Understanding panel conditioning: an examination of social desirability bias in self-reported height and weight in panel surveys using experimental data

SC Noah Uhrig

University of Essex

scnuhrig@essex.ac.uk

Abstract

Typically reliant on self-reports from panel data, a growing body of literature suggests that relative body weight can have negative effects on labour market outcomes. Given the interest in the effects of relative weight in the social sciences, this paper addresses the question of whether repeated interviewing affects the quality of these data. A theory that focuses on the sensitivity of the questions rather than the survey context is proposed. Examining experimental panel data from Understanding Society using quantile-regression, the findings for women are consistent with the argument that conditioning reduces social desirability effects. The ameliorative effects of panel conditioning on social desirability bias in self-reported height and bodyweight appear to strengthen the association between relative weight and employment for men, but not women, however.

Keywords

1. Introduction

Since the early 1980s, researchers have examined the extent to which obesity is penalised in labour market outcomes for both men and women (McLean and Moon 1980; Puhl and Heuer 2009). In particular, obesity seems to affect employability, especially for women (Morris 2007; Sarlio-Lähteenkorva and Lahelma 1999; Viner and Cole 2005). Obesity also seems to negatively impact wages for those who *are* employed (Baum and Ford 2004; Cawley 2004; Cawley, Grabka and Lillard 2005; Morris 2006; Pagán and Dávila 1997).

This literature mainly relies on self-reported height and weight, yet validation studies of self-reported height and bodyweight usually find a bias towards cultural ideals (Bostöm and Diderichsen 1997;

Spencer, Appleby, Davey and Key 2002). For many longitudinal surveys of general use, obtaining self-reported height and bodyweight is the most efficient and sustainable method of obtaining these data.

Panel surveys are a central means of understanding and explaining causal processes in human belief and action. However, the experience of being surveyed can limit valid inference about such processes in the social sciences (Kalton and Citro 2000; Sturgis, Allum and Brunton-Smith 2009). Humans are reflexive; they can integrate new information and experiences into their self-concept, their beliefs and opinions, and their behaviour. Humans can also be reactive since they often respond to situations and contexts in specifically patterned ways. Respondent reflexivity and reactivity can contribute to measurement error of

which this paper concerns itself with two flavours. First, panel conditioning is a systematic effect of participation in a previous interview on either response behaviour or on the behaviour targeted by survey measurement (Waterton and Lievesley 1989). Human reflexivity can lead to a systematic change in what is being measured or how measures work at later waves. Second, social desirability bias is a systematic effect of editing responses before answering survey questions. Respondents may be embarrassed, want to keep information private, or may want to avoid negative feelings or distress, and so they lie or otherwise misreport about something a researcher would like to know (DeMaio 1984; DePaulo, Kashy, Kirkendol, Wyer and Epstein 1996; Foddy 1993).

This paper concerns itself with the nexus between panel conditioning and social desirability within the context of self-reported height and bodyweight. Although validation has exposed bias in self-reports of height and bodyweight, little is known about their longitudinal measurement properties. Moreover, strong interest amongst both economists and sociologists in the effects of obesity on labour market outcomes, suggests that evaluating the effects of these biases on substantive relationships may be warranted. Therefore, this research aims to examine the specific question of whether biases in self-reported height and bodyweight are ameliorated by repeated interviewing. Furthermore, it explores whether these biases impact the relationship between obesity and employment.

2. The sensitivity of self-reported height and bodyweight

Questions which suffer from over- or under-reporting that might not be attributable to deficiencies in question comprehension, information retrieval or formatting are likely to be sensitive or threatening questions and hence suffer from social desirability bias (Bradburn and Sudman 1979; see also Tourangeau, Rips and Rasinski 2000). Cross-sectional validation of self-reported height and bodyweight normally finds systematic mis-reporting suggestive of social desirability bias. Bodyweight is typically under-reported, often in the range of 1/4 to 1/2 stone or about 1.5 to 3.5 kg (Borkan, Hults and Glynn 1983; Dekkers, van Wier, Hendriksen, Twisk and van Mechelen 2008; Spencer, Appleby, Davey and Key 2002). Under-

