Exploring effects of virtual interviewers on self-reported personal attributes

Jan Karem Höhne
German Centre for Higher Education Research and Science Studies (DZHW)
Leibniz University Hannover

Frederick G. Conrad
University of Michigan

Joshua Claassen
German Centre for Higher Education Research and Science Studies (DZHW)
Leibniz University Hannover

Abstract

Inexpensive and time-efficient web surveys have increasingly replaced other survey modes, especially in-person interviews. Even well-known social surveys follow this trend. However, web surveys suffer from low response rates and frequently struggle to assure that the data are of high quality. New advances in communication and video technology make it possible to introduce new approaches to web survey data collection. Building on these advances, we investigate life-like virtual interviewers (VIs) asking respondents questions. More specifically, we investigate whether and to what extent the gender of the VI affects respondents' answers to questions about their self-reported body height and weight. For this purpose, we recruited respondents from social media into a web survey (N = 1,512) and randomly allocated them to one of three experimental conditions: 1) assigned male VI, 2) assigned female VI, or 3) selected VI (i.e., respondents decided between the male and female VI). The results reveal that item nonresponse is low (about 5%) across all three experimental conditions. In line with previous research, female respondents favor female VIs, whereas male respondents do not show a clear preference. However, in contrast to previous research, we found that respondents tend to underreport weight if their gender matches that of the VI.

Keywords: Response behavior, Sensitive questions, Social media recruitment, Virtual interviewers (VIs), Web survey

Introduction and research questions

Survey data collection has fundamentally changed. Cost-efficient and streamlined web surveys continue to replace other survey modes, especially in-person interviews (Callegaro et al., 2015, p. 240). Even large-scale social surveys that typically relied on in-person interviews, such as the European Social Survey (ESS), now routinely collect data via web surveys (Fitzgerald & Aizpurua, 2024). However, web surveys are not a panacea. First, they generally produce lower response rates than other established survey modes, such as in-person and telephone interviews (Daikeler et al, 2020; Lozar Manfreda et al., 2008). Second, they are commonly text-based and thus they impede participation by people who are not skilled readers of the survey language

This document is a preprint and thus it may differ from the final version

(Höhne, 2023). Third, they may not produce data whose quality is as high as the quality of data collected in other modes (Callegaro et al., 2015). Cibelli Hibben et al. (2022) argue that the absence of interviewers complicates the creation of engagement and commitment. However, the authors found evidence that simply asking respondents to commit improves data quality in a text-based web survey.

To explore the consequences of re-introducing some features of interviewer administration, various studies have implemented pre-recorded videos of human interviewers reading questions to respondents in web surveys. Most frequently, these studies asked questions on sensitive topics, such as sexual behavior. Compared to text-based web surveys, some of these studies found no disclosure differences (Fuchs & Funke, 2009), but some other studies have (Fuchs, 2009). For example, Fuchs (2009) found that male and female respondents report sexually transmitted diseases less often when asked by an opposite-gender interviewer (treating gender as binary). He also found that male respondents less frequently report that they never had sex when asked by a male (same-gender) than a female (different-gender) interviewer. Conrad et al. (2023) generally reported more disclosure of sensitive information than in a live (two-way) video interview, indicating less socially desirable answering. In addition, about 70% of the respondents stated that they felt connected to the pre-recorded interviewer. Haan et al. (2017) also found that pre-recorded video interviewing produces less socially desirable responding than live interviewing (in-person and telephone). Recency effects (i.e., tendency to select answer options at the end of the scale) were less common in pre-recorded video interviewing than in telephone interviews. This finding corresponds to findings reported by Conrad et al. (2023). The authors observed significantly less instances of rounding (fewer numerical answers that are divisible by five or ten, considered a type of survey satisficing), compared to a live video interview. Although not statistically significant, there were also less instances of rounding compared to a text-based web survey.

Recent technological advances allow researchers to quickly and easily create high-quality video content to generate life-like virtual interviewers (VIs). VIs can vary with respect to a series of human (e.g., physical appearance including gender) and speech (e.g., timbre and accent) attributes that cannot be easily varied with human interviewers. VIs potentially create a kind of social presence for respondents in web surveys, introducing some benefits of interviewer-administration to web surveys that are typically self-administered (e.g., stimulating thoughtful answering). VIs may be experienced by respondents as less intrusive than human interviewers when they ask sensitive questions and may create a higher level of comfort and privacy for respondents, reducing the prevalence of socially desirable answers. As illustrated by Conrad et al. (2020), allowing respondents to choose a VI from a gallery of thumbnail interviewer videos to conduct a hypothetical future survey seemed to result in increased matching of respondent and interviewer race, presumably promoting rapport and disclosure of sensitive information.

Lind et al. (2013) investigated the disclosure of sensitive information in a lab experiment comparing data collected via audio computer-assisted self-interviewing (ACASI), in-person interviews, and interviews conducted by VIs (with low and high facial animation). The authors reported that disclosure for most questions was higher under ACASI than in any of the other modes, indicating that the presence of a face – even one that is not visually realistic – can deter disclosure.

Conrad et al. (2015) investigated the effects of a VI's (simulated) dialog capability (low and high) and facial animation (low and high) on respondents' answer behavior in a lab experiment. The authors reported higher response accuracy for the high-dialog-capability interviewer, including more clarification requests for ambiguous scenarios. They found more displays of engagement for interviewers with higher animation.

