Economics and Human Biology 45 (2022) 101118

Contents lists available at ScienceDirect

Economics and Human Biology

journal homepage: www.elsevier.com/locate/ehb

The reliability of adult self-reported height: The role of interviewers

Lukas Olbrich [a,b,*], Yuliya Kosyakova [a,c], Joseph W. Sakshaug [a,b,d]

[a] Institute for Employment Research (IAB), Nuremberg, Germany
[b] Ludwig-Maximilian University of Munich, Munich, Germany
[c] Otto-Friedrich University of Bamberg, Bamberg, Germany
[d] University of Mannheim, Mannheim, Germany

ARTICLE INFO

Keywords:
Self-reported height
Misreporting
Interviewer effects
Interview behavior
Measurement error

ABSTRACT

Surveys serve as an important source of information on key anthropometric characteristics such as body height or weight in the population. Such data are often obtained by directly asking respondents to report those values. Numerous studies have examined measurement errors in this context by comparing reported to measured values. However, little is known on the role of interviewers on the prevalence of irregularities in anthropometric survey data. In this study, we explore such interviewer effects in two ways. First, we use data from the US National Health and Nutrition Examination Survey and the UK Household Longitudinal Study to evaluate whether differences between reported and measured values are clustered within interviewers. Second, we investigate changes in adult self-reported height over survey waves in two German large-scale panel surveys. Here, we exploit that height should be constant over time for the majority of adult age groups. In both analyses, we use multilevel location-scale models to identify interviewers who enhance reporting errors and interviewers for whom unlikely height changes over waves occur frequently. Our results reveal that interviewers can play a prominent role in differences between reported and measured height values and changes in reported height over survey waves. We further provide an analysis of the consequences of height misreporting on substantive regression coefficients where we especially focus on the role of interviewers who reinforce reporting errors and unlikely height changes.

1. Introduction

Anthropometric measures are frequently used in the health and social sciences. For instance, multiple studies have investigated the associations between physical height and a variety of outcomes such as labor market success, well-being, and health (e.g., Batty et al., 2009; Case and Paxson, 2008; Deaton and Arora, 2009; Persico et al., 2004). Given a lack of administrative data on the general population, this information is frequently collected in surveys, either by taking physical measures or by asking the respondents for the respective values. While it is both simple and inexpensive to add items on height and weight to questionnaires, self-reported anthropometric measures are subject to measurement errors that can potentially affect substantive research results (Burke and Carman, 2017; Cawley, 2004). The prevalence of misreporting in self-reported anthropometric measures is well-established in the literature (for a review, see Gorber et al., 2007), though the magnitude of reporting errors seems to vary across studies. The risks of measurement errors in anthropometric data could be minimized by collecting physical measurements, however, this strategy is more costly and time-consuming as special training and measurement equipment are required (Burke and Carman, 2017).

While evidence on respondent reporting errors is abundant, the role of interviewers for the quality of self-reported anthropometric measures and their influence on substantive results is less well understood. Interviewers play a prominent role in (face-to-face) surveys due to their tasks such as establishing contact with the target respondent, gaining their cooperation, asking the survey questions, and recording the answers (West and Blom, 2017). Across those tasks, interviewers are prone to making errors of unintentional (i.e., accidental typos) and intentional (i.e., fabricating parts of the interview) nature that may affect the measurement of anthropometric variables (Groves, 2004). Intentional errors – known as interviewer falsification (AAPOR, 2003) – are of particularly high significance. Finn and Ranchhod (2017) identified fabricating interviewers by analyzing relative changes in adult height measured over two waves of a South African panel study. Their analysis indicated that interviewers influence anthropometric measures and that

* Correspondence to: Institute for Employment Research (IAB), Regensburger Str. 104, DE-90478 Nuremberg, Germany.
E-mail address: lukas.olbrich@iab.de (L. Olbrich).

https://doi.org/10.1016/j.ehb.2022.101118
Received 27 July 2021; Received in revised form 22 February 2022; Accepted 22 February 2022
Available online 28 February 2022

investigating presumably stable measures over panel waves can provide valuable insights into data quality.

The present study complements the existing literature on interviewer effects in three particular ways. First, we analyze the effect of interviewers on differences between reported and measured height values.[1] To do so, we use data from two large surveys from the US and UK that contain self-reported and measured height. Second, using data from two further large-scale German panel surveys, we investigate the interviewers' role in collecting self-reported anthropometric data with a focus on the development of respondents' height over panel waves. Here, our main proposition is that body height should be stable over survey years when younger and older respondent age groups are excluded. One of the panel surveys contains verified deviant interviewers, which allows for testing whether unlikely height changes are more likely to occur among deviant interviewers. Third, we investigate the extent to which interviewers can distort empirical results of substantive analyses involving self-reported height data.

In a nutshell, our results show that differences between reported and measured height are subject to interviewer effects, with some interviewers being particularly error-prone. For the panel surveys, our findings reveal that numerous respondents exhibit substantial and unlikely variation in reported height. Moreover, these changes are clustered at the interviewer-level and accumulate for multiple interviewers. Using the dataset containing verified deviant interviewers, we find that height changes allow for identifying deviant interviewer behavior. Lastly, we show that the consequences of height reporting errors for substantive research results are not equally distributed across interviewers.

2. Respondent- and interviewer-related measurement errors in height

Both respondents and interviewers can be a source of measurement error that drives differences between reported and measured height, on the one hand, and unlikely changes in self-reported height over panel waves, on the other hand. Before describing reasons for interviewer-related errors, we review reasons for respondent-related errors in height reporting. For each reason, we discuss potential impacts on misreporting and on changes over panel waves. Note that the following discussion is restricted to respondents who are generally not in the age range for expected growth (younger age groups) or decline (older age groups). After outlining reasons for respondent- and interviewer-related errors, we provide a simple formal model for interviewer-related measurement error in reported height, and discuss consequences for substantive regression results.

2.1. Respondent measurement error

Respondent-related measurement error in height reporting can occur in several ways. First, respondents may not exactly know their height and thus misreport. This can also lead to changes in reported height over panel waves. For instance, in some years the respondent may report a height of 178 cm, and a height of 177 cm in other years, thus creating variation in height over waves. A related error is heaping (or rounding), which describes bunching at values ending with 5 or 0 (Heineck, 2006). While these behaviors can explain smaller errors and changes in reported height, deviations of 5 cm or larger are suspicious as respondents are expected to know the range of their height. For example, it is improbable that respondents report that they are 178 cm in one wave and report that they are 173 cm two years later.

Furthermore, social desirability bias – the tendency to provide responses in line with social norms (Tourangeau and Yan, 2007) – could contribute to misreporting of height and weight (Burke and Carman, 2017; Larson, 2000). While height is often subject to overreporting, weight tends to be under-reported (Gorber et al., 2007), particularly among women (Boström and Diderichsen, 1997; Gil and Mora, 2011). Due to increasing familiarity with the survey and the interviewer, the amount of social desirability bias might increase or decrease over multiple waves of a panel study – a phenomenon known as panel conditioning (Warren and Halpern-Manners, 2012) – which can produce variation in reported height over time; though, Uhrig (2012) found no effect of panel conditioning on social desirability bias in reported height in a conditioning experiment in the Understanding Society Innovation Panel. The effects of social desirability bias are also mediated by survey mode. For instance, Pinkston (2017) found that misreporting anthropometric measures poses a larger problem in phone surveys than in face-to-face surveys for the National Longitudinal Survey of Youth (NLSY), which suggests that visual interactions are a deterrent to providing erroneous responses in interviewer-administered surveys. In contrast, Kroh (2005) found that male respondents tend to report higher weight values in self-administered surveys than in face-to-face surveys in the German Socio-Economic Panel (GSOEP), indicating that social desirability bias is lessened in a self-administered setting. Note that in these studies the mode was not randomly assigned to respondents and thus selection effects may contribute to the differences.

Altogether, respondent measurement errors can play an important role in the collection of anthropometric data. Hence, we account for respondent characteristics associated with the knowledge of their height and social desirability bias in our analysis. For changes in reported height over panel waves, however, reporting biases driven by socially desirable responding are of lesser concern since height itself and its associated biases should remain constant over time, compared to other anthropometric measures such as weight, which tends to fluctuate throughout adulthood and, thus, is more susceptible to differential misreporting (Burke and Carman, 2017). Therefore, respondent errors are likely to have limited effects on changes in reported height over time.

2.2. Interviewer measurement error

As with respondent-related measurement error in height, there are several explanations for deviations in height that can be attributed to interviewers. A rather simple interviewer-driven explanation refers to typos when entering responses. For instance, interviewers may transpose or incorrectly enter an adjacent number, i.e., 176 instead of 167 or 172 instead of 182. Similar to respondent errors, such errors should be distributed evenly and only accumulate for the most careless interviewers. Although such errors are unintentional, they harm data quality and prone interviewers should be monitored.

Another reason for errors in reported height is proxy interviewing (Moore, 1988). With the help of an available household member, some interviewers might opt for filling-out the questionnaire for non-present respondents. As the proxy respondent is unlikely to know the exact height of the intended respondent, the value might be misreported. Such proxy interviewing also generates changes in the reported height over survey waves if the intended respondent was available and reported their height in previous waves. If individual interviewers repeatedly resort to proxy interviewing without reporting it as a proxy interview, such interviewers will exhibit frequent errors in the reported height and changes over panel waves.

Erroneous selection of respondents could represent a further interviewer-related source of height error (Eckman and Koch, 2019). For instance, a previously participating respondent may refuse participation, but a neighboring respondent is willing to do so. The interviewer could conduct the interview with the willing respondent and avoid the stress of returning for further conversion attempts with the intended respondent. Such behavior is classified as interviewer falsification

[1] Since a person's body weight is likely to vary between reporting and measurement if not taken immediately after each other, we focus only on respondent height.

(AAPOR, 2003). However, this behavior may also happen by mistake. Nonetheless, as the interview is not conducted with the intended person, the respondent's reported height would differ from the true height; hence, the respondent's height profile is likely to change over waves in panel studies. If interviewers systematically apply such erroneous selection, they will be characterized by frequent and large height errors or changes in respondents' reported height.