reporting is consistently greater amongst those who are heavier and by women (Palta, Prineas, Berman and Hannan 1982; Rowland 1990; Stewart, Jackson, Ford and Beaglehole 1987). Both Spencer et al (2002) and Rowland (1990) find that the extent of under-reporting of bodyweight increases with increasing respondent weight, more so for women than for men. The margin of error for women is typically twice that for men at the heaviest weights. Rowland also finds that underweight men over-report their bodyweight by roughly 1/3 stone or about 2.3 kg. Validation of self-reported height finds over-reporting, though often by less than half an inch (Dekkers et al 2008; Rowland 1990; Spencer, Appleby, Davey and Key 2002). As with weight, mis-reports are associated with both gender and the true value. Greater over-reports are observed amongst shorter men – about 1.5 cm, or a little more than 1/2 inch, amongst the shortest men (Rowland 1990). Though generally of a small magnitude, Spencer et al (2002) find that men's overestimates of height are nearly twice that of women. Given these validation findings, it is clear that self-reporting height and bodyweight is sensitive for certain respondents.

3. Panel conditioning and the sensitivity of self-reported height and weight

Whether self-reported height and bodyweight continue to be biased in on-going panels is an open question. In their statement on the nexus between panel conditioning and social desirability bias, Waterton and Lievesley (1989) postulate that conditioning can be expressed as trust that develops between respondents and the survey organisation. Increased familiarity with the survey organisation and survey procedures over waves of a panel, fosters increased respondent trust. This greater trust leads to less concern over privacy and confidentiality and a greater willingness to divulge potentially unflattering information, or the experience of less threat when asked to do so. They find that respondents were significantly more willing to report racial prejudice and also to report their income at a subsequent interview (Waterton and Lievesley 1989). Similarly, Brannen (1993) finds a number of women report greater feelings of ease and willingness to talk freely to interviewers at later waves of a three year longitudinal study of mothers returning to the labour market after child-birth. However, Pevalin (2000) examines whether social

desirability affects the reporting of symptoms used to construct an indicator of mental illness in annual waves of a long-running panel, but finds no effects. Moreover, certain highly sensitive behaviours, such as drug taking, seem to suffer from greater social desirability bias at later panel waves, contrary to what would be predicted by Waterton and Lievesley's model (Fendrich and Vaughn 1994; Mensch and Kandel 1988; Wagstaff, Kulis and Elek 2009). These findings suggest that social desirability effects in panels may depend on something other than the social context of being interviewed.

Sensitive questions themselves can vary in their psychic consequences for respondents, from mild embarrassment, to feelings of intense violations of privacy or even shame, associated with providing the true answer (DeMaio 1984; Schaeffer 2000). Although people with various non-normative heights and bodyweight may be stigmatised in many interactions (Puhl and Heuer 2009), we might reasonably assume that survey questions obtaining height and bodyweight are on the lighter end of social desirability. These are questions that might elicit embarrassment, but most likely not intensely private feelings of shame which strike at the core of human identity (Tangney and Fischer 1995). Unlike shame, embarrassment is a relatively short-lived negative emotion that rarely shatters the self-concept of most who experience it (Tangney and Fischer 1995; Tangney, Miller, Flicker and Barlow 1996). Thus, panel conditioning may operate slightly differently with height and bodyweight than for more severely threatening questions such as illicit drug taking, abortion or various sexual practices.

One strategy people use to avoid or manage feelings of embarrassment, is lying. DePaulo and colleagues (1996) find that lying is often unplanned and focused on achieving psychic rewards even amongst social intimates (see also Goffman 1963). However, lying tends also to produce negative psychic consequences, such as feelings of social distance and unpleasantness in interaction (DePaulo et al 1996; Goffman 1963). One could surmise that after the initial shock of being asked for one's height or bodyweight, people may lie as they attempt to 'pass' as having a more ideal body figure than they actually possess. In a panel survey, the second time the question is asked, providing an accurate answer may not be nearly as embarrassing. Having learned that the information can be given without horrid consequences, there may be less motivation to mis-report, and indeed, the added motivation of avoiding negative feelings associated with lying may lead respondents to provide more accurate answers. Respondents may reflect on their experience of being asked certain questions, particularly those to which their answer was not entirely accurate, and be motivated to tell the truth at subsequent waves. This approach relies only on the effects of questions themselves, not on the developing relationship between panel respondents and the survey organisation.