Finally, Conrad et al. (2020) investigated VIs varying in gender (male and female) and race (white and African American). In their web survey, the authors observed a reduction of socially desirable response behavior when the demographic attributes of respondents and VIs matched. More specifically, when respondents matched the VI on race, they were more likely to report being "slightly overweight," a socially undesirable and presumably more truthful answer, than when they did not match on this attribute. Disclosure was also increased when respondents matched the VI on gender, but this increase did not reach statistical significance. Both male and female respondents favored female interviewers for a hypothetical future survey. As shown by Catania et al. (1996) for telephone interviews, allowing respondents to select a human interviewer based on gender can increase disclosure of sensitive information. In that study, female respondents showed a clear preference for female interviewers, whereas male respondents showed no clear preference for male or female interviewers. One explanation is that allowing respondents to choose the interviewer increases their sense of control over the interview situation, thereby creating a kind of comfort and sense of rapport, reducing the threat of disclosing sensitive information. Allowing respondents to select a human interviewer is logistically and ethically complicated but allowing respondents to select a VI is much more straightforward (Conrad et al., 2020). The attributes of VIs (e.g., appearance and speech) can be easily varied; in principle, it is even possible to allow respondents to create their own VIs by, for example, selecting attributes from a list.

Building on the study by Conrad et al. (2020), we investigate socially desirable responding when life-like VIs ask potentially sensitive questions. The VIs' realism encompasses their general visual appearance, their facial expressions, and their speech. However, in this study, VIs do not have any dialog capabilities (i.e., they do not "understand" respondents' utterances and are not responsive) but simply read questions aloud to respondents (similar to pre-recorded human interviewers). More specifically, VIs ask respondents two open numeric questions about their body height and weight. Collecting information about respondents' height and weight makes it possible to determine their Body Mass Index (BMI). The BMI is an established and widely used measure for evaluating a person's physical condition (e.g., whether he or she is overweight; Elevelt et al., 2021; Prince et al., 2008). Thus, asking about height and weight can be seen as sensitive potentially promoting socially desirable answers. For example, Olbrich et al. (2022) showed that male respondents are more prone to overreport their height. In addition, Lipps and Lutz (2017) showed that males and females (but especially females) self-report lower weight when interviewed by an opposite-gender human interviewer. In this study, we investigate whether and to what extent the disclosure of sensitive information in the form of height and weight is affected by the gender of the VI. More specifically, we attempt to answer the following two research questions:

1) Are respondents more likely to disclose sensitive information (i.e., lower height and higher weight) when they match the gender of the VI?

2) Do respondents disclose more sensitive information (i.e., lower height and higher weight) when they select the VI themselves?

To investigate these two research questions and to better understand when respondents are likely to provide more and less socially desirable answers to questions administered by VIs, we conducted a web survey (N = 1,512) in which respondents were recruited through ads posted on the social media platform Facebook (Meta) and were asked questions by a VI. We employed a between-subject design: respondents were randomly allocated to 1) a male VI, 2) a female VI, or 3) asked to choose their VI.

Method

Data

We drew a convenience sample by recruiting respondents through Facebook ads displayed in the newsfeed. Data was collected in Germany from 5th February to 18th March 2024. We posted six Facebook ads that were tailored to the respective combination of age and gender (e.g., "middle-aged/male" or "young/female"). The quota design was based on the German Microcensus (a small population census in the form of an annual household survey of official statistics in Germany; DESTATIS, 2024). When running Facebook ads for recruiting web survey respondents, researchers can set a budget for the ad campaign and Facebook automatically removes ads once the limit of the budget is reached. In this study, we set a maximum budget of about 1,700€. Appendix A includes a screenshot of the Facebook ad and Appendix B reports basic benchmarks regarding the costs of the Facebook ad campaign.

The Facebook ads included information on the survey topic (i.e., same-gender partnerships), expected survey time (approx. 5 minutes)[2], incentives (i.e., raffle of 5€), and a web survey link. On the first web survey page, we informed respondents about the study procedure, the likelihood (10% of the respondents received 5€) of receiving an incentive payment (made through PayPal)[3], and that the study adhered to existing data protection laws and regulations. Data collection and study design were evaluated and funded through the research fund committee of the German Society for Online Research (DGOF). In addition, the study was approved by the ethics committee of the German Centre for Higher Education Research and Science Studies (DZHW).

We created videos of VIs reading questions to be played by respondents during web survey completion, using the online video generation platform HeyGen (https://www.heygen.com/) allowing us to vary VI attributes, such as gender and clothing. The resulting videos were embedded in the web survey as required by the experimental design. For web survey administration, we used the Unipark software (https://www.unipark.com/).

Sample

Overall, about 95,000 Facebook users were reached by the ads of the web survey, 3,960 clicked on the link and visited the first page of the web survey, and 1,512 completed the web survey

[2] The actual median completion time was about 10 minutes.
[3] Lotteries are a common incentive scheme, especially in social media recruited web surveys (Thornton et al., 2016).

(we only analyze complete cases). These respondents were between 20 and 95 years old[4] with a mean age of 50.6 years. In total, 46.1% of them were female, 50.9% male, and 3.0% gender diverse (non-binary; neither female nor male). In terms of school education, 29.8% had completed lower or intermediate secondary school (low to intermediate education) and 70.2% had completed college preparatory secondary school or more (higher education). About 46.4% of respondents participated with a computer, 4.9% with a tablet, and 48.7% with a smartphone.

Experiment

We randomized respondents to one out of three experimental conditions. The first condition (n = 381) was asked two questions by a male VI (assigned male VI). The second condition (n = 370) was asked the same two questions by a female VI (assigned female VI). The third condition (n = 761) was asked the same two questions, but respondents were instructed to select either the male VI or the female VI (selected VI). Importantly, we assigned twice as many respondents to the third condition to assure that there were about as many male and female VI web surveys in this condition as in the other two conditions. Figure 1 shows screenshots of the VIs and experimental conditions.