Interviewer-related measurement errors in height can also occur when the interviewer deliberately fabricates the interview either fully or partially (i.e., by administering only parts of the questionnaire with the respondent and filling the rest in by themselves; see De Haas and Winker, 2016). Respondent height is a natural candidate for partial fabrication as the interviewer can estimate the height instead of asking for it, thus saving time and effort. Since interviewers are often paid per completed interview, shortening the interview effectively increases the interviewer's hourly wage (Josten and Trappmann, 2016; Kosyakova et al., 2015). In panel studies with high degrees of interviewer continuity, interviewers may also try to keep the interview short to reduce the respondents' burden and ensure cooperation in future panel waves. Additionally, the interviewer may assume that such behavior will not be detected anyway as they actually visited the household, conducted the interview, and their estimate of the respondent's height is probably 'close enough'. However, such partial fabricators are unlikely to guess the respondent's height correctly and in panel studies they are unlikely to recall the height value (either guessed or respondent-provided) that was recorded in previous rounds. Thus, interviewers who fabricate respondent heights will be characterized by differences between reported and measured height. In panel surveys, such interviewers are expected to have frequent changes in respondent height over survey waves.

In cases of complete fabrication of interviews, interviewers may not even visit the household to conduct parts of the interview and not observe the intended respondent at all. Although complete fabrication is rare, multiple case studies reveal that interviewer fabrication can pose a substantial threat to survey data (Finn and Ranchhod, 2017; Schräpler and Wagner, 2005; Schwanhäuser et al., 2020). With regard to self-reported height, fabricators cannot know the respondent's height and therefore the reported value is entirely made up. Thus, while partial fabricators can at least infer the height from observing the respondent, complete fabricators instead generate artificial information about the respondent that will differ from the true height. In panel studies, the artificial value will likely differ if the intended respondent participated in earlier waves and reported their height.

Note that the described interviewer-related measurement errors are not associated with systematic under- or overreporting, but rather generate unsystematic error. Moreover, most of the reasons laid out above refer to the behavior of single error-prone interviewers.

2.3. A formal model of interviewer-related measurement error in reported height

We formalize interviewer-related measurement error in reported height following the model of Crossley et al. (2021):

$$height_{ij} = height_i^* + \pi_j w_i + u_j \quad (1)$$

The variable $height_{ij}$ is observed in the data, where i corresponds to the respondent and j to the interviewer. The true (or measured) values and their variances are denoted by $height_i^*$ and σ_ν^2. For the analysis of changes in self-reported height over panel waves, $height_i^*$ corresponds to the previously reported height. The terms $\pi_j w_i$ and u_j represent the measurement error and both depend on the interviewers. The classical interviewer error that indicates the impact of interviewers on the reported height is denoted by u_j and is distributed with mean zero and variance σ_u^2. This term captures that interviewers obtain lower or higher reported heights, on average. In the measurement error model, interviewers also mediate the individual respondent reporting errors w_i that are distributed with mean zero and variance σ_w^2. Large values ($\pi_j > 1$) of π_j imply that interviewers enhance errors, while smaller values ($\pi_j < 1$) correspond with a reduction in respondent reporting errors.

The main focus of the analysis is on parameter π_j. Differences between measured and reported values and observed changes in reported height over panel survey waves are interpreted as individual reporting errors. However, as argued above, such errors may actually be driven by interviewers through frequent typos, proxy interviewing, erroneous respondent selection, or fabricated responses. In particular, we investigate whether interviewers differ with regard to π_j and whether there are exceptional interviewers who consistently show large errors in reported height (i.e., large values in π_j).

2.4. Consequences of interviewer-related measurement error

Consequences of an individuals' height for labor market outcomes, health outcomes, or general well-being are frequently assessed in a variety of fields, and measurement error in height can affect these estimated effects. Therefore, we also explore the extent to which interviewers can influence regression results. Classical measurement error theory states that random errors in the explanatory variable in linear regressions will attenuate the estimated coefficient (Fuller, 1987). With regard to interviewer errors, Crossley et al. (2021) showed that – under the assumption of independence of interviewer errors, reporting errors, the residual, and the true values – the variance of the classical interviewer effect, the expected value of the interviewer effect on individual reporting errors, and the variance of the latter interviewer effect increase attenuation. However, previous research on reporting errors in height showed that reporting errors are non-random (i.e., negatively correlated with true height, see O'Neill and Sweetman, 2013), and therefore the direction of the bias in estimated coefficients induced by reporting errors is unknown to the empiricist. In this study, we evaluate the extent to which single interviewers can affect regression coefficients and whether these effects differ across interviewers.

3. Data

As we seek to analyze deviations in reported height both from measured values and from previously reported values, we use data from several surveys. In the following, we briefly discuss these surveys and their measurement of height.

3.1. National Health and Nutrition Examination Survey (NHANES) III

The National Health and Nutrition Examination Survey (NHANES) III is a cross-sectional survey conducted in the United States from 1988 to 1994 (National Center for Health Statistics, 1994).[2] Data was collected by mobile interviewer and health examination teams who traveled across 89 survey locations. The interviewers conducted face-to-face interviews and at the end of each interview respondents were informed about the health examination. If respondents agreed to participate (77% participation rate), the interviewers scheduled an appointment for the examination that took place in mobile examination centers. In the survey, respondents were asked for their height without shoes in feet and inches. In the examination center, professional technicians measured the respondent's height.[3] We use the differences between these values to evaluate interviewer effects on reporting errors. As interviewers were aware that measurements would be taken a few weeks

[2] More recent publicly available NHANES data does not contain the interviewer ID variable, which is essential for our analysis.

[3] To avoid recording errors in the measured height, the examiners took photographs of the height scale and the height was recorded based on this photograph and later on compared to the photograph in a quality check.

Table 1
Descriptive statistics for each sample.

	N	N of int.	Avg. error (cm)	Avg. abs. error (cm)	% Abs. error > 5 cm
NHANES					
Female	5507	60	0.407	2.168	7.554
Male	4770	60	0.872	2.348	8.386
Total	10,277	60	0.623	2.251	7.940
UKHLS					
Female	3331	208	0.756	1.995	6.244
Male	2326	208	1.789	2.593	11.436
Total	5657	208	1.181	2.241	8.379
GSOEP					
Female	3961	195	-0.037	0.874	3.509
Male	3140	195	-0.071	0.957	3.726
Total	7101	195	-0.052	0.910	3.605
PASS					
Female	1127	82	-0.170	0.815	2.662
Male	920	82	0.023	0.940	3.913
Total	2047	82	-0.084	0.871	3.224

Notes: The average error is the average of the reported height minus the measured height. The average absolute error is the average of the absolute difference between the reported and measured height.

after the interview, deliberate fabrications of height values seem rather unlikely. In addition, the NHANES team recontacted roughly 10% of the respondents for verification to ensure that interviews were adequately conducted, and each questionnaire was checked for "error patterns" (National Center for Health Statistics, 1994, p. 31).

3.2. UK Household Longitudinal Study (UKHLS)

The UK Household Longitudinal Study (UKHLS) is a yearly household panel study that started in 2008 (University of Essex: Institute for Social and Economic Research, 2021). In the second wave, respondents of its predecessor, the British Household Panel Study (BHPS), were integrated in the UKHLS. In the first UKHLS wave, face-to-face interviewers asked respondents for their height and weight. In wave 2, a subsample of the general population sample respondents was selected for separate nurse visits that took place roughly 6 months after the interview (University of Essex: Institute for Social and Economic Research and National Centre for Social Research, 2014).[4] Professional nurses visited the respondents and took a variety of biomeasures such as height and weight, blood pressure, or the collection of blood samples (see Buck and McFall, 2012, for an overview). In this study, the main analysis variables are the self-reported height in wave 1 and the measured height collected after the wave 2 interview. Assuming absence of measurement errors in the nurse measures and absence of height changes between reporting and measurement, we can compare the self-reported height to the true measured height. Furthermore, wave 1 interviewers were likely unaware of the biomeasure collection after the wave 2 interview and thus might have been more careless as verification of collected values was not imminent. Note, however, that the UKHLS interviewers were controlled using a sophisticated monitoring system (Boreham et al., 2012), deeming extreme deviant behavior very unlikely.

3.3. German Socio-Economic Panel (GSOEP)

The German Socio-Economic Panel (GSOEP, DOI:10.5684/soep-core.v35) is a nationally representative, multi-mode household panel survey launched in 1984 (Goebel et al., 2019).[5] Since 2002, the GSOEP collects self-reported height and weight every two years. Our main variable of interest is the reported height. The exact wording of the question is "How tall are you? If you don't know, please estimate."[6] We use data from the most recent pair of height reports in 2016 and 2018. Due to the repeated collection of respondent height, it is possible to examine changes in reported height over several years for individual respondents.

To counteract such deviations, the GSOEP also provides an imputed and edited version of the height variable in the HEALTH data, a dataset that is part of GSOEP-Core and specifically provided to ease the analysis of health modules (SOEP Group, 2020). Missing values are simply replaced by the most recent reported height. Furthermore, "[i]t is assumed that for a two-year-period a change of body height of more than 10 cm is implausible if the values of the other observation years differ only in a range of at most 2 cm. Thus the respective information is imputed by the average of the other values of the respondent" (SOEP Group, 2020, p. 13). However, such editing has only been applied to 43 cases over all survey years (2002–2018) and all age ranges.