4. Experimentation with panel content

Holt (1989) argues that most analytic designs used to examine panel conditioning are not ideal for disentangling the factors that affect survey data quality. Most frequently, researchers compare responses from some later wave of a panel to a temporally corresponding cross-sectional sample to which the same or similar questions were asked (Corder and Horvitz 1989; Menard and Elliott 1993; Osgood, O'Malley, Bachman and Johnston 1989; Silberstein and Jacobs 1989; Wilson and Howell 2005). Rotating panel designs provide a better alternative because comparisons can be made between identical survey designs, sample designs, procedures and identical questions (Bailar 1975; Cohen and Burt 1985; Ghangurde 1982; McCormick, Butler and Singh 1992; Pennell and Lepkowski 1992; Silberstein and Jacobs 1989). A third analytic strategy involves examination of a single sample and predicting specific effects, from theory, about panel conditioning (Pevalin 2000; Sturgis, Allum and Brunton-Smith 2009). In all of these approaches, the effects of conditioning are confounded with other aspect of the study design in some way. Only experimentation can pinpoint the specific nature of effects, by holding study design features constant across randomised experimental treatments. Experimentation with panel data is rare – typically limited to the effects of randomised sample selection on behavioural outcomes such as voting or obtaining health checks (Bartels 1999; Battaglia, Zell and Ching 1996; Clausen 1968; Kraut and McConahay 1973; Traugott and Katosh 1979; Wilson and Howell 2005; Yalch 1976). Bridge et al (1977) alone use an experimental design to examine conditioning on subsequent *survey* measures. Keeping all other survey conditions constant across a two- wave panel, the authors

experimentally varied survey content and their 'between wave' communication with respondents. They found that content alone can induce opinionation, particularly for topics seen as important and where opinions had not previously been formed.

Having a separate and unique sample and study design, the *Understanding Society* "Innovation Panel" (the IP) is dedicated to methodological research relevant for improving longitudinal research resources. At Waves 1 and 2, the IP contained an experimental content allocation that could be used to examine the effects of prior survey content on measurement. At Wave 1, a random half of households received survey questions to obtain self-reported height and bodyweight, while the other half of households received no such content. At Wave 2, the entire sample was asked to self-report height and bodyweight. All other survey conditions remained constant across this experimental treatment, thereby isolating the effect of prior question content in the panel.

5. Hypotheses

Considering the nature of biases in self-reported height and weight, the approach outlined in Section 3 suggests the following specific hypotheses, given the study design. Considering first the effects of conditioning on self-reported weight:

Hypothesis 1: Validation finds greater bias amongst heavier women, therefore conditioned women should report heavier weights than unconditioned women at the upper end of the bodyweight distribution.

Hypothesis 2: Validation finds that lighter men are likely to over-report their bodyweight, therefore conditioned men should report lower weights than unconditioned men at the lower end of the bodyweight distribution.

With respect to height, the effects are again likely to be gender specific:

Hypothesis 3: Validation finds that taller women under-report their height therefore conditioned women should report taller heights than unconditioned women at the upper end of the height distribution.

Hypothesis 4: Validation finds that shorter men over-report their height therefore conditioned men should report shorter heights than unconditioned men at the lower end of the height distribution.

Validation of self-reported height and bodyweight typically find that small differences in self-reports render the classification of respondents into categories of relative bodyweight significantly biased. For example, Spencer et al found that about 41 percent of obese men and 27 percent of obese women were mis-classified based on self-reports (2002). For this reason:

Hypothesis 5: Conditioned respondents of both sexes should exhibit higher body-mass index values than unconditioned respondents.