As shown in Table 1, the experimental conditions did not differ with respect to age, gender, education, and share of the three device types.

Table 1. Sample composition across the three experimental conditions

Variables	Assigned male VI condition	Assigned female VI condition	Selected VI condition	Test statistics
Age	51.7	49.1	50.7	$F(2,1480) = 2.84$, $p = 0.06$
Female	47.1	48.0	44.8	$\chi^2(4) = 1.21$, $p = 0.88$
Education				$\chi^2(2) = 3.17$, $p = 0.21$
Low to intermediate	28.9	26.7	31.8	
Higher	71.2	73.3	68.2	
Device types				$\chi^2(4) = 6.55$, $p = 0.16$
Computer	45.7	44.3	47.7	
Tablet	7.1	4.1	4.2	
Smartphone	47.2	51.6	48.1	

Note. We report means for age. For the remaining variables, we report percentages. Due to rounding, percentages may not add up to 100%.

Questions and VIs

We asked two open numeric questions on respondents' height in centimeters (cm) and weight in kilograms (kg). The two questions were adopted from the study by Elevelt et al. (2021). Each

[4] We excluded one respondent with an overly high age of 109 years for two reasons. On the one hand, this respondent represents an extreme outlier. On the other hand, we had concerns about the truthfulness of the self-reported age.

question was accompanied by an open input field (below the VI video) in which respondents could enter their answer. Respondents clicked a play button to start the interviewer video (the VI remained static until the respondent clicked the play button; see Figure 1). The two open numeric questions were formulated as follows:

1) How tall are you? Please enter your body height in centimeters in the open field.

2) How much do you weigh? Please enter your body weight in kilograms in the open field.

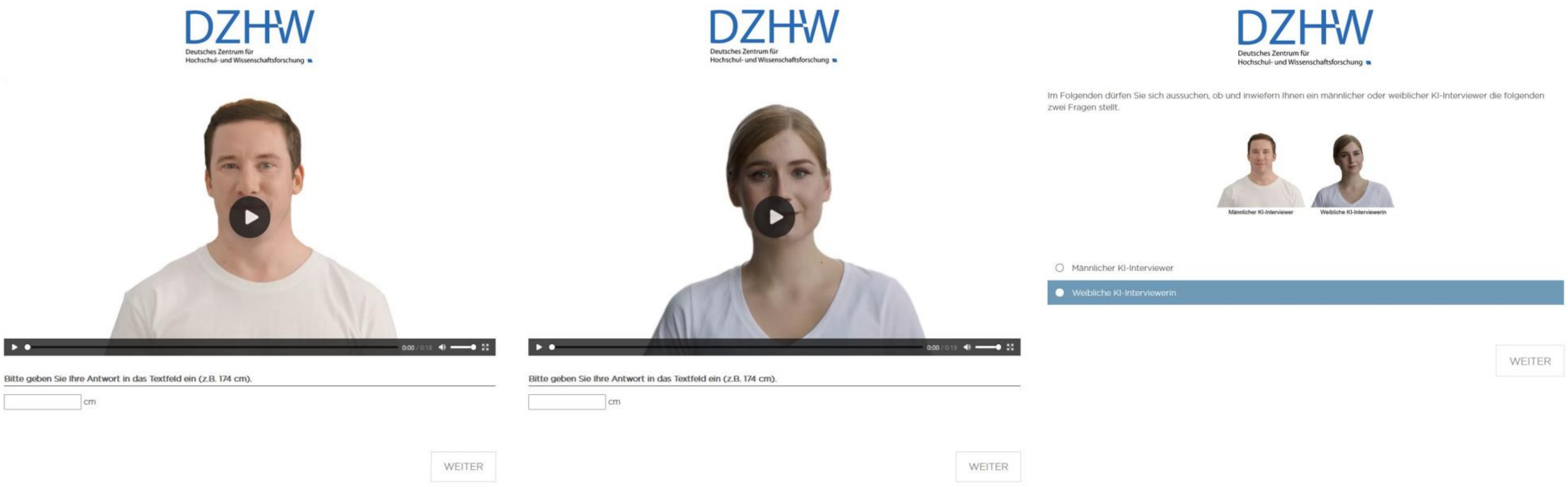
Before the respondents were exposed to the VIs, they received a short introduction telling them that a VI will ask them two questions on their height and weight (see Appendix C for an English translation). The VIs introduced themselves by saying their name (Alex in both cases) and thanking respondents for their participation. In the third experimental condition (selected VI), respondents were previously asked whether they would prefer to be surveyed by a male or female VI, showing pictures of them (see Appendix C for an English translation).

The two open numeric questions on respondents' height and weight were asked in the fourth quarter of the web survey. In total, 25 questions were administered before them. These questions were closed or open narrative questions with requests for text answers mostly dealing with same-gender partnerships. The data from most of those questions is not reported in the current article.

Analytical strategy

In a first step, we calculate summary statistics to shed light on respondents' answer behavior. More specifically, we report the shares of item nonresponse and mean values of respondents self-reported height and weight. Following Lenzner et al. (2024), we define item nonresponse as instances in which respondents provided no answer at all (complete item nonresponse; i.e., left the open input field blank) and instances in which respondents implicitly refused to answer by providing useless answers (soft item nonresponse; e.g., "123456," "nonsense," or "don't know"). Item nonresponse rates and mean values are reported for each open numeric question across the three experimental conditions: 1) assigned male VI, 2) assigned female VI, and 3) selected VI. We conduct chi-square tests for item nonresponse and one-way analyses of variance (ANOVAs) for mean values using the Bonferroni correction procedure.