3.4. Panel Labour Market and Social Security (PASS)

The last dataset is the Panel Study Labour Market and Social Security (PASS, DOI: 10.5164/IAB.PASS-SUF0619.de.en.v2, an annual survey of households that receive unemployment benefits and households of German residents launched in 2007 (Trappmann et al., 2013, 2019). Interviews are conducted with computer-assisted telephone or computer-assisted personal interviewing mode (CATI and CAPI, respectively). During the field period of wave 15, two interviewers suspicious of deviant interviewing were identified based on paradata and survey data analysis, audio recordings of the interviews, and recontact procedures (Beste et al., 2021). The first interviewer (ID 154) had extremely short interview durations and recontacts indicated that the interviewer completely fabricated some interviews. The second interviewer (ID 109) also had implausible interview durations but recontacts did not yield evidence of complete fabrications. Analyses of previous waves suggested that interviewer 109 acquired habits of speeding through the questionnaire and responding to questions instead of posing them to the respondent. Due to workload limits in the PASS, the consequences for survey results of these deviant interviewers are negligible. However, these cases provide a rare opportunity to assess whether unlikely changes in the reported height over panel waves accumulate for verified deviant interviewers. We use data from waves 9 and 12, in which both deviant interviewers were active and the PASS questionnaire contained an item on the respondent's height.

3.5. Sample description

Across all datasets, we apply the same sample restrictions. Respondents for which either reported height or measured/previously reported height is missing are excluded from the analysis.[7] With regard to the interview mode, the analysis is restricted to face-to-face interviews with the intended respondent. All telephone interviews, self-administered interviews without interviewer presence, and proxy interviews are excluded. Respondents who are younger than 21 years of age and older than 60 are dropped to avoid natural growth and shrinkage effects (Case and Paxson, 2008; Fernihough and McGovern, 2015). We exclude respondents who reported height values below 130 cm and respondents with height deviations of 30 cm or higher to ensure that our results are not driven by outliers. Lastly, we restrict the

[4] BHPS respondents were sampled for nurse visits in wave 3. As we do not have data on self-reported height for these respondents, these data were not included in our analysis.

[5] More information on the GSOEP is available at https://www.diw.de/soep.

[6] Note that this formulation explicitly allows for providing imprecise responses. However, as argued above, such imprecision should occur within a limited range and randomly vary across interviewers.

[7] Interviewers who systematically produce item nonresponse for the height question will therefore not be identified by our analysis.

samples to interviewers with more than 15 interviews to ensure that we obtain reliable measures of interviewer effects on the residual standard deviation. Fig. A1 in the appendix shows scatter plots of the reported height versus the measured or previously reported height for each sample.

Table 1 provides an overview of the resulting samples and the prevalence of reporting errors by gender. The NHANES data contain more than 10,000 observations who were interviewed by only 60 interviewers. The average error (reported minus measured height) is positive for both males and females which indicates that height is overreported. For approximately 7.9% of the sample the absolute error (absolute difference between reported and measured height) exceeds 5 cm. With regard to gender differences, errors tend to be larger for males, but the differences are minor. For the UKHLS, average interviewer cluster sizes are smaller (5657 observations distributed across 208 interviewers). Furthermore, gender seems to play a more important role in this sample. Males overreport their height by 1.8 cm, on average, while females overreport only by roughly 0.8 cm. For 11.4% of the males, the difference even exceeds 5 cm, whereas this share is only 6.2% for females. In the GSOEP, the average number of observations per interviewer is slightly higher than in the UKHLS. The average error (reported height minus height reported two years earlier) is close to zero and thus not as systematic as the errors found in the samples with measured height. The average absolute error is also substantially lower and below one cm. Approximately 3.6% of the sample have changes above 5 cm. On average, there are only minor differences between males and females. For the PASS, the patterns are similar to the GSOEP. The changes in height are non-systematic and below one cm in absolute terms. Differences between males and females are slightly larger than in the GSOEP, but still minor compared to the UKHLS.

Altogether, differences between reported and measured values are systematic, whereas changes in reporting across panel waves are smaller and not systematic. Note, however, that changes in reported height should be interpreted as an additional source of measurement error, rather than a substitute to deviations from measured values.

4. Analytical approach

4.1. Identification of interviewer-related measurement error

To examine interviewer effects on reporting errors, we rely on multilevel models. Multilevel models are the most commonly used method to investigate interviewer effects on survey outcomes as they allow the researcher to account for the clustering of respondents within interviewers (Schnell and Kreuter, 2005; West et al., 2013; West and Blom, 2017). In particular, we use multilevel location-scale models that provide a framework for modeling interviewer effects on the residual standard deviation (Brunton-Smith et al., 2017; Hedeker et al., 2008; Sturgis et al., 2021), and thus enable us to estimate whether interviewers vary with regard to π_j (see Eq. 1) and which interviewers especially enhance errors in the reported height.

The model with reported height observed for respondent i nested within interviewer j as the dependent variable is defined as:

$$height_{ij} = \beta_0 + \beta_1 height^*_{ij} + X\gamma + \theta_{1j} + \varepsilon_{ij} \quad (2)$$

$$\ln(\sigma_\varepsilon) = X\alpha + \theta_{2j}$$

The first line in Eq. (2) denotes the location equation. In the model, β_0 is a constant and β_1 is the coefficient of the measured height (for the panel studies the previously measured height), that is expected to be close to one. We also include a set of respondent characteristics (X) that could affect height reporting. In this framework, positive coefficients imply overreporting of height, while negative coefficients imply underreporting. The classical interviewer effects are denoted by θ_{1j} and are distributed with mean zero and standard deviation σ_{θ_1}. The residual is denoted by ε_{ij} and distributed with mean zero and standard deviation σ_ε. The second line in Eq. (2) denotes the scale equation. Here, the logarithm of the residual standard deviation is the dependent variable. The constant and control variables with corresponding coefficients are included in $X\alpha$. Positive coefficients imply error-enhancing characteristics, while negative coefficients imply the opposite. Interviewer effects on the residual standard deviation are denoted by θ_{2j} and distributed with mean zero and standard deviation σ_{θ_2}. Going back to Eq. (1), θ_{2j} is equivalent to π_j and allows us to infer whether specific interviewers operate as error-enhancing or error-reducing. Furthermore, σ_{θ_2} signifies the extent to which differences in errors across interviewers actually play a role.

In summary, the multilevel location-scale model allows for estimating the effects of interviewers and respondent characteristics both on systematic over- or underreporting and on the prevalence of unsystematic errors. With regard to respondent-related characteristics, we control for gender to account for differences in social desirability bias. As previous literature shows that reporting errors correlate with the respondent's age, we also include an age variable. Note that this association seems to be most significant for respondents aged above 60 (see Davillas and Jones, 2021). To approximate potential language difficulties during the interview, a binary variable on citizenship (for instance, in Germany acquiring of which requires language proficiency of "independent user" level, B1, see Council of Europe, 2001) is included. Since citizenship was not available in NHANES III, we used a binary variable indicating whether the respondent was born in the US.[8] In addition, it is presumed that knowledge of height correlates with respondent education, hence, we control for years of education. We account for missings in the control variables by including corresponding dummy indicators.[9]

4.2. Analysis of the consequences of interviewer-related measurement error in reported height on regression coefficients

With regard to the effects of measurement error on substantive research results, we do not evaluate the overall impact on regression coefficients, but rather analyze the extent to which interviewers differ in their effect on regression coefficients. To do so, we rely on so-called corrective equations with the measured or previously reported height as dependent variable and the current reported height and further control variables as independent variables (Bound et al., 2001; Cawley, 2004; Davillas and Jones, 2021; Lee and Sepanski, 1995). Such models are frequently used to correct for measurement error in datasets where the variable of interest is measured with error and thus coefficients based on this variable would be attenuated (in the absence of differential measurement error). In that case, the corrective equation is estimated in an auxiliary dataset where both true and erroneously measured values (i.e., measured and reported height) are available. Following this step, the estimated coefficients are used in the initial dataset to predict presumably error-free values which are then used for further analyses. Absent any measurement errors, the coefficient of the reported height in the corrective equation would be one and the coefficients of the control variables zero. With increasing measurement error, the coefficient will increasingly differ from one, which also means that using the reported instead of the measured variable as an explanatory variable in regression analyses will result in larger biases in the estimated coefficient.

To evaluate the impact of single interviewers, we first estimate the corrective equation described above. Second, we replace the reported

[8] Being born in the US is a more general measure than citizenship and may proxy further respondent characteristics besides language difficulties. Thus, we are cautious with interpreting the coefficient as the consequence of language problems and comparing coefficients across samples.

[9] Citizenship was missing for only one observation in the UKHLS sample. This observation was excluded from the analysis.

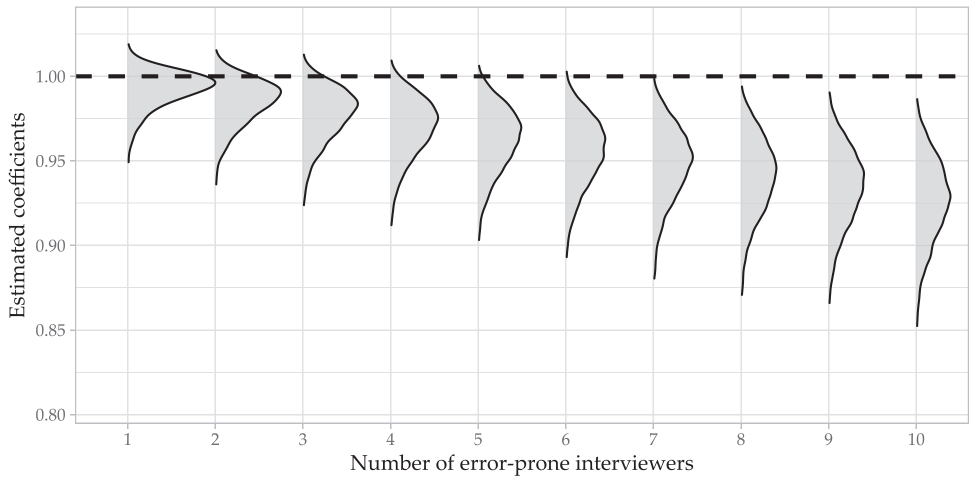

Fig. 1. Simulation results.

height of the interviewer under investigation with the measured or previously reported values while leaving the values of all other interviewers as is and fit the corrective equation again. Next, we calculate the difference between the resulting height coefficient and the height coefficient for the data without replacement. Lastly, as the change in the coefficient is also dependent on the interviewer's workload, we divide the difference by the workload. We repeat this for each interviewer separately. As a result, we obtain a measure of each interviewer's relative effect on the bias in regression coefficients.