Hypothesis 6: Conditioned respondents of both sexes should be more likely to be classified as overweight or obese as compared to unconditioned respondents.

The discussion in Section 3 and these hypotheses speak to the likelihood that social desirability is expressed through lying only. However, wilfully providing inaccurate information is not the only strategy of socially desirable responding (Tourangeau, Rasinski, Jobe, Smith and Pratt 1997; Tourangeau, Rips and Rasinski 2000). Item non-response is a common method to avoid providing information which is unflattering or otherwise highly sensitive (Kennickell 1996; Moore, Stinson and Welniak 1999). Indeed, non-response often decreases over waves of data collection (Bailar 1989; Cantor 1989; Porst and Zeifang 1987; Sturgis, Allum and Brunton-Smith 2009; Traugott and Katosh 1979; Waterton and Lievesley 1989). Between height and bodyweight, bodyweight suffers from greater non-response (Rowland 1990). Respondents may choose not to respond at all. Thus:

Hypothesis 7: Both men and women who are conditioned should have lower levels of item non-response for bodyweight than unconditioned men and women.

Another strategy of socially desirable responding is to provide round values. Rounding has been shown to be problematic in survey reporting of many types of information (see e.g. Roberts and Brewer 2001). In his study of U.S. men and women, Rowland (1990) found that rounding was common in self-reported bodyweight but not in self-reported height. He found that 60 percent of respondents rounded bodyweight to a numeric value ending in 0 or 5, such as 160 lbs or 185 lbs. Rounding was more common among women and heavier respondents, and those who provided rounded values for bodyweight were significantly less accurate than those who did not:

Hypothesis 8: If conditioning affects the social desirability of reported bodyweight, then conditioned respondents should be less likely to provide rounded values for bodyweight compared to unconditioned respondents.

What are the consequences of social desirability bias for the association between body-mass and employment outcomes? Across a range of settings and time periods, research on the link between obesity and employment generally finds a negative effect, which is stronger for women than for men (Puhl and Heuer 2009). This literature relies on self-reported height and bodyweight to calculate body-mass, though recent work by Morris using anthropometric measures from the mid-1990s, finds consistently negative, but small, effects for both men and women on employment (Morris 2007). We might expect, then, that social desirability bias produces an under-estimate of the relationship between obesity and employment:

Hypothesis 9: If conditioning reveals a relationship closer to that found from anthropometrics, then the relationship between obesity and employment should be strengthened, particularly for women.

6. Data

The IP sample was recruited in January 2008 using a stratified and clustered design. In total 2,760 addresses in 120 areas of Great Britain were selected from the Postcode Address File using post-code sectors as Primary Sampling Units. Areas north of the Caledonian Canal and all of Northern Ireland were excluded from sampling. Interviews were achieved in 1,489 households for a household response rate of 59.5 percent, not counting ineligible addresses. A total of 2,393 individual interviews with adult household members aged 16+ were obtained. All people resident at the address, including children, were defined as original sample members to be followed throughout the life of the study. The second round of interviews occurred in March 2009, where interviews were conducted in 1,122 households including approximately 72 new ones due to splits from original sampled households. Including full, partial or proxy interviews, Wave 2 data was gathered for 1,870 adults. The analysis proceeds with a balanced panel of respondents interviewed at both Wave 1 and Wave 2 of the IP. Thus, 168 new respondents at Wave 2 were excluded. Since self-reported bodyweight for pregnant women is subject to

different types of biases, the analysis excludes 22 women who were pregnant at Wave 2. Note, since the actual bodyweight provided at Wave 1 is irrelevant for this analysis, any pregnant women at Wave 1 were retained. Finally, cases with any missing data on any response variable or covariates were excluded.