We then report respondents' preferences of VIs varying in terms of gender (male and female). Importantly, we now only focus on respondents that were randomly assigned to the third experimental condition (selected VI) in which they could select a VI. More specifically, we investigate the percentage of male respondents selecting male and female VIs as well as the percentage of female respondents selecting male and female VIs. Finally, we conduct a chi-square test to evaluate whether male and female respondents selected male and female VIs in different proportions. Importantly, we excluded respondents selecting the answer option "diverse" when asked for their gender from the analyses because the proportion was very small (n = 44) so that separate analysis was not possible.



Assigned male VI condition **Assigned female VI condition** **Selected VI condition**

Figure 1. Screenshots of the VIs and experimental conditions
Note. Assigned male VI on the left (first condition), assigned female VI in the middle (second condition), and selected VI question on the right (third condition). Respondents in the third condition could select between the male VI from the first condition and the female VI from the second condition.

Turning to our main results and two research questions, we estimate linear regression models. Remember that height was measured in centimeters (cm) and weight was measured in kilograms (kg). When it comes to height, we follow the notion that social norms generally favor taller individuals. For example, height is positively associated with labor market success; this association applies to both males and females (Thompson et al., 2023). Accordingly, we expect respondents whose height is low to consider the question to be asking for sensitive information. We estimate three sequential linear regression models with height (in cm) as dependent variable. In the first model, we include interviewed by female VI (1 = "Yes") with male VI as reference and selected VI (1 = "Yes") with assigned VI as reference as independent variables. In the second model, we include respondents' gender in terms of female (1 = "Yes") with male as reference, and we also include age (in years), higher education (1 = "Yes") with low to intermediate education as reference, and smartphone participation (1 = "Yes") with computer and tablet as reference as independent variables. In the third model, we then additionally include matching gender between respondent and VI (1 = "Yes") as independent variable. We use mismatching gender between respondent and VI as reference. Finally, in the fourth model, we include an interaction term between interviewed by female VI and selected VI.

With respect to weight, we follow the notion that heavy weight is generally stigmatized. Heavy individuals frequently report that they are teased or treated unfairly because of their weight (Puhl et al., 2021). We therefore expect respondents to perceive high weight as sensitive (or undesirable) information. Again, we estimate four sequential linear regression models with weight (in kg) as the dependent variable. In doing so, we use the same model specifications as for height.[5]

Similar to sensitive information, item nonresponse could shed light on our two research questions: rather than shade the truth some respondents might skip the question so missing data may be a proxy for socially desirable responding. However, as shown in Table 2, the overall item nonresponse rate was relatively low, reducing the robustness of our regression models. More specifically, if the dichotomous dependent variable in logistic regressions is heavily unbalanced (e.g., about 5% of cases or less in one of the two categories), coefficients might be biased (King & Zeng, 2021). Data analysis was conducted with Stata (version SE 18.0).

Results

Summary statistics

In a first step, we look at missing data: complete (i.e., providing no answer at all) and soft item nonresponse (providing a useless answer). Table 2 reveals that item nonresponse is low. This applies to both questions and all three experimental conditions. For the question on height, item nonresponse varies between 2.4% (selected VI condition) and 4.5% (assigned male VI condition). The result of the chi-square test reveals no statistically significant differences between the conditions ($\chi^2(2) = 0.50$, $p = 0.11$). Even though item nonresponse is slightly higher for the question on weight than for the question on height, the overall item nonresponse rate is also low (less than 6%). For weight, item nonresponse varies between 4.2% (selected VI

[5] According to the German Microcensus, the average height of males in Germany in 2021 was 178.9 cm and the average weight was 85.8 kg. The average height of females in Germany in 2021 was 165.8 cm and the average weight was 69.2 kg (DESTATIS, 2023). The Microcensus information is based on self-reports so that weight is often underestimated compared to objectively measured values, whereas height is often overestimated.

condition) and 5.4% (assigned female VI condition). Again, the result of the chi-square test reveals no statistically significant differences between the conditions ($\chi^2(2) = 1.16$, $p = 0.56$).

Looking at our two key variables, there are only slight differences between the three experimental conditions when it comes to the average height and weight. Height varies between 173.8 cm (assigned male VI condition) and 174.4 cm (assigned female VI condition). The result of the one-way analysis of variance (ANOVA) reveals no statistically significant differences between the conditions ($F(2,1444) = 0.44$, $p = 0.65$). We find a similar picture for weight. More specifically, weight varies between 82.3 kg (selected VI condition) and 83.6 kg (assigned male VI condition). Again, the result of the ANOVA reveals no statistically significant differences between the conditions ($F(2,1424) = 0.57$, $p = 0.57$).

Height and weight can be used to determine the Body Mass Index (BMI) of a person. The BMI is calculated by dividing weight in kilograms (kg) by the square of height in meters (m). The mean values of BMI across the experimental conditions are presented in Appendix D.

In a next step, we look at VI selection by respondents. For this purpose, we exclusively consider respondents that were randomly allocated to the third experimental condition (selected VI). Table 3 presents the results of the VI selection. Female respondents show a clear preference for the female VI. Only about 20% of female respondents select a male VI. Male respondents do not show a clear preference for the male or female VIs. Similar to the findings by Catania et al. (1996) for telephone interviews, about 52% of male respondents select a male VI and about 48% of male respondents select a female VI. The result of the chi-square test reveals a statistically significant difference ($\chi^2(1) = 68.94$, $p < 0.001$). This indicates that male and female respondents have differing preferences when it comes to the gender of VIs.

