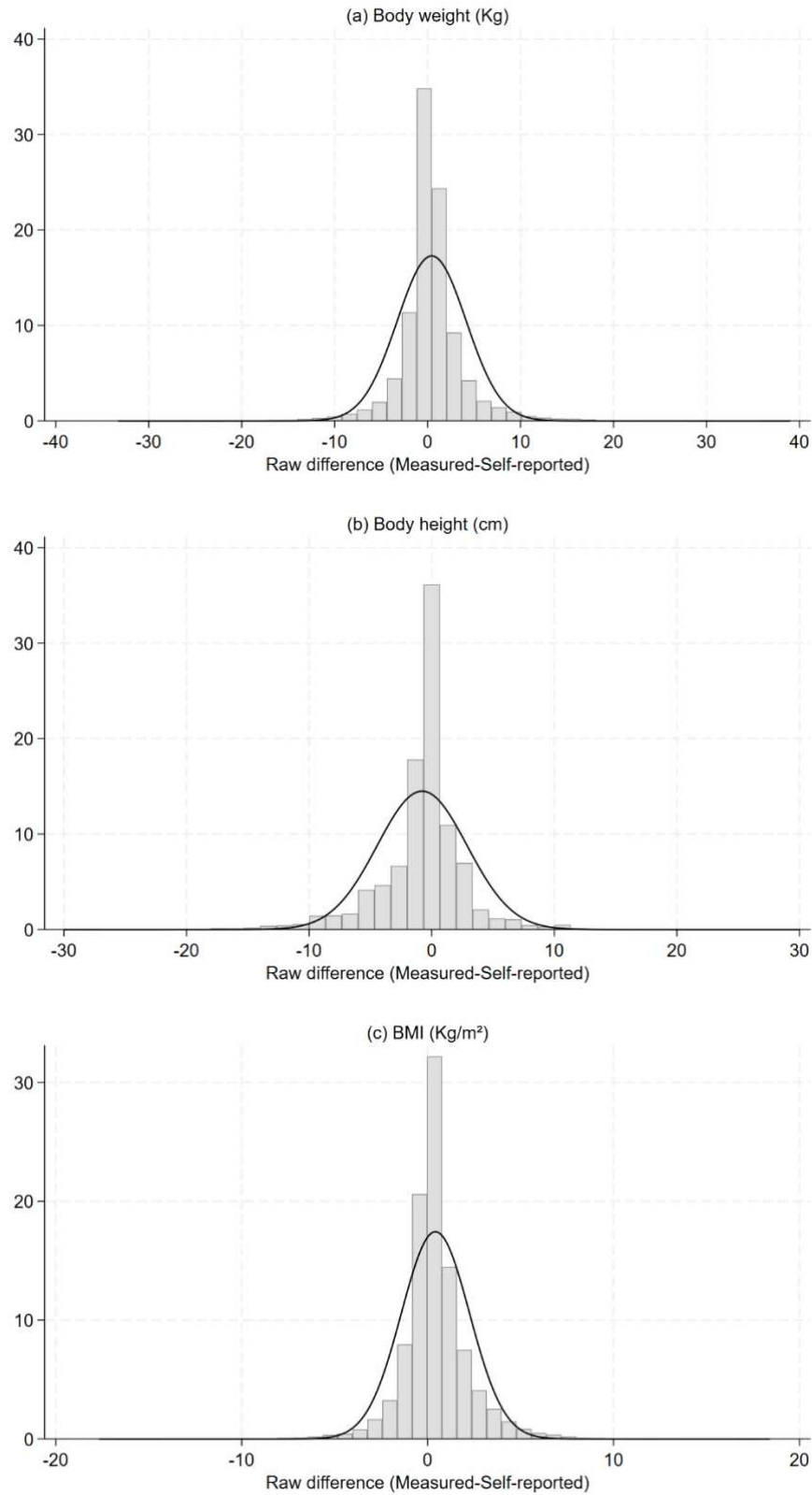


Note: The normal density curve is overlayed to each histogram.

Overall, across the graphs for body weight, height and BMI, the distribution of the raw difference between reported and measured values deviates from the normal distribution. Specifically, these is more mass around zero for the raw difference, as shown by the histograms, as opposed to the normal distribution. Compared to the normal distribution, less mass is observed with moderate and larger raw differences for all anthropometrics, while there is more mass at very high raw differences. Finally, it seems that the distribution for the raw body height difference is more skewed.

Descriptive statistics for the self-reported and measured weight and height data as well as for BMI measures are presented in Table 3.[13] The mean self-reported weight (71.5Kg) is slightly smaller than the mean measured weight (72Kg). Mean self-reported height is 0.8cm higher than measured height. Table 3 also shows that the mean absolute difference between the self-reported and measured data (expressed in terms of percentage of the measured values) is about 3% for body weight, 1% for height and 4.5% for the derived BMI measure.

Table 3. Descriptive statistics and (raw and *absolute*) difference between measured and self-reported data.

| Measure | Weight (Kg) | | Height (cm) | | BMI (Kg/m²) | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Self-reported | 71.5 | 14.6 | 165.2 | 9.5 | 26.2 | 4.7 |
| Measured | 72.0 | 15.0 | 164.4 | 9.5 | 26.6 | 4.9 |
| Raw difference[†] (measured–self-reported) | 0.4 | 3.8 | -0.8 | 3.7 | 0.4 | 1.8 |
| Absolute difference[††] | 2.2 | 3.1 | 2.2 | 3.1 | 1.2 | 1.5 |
| Absolute difference (% measured)[††] | 3.1 | 4.4 | 1.4 | 1.9 | 4.5 | 5.3 |

[†] The raw difference is calculated as the difference of measured from self-reported data.
[††] The absolute difference takes the *absolute* value of the difference between measured and self-reported data; in other words, the *absolute* difference is only positive numbers as it removes any negative signs.

---

[13] The corresponding kernel density distributions for self-reported and measured body weight, height and BMI are presented in Figure A2 (Appendix). It seems that both self-reported and measured body height data have approximately normally shaped distributions, although right skewed distributions are observed for the case of body weight and BMI. This is important as our model assumes normality for the factor distributions and identification of the components of the mixture of normals stems from non-normality in the (joint) distribution of observed outcomes.

Existing literature argues that reporting error in body weight self-reports may be mean reverting, when compared with measured anthropometric data (Cawley et al. 2015); respondents with high (low) values of measured body weight data tend to under-report (over-report) their body weight in self-reports. To provide some preliminary evidence of this, under the assumption that measured data are not subject to measurement error (an assumption we will relax later), Figure 2 shows the mean raw difference (measured–self-reported) in body weight and height data across deciles of the measured anthropometrics. Our results for body weight, show that the mean raw reporting error becomes less negative moving across the first three groups. This indicates that, on average, the self-reported weight is higher than measured weight for those with the lowest measured weight data. For the higher deciles of measured body weight there is a progressively increasing positive raw error indicating that measured weight is higher than the self-reports, with the under-reporting becoming more evident for those with higher measured weight.

Figure 2 also displays the mean raw differences for height. There is a progressively less negative mean raw difference moving to those of higher measured height up to the 80th percentile of measured height; i.e., self-reports of height are higher than measured data on average, with the over-reporting (almost) monotonically reducing in magnitude for those with higher measured height. For the two tallest deciles the mean raw reporting error is positive, suggesting that those of very high measured height tend to under-report their height. Overall, and despite the observed differences between weight and height, these results show that respondents with high (very high) measured body weight (height) tend to under-report, while over-reporting is evident for those of lower measured values. These summary statistics provide initial evidence on the presence of mean reverting error in self-reported anthropometrics under the assumption that measured data are not subject to measurement error. Although this is an assumption that we relax in our factor mixture models, this motivates accounting for mean reversion (or mean divergence) in the measurement error models.

Figure 2: Differences between measured and self-reported weight/height data by decile groups of measured anthropometrics

(a) Weight (Kg)