4.2.1. Simulation analysis

To demonstrate the potential impact of error-prone interviewers on regression estimates, we provide a simple simulation analysis. To do so, we rely on a basic framework with measurement error in the explanatory variable. We assume a dataset consisting of 50 interviewers conducting 2500 interviews in total. The explanatory variable x_i is distributed with mean 167.75 and standard deviation 9.64 (measured height values taken from the NHANES sample) and the true regression equation is $y_i = 0.5 + 1x_i + \nu_i$, where ν_i follows a normal distribution with mean zero and a standard deviation of three.

We evaluate the impact of error-prone interviewers by simulating random error with mean zero in xi for single interviewers and re-estimating the regression. We do that for a varying number of error-prone interviewers (1 10). In every repetition, each interviewer's share of interviews is determined by drawing a value from a uniform distribution that is then rescaled to sum up to one across all interviewers. The standard deviation of the random error is drawn from a uniform distribution between one and ten for each error-prone interviewer (plausible values based on the multilevel analysis). These parameters depict that an interviewer's influence depends on the size of the random error and the interviewer's workload. For each number of error-prone interviewers we run 10,000 replications. The results are depicted in Fig. 1. The y-axis denotes the estimated coefficients, the x-axis denotes the number of error-prone interviewers, and the dashed horizontal line depicts the true coefficient.

The simulation results emphasize the uncertainty associated with the consequences of error-prone interviewers. For example, for the case of three error-prone interviewers, the bias can be as large as 7.5%, while it is below 2.5% for the majority of samples. With increasing numbers of error-prone interviewers, this uncertainty steadily increases. Hence, in some samples error-prone interviewers have little impact as their workload is small and their errors are negligible. However, in some cases interviewers with large workloads substantially exacerbate errors which can heavily impact regression coefficients.

5. Results

We fit the multilevel location-scale models presented in section Analytical approach using Markov Chain Monte Carlo Methods. To do so, we rely on the R-package brms (Bürkner, 2018) that was developed to use Stan (Carpenter et al., 2017) via R (R Core Team, 2020). We fit four chains with 16,000 iterations and a burn-in period of 8000 draws for each model. As priors we use the default non-informative priors for coefficients and half student-t priors with 3 degrees of freedom and a minimal scale parameter of 2.5 for the intercepts and the standard deviations of the interviewer effects. Convergence of the chains was assessed by $\widehat{R}$ (Gelman et al., 2013).

5.1. Correlates of reporting errors

Table 2 reports posterior means and 95% credible intervals of the covariates in the multilevel location-scale model for each sample. The location equation section denotes the extent to which the covariates contribute to systematic over- or underreporting of height. The scale equation denotes whether the covariates are associated with higher or lower residual variation.

With regard to the interviewers, the multilevel models indicate clustering effects both in the location and scale equations. Interviewer effects in the location equation are larger for the NHANES and UKHLS, which shows that interviewers play a lesser role for systematic differences between self-reported values. For the scale equation, we find evidence for a larger heterogeneity of interviewer scale effects in the GSOEP and PASS. Before investigating these effects in more detail in the next section, we briefly summarize the associations of respondent characteristics with height misreporting.

First, we discuss the results for the location equation. As expected, the coefficient for the measured or previously reported height is close to one across all samples. We also find that males are subject to more overreporting than females across all samples. In line with the descriptive results, this difference is more pronounced for the NHANES and the UKHLS. For the age groups, the results are less consistent. The coefficients provide evidence that compared to the reference group of 21–30 year-olds, the oldest age group is overreporting their height in the NHANES and UKHLS. For the GSOEP and PASS, age does not seem to play a role for reporting errors. Years of education are not associated with over- or underreporting across all samples. Lastly, citizenship or being born in the respective country correlates with underreporting in the NHANES, while we find no evidence for such a pattern in the UKHLS and GSOEP and only little evidence of overreporting in the PASS.

For the scale equation, positive coefficients denote higher residual

Table 2
Multilevel regression results. Dependent variable: Reported height.

	NHANES	UKHLS	GSOEP	PASS
Location equation				
Intercept	0.264	2.795	2.436	2.097
	[0.945, 1.474]	[1.091, 4.510]	[1.750, 3.136]	[0.720, 3.462]
Meas./prev. rep. height	1.003	0.989	0.985	0.984
	[0.995, 1.010]	[0.979, 0.999]	[0.980, 0.989]	[0.976, 0.993]
Male	0.820	1.191	0.188	0.346
	[0.677, 0.966]	[0.996, 1.390]	[0.108, 0.268]	[0.194, 0.497]
Age: 31–40	-0.196	0.107	0.011	0.127
	[0.328, 0.064]	[0.106, 0.320]	[0.106, 0.126]	[0.063, 0.316]
Age: 41–50	-0.031	0.061	0.056	0.146
	[0.172, 0.112]	[0.143, 0.265]	[0.058, 0.168]	[0.041, 0.334]
Age: 51–60	0.445	0.403	-0.027	0.060
	[0.285, 0.602]	[0.183, 0.622]	[0.143, 0.088]	[0.128, 0.242]
Education	-0.015	-0.004	0.007	0.004
	[0.034, 0.005]	[0.026, 0.018]	[0.003, 0.016]	[0.015, 0.023]
Citizen/Born in country	-0.364	-0.381	-0.020	0.274
	[0.551, 0.178]	[0.728, 0.035]	[0.133, 0.095]	[0.051, 0.596]
Scale equation				
Intercept	1.890	1.363	1.976	1.669
	[1.811, 1.969]	[1.228, 1.498]	[1.849, 2.102]	[1.413, 1.924]
Male	0.002	0.100	0.121	0.111
	[0.026, 0.030]	[0.059, 0.140]	[0.083, 0.160]	[0.035, 0.187]
Age: 31–40	-0.079	-0.038	-0.168	-0.081
	[0.115, 0.043]	[0.100, 0.023]	[0.232, 0.104]	[0.208, 0.044]
Age: 41–50	-0.126	-0.096	-0.194	-0.163
	[0.165, 0.086]	[0.155, 0.036]	[0.256, 0.133]	[0.290, 0.041]
Age: 51–60	-0.063	-0.082	-0.259	-0.115
	[0.106, 0.020]	[0.145, 0.019]	[0.323, 0.195]	[0.236, 0.003]
Education	-0.047	-0.023	-0.090	-0.065
	[0.052, 0.043]	[0.030, 0.017]	[0.098, 0.083]	[0.079, 0.050]
Citizen/Born in country	-0.319	-0.098	-0.365	-0.437
	[0.359, 0.279]	[0.192, 0.007]	[0.420, 0.312]	[0.576, 0.302]
sd(Location Intercept)	0.151	0.434	0.025	0.068
	[0.075, 0.232]	[0.331, 0.542]	[0.001, 0.066]	[0.003, 0.162]
sd(Scale Intercept)	0.169	0.198	0.507	0.481
	[0.133, 0.211]	[0.170, 0.228]	[0.455, 0.565]	[0.405, 0.572]
N of interviewers	60	208	195	82
N of observations	10,277	5657	7101	2047

Notes: 95% credible intervals in brackets. For the UKHLS and NHANES we use the measured height, for the GSOEP and PASS we use the previously reported height as explanatory variable. The reference category for age is 21–30. For the UKHLS, GSOEP, and PASS we use a binary variable on citizenship, for the NHANES we use a binary variable on whether the respondent was born in the US.

variance and thus reporting error, while negative coefficients imply the opposite. In the NHANES, there is no evidence for gender differences, whereas in the UKHLS, GSOEP, and PASS reporting errors are more prevalent for men. With regard to the respondents' age, reporting errors seem to be the largest problem for the reference group, the 21–30 year-olds. For the other age groups, heterogeneities are rather small and do not follow a consistent pattern across all samples. As hypothesized, education reduces reporting errors across all samples which indicates that the lower-educated are more likely to misreport or simply less knowledgeable of their height. We also find that respondents with citizenship or who were born in the respective country have lower reporting errors than non-citizens. This could be a consequence of language problems during the interview.

5.2. Error-prone interviewers

Having established that interviewers affect errors and changes in height measurements in all samples, we continue by using the multilevel location-scale model results to examine error-prone interviewers. For each sample, the five interviewers with the largest posterior mean scale effect are listed in Table 3. In addition, information such as the average reporting error or the share of interviews with errors exceeding 5 cm are reported. Note that the interviewer effects are conditional on the covariates and thus interviewers with large average absolute errors might be ranked rather low if their observations have error-prone characteristics (i.e., low-educated or subject to language problems). Graphical presentations of the interviewer effects are provided in Figs. B1 to B4 in the appendix. In the following, the results for each sample are discussed in more detail.

In the NHANES, the first-ranked interviewer has 121 interviews with an average absolute reporting error of 3.43 cm. For 21 of these interviews, the absolute error exceeds 5 cm. The second-ranked interviewer has 74 interviews with errors exceeding 5 cm for almost a third of the respondents and an average absolute error of more than 4 cm. For comparison, the second-to-last ranked interviewer (not displayed in the table) has 253 interviews with errors exceeding 5 cm for four respondents and an average absolute error of 1.6 cm. The last-ranked interviewer has 28 interviews with errors exceeding 5 cm in one interview. Hence, NHANES interviewers differ in their contribution to measurement error in height. This is particularly notable as the interviewers were aware that the respondents' height would be measured a few weeks after the interview which would have allowed for verification by supervisors.

Next, we discuss the results for the UKHLS where interviewers were not aware of height measurements in later waves. The first-ranked interviewer has 19 interviews with an average absolute error of 4.9 cm and errors exceeding 5 cm for more than 30%. For the lower-ranked interviewers these values are less severe, although some interviewers still have absolute average errors exceeding 3 cm or shares of errors exceeding 5 cm above 20%. As shown in Fig. B2, we find little evidence for interviewer effects for the majority of interviewers, except for the set of interviewers at the right-hand side in Fig. B2.