Table 2. Item nonresponse (in percentages) and mean values for height (in centimeters) and weight (in kilograms) across the three experimental conditions

	Assigned male VI condition	Assigned female VI condition	Selected VI condition
Height question			
Item nonresponse	4.5 (17)	4.3 (16)	2.4 (18)
Mean values	173.8 (364)	174.4 (354)	174.3 (729)
Weight question			
Item nonresponse	5.3 (20)	5.4 (20)	4.2 (31)
Mean values	83.6 (361)	82.7 (350)	82.3 (716)

Note. Case numbers are in parentheses.

Table 3. VI selection (in percentages) across male and female respondents

	Gender of VI	
	Male	Female
Respondent gender		
Male	51.8 (201)	48.2 (187)
Female	21.7 (72)	78.3 (260)

Note. Case numbers are in parentheses. We only included respondents that were randomly assigned to the third experimental condition (selected VI) in which they could select either the male VI from the first condition or the female VI from the second condition.

Research questions 1 and 2

We now investigate the disclosure of sensitive information when respondents and VI match in gender (first research question) and the disclosure of sensitive information when respondents select the VI themselves (second research question). To this end, we estimated four sequential linear regression models for height (in cm) as dependent variable and four sequential linear regression models for weight (in kg) as dependent variable. Tables 4 and 5 present the results of the regression models.

Starting with height (see Table 4) we observe that respondents that are asked the question by a female VI report a lower height (Model 1). VI selection, in contrast, is not associated with respondents' height. The adjusted R^2 value of Model 1 is very low. Turning to Model 2, it is evident that the coefficient of female VI is not associated with height anymore. The main reason for this finding is the inclusion of respondent gender in the model: reflecting real population differences, female respondents report being shorter (negative coefficient) than male respondents, irrespective of the gender of their VI. In the selected VI condition, about 80% of female respondents selected a female VI (see Table 3). In Model 1, it appears that female respondents confound the association between female gender of the VI and height as dependent variable (see Appendix E for the regression models without respondents in the selected VI condition). We also controlled respondent characteristics, including age, education, and device type (smartphone participation). While age is negatively associated with height, higher education is positively associated with height. Even though smartphones have some device characteristics, such as small screen sizes and on-screen keypads that shrink the viewing space for substantive web survey content, that can affect response behavior (Krebs & Höhne, 2021), smartphone participation is not associated with height at all. The adjusted R^2 value of Model 2 increases substantially. In Model 3, we additionally include matching gender between respondent and interviewer. However, the coefficient is not associated with height. The adjusted R^2 value in Model 3 remains high (same level as in Model 2). Finally, the interaction term in Model 4 is not statistically significant and does not result in an increased adjusted R^2 value.

With respect to weight, we find a somewhat different picture than for height. In model 1, neither interviewed by female VI nor selected VI are associated with respondents' weight. The adjusted R^2 value remains low. In Model 2, female is negatively associated with weight. In contrast to height, age is positively and higher education is negatively associated with weight. The adjusted R^2 value of Model 2 increases. Finally, in Model 3, we now find that matching gender between respondent and interviewer is negatively associated with weight. This implies that respondents report a lower weight if they are asked by a VI of the same gender. This result contradicts findings reported by Lipps and Lutz (2017) for telephone interviews as well as findings reported by Conrad et al. (2020) for choosing a VI to conduct a hypothetical future survey. The adjusted R^2 value in Model 3 remains on the same level as in Model 2. Finally, as in our analysis on height, the interaction term in Model 4 is not statistically significant and does not result in an increased adjusted R^2 value.

Table 4. Linear regression models with height (in cm) as dependent variable

	Model 1		Model 2		Model 3		Model 4	
	B	SE	B	SE	B	SE	B	SE
Intercept	175.01***	0.44	181.54***	0.91	181.61***	0.93	181.30***	0.96
Interviewed by female VI	-1.93***	0.51	0.15	0.39	0.13	0.39	0.60	0.54
Selected VI	0.37	0.51	-0.16	0.38	-0.13	0.39	0.44	0.60
Female respondent			-12.37***	0.40	-12.34***	0.40	-12.25***	0.41
Age			-0.05***	0.01	-0.05***	0.01	-0.05**	0.01
Higher education			1.17**	0.43	1.17**	0.43	1.17**	0.43
Smartphone participation			0.03	0.39	0.04	0.39	0.02	0.39
Matching gender between respondent and VI					-0.17	0.40	-0.22	0.40
Selected VI * Interviewed by female VI							-1.00	0.79
Observations	1,340		1,340		1,340		1,340	
Adjusted R^2	0.01		0.44		0.44		0.44	

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. B = Unstandardized coefficient. SE = Standard error. Listwise deletion of missing values. Independent variables: Interviewed by female VI (1 = “Yes”) with interviewed by male VI as reference, selected VI (1 = “Yes”) with assigned VI as reference, female respondent (1 = “Yes”) with male respondent as reference, age (in years), higher education (1 = “Yes”) with low to intermediate education as reference, smartphone participation (1 = “Yes”) with computer and tablet as reference, and matching gender between respondent and VI (1 = “Yes”) with mismatching gender between respondent and VI as reference.