(b) Height (cm)
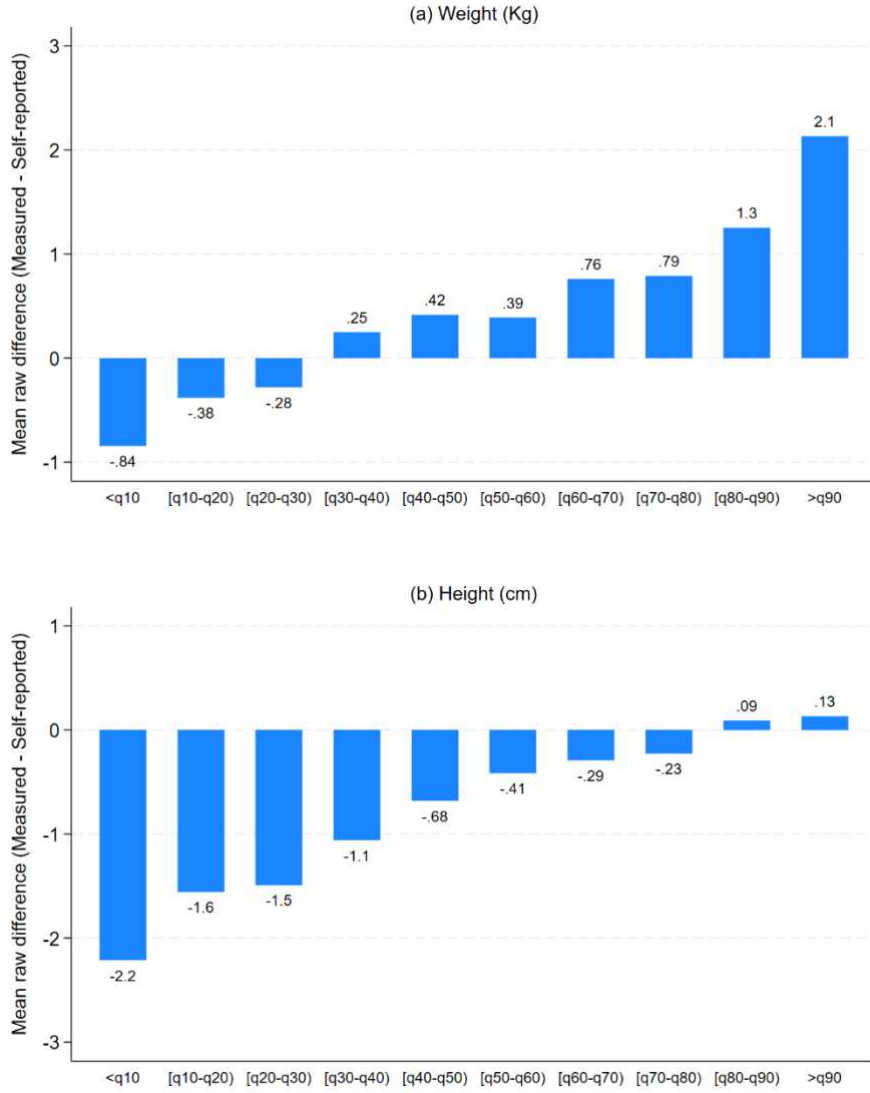
# 4. Results

### 4.1 Estimates of structural parameters: mixture model without covariates.

Table 4 presents the estimates for the KY model (expressed in their natural metrics). Following Jenkins and Rios-Avila (2020), the completely labelled observations are defined as those observations with $|r_i - s_i| \leq \delta$. Our model presented in Table 4 assumes $\delta = 0$, i.e., the completely labelled observations are only those with no differences between self-reported and measured values. Under this demanding requirement, given the differences in precision of the scales used for measured and the self-reported outcomes, the completely labelled cases represent just 10% and 23% of our observations for weight and height

respectively. Sensitivity analysis is also conducted to test the robustness of our results when this requirement is relaxed.

Table 4 shows that the mean of latent true body weight $(\mu_\xi)$ is 71.9Kg (with a standard deviation, $\sigma_\xi = 14.9$). The distribution of the latent true weight has a higher mean (by about 0.4Kg) than the mean of self-reported body weight (Table 3); the p-value for the difference in means is less than 0.01. The estimated mean of true body height is 164.5 cm (with a standard deviation $\sigma_\xi = 9.4$). This value is lower (by -0.7 cm) than the mean of the self-reported height (Table 3).

The probability $(\pi_r)$ that measured weight and height reflect the corresponding true values is high: 98.6% for weight and 96.7% for height. This indicates that the probability of error-prone measured body weight and height data occur with a low, but systematically different from zero, probability $(1 - \pi_r)$ of about 1.4% (p-value < 0.01) and 3.3% (p-value < 0.01), respectively. Error-prone measurement of body weight (reflecting the recording errors) leads to an estimated mean $(\mu_\zeta)$ of 78.9Kg for these erroneous observations, which is 7Kg (or almost 10%) higher than the estimated mean of true weight; data recording error in measured weight is also associated with a higher standard deviation $(\sigma_\zeta = 19.4)$ compared to the estimated true weight distribution $(\sigma_\xi = 14.9)$. Similarly, error-prone measured body height (that is subject to potential recording error) has an estimated mean $(\mu_\zeta)$ for the erroneous observations of 159.8cm, which is lower than the estimated mean of the true height (by about 4.7cm, i.e., 2.9% of the mean of the true height), as well as having a lower estimated standard deviation compared to the true height distribution $(\sigma_\zeta = 8.9$ compared to $\sigma_\xi = 9.4)$.

Turning to self-reported weight and height, the estimated probability $(\pi_s)$ that the self-reported anthropometrics equal the true body weight and height (i.e., they are free from any measurement error) are, as expected given the difference in precision of the two measures, relatively low at about 10% and 24%, respectively. Table 4 shows that mean reversion $(\rho)$ in case of both self-reported body weight and height data is small in magnitude (close to zero) although statistically significant at the 1% level. This indicates that after accounting for all other sources of measurement error in self-reported data, mean reversion seems to play a limited role. Error due to the reporting precision (precision error) in self-reported body weight and height data have mean values $(\mu_\eta)$; of -0.33Kg for weight and 0.4cm for height. The estimated probability of the Case S2 type of observations, $(1 - \pi_s)(1 - \pi_\omega)$, is about 62% for weight and 44% for height. Moreover, Table 4 shows that the probability $(1 - \pi_s)\pi_\omega$ that self-reported anthropometric data

contains additional measurement error, Case S3, is about 28% for self-reported weight and 31% for self-reported height.

Table 4 (Panel B) presents estimates of the membership probabilities for the six latent classes (as described in Table 2). The first latent class consists of error-free self-reported (S1) and measured (R1) anthropometric data with a probability of 10% for body weight and 23% for height. The probability that there are error-free measured anthropometrics and survey reporting error in self-reported anthropometrics is about 61% for weight and 43% for height $\left(Pr(R = 1, S = 2)\right)$. The probability of error free measured anthropometrics and additional reporting error in self-reported data, corresponding to the third latent class, is 27% for weight and 30% for height. Regarding the remaining latent classes, where there are recording errors in measured anthropometrics, we find small probabilities. For instance, the probability that weight and height observations contain error in the self-reported data and recording errors in the measured anthropometrics, corresponding to the fifth latent class $\left(Pr(R = 2, S = 2)\right)$, is 0.9% and 1.5% for weight and height, respectively. Overall, these results indicate that, although there are non-negligible recording errors in measured body weight and height data (about 7kg and 4.7cm difference on average as compared to true body weight and height, respectively), their probability of occurrence is small.

We conducted a sensitivity analysis, where measured body weight and height data are rounded to the nearest integer (Table A1, Appendix); this allows us to have the same scale in measured and reported data, but it masks the part of measurement error that is attributable to lack of precision in the recording of the self-reported data. There are differences in the six latent classes probabilities, reflecting the difference in the proportion of completely labelled cases $\left(Pr(R = 1, S = 1)\right)$. For instance, the increase in the probability of completely labelled cases as opposed to the case of our base case results (from 10% in the base case to 26.3% for the sensitivity analysis for weight; and, from 23.3% to 32.4% for height), is reflected in the reduction in the latent class probabilities for classes two and three (Table 4 vs Table A1).

Finally, we conducted a sensitivity analysis to explore whether our results presented in Table 4 are sensitive to using the average of the two weight and height measurements to define measured anthropometrics (for the mixture models). The corresponding parameter estimates, and latent class probabilities (Table A2, Appendix) are practically identical to those presented in Table 4.

Table 4: Estimates of factor mixture model for body weight and height.

| | Weight (Kg) | Height (cm) |
|---|---|---|
| **Panel A: Parameters** | | |
| $\mu_\xi$ | 71.911*** | 164.518*** |
| | (0.077) | (0.050) |
| $\sigma_\xi$ | 14.853*** | 9.448*** |
| | (0.055) | (0.035) |
| $\mu_\zeta$ | 78.892*** | 159.767*** |
| | (1.099) | (0.395) |
| $\sigma_\zeta$ | 19.395*** | 8.895*** |
| | (0.728) | (0.261) |
| $\mu_\eta$ | -0.328*** | 0.400*** |
| | (0.014) | (0.024) |
| $\sigma_\eta$ | 1.636*** | 1.837*** |
| | (0.018) | (0.027) |
| $\mu_\omega$ | -0.333*** | 1.185*** |
| | (0.067) | (0.070) |
| $\sigma_\omega$ | 5.127*** | 4.469*** |
| | (0.085) | (0.074) |
| $\pi_r$ | 0.986*** | 0.967*** |
| | (0.001) | (0.002) |
| $\pi_s$ | 0.101*** | 0.241*** |
| | (0.002) | (0.002) |
| $\pi_\omega$ | 0.306*** | 0.414*** |
| | (0.007) | (0.011) |
| $\rho$ | -0.024*** | -0.037*** |
| | (0.001) | (0.002) |
| **Panel B: Class probabilities** | | |
| $Pr(R = 1, S = 1)$ | 0.100*** | 0.233*** |
| | (0.002) | (0.002) |
| $Pr(R = 1, S = 2)$ | 0.615*** | 0.430*** |
| | (0.007) | (0.009) |
| $Pr(R = 1, S = 3)$ | 0.271*** | 0.304*** |
| | (0.007) | (0.008) |
| $Pr(R = 2, S = 1)$ | 0.001*** | 0.008*** |
| | (0.000) | (0.001) |
| $Pr(R = 2, S = 2)$ | 0.009*** | 0.015*** |
| | (0.001) | (0.001) |
| $Pr(R = 2, S = 3)$ | 0.004*** | 0.011*** |
| | (0.000) | (0.001) |
| Log-likelihood | -251,431 | -234,482 |
| Observations | 37,335 | 37,335 |

Note. The fraction of completely labelled observations (i.e., $|r_i - s_i| = 0$) is 10.0% for body weight, and 23.3% for body height.
*** p<0.01

## 4.2 Mixture model with covariates

Table 5 reports the AIC and BIC for the KY models, separately for body weight and height, with no covariates (i.e., our baseline model), and for the KY models that account for our set of covariates. Across all factor mixture models for body weight and height, those that account for covariates have lower AIC and BIC as opposed to the counterparts without covariates (baseline models). Overall, it seems that models with covariates perform better than our baseline models, suggesting that the former can be used to explore potential differential patterns in measurement error across individual characteristics.

**Table 5: AIC and BIC for the factor mixture models for body weight and height: models with and without covariates**

|  | No covariates | | With covariates | |
| --- | --- | --- | --- | --- |
|  | Weight | Height | Weight | Height |
| AIC | 502885.4 | 468987.5 | 495340.2 | 444261.2 |
| BIC | 502987.7 | 469089.9 | 496039.5 | 444960.5 |
| Observations | 37,335 | 37,335 | 37,335 | 37,335 |

Tables 6 and 7 report the estimates of our factor mixture model with covariates for body weight and height, respectively. We report the APMs for the full sample ('all') and for the specific groups of individuals based on the set of covariates we account for in the case of true latent body weight and height, precision error ($\eta_i$), additional random noise ($\omega_i$), and the mean-reverting error in self-reported anthropometrics (Panel A). Tables 6 and 7 also report the estimates of error probabilities and class probabilities (Panel B). Regarding the estimates of the membership probabilities for the six latent classes (Tables 6 and 7, Panel B), these are very similar to the corresponding results without covariates (Table 4).