Self-reported height was obtained with the following question "I would like to ask you about height and weight. There is interest in how people's weight, given their height, is associated with their health. How tall are you without shoes?" Self-reported bodyweight was obtained by asking "What is your current weight?" Although answers could be given in either imperial or metric units, responses were overwhelmingly provided in imperial therefore metric reports are excluded from the analysis. Three follow-up questions asked whether the respondent was fairly sure of their bodyweight or if it was an estimate, when they last weighed themselves with scales and for women, whether they were currently pregnant.

Survey data accuracy is best judged with validation data, yet such data are often a rare and prized commodity in methodological research. In the absence of validation data, an analysis of experimental data with well established biases in which we might observe improvement would be a viable alternative. If conditioning reduces socially desirable responding, the response distribution for both height and bodyweight should be affected in ways counter to observed biases. Thus, those respondents who were asked for their height and bodyweight at Wave 1, i.e. the "conditioned" treatment, should report heights and weights that are systematically opposed to the biases identified by validation.

7. Models and measures

Hypotheses 1 through 5 outlined in Section 5 address the effects of conditioning on specific social desirable responding at specific regions of the bodyweight, height and body-mass distributions. To examine these hypotheses, I use quantile regression. Where linear regression predicts mean values, quantile-regression focuses on the conditional response distribution. That is, quantile-regression can estimate a specified percentile or percentiles of a continuous response variable conditioned on a set of covariates (Hao and Naiman 2007; Koenker and Bassett 1978). For this reason, quantile-regression is more appropriate for

examining the effects of panel conditioning on social desirability bias, given that it affects specific points in the distributions of bodyweight, height and body-mass.

I specify quantile regression models for bodyweight, height and body-mass the same way. Taking weight, w_i for example, the model for the pth quantile, Q_p, is commonly written (see e.g. Abrevaya 2001; Hao and Naiman 2007) as:

$$Q_p(w_i|x_i) = \beta_{0,p} + x_{i,k}\beta_{k,p} + u_{i,p}, \quad (i = 1, \dots, n) \quad (1)$$

where $0 < p < 1$ indicates the proportion of the population having weights below the quantile at p and β_p represent the marginal effects of the k covariates $x_{i,k}$ on the quantile value. Equation (1) implies that the conditional p^{th} quantile is determined by the quantile-specific parameters $\beta_{0,p}$ and $\beta_{k,p}$, and the specific values of the covariates $x_{i,k}$. An error term $u_{i,p}$ for any particular conditional quantile is assumed to be zero. Since quantile-regression estimates points in the dependent variable's distribution, it requires no distributional assumptions about the dependent variable and is robust to any skewness that may be present (Arulampalam, Booth and Bryan 2007; Buchinsky 1994; Buchinsky 1998; Koenker and Hallock 2001). To test hypotheses about the upper and lower ends of the bodyweight, height and body-mass distributions, the 25th, 50th and 75th percentiles are estimated simultaneously using Stata 10 (StataCorp 2007). These points were chosen to reflect the main junctures in the distribution, i.e. median and inter-quartile range, whilst maintaining sample size at each point.

Estimation of coefficients in quantile-regression is implemented in Stata using the method of minimum absolute deviations via linear programming. This approach to estimation was first implemented by Wagner (1959) and details of Stata's implementation of the procedure can be found in the Stata documentation (StataCorp 2007). Standard errors of the estimates for quantile regression coefficients can be obtained by various means (for a review, see Buchinsky 1995). The standard errors presented in this paper are obtained by bootstrapping (Arulampalam, Booth and Bryan 2007; Hao and Naiman 2007; Koenker and Hallock 2001). Bootstrapping is preferred because the resulting standard errors are not affected by sample size when estimating

coefficients at the extreme quartiles of the distribution (Hao and Naiman 2007). To control for clustering in the sample design when estimating standard errors, primary sampling units were resampled as part of the bootstrapping procedure, rather than individuals (Arulampalam, Booth and Bryan 2007).