Table 5. Linear regression models with weight (in kg) as dependent variable

	Model 1		Model 2		Model 3		Model 4	
	B	SE	B	SE	B	SE	B	SE
Intercept	83.84***	0.91	85.09***	2.36	86.12***	2.39	86.53***	2.47
Interviewed by female VI	-1.61	1.06	0.67	1.01	0.35	1.01	-0.25	1.39
Selected VI	-0.46	1.06	-1.27	0.99	-0.86	1.00	-1.60	1.53
Female respondent			-12.18***	1.02	-11.83***	1.03	-11.95***	1.05
Age			0.13***	0.03	0.13***	0.03	0.13***	0.03
Higher education			-3.39**	1.10	-3.30**	1.10	-3.30**	1.10
Smartphone participation			-0.92	1.02	-0.76	1.02	-0.75	1.02
Matching gender between respondent and VI					-2.51*	1.02	-2.45*	1.02
Selected VI * Interviewed by female VI							1.29	2.03
Observations	1,318		1,318		1,318		1,318	
Adjusted R^2	0.00		0.13		0.13		0.13	

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. B = Unstandardized coefficient. SE = Standard error. Listwise deletion of missing values. Independent variables: Interviewed by female VI (1 = “Yes”) with interviewed by male VI as reference, selected VI (1 = “Yes”) with assigned VI as reference, female respondent (1 = “Yes”) with male respondent as reference, age (in years), higher education (1 = “Yes”) with low to intermediate education as reference, smartphone participation (1 = “Yes”) with computer and tablet as reference, and matching gender between respondent and VI (1 = “Yes”) with mismatching gender between respondent and VI as reference.

Discussion

The aim of this study was to investigate whether and to what extent respondents disclose more sensitive information if they match the gender of the VI (first research question) and are allowed to select the VI themselves (second research question). To this end, we selected two open numeric questions about respondents' height and weight following the assumption that social norms generally favor taller individuals (Thompson et al., 2023) and individuals who weigh less (Puhl et al., 2021). With respect to the first research question, our findings showed that respondents do not disclose more sensitive information (e.g., higher weight) when their gender matches that of the VI asking the questions. Our findings also shed light on our second research question as they showed that respondents do not disclose more sensitive information when they select the VI themselves.

One way for respondents to circumvent answering sensitive questions is to engage in item nonresponse by providing either no answer at all (complete item nonresponse) or a useless answer (soft item nonresponse). As shown in our results, item nonresponse rates were below 5% (question on height) and 6% (question on weight), respectively. These item nonresponse rates are slightly higher than those reported by previous studies. For example, Couper et al. (2011) reported item nonresponse rates of up to 3% for open numeric questions on non-sensitive topics. In addition, Conrad et al. (2023, p. 155) reported an overall item nonresponse rate of about 2% in the context of prerecorded video interviewing, including both sensitive closed and non-sensitive open numeric questions. Thus, the gender of our VIs did not affect the rate of missing data. The negligible differences in item nonresponse (about one percentage point) between our two open numeric questions also indicate that the questions under investigation have similar sensitivity levels. To put it differently, respondents did not seem to find one question more sensitive than the other.

While we found that a gender match did not influence respondents' self-reported height, it was negatively associated with respondents' self-reported weight. This finding contradicts the assumption that an attribute match between respondents and VIs generally fosters the creation of trust and rapport, increasing the disclosure of sensitive information. Relatedly, Conrad et al. (2023) found that respondents were more likely to report being "slightly overweight" when the VI matched their race. A similar but statistically non-significant increase was found when respondents matched the VI's gender. One explanation for the difference between our results and those of Conrad et al. (2023) is that we used open numeric questions without predefined answer options that constitute a "frame of reference" on which respondents can draw. More specifically, response options in closed questions, including their range, may provide respondents with cues on the average weight in the population, and these cues are used by them as a benchmark when selecting an option. Open numeric questions, in contrast, do not provide such a frame to respondents. In future studies, we therefore recommend experimentally varying the question type (closed versus open numeric) when respondents are asked to report their weight to VIs with different genders.

Turning to our second research question on respondents' disclosure when they self-select the VI, we found no association between the selected VI condition and our dependent variables. This similarly applies to height and weight. Thus, being able to select a male or female VI does not necessarily provide respondents with a sense of control over the web survey situation that generates comfort and rapport. In this study, we used VIs that appeared life-like due to their visual appearance, facial expression, and speech. However, the VIs were not autonomous or responsive as they remained static until the respondent clicked the play button, and even then, they simply read questions aloud. These design features may have hindered the development of a connection between respondents and the VIs, which would potentially have increased disclosure. To shed light on respondents' perceptions of the VIs, it appears worthwhile to include evaluative questions allowing respondents to, for example, evaluate the VIs in terms of naturalness, authenticity, and rapport (see Conrad et al., 2015). Even though allowing respondents to select a VI can be seen as a nice design feature that is much less complicated to implement than for human interviewers, our results did not provide any evidence that it is methodologically beneficial. For now, we therefore argue that it does not appear necessary to give respondents a choice when asking sensitive questions in web surveys through VIs that differ in gender.

Interestingly, we found that the variable interviewed by female VI obtained a lower height from respondents (see Model 1 in Table 4), but the negative association vanished when we controlled for further variables. More specifically, it seemed what initially appeared to be an effect of VI gender on self-reported height can actually be explained by the preponderance of female respondents choosing the female VI in the selected VI condition. Remember that 78% of female respondents selected a female VI and that they self-reported a lower height. In addition, the increase in the adjusted R^2 value (from Model 1 to 2 in Table 4) suggested that there simply is no evidence that gender of the VI (and its match with respondents' gender) affected respondents' self-reported height and weight. This claim is further supported by our additional analysis on height, excluding respondents allocated to the selected VI condition (see Appendix E). Taken together, a gender match between respondents and VIs did not increase the disclosure of sensitive information.