Table 3
Interviewers with highest scale effects.

Rank	ID	N	Scale effect	Avg. abs. error (cm)	Max. error (cm)	% Abs. error > 5 cm
NHANES						
1	78	121	0.4036	3.43	26.38	17.36
2	82	74	0.3574	4.15	22.12	29.73
3	17	141	0.2544	2.86	17.64	14.89
4	86	71	0.2414	3.13	15.08	18.31
5	31	65	0.2380	2.70	15.52	13.85
UKHLS						
1	50	19	0.5271	4.85	21.90	31.58
2	584	20	0.5028	3.34	26.74	5.00
3	303	16	0.4734	3.32	18.10	18.75
4	680	33	0.4406	3.73	15.02	24.24
5	537	49	0.4397	2.22	25.04	6.12
GSOEP						
1	408	18	1.2960	2.67	22.00	11.11
2	397	22	1.2700	1.27	20.00	4.55
3	331	47	1.1943	4.09	22.00	27.66
4	97	32	1.0686	1.53	20.00	9.38
5	416	30	1.0277	1.10	10.00	10.00
PASS						
1	217	23	0.8742	1.65	10.00	8.70
2	154	20	0.8613	3.40	10.00	35.00
3	88	16	0.8448	2.38	11.00	18.75
4	113	20	0.7553	1.25	14.00	10.00
5	241	17	0.7486	1.71	13.00	11.76

Notes: The scale effects denote the mean posteriors of the interviewer scale effects estimated in the multilevel location-scale models.

Table 4
Regression results. Dependent variable: Measured/previously reported height.

	NHANES	UKHLS	GSOEP	PASS
Intercept	26.767 ***	21.538 ***	9.155 ***	7.284 ***
	(0.909)	(1.170)	(0.812)	(1.030)
Reported height	0.821 ***	0.859 ***	0.944 ***	0.958 ***
	(0.006)	(0.007)	(0.005)	(0.006)
Male	2.029 ***	1.004 ***	0.786 ***	0.367 ***
	(0.105)	(0.127)	(0.089)	(0.123)
Age: 31–40	0.150 **	-0.256 **	-0.032	-0.152
	(0.072)	(0.110)	(0.098)	(0.157)
Age: 41–50	-0.078	-0.274 ***	-0.034	-0.290 *
	(0.078)	(0.104)	(0.097)	(0.154)
Age: 51–60	-0.518 ***	-0.796 ***	-0.096	-0.172
	(0.084)	(0.115)	(0.096)	(0.151)
Education	0.085 ***	0.049 ***	0.013	0.031 *
	(0.011)	(0.012)	(0.009)	(0.016)
Citizen/Born in country	1.104 ***	0.548 ***	0.185 **	-0.338
	(0.090)	(0.180)	(0.086)	(0.207)
Observations	10,277	5657	7101	2047
Adjusted R^2	0.914	0.925	0.948	0.959

Notes: Heteroskedasticity-robust standard-errors in parentheses.
Signif. Codes: *** : 0.01, **: 0.05, * : 0.1.

In the GSOEP, the heterogeneity across interviewers is substantially larger. For 85 of 195 interviewers, the absolute error never exceeds 5 cm and for 135 interviewers the average absolute error is below 1 cm. In that case, interviewers with rather large errors are particularly worrisome. The first and second-ranked interviewers' effect is mainly driven by extreme reporting errors. For the first-ranked interviewer 2 of 18 observations have absolute errors of 22 and 19 cm, respectively. For the second-ranked interviewer a single respondent with an absolute error of 20 cm drives the effect. While such errors are of course problematic, such isolated incidents may as well be driven by simple data entry errors. For the third-ranked interviewer this is not the case. The average absolute error is above 4 cm and the absolute error exceeds 5 cm for more than 25%. In the full sample these values are 0.9 cm and 3.6%. These differences suggest that interviewer 331 systematically drives unlikely changes in reported height.

Similar to the GSOEP, a large share of the PASS interviewers have no absolute errors exceeding 5 cm (N = 43; 52.4%) and average absolute errors below 1 cm (N = 53; 64.6%). We are particularly interested in the verified deviant interviewers 154 and 109. Interviewer 154 is ranked second and the height changes exceed 5 cm in 35% of the cases, and the average absolute error is 3.4 cm. These values clearly deviate from the sample averages. Thus, deviant interviewer 154 is characterized by frequent height changes which is in line with the hypothesis that (partial) falsifiers are unable to recover the height reported in previous waves. To the contrary, interviewer 109 is ranked 72nd with an average absolute error below 0.6 cm. Given that this interviewer presumably reinforced deviant habits over time, this indicates that the height question was not subject to deviant behavior in waves 9 and 12. For the other high-ranked interviewers, both the shares of large errors and the average errors are lower than for interviewer 154 and thus a lesser reason for concern.

In summary, the analysis of the NHANES and UKHLS data shows that there are several interviewers who are particularly prone to errors in height reporting. Furthermore, the analysis of the GSOEP and PASS data demonstrated that single interviewers also exacerbate changes in height reporting that should not occur. For the PASS data, we even find that deviant interviewer behavior can lead to frequent changes in height and that deviant interviewers can even be detected by such behavior. With regard to the analyses of the other samples, this implies that inadequate interviewer behavior might explain misreporting or frequent changes in reported height.

5.3. *Effects of height errors on substantive regression estimates*

In this section, we evaluate whether interviewers differ in their contribution to bias in regression coefficients using corrective equations. Table 4 reports the results for the corrective equations for each sample. The estimated coefficients show that the height coefficients are attenuated across all samples. This attenuation is larger for the NHANES and UKHLS. Several of the control variables associated with reporting errors are statistically significant (see Section 5.1). The results of the analysis procedure described in Section 4.2 are depicted in Fig. 2. The x-axis denotes the rank of the interviewers by their scale effect taken from the results of the multilevel location-scale models (in ascending order). The y-axis shows the average change per respondent in the coefficient of the reported height when the respective interviewer's values are replaced by the measured or previously reported height, i.e., what happens on average when the reported value of one observation of the respective interviewer is replaced with the measured value. For example, for the NHANES this corresponds to the change in the coefficient when one out of 10,277 values is replaced. Positive values depict a reduction in attenuation. A local linear regression line was added to the graphs to show whether the relative coefficient change varies across the interviewer ranks.

In the NHANES, the replacement of reported with measured values reduces attenuation for every interviewer. However, the reduction is not constant across all interviewers. For the interviewers with ranks above 40, the average changes in the coefficient are steadily increasing, which is also denoted by the positive slope of the local linear regression line. At the same time, the variation of average changes is increasing, indicating that the higher-ranked interviewers are heterogeneous in their error pattern. For the UKHLS, several interviewers have slightly negative values, but for the vast majority of interviewers the changes are positive. Similar to the NHANES, the average coefficient change is homogeneously distributed for most of the interviewers and substantially increases for the highest-ranked interviewers. The steep slope of the regression line shows that the highest-ranked interviewers have larger effects on the coefficient than the remaining interviewers. This provides evidence that some interviewers enhance bias in regression coefficients induced by measurement error. Next, we turn to the GSOEP and PASS, where reported height values are replaced by previously reported

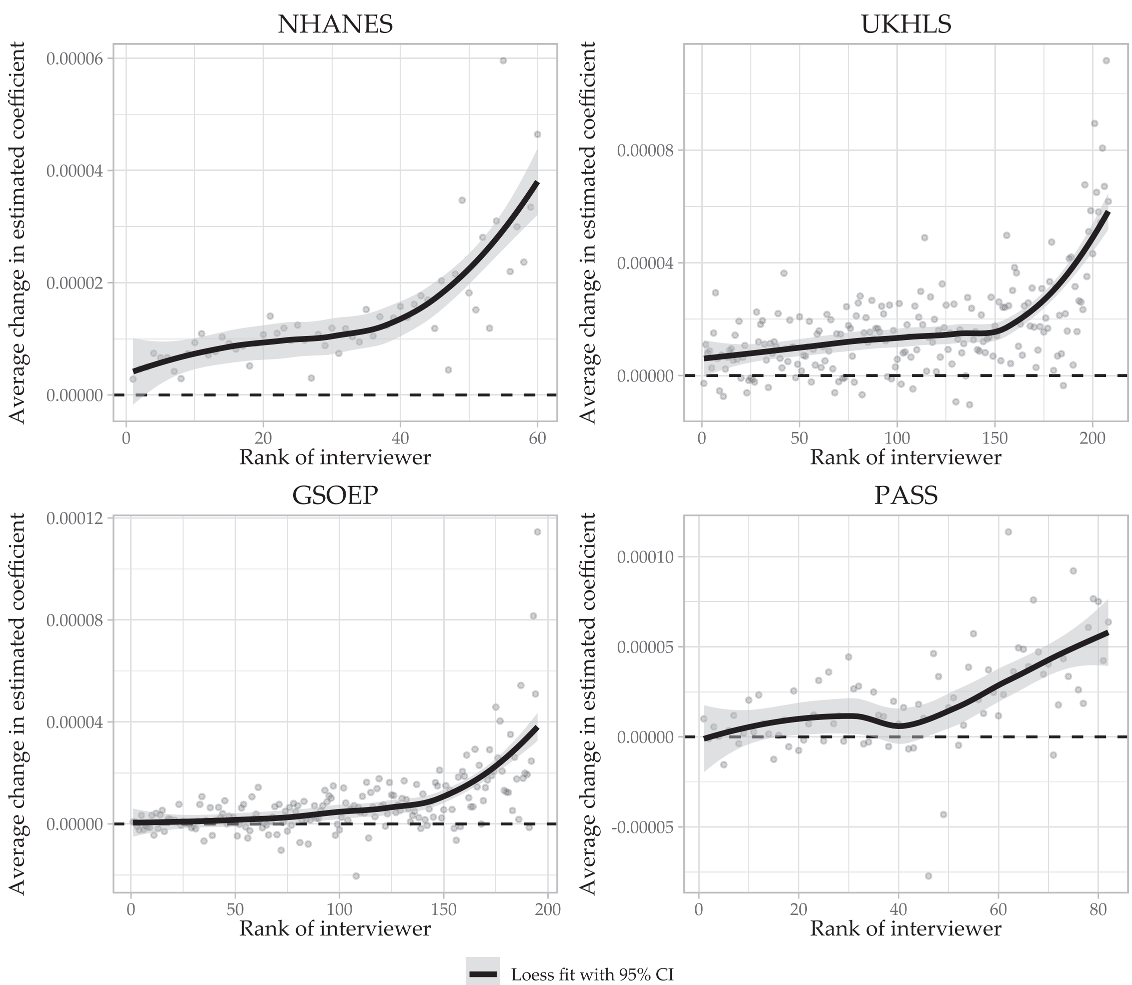

Fig. 2. Interviewer effects on regression estimates.

values. In the GSOEP, we find a similar pattern as for the NHANES and UKHLS. For the first 150 interviewers, the changes are relatively small and distributed around zero. For the higher-ranked interviewers, the changes and their variation increase. For the PASS, we also observe an increase in the average change and the heterogeneity of changes. Hence, the coefficient attenuation driven by changes in height reporting is not equally distributed across interviewers.