*True anthropometrics*: The estimated mean of true body weight is 71.9 and SD 13.8 ('all' estimates); as in the case of our model without covariates (Table 4), these results show that the distribution of the latent true weight has a higher mean (by about 0.4kg) than the mean of self-reported body weight (p-value of the difference in means <0.001). Differences across individual characteristics are as we expect. In line with existing findings (Fryar et al., 2021), men have greater average and more dispersed body weight than women; gender differences in APM are highly statistically significant (as shown by "+++" reflecting the statistical significance of the pairwise comparisons of APMs by gender). Moreover, there is an inverted U-shaped association between mean true body

weight and age (Baum, 2007); variations in the dispersion of the true weight distribution are also observed across age groups. We also observe systematic regional differences in mean latent weight (with "+++" in Table 6, reflecting the joint significance of between-region groups differences in APM); the APM for the mean true body weight is higher in the South and the Southeast as opposed to other regions; this confirms existing literature about higher obesity rates for these regions in Brazil (Rimes-Dias et al., 2022).

Turning to the latent true height (Table 7), the estimated mean is 164.4 (with a standard deviation $\sigma_\xi$=6.848); in line with the corresponding values from our model without covariates (Table 4), the estimated true mean height is lower (by -0.8 cm) than the mean of self-reported height (Table 3). Higher mean true height, but also a larger dispersion in the relevant distribution, is observed for males and younger individuals (Fryar et al., 2021). Regional variations in the mean and standard deviation of the true height distribution are also evident in Table 7.

*Measured anthropometrics*: We find that the mean and the standard deviation of measured body weight and height ($\mu_\zeta, \sigma_\zeta$) are comparable to the corresponding parameters from our baseline models without covariates (Tables 6 and 7 versus Table 4) and, thus, estimating models with covariates does not change the conclusions of our analysis. Note that, given absence of interviewer-level data, we do not condition these parameters on covariates when estimating the factor mixture models presented in Tables 6 and 7. Specifically, the error-prone measured body weight has a mean and standard deviation ($\mu_\zeta = 78.792, \sigma_\zeta = 18.952$), which are higher than the corresponding values for the true weight distribution; in line with our results from our baseline model without covariates, the estimated mean of measured body weight for those cases that are subject to (intentional or unintentional) recording error is around 7kg (or almost 10%) higher than the estimated mean of true weight.

Regarding height, the estimated mean of the measured height for those cases that are subject to error (error-prone measured body height) is around 163.0cm; this is very close to the corresponding estimated mean from the baseline model without covariates (around 160cm in Table 4), and, confirm our baseline results suggesting that the mean of the error-prone measured height is lower than the estimated mean of true height. In line with our baseline models, our analysis shows that errors in measured anthropometrics occur with a probability ($1 - \pi_r$) that is very low in magnitude (about 1.5% for body weight and height) but systematically different from zero (p-values < 0.01).

*Measurement errors heterogeneity in self-reported weight*: The estimated mean value of precision error in self-reported body weight $(\mu_\eta)$ is -0.29 (with a standard deviation, $\sigma_\eta = 1.60$). Taking these values as a benchmark ("All" estimates, Table 6), there are systematic differences in the precision error distribution across population groups. Mean of the reporting precision error is positive for males (0.057), while it is negative for females (-0.60); gender differences in APM are highly statistically significant (as shown by "+++" reflecting the statistical significance of the pairwise comparisons of APMs by gender). Moreover, the standard deviation of the imprecision error is higher for men. In other words, in seems that men and women have different patterns of reporting for self-reported weight; there is systematic mean upward bias for males with a higher dispersion, while a downward bias with lower dispersion is observed for females. Turning to age groups, the estimated APM of mean imprecision error is negative for all age groups, with variations in both mean imprecision error in body weight (and its dispersion) across age groups. There are systematic regional differences in mean imprecision error (as evident by the "+++" in Table 6 reflecting the joint significance of between-region groups differences in APM); APM for the mean impression error is more negative for the Southeast (-0.347) and Northeast (-0.292), while the South has an APM for mean imprecision error of -0.180.

The additional random error in self-reported body weight data has a negative mean $(\mu_\omega)$, -0.460, and a standard deviation of $\sigma_\omega = 4.97$. As in the case of the imprecision error, there is a gender differences, with random error being much more negative and highly significant for females (-0.713) than males (-0.178). There are also systematic age variations (with the joint test for between-age groups differences in APMs being statistically significant at the 1% level) — notably the APM for the mean random error is negative and (non-monotonically) increasing in absolute terms for older respondents, while it is positive and higher in (absolute) magnitude for the oldest age group (70+); the oldest age group has the highest dispersion (APM for standard deviation is 5.7). There are negative and statistically significant APMs for mean random error for all regions and between-region differences in the mean random errors and their dispersion.

Turning to the APM for mean reversion $(\rho)$, we confirm our results from the models without covariates suggesting the mean reversion is small in magnitude (close to zero) but statistically significant ("all" APM in Table 6). There are gender, between-age group and between-region differences in mean reverting patterns.

# Table 6: Estimates of factor mixture model for body weight with covariates.

<div align="center">Panel A: Parameters</div>

| | | APM | (SE) | | | APM | (SE) |
|---|---|---|---|---|---|---|---|
| All | $\mu_\xi$ | 71.905*** | (0.072) | | $\sigma_\xi$ | 13.811*** | (0.051) |
| Female | | 67.336*** | (0.097) | | | 13.488*** | (0.069) |
| Male | | 77.009*** | (0.107)+++ | | | 14.173*** | (0.076)+++ |
| Age ≤ 29 | | 69.488*** | (0.156) | | | 13.733*** | (0.111) |
| Age 30-39 | | 73.093*** | (0.143) | | | 14.107*** | (0.102) |
| Age 40-49 | | 73.723*** | (0.158) | | | 13.924*** | (0.112) |
| Age 50-59 | | 72.821*** | (0.181) | | | 13.824*** | (0.128) |
| Age 60-69 | | 71.197*** | (0.226) | | | 13.686*** | (0.160) |
| Age ≥ 70 | | 67.625*** | (0.267)+++ | | | 12.616*** | (0.189)+++ |
| North | | 69.854*** | (0.162) | | | 13.471*** | (0.115) |
| Northeast | | 70.413*** | (0.135) | | | 13.387*** | (0.096) |
| Southeast | | 73.128*** | (0.145) | | | 14.132*** | (0.104) |
| South | | 74.190*** | (0.191) | | | 14.209*** | (0.136) |
| Center-west | | 72.832*** | (0.194)+++ | | | 14.073*** | (0.138)+++ |
| | | | | | | | |
| All | $\mu_\zeta$ | 78.792*** | (1.041) | | $\sigma_\zeta$ | 18.952*** | (0.686) |
| | | | | | | | |
| All | $\mu_\eta$ | -0.289*** | (0.014) | | $\sigma_\eta$ | 1.600*** | (0.017) |
| Female | | -0.598*** | (0.017) | | | 1.429** | (0.018) |
| Male | | 0.057** | (0.023)+++ | | | 1.790** | (0.023)+++ |
| Age ≤ 29 | | -0.159*** | (0.030) | | | 1.541*** | (0.027) |
| Age 30-39 | | -0.287*** | (0.027) | | | 1.593*** | (0.025) |
| Age 40-49 | | -0.430*** | (0.030) | | | 1.636*** | (0.027) |
| Age 50-59 | | -0.267*** | (0.035) | | | 1.605*** | (0.032) |
| Age 60-69 | | -0.355*** | (0.044) | | | 1.610*** | (0.039) |
| Age ≥ 70 | | -0.199*** | (0.058)+++ | | | 1.672*** | (0.049)+++ |
| North | | -0.251*** | (0.031) | | | 1.579*** | (0.029) |
| Northeast | | -0.292*** | (0.026) | | | 1.584*** | (0.025) |
| Southeast | | -0.347*** | (0.029) | | | 1.652*** | (0.027) |
| South | | -0.180*** | (0.036) | | | 1.576*** | (0.032) |
| Center-west | | -0.342*** | (0.036)+++ | | | 1.584*** | (0.032)+++ |
| | | | | | | | |
| All | $\mu_\omega$ | -0.460*** | (0.066) | | $\sigma_\omega$ | 4.966*** | (0.082) |
| Female | | -0.713*** | (0.086) | | | 4.786* | (0.093) |
| Male | | -0.178* | (0.100)+++ | | | 5.167* | (0.100)+++ |
| Age ≤ 29 | | -0.307** | (0.134) | | | 4.672*** | (0.118) |
| Age 30-39 | | -0.794*** | (0.126) | | | 4.828*** | (0.122) |
| Age 40-49 | | -0.782*** | (0.146) | | | 5.170*** | (0.132) |
| Age 50-59 | | -0.672*** | (0.164) | | | 5.121*** | (0.145) |
| Age 60-69 | | 0.081 | (0.201) | | | 4.829*** | (0.171) |
| Age ≥ 70 | | 1.211*** | (0.291)+++ | | | 5.676*** | (0.252)+++ |
| North | | -0.277* | (0.146) | | | 5.017*** | (0.131) |
| Northeast | | -0.452*** | (0.122) | | | 4.963*** | (0.110) |
| Southeast | | -0.532*** | (0.137) | | | 5.168*** | (0.130) |
| South | | -0.623*** | (0.165) | | | 4.690*** | (0.146) |
| Center-west | | -0.420** | (0.169)+++ | | | 4.827*** | (0.156)+++ |
| | | | | | | | |
| All | $\rho$ | -0.038*** | (0.001) | | | | |
| Female | | -0.030*** | (0.001) | | | | |
| Male | | -0.046*** | (0.002)+++ | | | | |
| Age ≤ 29 | | -0.042*** | (0.002) | | | | |
| Age 30-39 | | -0.036*** | (0.002) | | | | |
| Age 40-49 | | -0.033*** | (0.002) | | | | |
| Age 50-59 | | -0.036*** | (0.002) | | | | |
| Age 60-69 | | -0.039*** | (0.003) | | | | |
| Age ≥ 70 | | -0.050*** | (0.004)+++ | | | | |
| North | | -0.037*** | (0.002) | | | | |
| Northeast | | -0.037*** | (0.002) | | | | |
| Southeast | | -0.038*** | (0.002) | | | | |