Hypotheses 7, 8 and 9 outlined in Section 5 suggest the use of models for dichotomous outcomes. Here, I used logistic regression to examine these hypotheses. Assume some event, *Y*, such as non-response to a bodyweight question. The dependent variable is expressed as a log transformation of the odds, ϑ, of the event:

$$\vartheta(Y = 1) = \frac{\Pr(Y=1)}{1-\Pr(Y=1)} \quad (2)$$

Given that probabilities range from 0 to 1, the odds can range from 0, when Pr(Y = 1) = 0, to infinity when the Pr(Y = 1) = 1. By taking the natural logarithm of the odds, we obtain the *logit*:

$$L = log_e\vartheta \quad (3)$$

This *logit* transformation of the probability is then modelled as a linear function of covariates (Hanuchek and Jackson 1977):

$$L_i = \beta_0 + \beta_k x_{i,k} + u_i, \quad (i = 1, \dots. n) \quad (4)$$

Hypothesis 6 requires ascertaining whether conditioning effects on social desirable responding results in different classification into categories of relative weight. To test this hypothesis, I use a multinomial logistic model. This model is a generalisation of logistic regression that allows for more than two discrete outcomes. Suppose that there are *k* categorical outcomes with some base outcome as being category 1. The probability that the response for some i^{th} observation is:

$$p_{i,j} = \Pr(y_j = i)$$
$$= \begin{cases} \dfrac{1}{1 + \sum_{m=2}^{k} \exp{(x_j\beta_m)}}, if \ i = 1 \\ \dfrac{\exp{(x_j\beta_i)}}{1 + \sum_{m=2}^{k} \exp{(x_j\beta_m)}}, if \ i > 1 \end{cases} \quad (5)$$

where x_j is the row vector of observed values of the independent variables for the jth observation and β_m is the coefficient vector for outcome *m*. Estimates for both logistic regression and

multinomial logistic regression were obtained by maximum likelihood using Stata 10 (StataCorp 2007). Standard errors in both logistic regression and multinomial logistic regression models were corrected for clustering in the sample design using established methods (Kreuter and Valiant 2007).

I examine three distinct dependent variables – self-reported height, self-reported bodyweight and calculated relative bodyweight – all measured at Wave 2. Table 1 contains descriptive statistics for key variables used in the analyses presented in this paper. Body-mass is measured by relative weight, or bodyweight in kilograms divided by height in metres-squared (Morris 2007; Morris 2006). Known as the "body-mass index" (BMI), values are ordinarily grouped into categories of underweight (BMI < 18.5), normal weight (18.5 ≤ BMI < 25), overweight (25≤ BMI < 30) and obese (30 ≤ BMI). The main covariate of interest is assignment to experimental treatment – that is, whether the height and bodyweight questions were asked of the respondent at Wave 1 or not. All respondents receiving the height and bodyweight questions at Wave 1 are treated as being "conditioned" whereas those who did not receive these questions were in the "unconditioned" group.

Table 1. Descriptive statistics of variables used in analyses

	N	Mean/Pct	St. Dev			
Age	1724	50.8	17.5			
Conditioned	1724	49.8%				
Female	1724	54.0%				
First Degree, or higher	1714	22.1%				
School degree	1714	32.8%				
Other degree	1714	27.6%				
No degree	1714	17.5%				
Recently weighed	1648	42.8%				
Rounded weight response	1386	58.4%				
Weight non-response	1416	2.1%				
Men	N	Mean / Pct	St. Dev	25th-	50th-%	75th-%
Weight (in pounds)	639	182.2	31.2	161	180	199
Height (in inches)	725	69.7	2.9	68	70	72
Calculated BMI	632	26.4	4.1	23.7	25.9	28.8
Underweight	632	1.0%				
Normal weight	632	37.3%				
Overweight	632	44.9%				
Obese	632	16.7%				
Employed	793	58.4%				
Women	N	Mean / Pct	St. Dev	25th-	50th-%	75th-%
Weight (in pounds)	747	151.3	30.7	131	147	168
Height (in inches)	862	63.9	2.7	62	64	66
Calculated BMI	739	26.1	5.1	22.6	25.5	28.5
Underweight	739	2.7%				
Normal weight	739	44.4%				
Overweight	739	34.8%				
Obese	739	18.1%				
Employed	931	51.2%				
Responsible for a child <	931	21.3%				

All models also control for respondent age and education. Age and education can be assumed to be proxies for cognitive abilities, which in turn have consequences for measurement error (Alwin 2007). Education was measured in terms of highest qualifications obtained, categorised into four groups: University degree or higher, or an equivalent, which I label "First Degree or higher"; completion of compulsory schooling or its equivalent, including those staying on until age 18, which I call "School Degree"; all other qualifications not elsewhere classified, including foreign degrees, which I call "Other qualifications"; and no reported qualifications at all.