Our results imply that particularly error-prone interviewers can have – compared to the remaining interviewers – larger effects on regression coefficients in all samples. By depicting the relevance of the interviewer ranking derived from the multilevel location-scale models, this analysis also shows that the multilevel results can be used to target suspicious interviewers for further controls to limit their effects on research results.

6. Discussion

Anthropometric data are of crucial importance in the health and social sciences. Given the lack of administrative data for the general population, anthropometric data are often collected in surveys by directly asking respondents to report their values. While respondent misreporting of such data is well-documented in the literature, the role of the interviewers who collect these measures has not been extensively examined. Both sources of error could lead to false conclusions in health and social research, for instance, with regard to the impact of anthropometric measures on individual life-course outcomes. Using data from four large-scale surveys (NHANES, UKHLS, GSOEP, PASS) and multilevel models, we examined two types of errors: differences between reported and measured height and unlikely changes in adult self-reported height over time. After providing multiple theoretical explanations for the causes of interviewer effects on erroneous height reporting, we first assessed empirically whether these errors are associated with interviewers. Using multilevel model results, we also examined whether height reporting errors allow for identifying particularly error-prone interviewers. Lastly, we evaluated whether the effects of height reporting errors on substantive research results are equally distributed across interviewers. This is a novel contribution and important in understanding the role of interviewer-related measurement error on self-reported anthropometric data and the consequent effect of bias in substantive data analyses.

Our results revealed that interviewers affect both differences between reported and measured height and unlikely changes in adult self-reported height over panel survey waves. Our results further indicated that these errors accumulate for several interviewers. As one of the datasets contained falsified interviews, we were able to demonstrate that height changes can accumulate for deviant interviewers. Moreover, we found that the effects of height reporting errors on regression coefficients are not equally distributed across interviewers. A simulation analysis of the effects of error-prone interviewers showed that in many cases error-prone interviewers will have little consequence on regression coefficients as their workloads and error sizes are small. However, if the error-prone interviewers' share of interviews increases or the size of their errors increases, substantive research results will be increasingly biased. In practice, interviewer effects on additional analysis variables might further enhance bias in estimated parameters. This is particularly relevant for studies using the body mass index, where both components – height and weight – can be subject to interviewer effects.

All surveys in our analysis had interviewer monitoring procedures in place. In surveys with no (or poor) monitoring procedures, the effects and consequences of interviewers documented in this article would likely be reinforced. This calls for putting limits on interviewer workloads and thorough interviewer training and monitoring to minimize the prevalence and impact of error-prone interviewing. Note that monitoring systems can be quite inexpensive and easily automated (for instance by comparing previously reported height to the reported height in the current wave), and thus the benefits of monitoring interviewers will outweigh costs in most cases. Furthermore, data producers should provide interviewer IDs to researchers to enable testing and accounting for interviewer effects in their analyses.

Our analysis is subject to several limitations. While our approach flagged a verified deviant interviewer, we were unable to identify the exact reasons why frequent height errors occurred for other error-prone interviewers. Practitioners may shed more light on this issue by applying the analysis during the fieldwork period and following up with flagged interviewers. Second, we were unable to fully disentangle respondent and interviewer effects. It is possible that some interviewers were predominantly assigned to presumably difficult respondents characterized by erroneous height reporting. However, interviewers usually work in geographic clusters and are not assigned to respondents solely based on their skills and the target respondents' interview difficulty. Thus, it seems unlikely that the patterns observed for error-prone interviewers were fully driven by respondents. Lastly, homogeneity of respondents within and heterogeneity across geographic regions regarding characteristics shaping reporting errors might affect our results. However, in our case, respondents from specific regions where error-prone interviewers work would have to be particularly poor at reporting their height. Given that we control for several socio-demographics and that previous research found minor effects of regions on data quality measures or reporting error (e.g., Meyer and Mittag, 2019; Sturgis et al., 2021), regional effects are unlikely to explain the prevalence of error-prone interviewers.

Our analysis exhibited a straight-forward approach to exploit height reporting to conduct data quality control procedures. The multilevel location-scale model may also be applied to further survey variables which can be compared to validation data such as register data or recontact interviews. Future research on interviewers' role in panel studies may delve into whether interviewer effects drive variation in less-stable items and whether changes in interviewer assignment increase response variation over waves. Developing further tools to ensure high data quality of panel surveys is especially relevant as such databases have consistently large numbers of users in a variety of research fields.

CRediT authorship contribution statement

Lukas Olbrich has completed the data analyses, drafted the paper, and revised the paper based upon Yuliya Kosyakova's and Joseph W. Sakshaug's comments. Yuliya Kosyakova has supervised the process of data analyses, reviewed, edited, and commented on various versions of the paper. Joseph W. Sakshaug has supervised the process of data analyses, reviewed, edited, and commented on various versions of the paper.

Acknowledgements

Lukas Olbrich gratefully acknowledges financial support from the Graduate Programme of the Institute for Employment Research (IAB) and the University of Erlangen-Nuremberg (GradAB). Joseph W. Sakshaug gratefully acknowledges financial support from the Ludwig-Maximilian University of Munich. We are grateful to two anonymous reviewers and the editor for very helpful comments.

Appendix

See Fig. A1 and Fig. B1, Fig. B2, Fig. B3, Fig. B4.

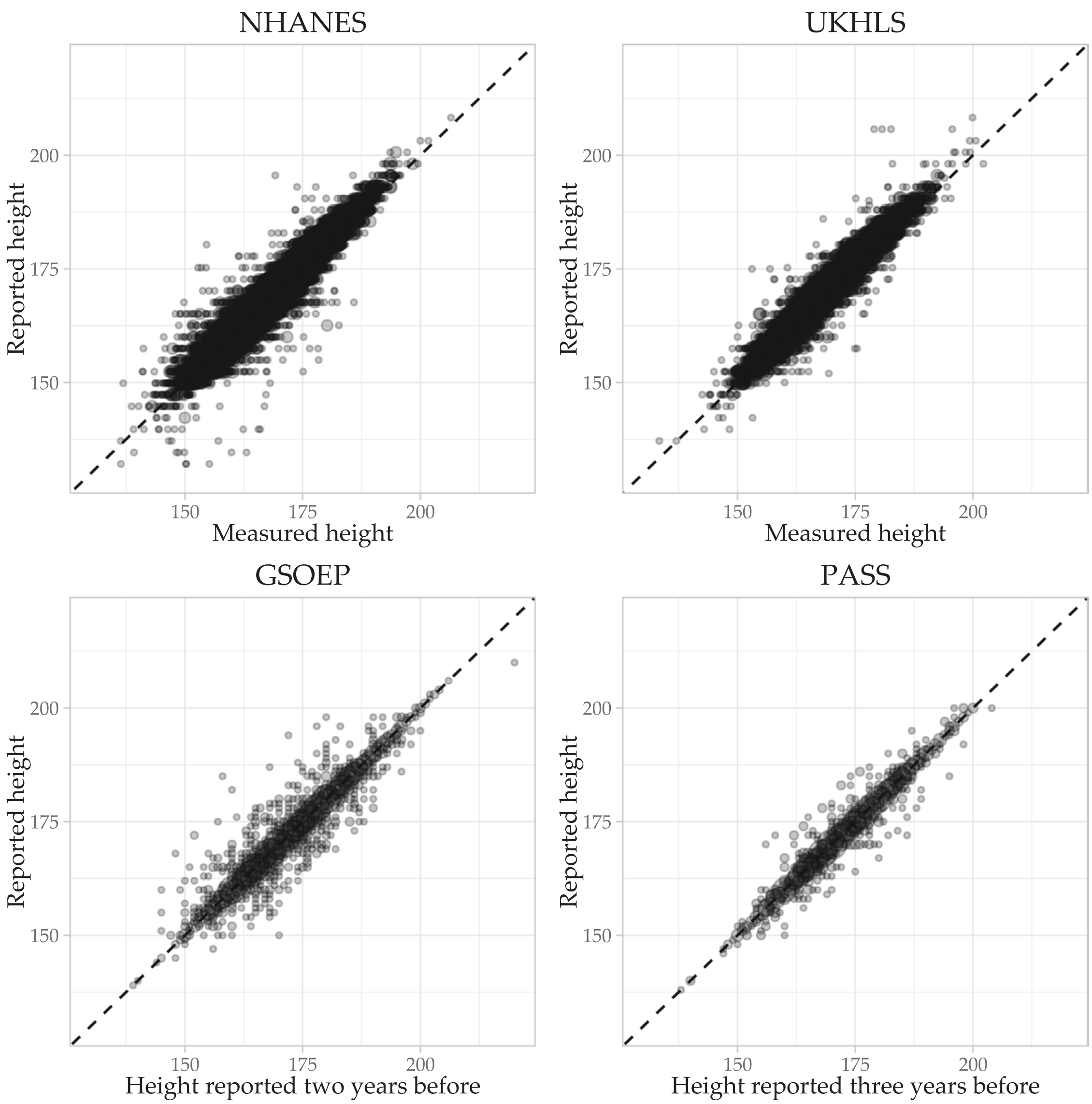

Fig. A1. Reported height versus measured/previously reported height (in cm).