| | | | | | |
|---|---|---|---|---|---|
| South | -0.037*** | (0.002) | | | |
| Center-west | -0.039*** | (0.003)+++ | | | |

| Panel B: Probabilities | | | | | |
|---|---|---|---|---|---|
| $\pi_r$ | 0.985*** | (0.001) | $Pr(R=1, S=1)$ | 0.100*** | (0.002) |
| $\pi_s$ | 0.101*** | (0.002) | $Pr(R=1, S=2)$ | 0.619*** | (0.007) |
| $\pi_\omega$ | 0.300*** | (0.007) | $Pr(R=1, S=3)$ | 0.266*** | (0.006) |
| | | | $Pr(R=2, S=1)$ | 0.002*** | (0.000) |
| | | | $Pr(R=2, S=2)$ | 0.010*** | (0.001) |
| | | | $Pr(R=2, S=3)$ | 0.004*** | (0.000) |

Note. The fraction of completely labelled observations (i.e., $|r_i - s_i| = 0$) is 10.0% for body weight.
*** p<0.01, ** p<0.05, and * p<0.10; Significance levels for tests of APM=0.
+++ p<0.01, ++ p<0.05, and + p<0.10; Significance level for tests for pairwise comparisons of APMs by gender and for the joint significance of between age groups and regional differences in APMs.

*Measurement error heterogeneity in self-reported height.* The estimated mean and standard deviation of precision error in self-reported body height $(\mu_\eta, \sigma_\eta)$ are 0.50 and 1.92 ("All" estimates, Table 7), respectively. Taking these values as a benchmark, we observe differences in the precision error distribution across covariates (Table 7). The mean of the reporting precision error does not differ systematically by gender. The imprecision error in body height is almost monotonically increasing across age groups, with a similar pattern for the dispersion of the imprecision reporting error distribution for older age groups. There are systematic regional variations for the mean imprecision error; the APM for the mean impression error which is more positive for the South (0.70), while the Northeast has the lowest estimated APM for mean imprecision error (0.39).

The additional random error in self-reported body height has a positive mean value $(\mu_\omega)$, 1.28, and a standard deviation of $\sigma_\omega = 4.96$ ("All" estimates in Table 7). Mean values of the additional random noise are higher for females as opposed to males (2.34 versus 0.09). There are also systematic age variations (as shown by "+++" in Table 7) — notably the APM for the mean random error is positive and (mostly) increasing with age; the oldest age groups (60-69 and ≥70) have the highest mean of the random error and the highest dispersion. Systematic regional differences in APM for mean random error as well as for the dispersion of the distribution of the errors are also evident.

As in the baseline model without covariates (Table 4), the mean reversion $(\rho)$ for the case of self-reported height is small in magnitude (close to zero) but statistically significant ("all" APM in Table 7).

Table 7: Estimates of factor mixture model for body height with covariates.

<div align="center">Panel A: Parameters</div>

| | | APM | SE | | APM | SE |
|---|---|---|---|---|---|---|
| All | $\mu_\xi$ | 164.427*** | (0.036) | $\sigma_\xi$ | 6.848*** | (0.026) |
| Female | | 158.649*** | (0.048) | | 6.641*** | (0.034) |
| Male | | 170.881*** | (0.054)+++ | | 7.081*** | (0.039)+++ |
| Age ≤ 29 | | 166.843*** | (0.079) | | 6.874*** | (0.056) |
| Age 30-39 | | 165.781*** | (0.070) | | 6.870*** | (0.050) |
| Age 40-49 | | 164.467*** | (0.077) | | 6.787*** | (0.055) |
| Age 50-59 | | 162.817*** | (0.091) | | 6.887*** | (0.065) |
| Age 60-69 | | 161.216*** | (0.112) | | 6.762*** | (0.080) |
| Age ≥ 70 | | 159.667*** | (0.147)+++ | | 6.922*** | (0.105)+++ |
| North | | 162.041*** | (0.082) | | 6.776*** | (0.059) |
| Northeast | | 163.077*** | (0.070) | | 6.886*** | (0.050) |
| Southeast | | 166.060*** | (0.072) | | 6.935*** | (0.051) |
| South | | 166.360*** | (0.093) | | 6.852*** | (0.066) |
| Center-west | | 165.158*** | (0.093)+++ | | 6.713*** | (0.067)+++ |
| | | | | | | |
| All | $\mu_\zeta$ | 162.984*** | (0.585) | $\sigma_\zeta$ | 9.172*** | (0.385) |
| | | | | | | |
| All | $\mu_\eta$ | 0.503*** | (0.022) | $\sigma_\eta$ | 1.917*** | (0.026) |
| Female | | 0.486*** | (0.031) | | 1.943*** | (0.032) |
| Male | | 0.521*** | (0.029) | | 1.889*** | (0.030)+ |
| Age ≤ 29 | | 0.195*** | (0.040) | | 1.730*** | (0.036) |
| Age 30-39 | | 0.236*** | (0.037) | | 1.800*** | (0.034) |
| Age 40-49 | | 0.405*** | (0.044) | | 1.883*** | (0.041) |
| Age 50-59 | | 0.654*** | (0.056) | | 2.024*** | (0.049) |
| Age 60-69 | | 1.078*** | (0.076) | | 2.166*** | (0.070) |
| Age ≥ 70 | | 1.697*** | (0.108)+++ | | 2.525*** | (0.090)+++ |
| North | | 0.427*** | (0.051) | | 2.111*** | (0.048) |
| Northeast | | 0.393*** | (0.040) | | 1.958*** | (0.040) |
| Southeast | | 0.488*** | (0.040) | | 1.779*** | (0.035) |
| South | | 0.697*** | (0.050) | | 1.788*** | (0.044) |
| Center-west | | 0.633*** | (0.056)+++ | | 1.983*** | (0.048)+++ |
| | | | | | | |
| All | $\mu_\omega$ | 1.276*** | (0.072) | $\sigma_\omega$ | 4.958*** | (0.078) |
| Female | | 2.342*** | (0.104) | | 5.081*** | (0.091) |
| Male | | 0.086*** | (0.101)+++ | | 4.820*** | (0.098)++ |
| Age ≤ 29 | | 0.250*** | (0.146) | | 4.704*** | (0.129) |
| Age 30-39 | | 1.063*** | (0.132) | | 4.771*** | (0.115) |
| Age 40-49 | | 1.459*** | (0.150) | | 4.919*** | (0.115) |
| Age 50-59 | | 1.571*** | (0.183) | | 5.087*** | (0.160) |
| Age 60-69 | | 2.471*** | (0.239) | | 5.158*** | (0.172) |
| Age ≥ 70 | | 2.352*** | (0.344)+++ | | 6.128*** | (0.238)+++ |
| North | | 1.664*** | (0.170) | | 5.220*** | (0.139) |
| Northeast | | 1.826*** | (0.136) | | 5.410*** | (0.112) |
| Southeast | | 1.083*** | (0.137) | | 4.615*** | (0.124) |
| South | | 0.627*** | (0.173) | | 4.692*** | (0.138) |
| Center-west | | 0.756*** | (0.182)+++ | | 4.672*** | (0.139)+++ |
| | | | | | | |
| All | $\rho$ | -0.043*** | (0.003) | | | |
| Female | | -0.034*** | (0.004) | | | |
| Male | | -0.053*** | (0.004)+++ | | | |
| Age ≤ 29 | | -0.025*** | (0.005) | | | |
| Age 30-39 | | -0.030*** | (0.005) | | | |
| Age 40-49 | | -0.034*** | (0.005) | | | |
| Age 50-59 | | -0.048*** | (0.007) | | | |
| Age 60-69 | | -0.081*** | (0.010) | | | |
| Age ≥ 70 | | -0.112*** | (0.013)+++ | | | |
| North | | -0.043*** | (0.006) | | | |
| Northeast | | -0.040*** | (0.005) | | | |
| Southeast | | -0.046*** | (0.005) | | | |
| South | | -0.044*** | (0.006) | | | |

| | | | Panel B: Probabilities | | |
|---|---|---|---|---|---|
| Center-west | -0.041*** | (0.007)+++ | | | |
| $\pi_r$ | 0.984*** | (0.001) | $Pr(R=1,S=1)$ | 0.232*** | (0.002) |
| $\pi_s$ | 0.236*** | (0.002) | $Pr(R=1,S=2)$ | 0.473*** | (0.008) |
| $\pi_\omega$ | 0.371*** | (0.010) | $Pr(R=1,S=3)$ | 0.279*** | (0.008) |
| | | | $Pr(R=2,S=1)$ | 0.004*** | (0.000) |
| | | | $Pr(R=2,S=2)$ | 0.008*** | (0.001) |
| | | | $Pr(R=2,S=3)$ | 0.005*** | (0.000) |

Note. The fraction of completely labelled observations (i.e., $|r_i - s_i| = 0$) is 23.3% for body height.
*** p<0.01, ** p<0.05, and * p<0.10 (Significance levels for tests of APM=0).
+++ p<0.01, ++ p<0.05, and + p<0.10; Significance level for tests for pairwise comparisons of APMs by gender and for the joint significance of between age groups and regional differences in APMs.