Two other variables of interest include whether the reported bodyweight was rounded and when the respondent most recently weighed her/himself. Rounding was measured using "digit preference" which was indicated if the report was of a whole or half-stone, e.g. 9½ stone rather than 9 stone 8 ounces. Respondents were asked when they most recently weighed themselves, in order to provide some sense of how accurate the response might actually be. Whether the respondent "Recently weighed" was indicated by whether the respondent reported weighing themselves within 4 weeks of the interview.

Finally, employment is measured as whether the respondent did any work for pay in the seven days ending the Sunday prior to interview. If the respondent did no work, but had a job from which they were temporarily away, they were classified as being employed. Motherhood is one limiting factor in determining whether women participate in the labour market. The models for female employment, therefore, also includes a variable for whether she is the responsible adult for a child under age 10 in the household.

8. Results

Table 2 contains results of quantile-regressions of self-reported bodyweight on conditioning treatment predicting the 25th, 50th and 75th percentiles. Hypothesis 1 implies that the 75th percentile for conditioned women would be higher than for unconditioned women. The results in column IV indicate that there is very little effect of conditioning on the 25th and 50th percentile, but a coefficient of 5.27 ($p < 0.10$) suggests that the 75th percentile for conditioned women is about 5¼ pounds higher than for unconditioned women. This effect strengthens, once rounding and recent weighing are controlled in the models. The 75th percentile for conditioned women is about half a stone higher than for unconditioned women. Note that rounding tends to increase the reported bodyweight – that is, rounding tends to increase the distributional points by about 5 pounds. Hypothesis 2 suggests that the 25th percentile of bodyweight for conditioned men should be lower than for unconditioned men. The findings shown in columns I, II and III are consistent with this hypothesis but are not significant.   It should be noted that conditioning seems to yield lower median and 75th percentiles for men's bodyweight as well, though the effects are not significant.

Table 2. Results from simultaneous quantile-regressions of conditioning, rounding and recent weighing on 25th, 50th, and 75th percentiles of self-reported bodyweight

		Men			Women		
		I	II	III	IV	V	VI
25th %	Conditioning	-1.33	-1.42	-1.33	-1.23	-1.18	-1.51
		(3.11)	(3.21)	(3.04)	(3.09)	(3.09)	(3.04)
	Rounding		0.09	3.04		4.66*	3.61
			(2.69)	(2.95)		(2.97)	(2.97)
	Recent weighing			7.33**			0.82
				(2.03)			(2.59)
50th %	Conditioning	-3.29	-2.66	-3.11	0.27	-0.32	-0.63
		(2.77)	(2.88)	(2.90)	(2.45)	(2.13)	(2.28)
	Rounding		-1.76	0.51		5.44**	5.53**
			(3.17)	(3.52)		(2.26)	(2.41)
	Recent weighing			4.31			-0.29
				(3.40)			(2.27)
75th %	Conditioning	-3.80	-3.71	-4.48	5.27*	7.07**	7.04**
		(3.56)	(3.52)	(3.42)	(2.97)	(3.04)	(3.17)
	Rounding		-0.98	0.40		5.35	5.26
			(3.44)	(3.20)		(3.29)	(3.24)
	Recent weighing			5.43*			-0.78
				(3.18)			(2.59)
	N	635	635	629	744	744	741

* p < 0.10, ** p < 0.05

Notes. Shown are coefficients. Standard errors, shown in parentheses, are obtained by bootstrap methods (500 replications) and adjusted for clustering in sample design. Respondent age and education are included in the models but not shown.