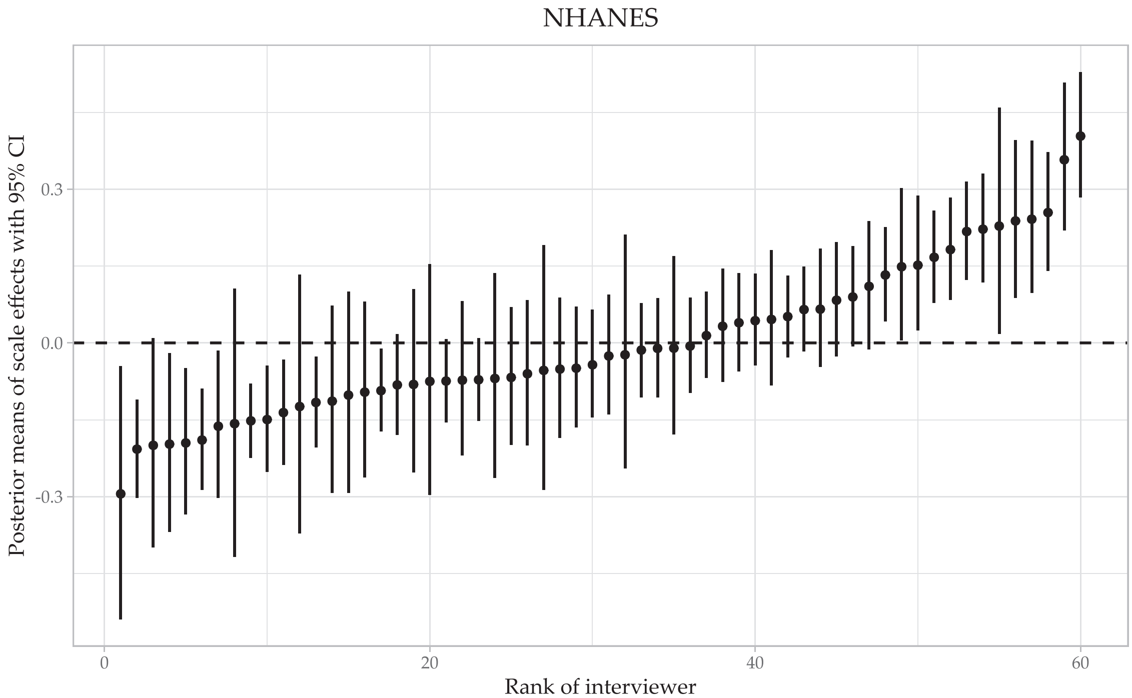

Fig. B1. Interviewer effects on scale, NHANES.

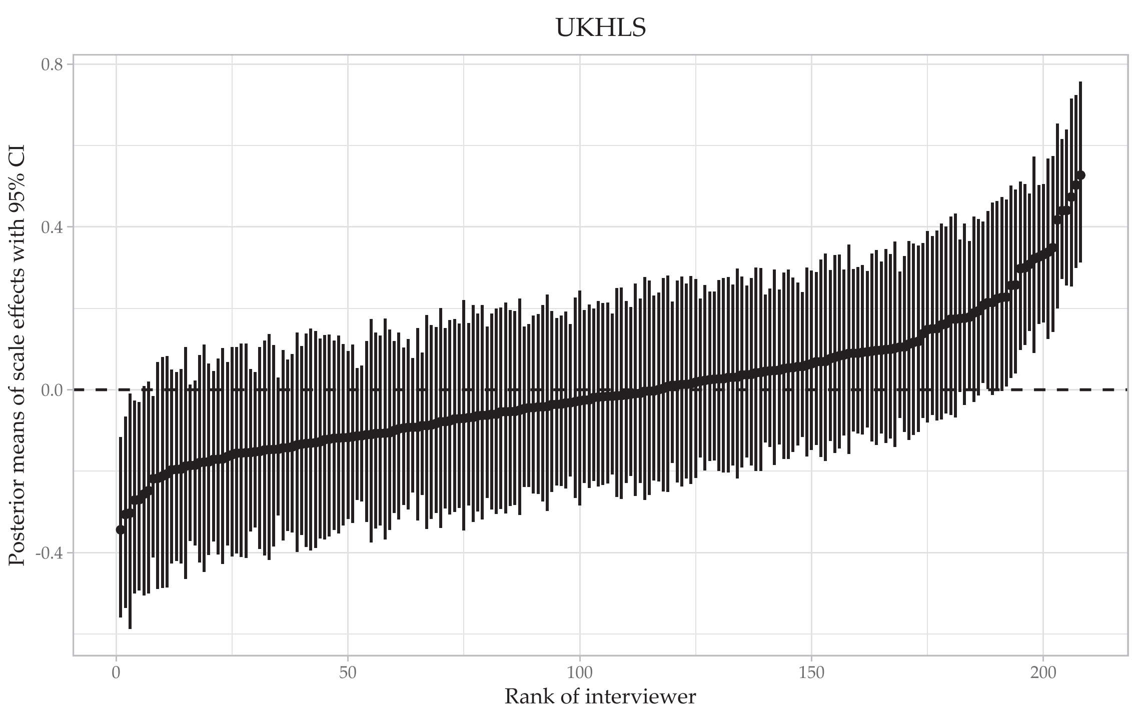

Fig. B2. Interviewer effects on scale, UKHLS.

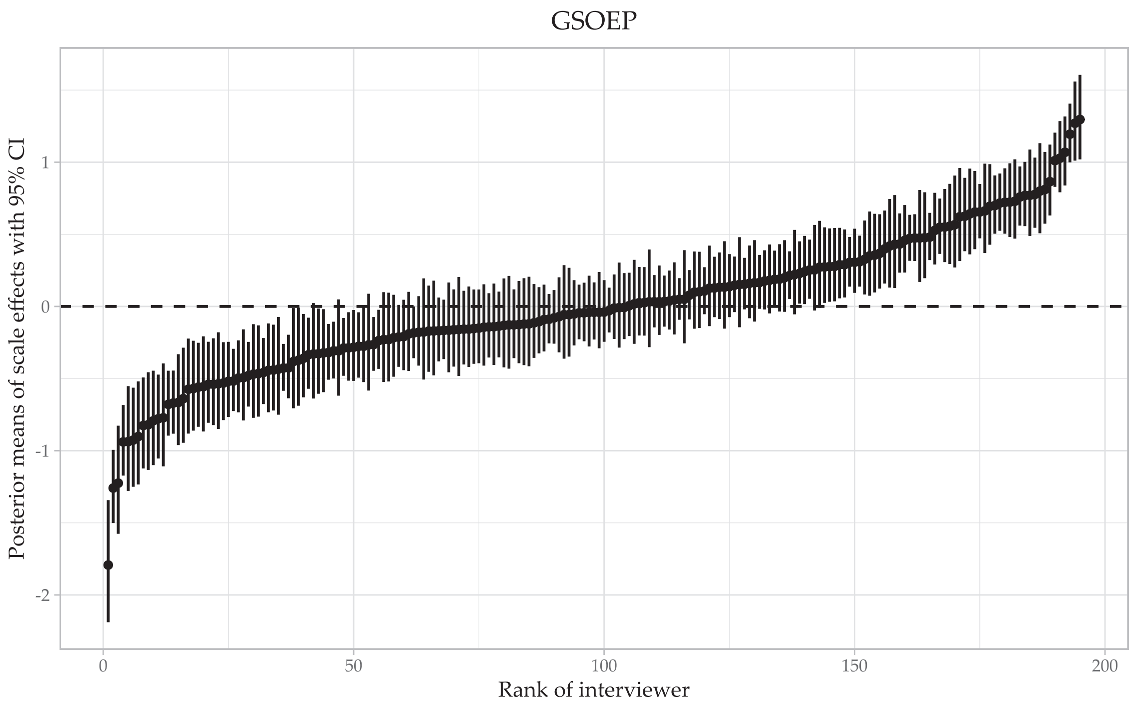

Fig. B3. Interviewer effects on scale, GSOEP.

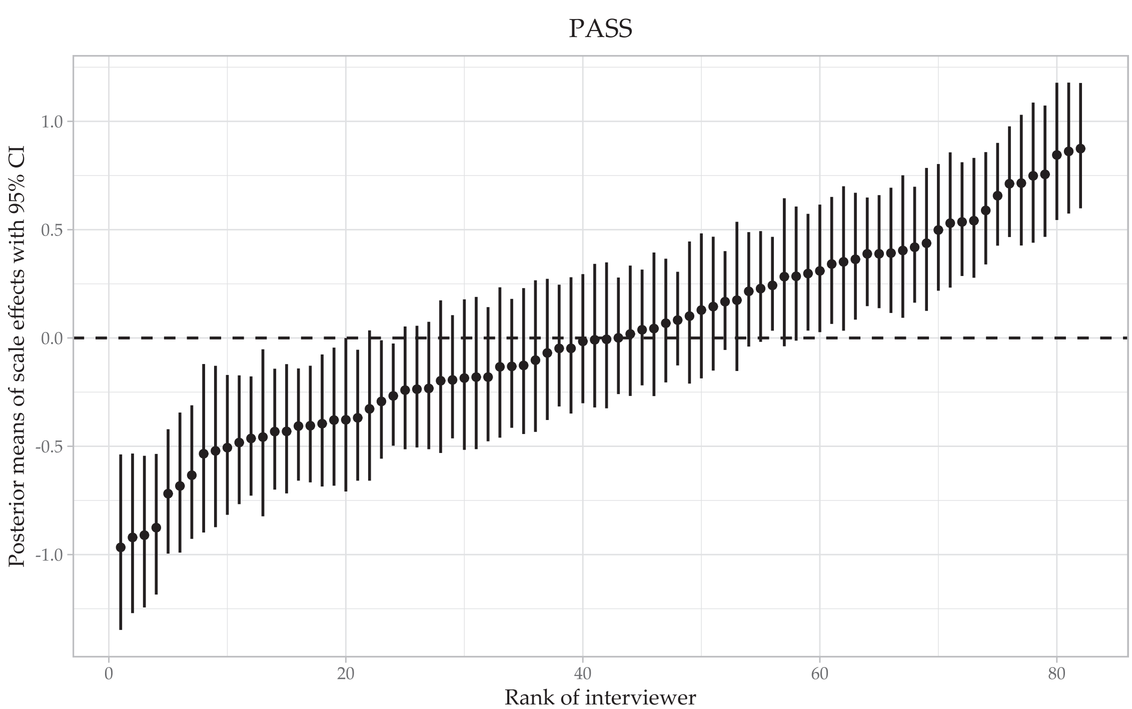

Fig. B4. Interviewer effects on scale, PASS.