## 4.3 Post-estimation analysis

Table 8 shows the precision of the seven types of "hybrid" predictions for body weight, for our factor mixture models with and without covariates, using simulations with 1,000 replications. The results for height are shown in Table 9. Our first measure of reliability is analogous to the slope coefficient from a (hypothetical) regression of true anthropometrics on the observed anthropometrics measure; higher values correspond to greater reliability and a value greater than one indicates mean reversion. Reliability 2 represents the squared correlation between true anthropometrics and observed anthropometrics measure. These reliability measures should only be used to assess how close a given measure is to the relevant true value and should not be compared across model specifications. For the models of body weight and height, with and without covariates, all hybrid measures provide very large reliability coefficients. A closer look at Tables 8 and 9 shows that the smallest MSE is found for the weighted (conditional) prediction for both anthropometric measures and across models with and without covariates. This indicates that these predictors perform better, as shown by the MSE using out-of-sample simulations, and, thus, the weighted (conditional) prediction is our preferred "hybrid" prediction for both weight and height.

**Table 8: Precision of "hybrid" body weight predictions.**

| | No covariates | | | With covariates | | |
|---|---|---|---|---|---|---|
| | Rel1 | Rel2 | MSE | Rel1 | Rel2 | MSE |
| Measured body weight ($r$) | 0.973 | 0.959 | 9.242 | 0.968 | 0.953 | 9.191 |
| Self-reported body weight ($s$) | 0.977 | 0.956 | 9.973 | 0.985 | 0.952 | 9.425 |
| *Hybrid body weight predictors* | | | | | | |
| 1. Weighted (unconditional) | 0.978 | 0.964 | 8.142 | 0.975 | 0.959 | 8.080 |
| 2. Weighted (unconditional) unbiased | 0.978 | 0.964 | 8.139 | 0.974 | 0.959 | 8.077 |
| 3. Weighted (conditional) | 1.000 | 0.997 | 0.697 | 1.000 | 0.996 | 0.709 |
| 4. Weighted (conditional) unbiased | 1.000 | 0.997 | 0.701 | 0.999 | 0.996 | 0.713 |
| 5. Two-stage | 0.998 | 0.996 | 0.867 | 0.998 | 0.995 | 0.878 |
| 6. Two-stage, unbiased | 0.998 | 0.996 | 0.869 | 0.997 | 0.995 | 0.881 |
| 7. System-wide linear | 1.000 | 0.978 | 4.791 | 1.000 | 0.976 | 4.665 |

**Table 9: Precision of "hybrid" body height predictions.**

| | No covariates | | | With covariates | | |
|---|---|---|---|---|---|---|
| | Rel1 | Rel2 | MSE | Rel1 | Rel2 | MSE |
| Measured body height ($r$) | 0.962 | 0.930 | 6.442 | 0.971 | 0.956 | 2.126 |
| Self-reported body height ($s$) | 0.927 | 0.901 | 9.804 | 0.838 | 0.811 | 10.877 |
| *Hybrid body height predictors* | | | | | | |
| 1. Weighted (unconditional) | 0.980 | 0.946 | 4.881 | 0.982 | 0.962 | 1.790 |
| 2. Weighted (unconditional) unbiased | 0.977 | 0.946 | 4.872 | 0.980 | 0.963 | 1.792 |
| 3. Weighted (conditional) | 1.000 | 0.985 | 1.338 | 1.000 | 0.984 | 0.777 |
| 4. Weighted (conditional) unbiased | 0.997 | 0.985 | 1.355 | 0.998 | 0.983 | 0.790 |
| 5. Two-stage | 0.991 | 0.980 | 1.828 | 0.989 | 0.979 | 1.002 |
| 6. Two-stage, unbiased | 0.989 | 0.980 | 1.839 | 0.988 | 0.979 | 1.017 |
| 7. System-wide linear | 1.000 | 0.957 | 3.838 | 1.000 | 0.963 | 1.753 |

Simulation analysis helps to identify the preferred predictors for the latent true body weight and height. After choosing the preferred prediction, the sample data are used to compute the true latent anthropometric measures in order to calculate a BMI measure that aims to approximate the true BMI distribution. Table 10 provides descriptive statistics of this preferred "hybrid" BMI measure obtained from models with and without covariates. In Table 10 presents descriptive statistics for the "corrected self-reported BMI" measure — a frequently used measure in studies that do not have access to measured anthropometric data (Cawley 2015).[14] Although the "corrected self-reported BMI" is not

---

[14] Existing studies in the economics of obesity literature that rely on self-reported anthropometrics often estimate corrective equations (or utilize the coefficients from existing equations) based on the relationship between measured and self-reported body weight and height data from alternative data sources (Cawley 2015). To mimic correction procedures for self-reported anthropometrics in the existing studies, we estimate analogous "corrective" equations by regressing measured weight and height data on self-reports and a vector of demographics (results from these equations are

driven by the mixture measurement error models, adding BMI measures in Table 10 that are based on "corrective" equations allows comparisons with a popular measure used in the existing literature in the absence of measured BMI data.

The "hybrid" BMI measures, with and without covariates, and the BMI based on measured data are very close both at the mean and across their distribution. It seems however that at the right tails, the q75 and q90 are slightly lower for the "hybrid" measures (both with and without covariates) as opposed to the BMI based on measured data; the latter reflected in the difference in inter-quantiles ranges between the "hybrid" measures and measured BMI. On the other hand, BMI values that are based on self-reported data are always lower at the mean level and across quantiles of the distribution as well as with a lower dispersion compared to both the "hybrid" measures (obtained from models with and without covariates) and the BMI based on measured data. Statistics for the "corrected self-reported BMI" depart from the self-reported BMI values; the "corrected self-reported BMI" has a higher mean and quantiles, up to median of the distribution, when compared to the "hybrid" measures, while the opposite is the case at higher quantiles (p75 and p90).

Overall, these results suggest that similar "hybrid" measures are obtained from the models with and without covariates and that these measures are very close to the BMI measure based on the measured weight/height data. On the other hand, the BMI based on self-reported data underestimate the "true" values. This indicates that the recording error in measured anthropometrics does not translate into major differences between the "hybrid" and the measured anthropometrics as a result of their small likelihood of occurrence in our sample. Finally, the distribution of the popular "corrected self-reported BMI", does not perform well as an alternative to measured BMI or our "hybrid" BMI measures.

---

available in the Appendix, Table A3). The predictions from these equations are used to calculate self-reports of body weight and height that are corrected for reporting error – these results from our "corrected self-reported BMI" measure as presented in Tables 10 and 11.

Table 10. Distributions of BMI based on preferred "hybrid" anthropometric predictions, BMI based on self-reported, "corrected self-reported" and measured body weight/height data.

| | Self-reported BMI | Corrected self-reported BMI | Measured BMI | Hybrid BMI | |
| --- | --- | --- | --- | --- | --- |
| | | | | No covariates | With covariates |
| Statistics | | | | | |
| Mean | 26.158 | 26.572 | 26.580 | 26.526 | 26.555 |
| p10 | 20.776 | 21.142 | 20.911 | 20.936 | 20.942 |
| p25 | 22.942 | 23.322 | 23.147 | 23.131 | 23.152 |
| p50 | 25.510 | 25.939 | 25.960 | 25.914 | 25.950 |
| p75 | 28.720 | 29.212 | 29.320 | 29.237 | 29.278 |
| p90 | 32.242 | 32.748 | 32.941 | 32.829 | 32.863 |
| Inter-quantile rages | | | | | |
| p75 - p25 | 5.778 | 5.890 | 6.173 | 6.106 | 6.125 |
| p50 - p10 | 4.735 | 4.798 | 5.049 | 4.977 | 5.009 |
| p90 - p50 | 6.732 | 6.809 | 6.981 | 6.916 | 6.912 |
| p90 - p10 | 11.466 | 11.607 | 12.030 | 11.893 | 11.921 |

## 4.4 Implications for the association between BMI and health care utilization

In this sub-section, we provide evidence to test the sensitivity of econometric analyses where BMI is used as an explanatory variable. We compare results with the "hybrid" BMI measures, estimated from our factor mixture models, with those based on self-reported, "corrected self-reported" and measured anthropometrics.

We estimate linear regression models to measure the association between BMI and the frequency of hospital admissions in the previous 12 months (Table 11). To facilitate interpretation of specifications that use polynomials in BMI, Figure 3 presents the adjusted predictions at representative values (APRs), i.e., the predicted health care use across selected BMI values with all the other variables kept at their mean values (based on the models presented in Table 11). As shown in Figure 3, the APRs for health care use are practically identical for the "hybrid" measures of BMI. Although the APRs for health care use for the measured and "hybrid" BMI measures are similar across their distribution, they differ at the very high tails of the BMI distribution (above BMI values of 41.5 kg/m²). Turning to self-reported BMI, the relevant results depart from those of our "hybrid" BMI measures (and measured BMI) especially at the lower (BMI values below 23.5 kg/m²) and higher tails (BMI values above 37 kg/m²) of the BMI distribution. The APRs for health care use for the corrected self-reported BMI lie between the corresponding results for self-reported and measured anthropometrics, lying much closer to those for BMI based on measured data.