Considerations on the empirical findings

In contrast to previous research, we did not find systematic effects of VI gender in our study. From our perspective, the reasons for our findings may be attributable to various substantive and design-related aspects. First, questions on height and weight may not have been sufficiently sensitive to elicit differential disclosure behavior by respondents. In particular, in self-administered web surveys, respondents are generally more accustomed to answering sensitive questions than in interviewer-administered surveys, thereby possibly perceiving questions on height and weight as less sensitive. It also appears plausible that the strength of social desirability related to reporting height and weight does not differ meaningfully by respondent gender. This limits the potential for gender of VI effects to emerge. As respondents encountered only two VI-based questions after completing various survey questions on another, perhaps even more sensitive topic (i.e., same-gender partnerships), the overall salience of the treatment may have been reduced. Consequently, the VIs may have failed to establish any form of social presence, so that respondents do not attend closely to VI attributes in the form of gender. In addition, the appearance of both VIs was relatively casual as they wore white shirts. A more formal outfit, such as a business casual one with a button-down shirt (male VI) or a blouse (female VI), might have been recognized differently by respondents and resulted in greater effects.

Limitations and future research opportunities

This study has some methodological limitations that provide new avenues for future research. First, we only looked at two open numeric questions dealing with sensitive topics in the form of height and weight. Future research could investigate a more diverse set of questions that varies in terms of sensitivity and topics, and it could assess the level of sensitivity in a previously conducted norming study by human judges (see, for example, Conrad et al., 2023). In addition, it might be worthwhile investigating further VI attributes, such as appearance and clothing.

Second, we implemented an experimental study design with two different VIs (male and female) and a selected VI condition. However, to better evaluate our findings, it would have been methodologically beneficial to include a text-based web survey interface. A text-based interface without VI could have served as a standard of comparison in our analysis.

Third, our sample was recruited through Meta's Facebook with comparatively highly educated respondents. Over 70% of the respondents indicated that they completed college preparatory secondary school or more. However, VIs reading questions to respondents may be especially well suited for respondents with low education and/or literacy. We therefore urge future studies to specifically focus on these respondent groups when investigating the methodological merits and limits of VIs.

Finally, similar to other studies, we conducted our study in one single country (Germany). Considering the survey literature, most studies have been carried out in the US so far, limiting the conclusions that can be drawn about cross-national or cross-cultural settings. We therefore encourage future studies to be conducted in further countries or even across countries.

Conclusion

By shedding light on the gender match of respondents and VIs and the disclosure of sensitive information, this study provides novel insights on open numeric questions administered in web surveys. Our study indicates that the gender of VIs should be considered carefully because it may influence disclosure of sensitive information (e.g., weight). Research on web surveys and VIs (especially those having a life-like appearance) is still in its infancy and their underlying technologies, such as image diffusion, text-to-speech synthesis, and video animation, are rapidly evolving. Thus, VIs in web surveys certainly warrant further investigation. This especially applies to VIs that are (more) autonomous or responsive than the ones tested in this study.

References

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). Web survey methodology. Sage. https://doi.org/10.4135/9781529799651

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. Public Opinion Quarterly, 60, 345–375. https://doi.org/10.1086/297758

Cibelli Hibben, K.C., Felderer, B., & Conrad, F.G. (2022). Respondent commitment: applying techniques from face-to-face interviewing to online collection of employment data.

International Journal of Social Research Methodology, 25, 15–27. https://doi.org/10.1080/13645579.2020.1826647
Conrad, F. G., Schober, M. F., Nielsen, D., & Reichert, H. (2020). Virtual interviewers, social identities, and survey measurement error. In K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), Interviewer effects from a total survey error perspective (pp. 149–164). CRC Press.
Conrad, F. G., Schober, M. F., Hupp, A. L., West, B. T., Larsen, K. M., Ong, A. R., & Wang, T. (2023). Video in survey interviews: Effects on data quality and respondent experience. Methods, Data, Analyses. https://doi.org/10.12758/mda.2022.13
Conrad, F. G., Schober, M. F., Jans, M., Orlowski, R. A., Nielsen, D., & Levenstein, R. (2015). Comprehension and engagement in survey interviews with virtual agents. Frontiers in Psychology, 6, 1578. https://doi.org/10.3389/fpsyg.2015.01578
Couper, M. P., Kennedy, C., Conrad, F. G., & Tourangeau, R. (2011). Designing input fields for non-narrative open-ended responses in web surveys. Journal of Official Statistics, 27, 65–85.
Daikeler, J., Bosnjak, M., & Lozar Manfreda, K. (2020). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. Journal of Survey Statistics and Methodology, 8, 513-539. https://doi.org/10.1093/jssam/smz008
DESTATIS (Federal Statistical Office of Germany). (2023). Microcensus 2021. https://www.destatis.de/DE/Home/_inhalt.html
DESTATIS (Federal Statistical Office of Germany). (2024). Microcensus 2023. https://www.destatis.de/DE/Home/_inhalt.html
Elevelt, A., Höhne, J.K., & Blom, A. (2021). Squats in surveys: Investigating the feasibility of, compliance with, and respondents' performance on fitness tasks in self-administered smartphone surveys using acceleration data. Frontiers in Public Health, 9. https://doi.org/10.3389/fpubh.2021.627509
Fitzgerald, R., & Aizpurua, E. (2024). 20 Years of the European Social Survey: An Interview with Rory Fitzgerald. Survey Practice, 17. https://doi.org/10.29115/SP-2024-0025
Fuchs, M. (2009). Gender-of-interviewer effects in a video-enhanced web survey: Results from a randomized field experiment. Social Psychology, 40(1), 37–42. https://doi.org/10.1027/1864-9335.40.1.37
Fuchs, M., & Funke, F. (2009). Die video-unterstützte Online-Befragung: Soziale Präsenz, soziale Erwünschtheit und Underreporting sensitiver Informationen. In N. Jackob, H. Schoen, & T. Zerback (Eds.), Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung (pp. 159–180). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91791-7_10
Haan, M., Ongena, Y. P., Vannieuwenhuyze, J. T. A., & de Glopper, K. (2017). Response behavior in a video-web survey: A mode comparison study. Journal of Surveys Statistics and Methodology, 5(1), 48–69. https://doi.org/10.1093/jssam/smw023
Höhne, J.K. (2023). Are respondents ready for audio and voice communication channels in online surveys? International Journal for Social Research Methodology, 26, 335-342. https://doi.org/10.1080/13645579.2021.1987121
King, G., & Zeng, L. (2001). Logistic regression in rare events data. Political Analysis, 9, 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Krebs, D., & Höhne, J.K. (2021). Exploring scale direction effects and response behavior across PC and smartphone surveys. Journal of Survey Statistics and Methodology, 9, 477-495. https://doi.org/10.1093/jssam/smz058