References

AAPOR, 2003. Interviewer falsification in survey research: Current best methods for prevention, detection, and repair of its effects. ⟨https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/falsification.pdf⟩.

Batty, G.D., Shipley, M.J., Gunnell, D., Huxley, R., Kivimaki, M., Woodward, M., Lee, C.M.Y., Smith, G.D., 2009. Height, wealth, and health: An overview with new data from three longitudinal studies. Econ. Hum. Biol. 7 (2), 137–152.

Beste, J., Olbrich, L., & Schwanhäuser, S. (2021). Interviewer:innenkontrolle im Panel Arbeitsmarkt und soziale Sicherung (PASS). FDZ-Methodenbericht 4.

Boreham, R., Boldysevaite, D., Killpack, C., 2012. UKHLS: Wave 1 Technical Report.

Boström, G., Diderichsen, F., 1997. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. Int. J. Epidemiol. 26 (4), 860–866.

Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, fifth ed. Elsevier Science B.V, pp. 3705–3843.

Brunton-Smith, I., Sturgis, P., Leckie, G., 2017. Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location–scale model. J. R. Stat. Soc. Ser. A Stat. Soc. 180 (2), 551–568.

Buck, N., McFall, S., 2012. Understanding Society: design overview. Longitud. Life Course Stud. 3 (1), 5–17.

Burke, M.A., Carman, K.G., 2017. You can be too thin (but not too tall): Social desirability bias in self-reports of weight and height. Econ. Hum. Biol. 27, 198–222.

Bürkner, P.C., 2018. Advanced Bayesian multilevel modeling with the R package brms. R J. 10 (1), 395–411.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: A probabilistic programming language. J. Stat. Softw. 76 (1), 1–32.

Case, A., Paxson, C., 2008. Stature and status: height, ability, and labor market outcomes. J. Political Econ. 116 (3), 499–532.

Cawley, J., 2004. The impact of obesity on wages. J. Hum. Resour. 39 (2), 451–474.

Council of Europe, 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge University Press.

Crossley, T.F., Schmidt, T., Tzamourani, P., Winter, J.K., 2021. Interviewer effects and the measurement of financial literacy. J. R. Stat. Soc. Ser. A: Stat. Soc. 184 (1), 150–178.

Davillas, A., Jones, A.M., 2021. The implications of self-reported body weight and height for measurement error in BMI. Econ. Lett. 209, 110101.

De Haas, S., Winker, P., 2016. Detecting fraudulent interviewers by improved clustering methods – the case of falsifications of answers to parts of a questionnaire. J. Off. Stat. 32 (3), 643–660.

Deaton, A., Arora, R., 2009. Life at the top: the benefits of height. Econ. Hum. Biol. 7 (2), 133–136.

Eckman, S., Koch, A., 2019. Interviewer involvement in sample selection shapes the relationship between response rates and data quality. Public Opin. Q. 83 (2), 313–337.

Fernihough, A., McGovern, M.E., 2015. Physical stature decline and the health status of the elderly population in England. Econ. Hum. Biol. 16, 30–44.

Finn, A., Ranchhod, V., 2017. Genuine fakes: The prevalence and implications of data fabrication in a large South African survey. World Bank Econ. Rev. 31 (1), 129–157.

Fuller, W.A., 1987. Measurement Error Models. John Wiley & Sons, Inc.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, third ed. CRC Press.

Gil, J., Mora, T., 2011. The determinants of misreporting weight and height: the role of social norms. Econ. Hum. Biol. 9 (1), 78–91.

Goebel, J., Grabka, M.M., Liebig, S., Kroh, M., 2019. The german socio-economic panel (SOEP). Jahrb. Natl. Stat. 239 (2), 345–360.

Gorber, S.C., 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. Obes. Rev. 8 (4), 307–326.

Groves, R.M., 2004. Survey Errors and Survey Costs, 2nd ed..,. John Wiley & Sons, Inc,.

Hedeker, D., Mermelstein, R.J., Demirtas, H., 2008. An application of a mixed-effects location scale model for analysis of Ecological Momentary Assessment (EMA) data. Biometrics 64 (2), 627–634.

Heineck, G., 2006. Height and weight in Germany, evidence from the German Socio-Economic Panel, 2002. Econ. Hum. Biol. 4 (3), 359–382.

Josten, M., Trappmann, M., 2016. Interviewer effects on a network-size filter question. J. Off. Stat. 32 (2), 349–373.

Kosyakova, Y., Skopek, J., Eckman, S., 2015. Do interviewers manipulate responses to filter questions? Evidence from a multilevel approach. Int. J. Public Opin. Res. 27 (3), 417–431.

Kroh, M., 2005. Intervieweffekte bei der Erhebung des Körpergewichts in Bevölkerungsumfragen. Gesundheitswesen 67 (8–9), 646–655.

Larson, M.R., 2000. Social desirability and self-reported weight and height. Int. J. Obes. 24 (5), 663–665.

Lee, L.F., Sepanski, J.H., 1995. Estimation of linear and nonlinear errors-in-variables models using validation data. J. Am. Stat. Assoc. 90 (429), 130–140.

Meyer, B.D., Mittag, N., 2019. Misreporting of government transfers: how important are survey design and geography? South. Econ. J. 86 (1), 230–253.

Moore, C.J., 1988. Self/proxy response status and survey response quality. J. Off. Stat. 4 (2), 155–172.

National Center for Health Statistics, 1994. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. Vital.-. Health Stat. 1, 32.

O'Neill, D., Sweetman, O., 2013. The consequences of measurement error when estimating the impact of obesity on income. IZA J. Labor Econ. 2 (1), 1–20.

Persico, N., Postlewaite, A., Silverman, D., 2004. The effect of adolescent experience on labor market outcomes: the case of height. J. Political Econ. 112 (5), 1019–1053.

Pinkston, J.C., 2017. The dynamic effects of obesity on the wages of young workers. Econ. Hum. Biol. 27, 154–166.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. ⟨https://www.r-project.org/⟩.

Schnell, R., Kreuter, F., 2005. Separating interviewer and sampling-point effects. J. Off. Stat. 21 (3), 389–410.

Schräpler, J.-P., Wagner, G.G., 2005. Identification, characteristics and impact of faked interviews in surveys: an analysis by means of genuine fakes in the raw data of SOEP. Allg. Stat. Arch. 89 (1), 7–20.

Schwanhäuser, S., Sakshaug, J.W., Kosyakova, Y., Kreuter, F., 2020. Statistical identification of fraudulent interviews in surveys - improving interviewer controls. In: Olson, K., Smyth, J.D., Dykema, J., Holbrook, A., Kreuter, F., West, B.T. (Eds.), Interviewer Effects from a Total Survey Error Perspective. CRC Press, pp. 91–106.

SOEP Group, 2020. SOEP-Core v35 – HEALTH. SOEP Survey Papers 830: Series D – Variable Descriptions and Coding.

Sturgis, P., Maslovskaya, O., Durrant, G., Brunton-Smith, I., 2021. The interviewer contribution to variability in response times in face-to-face interview surveys. J. Surv. Stat. Methodol. 9 (4), 701–721.

Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. Psychol. Bull. 133 (5), 859–883.

Trappmann, M., Bähr, S., Beste, J., Eberl, A., Frodermann, C., Gundert, S., Schwarz, S., Teichler, N., Unger, S., Wenzig, C., 2019. Data resource profile: panel study labour market and social security (PASS). Int. J. Epidemiol. 48 (5), 1411–1411 G.

Trappmann, M., Beste, J., Bethmann, A., Müller, G., 2013. The PASS panel survey after six waves. J. Labour Mark. Res. 46 (4), 275–281.

Uhrig, S.C.N., 2012. Understanding panel conditioning: an examination of social desirability bias in self-reported height and weight in panel surveys using experimental data. Longitud. Life Course Stud. 3 (1), 120–136.

University of Essex: Institute for Social and Economic Research, 2021. Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009, fourteenth ed. UK Data Service,. https://doi.org/10.5255/UKDA-SN-6614-15.

University of Essex: Institute for Social and Economic Research and National Centre for Social Research, 2014. Understanding Society: Waves 2 and 3 Nurse Health Assessment, 2010- 2012 [data collection], third e. UK Data Service,. https://doi.org/10.5255/UKDA-SN-7251-3.

Warren, J.R., Halpern-Manners, A., 2012. Panel conditioning in longitudinal social science surveys. Sociol. Methods Res. 41 (4), 491–534.

West, B.T., Blom, A.G., 2017. Explaining interviewer effects: a research synthesis. J. Surv. Stat. Methodol. 5 (2), 175–211.

West, B.T., Kreuter, F., Jaenichen, U., 2013. "Interviewer" effects in face-to-face surveys: a function of sampling, measurement error, or nonresponse? J. Off. Stat. 29 (2), 277–297.