Table 11: Linear regression models of healthcare utilization in the last 12 months on BMI measures.

| | Self-reported BMI | Corrected Self-reported BMI | Measured BMI | Hybrid BMI | |
|---|---|---|---|---|---|
| | | | | No covariates | With covariates |
| BMI | -5.994*** | -5.908*** | -6.281*** | -5.991** | -6.301*** |
| | (2.165) | (2.250) | (2.378) | (2.374) | (2.472) |
| BMI² | 19.951*** | 19.211** | 19.791** | 18.796** | 19.712** |
| | (7.375) | (7.609) | (8.096) | (8.101) | (8. 445) |
| BMI³ | -20.484** | -19.224** | -19.101** | -17.997** | -18.821** |
| | (8.137) | (8.348) | (8.929) | (8.967) | (9.352) |
| Observations | 37,335 | 37,335 | 37,335 | 37,335 | 37,335 |

† BMI is divided by 100.
Notes: Standard errors robust to heteroscedasticity in parentheses. Our models account for age, gender, ethnicity and geographic region fixed effects.
** p<0.05, *** p<0.01

## Figure 3: Predicted health care use across selected BMI values (based on OLS models in Table 11)
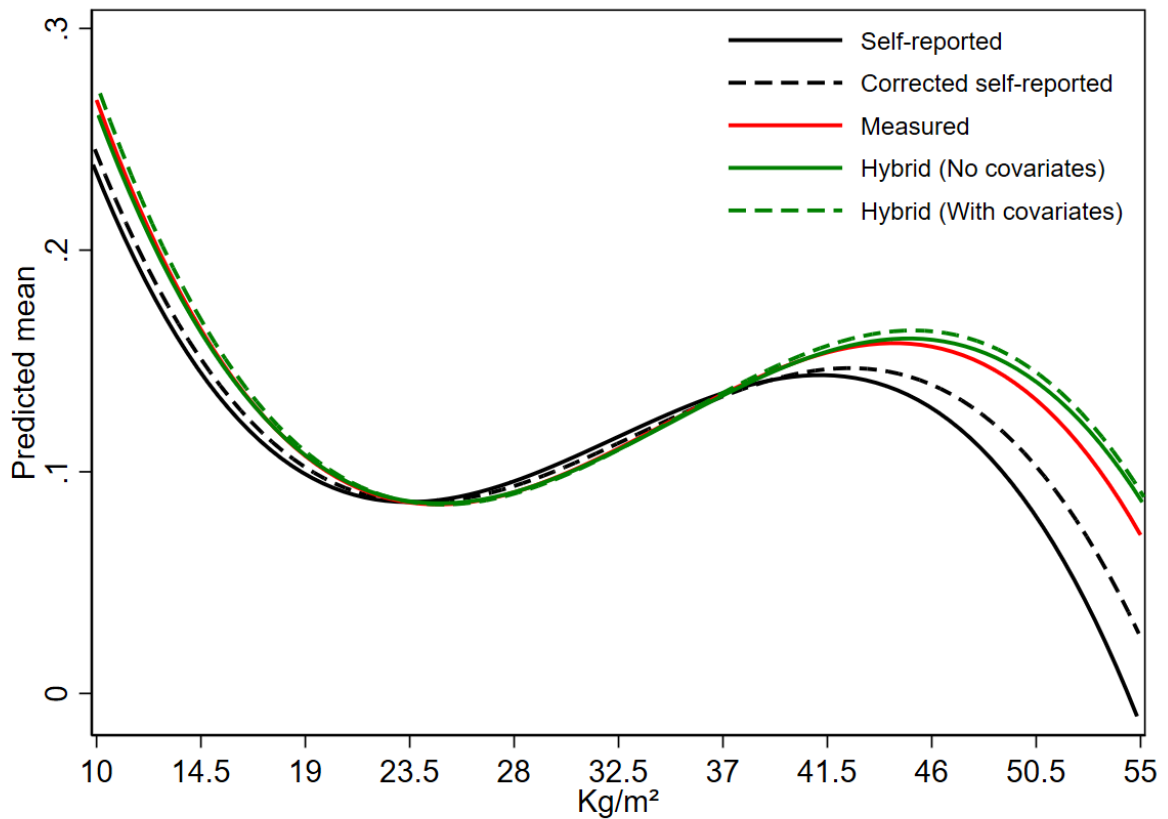
# 5. Conclusion

Existing research in the economics of obesity shows that self-reported data are subject to measurement error, which can lead to biased estimates in empirical research that relies on self-reported anthropometrics (e.g., Cawley 2015; Cawley et al. 2015; Davillas and Jones 2021; Gil and Mora 2011; O'Neill and Sweetman 2013). These analyses, however, explicitly assume that measured anthropometrics are error-free as they are treated as "gold standards" when compared to self-reported data. The literature provides little discussion of the potential measurement errors that measured anthropometrics may entail. The latter is of particular relevance given developments in large-scale social surveys that involve the integration of physical health measurements, in addition to traditional self-reported measures. To fill this gap in the literature, we use the KY factor mixture model (Kapteyn and Ypma, 2007) to analyse and characterize measurement error in both self-reported and measured anthropometrics with national representative data from the 2013 National Health Survey in Brazil.

We find that a very small, but statistically significant fraction of measured anthropometrics, may contain recording errors. The estimated probability that the self-reported anthropometrics are free from any measurement error are, as expected, relatively low at about 10% and 23% for body weight and height data — these results remain robust with and without accounting for covariates in our factor mixture models. This highlights that respondent's lack of awareness of their true anthropometrics in combination with the lack of precision of the self-reported questionnaires may be sources of the observed measurement error. For example, it has been argued that enhancing people's knowledge of their exact anthropometric values (by monitoring interventions) may indeed improve their ability to accurately report their anthropometric values (Sherry et al. 2007).

Of particular interest, our analysis reveals that mean reverting errors in self-reported anthropometrics are low in magnitude, after accounting for other sources of errors in self-reported data. These findings contradict the existing literature that compares self-reported with measured anthropometrics, arguing there are strong mean reverting patterns at least for body weight (e.g., Cawley et al. 2015). However, unlike our analysis, it should be noted that most studies that compare self-reported with measured anthropometrics assume measured anthropometrics are error-free and do not account for other potential sources of measurement errors. Our study is, therefore, potentially useful

for exploring the sources of measurement errors that may affect both self-reported and measured anthropometrics and the magnitude of bias that each source of error may cause.

Factor mixture models that account for covariates are used to explore the potential heterogeneity of reporting errors in self-reported as well as the true latent anthropometrics across population groups. A limitation of our analysis is the absence of interviewer-level data. This necessarily limits our factor mixture models, with no covariates accounted for in the measurement error in measured anthropometrics. If we had had the opportunity to provide insights on how measurement error in the physical measurements varies across interviewer characteristics, we would have been able to provide relevant recommendations to survey data teams to improve measurement protocols.

Latent true anthropometrics vary across age groups, by gender and across regions broadly in line with the existing literature (e.g., Arntsen et al., 2023; Baum and Ruhm, 2009; Davillas and Jones, 2020; Fryar et al., 2021; Rimes-Dias et al., 2022). Males have higher true mean body weight and height; mean true body weight is associated with an inverted U-shaped relationship with age; mean true height is monotonically decreasing with age, reflecting birth-cohort effects and loss of height as people become older (e.g., Arntsen et al., 2023).

Overall, we find that being older is associated with higher reporting errors (due to imprecision error or other random errors) in self-reported body weight and height. These results highlight the role of age-related changes in cognitive and communicative functioning on self-reported data (Knäuper et al. 2016); it has been shown than age-related impacts on cognitive ability, question interpretation as well as memory retrieval have impacts of people's self-reports as they become older. This is of particular relevance in the case of self-reports of body weight and height as it involves respondent's cognitive ability, memory and ability to process information. Moreover, we find the presence of systematic gender differences in measurement error in self-reported weight and height data, with women reporting with more errors than men. These results are broadly in line with the existing literature (e.g., Cawley et al. 2015, Gil and Mora 2011). For example, it may be the case that women experience greater social stigma than men for having excess weight or that other related social norm pathways affecting reporting behaviour in anthropometrics may be the relevant underlying mechanisms (Gil and Mora 2011; Puhl and Heuer 2009; Sattler et al. 2018). The observed regional differences in self-reported measurement error may reflect cultural, socioeconomic and demographic differences across Brazilian regions; however, it is impossible to disentangle the role of particular

characteristics on shaping these results given the aggregate regional variations we employ in our analysis.

To explore the practical implications of our measurement error results, post-estimation analysis and out of the sample simulations are employed to estimate hybrid anthropometric predictions that best approximate the true body weight and height distribution. Our proxies of true BMI distribution are very close to the distribution of BMI based on measured anthropometrics. On the other hand, BMI based on self-reported data seems to under-estimate the true BMI distribution. "Corrected self-reported BMI" measures, based on conventional methods to mitigate reporting error in self-reports using predictions from corrective equations, do not perform as well as our "hybrid" BMI measures.

We implement analysis on the potential implications of the measurement error when different BMI measures are used as explanatory variable in econometric models on health care utilization. We find similar econometric results when they are based on measured data or on our hybrid BMI measures across most of the BMI distribution, while only small differences are observed at the very high tails of the distribution (above BMI values of 41.5 kg/m$^2$). Differences are also observed in the econometric results based on self-reported or "corrected self-reported" data when compared to our hybrid BMI measures at the lower and higher BMI tails. Our findings further confirm existing evidence suggesting that BMI based on self-reported data may bias econometric results when BMI is used as an explanatory variable (e.g., Cawley et al. 2015), and suggest that conventional ways to correct self-reported anthropometrics may not provide mitigation.

Measured anthropometrics may encompass some systematic measurement error, but our estimates suggest a very low prevalence of errors and this is reflected in the presence of only small differences, concentrated at the right tail of the distribution, compared to our proxies of true BMI. Nevertheless, the possibility of errors in measured anthropometrics should be acknowledged when searching for an error-free adiposity measure, especially when focusing on the extreme right tail of the distribution of BMI.

# References

Arntsen, S. H., Borch, K. B., Wilsgaard, T., Njølstad, I., Hansen, A. H. (2023) Time trends in body height according to educational level. A descriptive study from the Tromsø Study 1979–2016. PLoS One, 18(1), e0279965.

Baum II, C.L., Ruhm, C. J. (2009) Age, socioeconomic status and obesity growth. Journal of Health Economics, 28(3), 635-648.

Baum, C.L. (2007) The effects of race, ethnicity, and age on obesity. Journal of Population Economics, 20, 687-705.

Bilger M, Kruger EJ, Finkelstein EA (2017) Measuring socioeconomic inequality in obesity: looking beyond the obesity threshold. Health Economics 26:1052–1066.