Lenzner, T., Höhne, J.K., & Gavras, K. (2024). Innovating web probing: Comparing written and oral answers to open-ended probing questions in a smartphone survey. Journal of Survey Statistics and Methodology. https://doi.org/10.1093/jssam/smae031

Lind, L. H., Schober, M. F., Conrad, F. G., & Reichert, H. (2013). Why do survey respondents disclose more when computers ask the questions? Public Opinion Quarterly, 77, 888–935. https://doi.org/10.1093/poq/nft038

Lipps, O., & Lutz, G. (2017). Gender of interviewer effects in a multitopic centralized CATI panel survey. methods, data, analyses, 11, 67–86. https://doi.org/10.12758/mda.2016.009

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: a meta-analysis comparing response rates. Journal of the Market Research Society, 50, 79–104. https://doi.org/10.1177/147078530805000107

Olbrich, L., Kosyakova, Y, & Sakshaug, J.W. (2022). The reliability of adult self-reported height: The role of interviewers. Economics & Human Biology, 45, 101118. https://doi.org/10.1016/j.ehb.2022.101118

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. International Journal of Behavioral Nutrition and Physical Activity, 5, 56. https://doi.org/10.1186/1479-5868-5-56

Puhl, R. M., Lessard, L. M., Pearl, R. L., Himmelstein, M. S., & Foster, G. D. (2021). International comparisons of weight stigma: Addressing a void in the field. International

Thompson, K., Portrait, F., & Schoonmade, L. (2023). The height premium: A systematic review and meta-analysis. Economics and Human Biology, 50, Article 101273. https://doi.org/10.1016/j.ehb.2023.101273

Thornton, L., Batterham, P. J., Fassnacht, D. B., Kay-Lambkin, F., Calear, A. L., & Hunt, S. (2016). Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interventions, 4, 72–81. https://doi.org/10.1016/j.invent.2016.02.001

Appendix A

Figure A1. Screenshot of the Facebook ad for recruiting web survey respondents

Appendix B

Table B1. Basic benchmarks regarding the costs of the Facebook ad campaign

Age and gender combination	Average price for web survey completion
Young/male	1.17€
Middle-aged/male	0.64€
Old/male	0.84€
Young/female	1.44€
Middle-aged/female	0.84€
Old/female	2.09€

Note. The price calculation is based on the number of web survey completions related to each ad (not on self-reported age and gender in the web survey). The average price per click on the web survey link posted on Facebook was 0.42€ and the average price per web survey completion was 1.11€.

Appendix C

English translations of the virtual interviewer introduction and the virtual interviewer selection question

Introduction of the VIs
In the following, a virtual interviewer will ask you two questions via video. You can play the videos as often as you like. To play the videos, simply click on the play button on the video.

In the first question, the virtual interviewer asks you to state your body height and in the second question, the virtual interviewer asks you to state your body weight.

You can enter your answers in an open field below the video.

Selected VI question (third experimental condition only)
Below you can select whether a male or female virtual interviewer asks you two questions.

[Picture of the male and female VIs; see also Figure 1]

Answer options: 1 "Male interviewer" or 2 "Female interviewer"

Appendix D

Table D1. Mean values of the Body Mass Index (BMI) across the three experimental conditions

	Assigned male VI condition	Assigned female VI condition	Selected VI condition
Respondents			
Males	27.1 (181)	27.8 (173)	27.2 (375)
Females	28.0 (164)	26.3 (162)	26.7 (313)
Overall	27.5 (345)	27.1 (335)	27.0 (688)

Note. Case numbers are in parentheses. The BMI is calculated by dividing weight in kilograms (kg) by the square of height in meters (m). Values between 18.5 and 24.9 fall into the healthy range.

Appendix E

Table E1. Linear regression models with height (in cm) as dependent variable without respondents in the selected VI condition

	Model 1		Model 2		Model 3	
	B	SE	B	SE	B	SE
Intercept	173.78***	0.50	180.94***	1.31	181.05***	1.33
Interviewed by female VI	0.55	0.71	0.59	0.54	0.58	0.54
Female respondent			-12.11***	0.55	-12.11***	0.55
Age			-0.05**	0.02	-0.05**	0.02
Higher education			1.92**	0.62	1.93**	0.62
Smartphone participation			-0.26	0.55	-0.26	0.55
Matching gender between respondent and VI					-0.24	0.53
Observations	667		667		667	
Adjusted R^2	0.00		0.44		0.44	

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. B = Unstandardized coefficient. SE = Standard error. Listwise deletion of missing values. Independent variables: Interviewed by female VI (1 = "Yes") with interviewed by male VI as reference, female respondent (1 = "Yes") with male respondent as reference, age (in years), higher education (1 = "Yes") with low to intermediate education as reference, smartphone participation (1 = "Yes") with computer and tablet as reference, and matching gender between respondent and VI (1 = "Yes") with mismatching gender between respondent and VI as reference.