Bowling A (2005) Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health 27(3):281-291.

Cawley J (2015) An economy of scales: a selective review of obesity's economic causes, consequences, and solutions. Journal of Health Economics 43:244–268.

Cawley J, Meyerhoefer C (2012) The medical care costs of obesity: an instrumental variables approach. Journal of Health Economics 31(1):219-230.

Cawley J (2004) The impact of obesity on wages. Journal of Human Resources 39(2):451-474.

Cawley J (2015) An economy of scales: a selective review of obesity's economic causes, consequences, and solutions. Journal of Health Economics 43:244–268.

Cawley J, Maclean JC, Hammer M, Wintfeld N (2015) Reporting error in weight and its implications for bias in economic models. Economics & Human Biology 19:27-44.

Damacena GN, Szwarcwald CL, Malta DC, et al (2015) The development of the National Health survey in Brazil, 2013. Epidemiologia e Serviços de Saúde 24:197-206.

Davillas A, Benzeval M (2016) Alternative measures to BMI: Exploring income-related inequalities in adiposity in Great Britain. Social Science & Medicine 166:223-232.

Davillas, A., Jones, A.M. (2020) Regional inequalities in adiposity in England: distributional analysis of the contribution of individual-level characteristics and the small area obesogenic environment. Economics & Human Biology, *38*, 100887.

Davillas A, Jones A.M. (2021) The implications of self-reported body weight and height for measurement error in BMI. Economics Letters 209:110101.

Davillas A, Pudney S (2017) Concordance of health states in couples: Analysis of self-reported, nurse administered and blood-based biomarker data in the UK Understanding Society panel. Journal of Health Economics 56:87-102.

Davillas A, Pudney S (2020a) Biomarkers as precursors of disability. Economics & Human Biology 36:100814.

Davillas A, Pudney S (2020b) Biomarkers, disability and health care demand. Economics & Human Biology 39:100929.

Engstrom JL, Paterson SA, Doherty A, et al (2003) Accuracy of self-reported height and weight in women: an integrative review of the literature. Journal of Midwifery & Women's Health 48(5):338-345.

Finn A, Ranchhod V (2017) Genuine fakes: the prevalence and implications of data fabrication in a large South African survey. World Bank Economic Review 31(1):129-157.

Fryar, C. D., Carroll, M. D., Gu, Q., Afful, J., Ogden, C.L. (2021) Anthropometric reference data for children and adults. U.S. Department of Health & Human Services, National Centre of Health Statistics, United States.

Gil J, Mora T (2011) The determinants of misreporting weight and height: the role of social norms. Economics & Human Biology 9:78–91.

Gorber SC, Tremblay M, Moher D, Gorber B (2007) A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obesity Reviews 8(4):307-326.

Groves RM (2005) Survey errors and survey costs. John Wiley & Sons.

Jenkins SP, Rios-Avila F (2020) Modelling errors in survey and administrative data on employment earnings: Sensitivity to the fraction assumed to have error-free earnings. Economics Letters 192:109253.

Jenkins SP, Rios-Avila, F (2021) Measurement error in earnings data: Replication of Meijer, Rohwedder, and Wansbeek's mixture model approach to combining survey and register data. Journal of Applied Econometrics 36(4):474-483.

Jenkins SP, Rios-Avila, F (2023a) Reconciling reports: modelling employment earnings and measurement errors using linked survey and administrative data. Journal of the Royal Statistical Society Series A: Statistics in Society 186(1):110-136.

Jenkins SP, Rios-Avila, F (2023b) Finite mixture models for linked survey and administrative data: estimation and post-estimation. The Stata Journal 23(1):53-85.

Johnston DW, Propper C, Shields MA (2009) Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. Journal of Health Economics 28(3), 540-552.

Kapteyn A, Ypma JY (2007) Measurement error and misclassification: a comparison of survey and administrative data. Journal of Labor Economics 25:513–551.

Keith SW, Fontaine KR, Pajewski NM, Mehta T, Allison DB (2011) Use of self-reported height and weight biases the body mass index–mortality association. International Journal of Obesity 35(3):401-408.

Knäuper B, Carrière K, Chamandy M, Xu Z, Schwarz N, Rosen NO (2016) How aging affects self-reports. European Journal of Ageing 13:185-193.

Li J, Simon G, Castro MR., Kumar V, Steinbach MS, Caraballo PJ (2021) Association of BMI, comorbidities and all-cause mortality by using a baseline mortality risk model. PLoS ONE 16(7):e0253696.

Lin X, Xu Y, Xu J et all (2020) Global burden of noncommunicable disease attributable to high body mass index in 195 countries and territories, 1990-2017. Endocrine 69(2):310-320.

Ljungvall Å, Gerdtham UG, Lindblad U (2015) Misreporting and misclassification: implications for socioeconomic disparities in body-mass index and obesity. The European Journal of Health Economics 16:5-20.

Lobstein T, Leach RJ (2007) Tackling obesities: future choices—international comparisons of obesity trends, determinants and responses—evidence review. *London: Foresight Programme of the Government Office for Science.*

Meijer E, Rohwedder S, Wansbeek T (2012) Measurement error in earnings data: using a mixture model approach to combine survey and register data. Journal of Business & Economic Statistics 30:191–201.

Ng M, Fleming T, Robinson M et al (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 384(9945):766-781.

O'Neill D, Sweetman O (2013) The consequences of measurement error when estimating the impact of obesity on income. IZA Journal of Labor Economics 2(1):1-20.

Olbrich L, Kosyakova Y, Sakshaug JW (2022) The reliability of adult self-reported height: the role of interviewers. Economics & Human Biology. https://doi.org/ 10.1016/j.ehb.2022.101118.

PNS (2013) Pesquisa Nacional de Saúde 2013 – Manual de Antropometria. Instituto Brasileiro de Geografia e Estatistica. Rio de Janeiro. [Available at: https://biblioteca.ibge.gov.br/visualizacao/instrumentos_de_coleta/doc3426.pdf]

Prospective Studies Collaboration, Whitlock G, Lewington S, et al (2009) Body-mass index and cause-specific mortality in 900000 adults: collaborative analyses of 57 prospective studies. The Lancet 373(9669):1083-1096.

Puhl RM, Heuer CA (2009) The stigma of obesity: a review and update. Obesity 17(5):941-964.

Reis RCPD, Duncan BB, Malta DC et al (2022) Evolution of diabetes in Brazil: prevalence data from the 2013 and 2019 Brazilian National Health Survey. Cadernos de Saúde Pública 38:e00149321. https://doi.org/10.1590/0102-311X00149321

Rimes-Dias, K.A., Costa, J.C., Canella, D. S. (2022). Obesity and health service utilization in Brazil: data from the National Health Survey. *BMC Public Health*, *22*(1), 1474.

Rooth DO (2009) Obesity, attractiveness, and differential treatment in hiring a field experiment. Journal of Human Resources 44(3):710-735.

Rtveladze K, Marsh T, Webber L, et al (2013) Health and Economic Burden of Obesity in Brazil. PLoS ONE 8(7):e68785. https://doi.org/10.1371/journal.pone.0068785

Sattler KM, Deane FP, Tapsell L, Kelly PJ (2018) Gender differences in the relationship of weight-based stigmatisation with motivation to exercise and physical activity in overweight individuals. Health Psychology Open 5(1):2055102918759691. https://doi.org/10.1177/2055102918759691

Sherry B, Jefferds ME, Grummer-Strawn LM (2007) Accuracy of adolescent self-report of height and weight in assessing overweight status: a literature review. Archives of Pediatrics & Adolescent Medicine 161(12):1154-1161.

Szwarcwald CL, Malta DC, Pereira CA, et al (2014). Pesquisa Nacional de Saúde no Brasil: concepção e metodologia de aplicação. Ciência & Saúde Coletiva 19(2):333-342.

Triaca LM, Jacinto PA, França MTA, Tejada CAO (2020) Does greater unemployment make people thinner in Brazil? Health Economics 29:1279-1288.

U.S.D.H.H.S. (2010) The Surgeon General's Vision for a Healthy and Fit Nation. U.S. Department of Health and Human Services, Office of the Surgeon General, Rockville, MD

Zelenytė V, Valius L, Domeikienė A, et al (2021) Body size perception, knowledge about obesity and factors associated with lifestyle change among patients, health care professionals and public health experts. BMC Family Practice 22(1):1-13.

Zhang Q, Wang Y (2004) Socioeconomic inequality of obesity in the United States: do gender, age, and ethnicity matter? Social Science & Medicine 58(6):1171-1180.

# Appendix

Figure A1: Kernel densities for the absolute differences between the 1st and 2nd body weight and height physical measurement
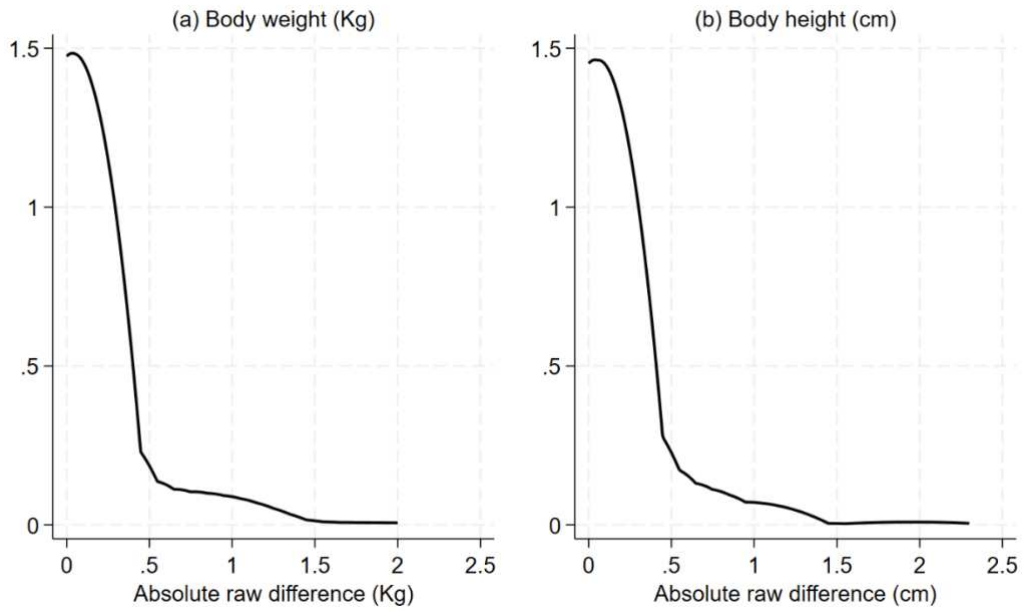
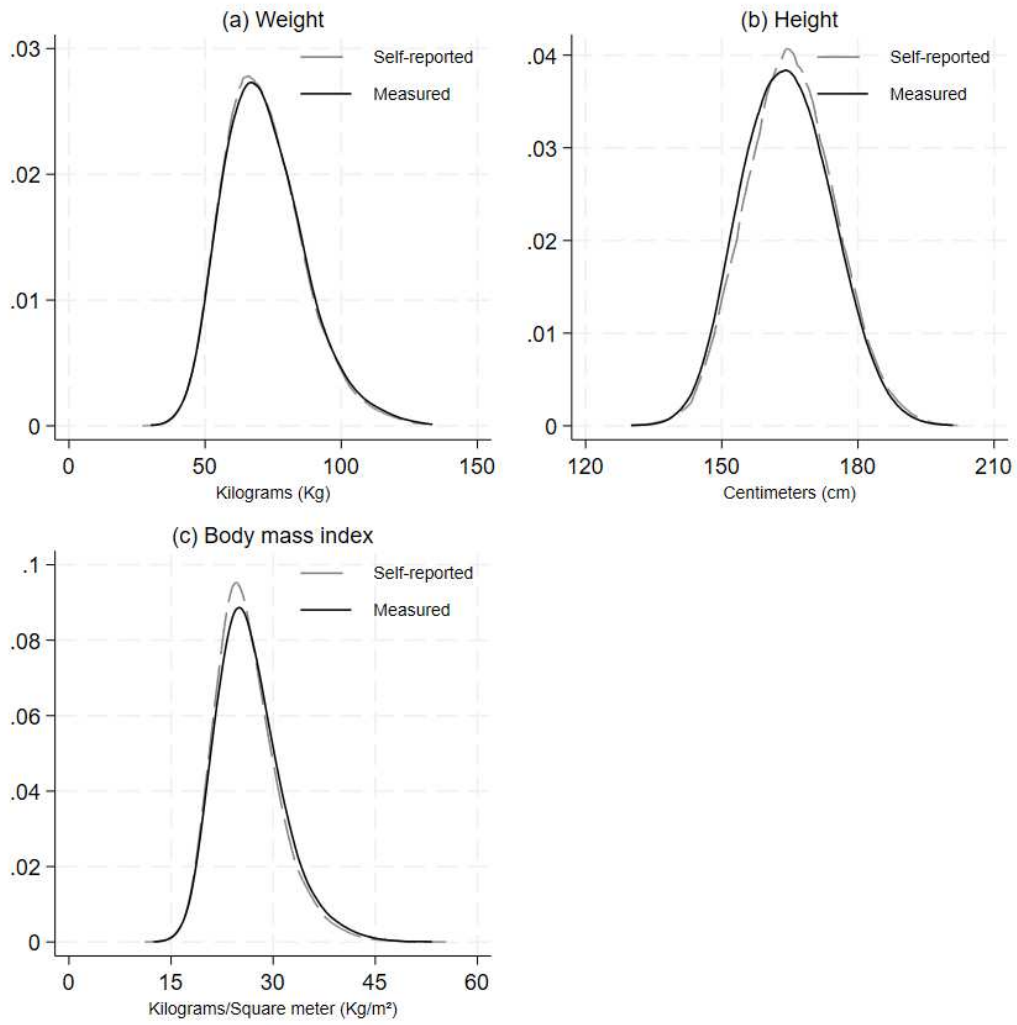# Figure A2. Kernel densities: body weight, height, and BMI.

Table A1: Estimation of factor mixture model for body weight and height – measured weight/height data are rounded at the nearest integer.

| Parameter | Weight (kg) | Height (cm) |
|---|---|---|
| $\mu_\xi$ | 71.571*** | 164.338*** |
| | (0.077) | (0.050) |
| $\sigma_\xi$ | 14.883*** | 9.459*** |
| | (0.055) | (0.035) |
| $\mu_\zeta$ | 79.543*** | 158.978*** |
| | (1.457) | (0.450) |
| $\sigma_\zeta$ | 20.207*** | 8.591*** |
| | (0.949) | (0.307) |
| $\mu_\eta$ | 0.080*** | 1.065*** |
| | (0.021) | (0.029) |
| $\sigma_\eta$ | 2.287*** | 2.306*** |
| | (0.023) | (0.037) |
| $\mu_\omega$ | -0.511*** | 0.544*** |
| | (0.101) | (0.094) |
| $\sigma_\omega$ | 5.936*** | 4.757*** |
| | (0.127) | (0.101) |
| $\pi_r$ | 0.990*** | 0.970*** |
| | (0.001) | (0.002) |
| $\pi_s$ | 0.265*** | 0.334*** |
| | (0.002) | (0.003) |
| $\pi_\omega$ | 0.252*** | 0.379*** |
| | (0.009) | (0.016) |
| $\rho$ | -0.041*** | -0.056*** |
| | (0.001) | (0.002) |
| **Class probabilities** | | |
| $Pr(R=1, S=1)$ | 0.263*** | 0.324*** |
| | (0.002) | (0.002) |
| $Pr(R=1, S=2)$ | 0.544*** | 0.401*** |
| | (0.007) | (0.011) |
| $Pr(R=1, S=3)$ | 0.183*** | 0.245*** |
| | (0.006) | (0.011) |
| $Pr(R=2, S=1)$ | 0.003*** | 0.010*** |
| | (0.000) | (0.001) |
| $Pr(R=2, S=2)$ | 0.005*** | 0.012*** |
| | (0.001) | (0.001) |
| $Pr(R=2, S=3)$ | 0.002*** | 0.008*** |
| | (0.000) | (0.001) |
| Log-likelihood | -249,954.7 | -230,849.5 |
| Observations | 37,335 | 37,335 |

Notes: The fraction of labelled observations (i.e., $|r_i - s_i| = 0$) is 26.3% for weight, and 32.4% relative to height.
*** p<0.01

Table A2: Estimation of factor mixture model for body weight and height (measured data: average between 1$^{st}$ and 2$^{nd}$ measurement).

| Parameter | Weight (kg) | Height (cm) |
|---|---|---|
| $\mu_\xi$ | 71.936*** | 164.519*** |
| | (0.077) | (0.050) |
| $\sigma_\xi$ | 14.848*** | 9.443*** |
| | (0.055) | (0.035) |
| $\mu_\zeta$ | 79.557*** | 159.380*** |
| | (1.208) | (0.391) |
| $\sigma_\zeta$ | 19.751*** | 8.716*** |
| | (0.793) | (0.260) |
| $\mu_\eta$ | -0.342*** | 0.333*** |
| | (0.014) | (0.022) |
| $\sigma_\eta$ | 1.600*** | 1.695*** |
| | (0.017) | (0.026) |
| $\mu_\omega$ | -0.351*** | 1.216*** |
| | (0.065) | (0.064) |
| $\sigma_\omega$ | 5.097*** | 4.320*** |
| | (0.082) | (0.066) |
| $\pi_r$ | 0.988*** | 0.967*** |
| | (0.001) | (0.002) |
| $\pi_s$ | 0.088*** | 0.217*** |
| | (0.001) | (0.002) |
| $\pi_\omega$ | 0.309*** | 0.437*** |
| | (0.007) | (0.011) |
| $\rho$ | -0.023*** | -0.032*** |
| | (0.001) | (0.002) |
| **Class probabilities** | | |
| $Pr(R=1, S=1)$ | 0.087*** | 0.210*** |
| | (0.001) | (0.002) |
| $Pr(R=1, S=2)$ | 0.622*** | 0.426*** |
| | (0.007) | (0.009) |
| $Pr(R=1, S=3)$ | 0.278*** | 0.331*** |
| | (0.007) | (0.008) |
| $Pr(R=2, S=1)$ | 0.001*** | 0.007*** |
| | (0.000) | (0.000) |
| $Pr(R=2, S=2)$ | 0.008*** | 0.015*** |
| | (0.001) | (0.001) |
| $Pr(R=2, S=3)$ | 0.004*** | 0.011*** |
| | (0.000) | (0.001) |
| Log-likelihood | -250,889.2 | -234,764.3 |
| Observations | 37,335 | 37,335 |

Notes: Robust standard errors to heteroscedasticity in parentheses. The fraction of labelled observations (i.e., $|r_i - s_i| = 0$) is 8.7% for weight, and 21.0% relative to height.
*** p<0.01

Table A3: "Corrective equations" for body weight and height.

| | Measured weight | Measured height |
|---|---|---|
| Self-reported weight | 1.0015*** | - |
| | (0.0017) | |
| Self-reported height | - | 0.8344*** |
| | | (0.0035) |
| Male | -0.5494* | 2.4708*** |
| | (0.3314) | (0.0575) |
| Age | 0.0718*** | -0.0184** |
| | (0.0105) | (0.0072) |
| Age² | -0.0008*** | -0.0003*** |
| | (0.0001) | (0.0001) |
| Male x Age | -0.0226 | - |
| | (0.0149) | |
| Male x Age² | 0.0003** | - |
| | (0.0002) | |
| Adj. R² | 0.9366 | 0.8679 |
| Observations | 37,335 | 37,335 |

Note: Standard errors robust to heteroscedasticity in parentheses.
*** p<0.01, ** p<0.05, and *p<0.1.