Results in Table 3 evaluate Hypotheses 3, 4 and 5. Hypothesis 3 implies that the conditioning effect on the 75th percentile of the height distribution for women should be positive. The results in Table 3 for women show that across the percentiles estimated, conditioning seems to increase the point in the distribution estimated, though at greater amounts in the lower end of the distribution than at the upper end of the distribution. However, the conditioning effect for women is not significant across all percentiles.   Hypothesis 4 implies a negative effect of conditioning on the 25th percentile estimate of conditioned men's height. Instead, the results show a positive effect for conditioning.   In fact, the 75th percentile for conditioned men is about 2/3 of an inch higher than for unconditioned men and this effect is statistically significant (*p* < 0.05), though the conditioning effects on the 25th and 50th percentiles are not.

Hypothesis 5 implies that conditioning should positively affect the body-mass distribution at all points for both men and women. For men, the effects of conditioning are negative across the estimated quartile points and all are generally of the same effect size. However, none of these effects are statistically significant. For women, the coefficients for conditioning are negative on the 25th and 50th percentiles while the effect is positive on the 75th percentile.   The coefficient for conditioning on the 75th percentile of .93, means that conditioned women are almost one full point higher on the body-mass index scale compared to unconditioned women. This would be expected if conditioning induced women to report less biased amounts of bodyweight and height. However, all coefficients in these quantile regressions of women's BMI percentiles are not significant.

Table 3. Results from simultaneous quantile-regressions of conditioning, rounding and recent weighing on 25th, 50th, and 75th percentiles of self-reported height, and calculated body-mass index (BMI) for men and women

	Men		*Women*	
	Height	**BMI**	**Height**	**BMI**
25th %	0.16	-0.41	0.16	-0.11
	(0.27)	(0.40)	(0.23)	(0.39)
50th %	0.38	-0.54	0.13	-0.37
	(0.28)	(0.36)	(0.26)	(0.49)
75th %	0.62***	-0.42	0.10	0.93
	(0.24)	(0.43)	(0.23)	(0.66)
N	721	628	859	736

*** $p < 0.01$

Notes.   Shown are coefficients for conditioning only. Standard errors, shown in parentheses, are obtained by bootstrap methods (500 replications) and are adjusted for clustering in the sample design. Respondent age and education are included in the model, but not shown in the table.

Table 4 presents results from a multinomial logistic regression of conditioning on categorisation of body-mass. Hypothesis 6 implies that the conditioned sample should be classified into heavier categories of body-mass than the unconditioned sample. The results in Table 4 show that this is not the case. Note that the values of relative bodyweight, used to categorise people into underweight, normal weight, overweight and obese, do not conform to the 25th, 50th and 75th percentile of the body-mass index distribution. This classification pattern is meant to reflect the health effects of bodyweight given height. Thus, the nearly 1 point increase observed in body-mass at the 75th percentile for conditioned women shown in Table 3 does not necessarily translate into a re-categorisation. For men, conditioning seems to increase the likelihood of being classified as underweight relative to normal bodyweight, with small negative effects for conditioning on being overweight or obese relative to normal. However, none of these effects are statistically significant. For women, conditioning seems to reduce the odds of being classified as underweight or overweight relative to normal, with a small positive effect on the odds of classification as obese, though none of these effects are statistically significant.

Table 4. Results from multinomial logistic regression of conditioning on categorisation of body-mass

	Men	*Women*
Underweight	1.25	-0.32
	(0.95)	(0.54)
Normal weight (omitted)	--	--
Overweight	-0.08	-0.16
	(0.22)	(0.21)
Obese	-0.09	0.03
	(0.28)	(0.23)
N	672	761

Notes. Shown are coefficients for conditioning only, age and education are also included in these models but are not shown. Standard errors, which are adjusted for clustering in sample design, are shown in parentheses. Normal weight is the omitted category.

Table 5. Results from logistic regression of conditioning and recent weighing on various indicators of survey response quality

Men	Rounding		Weight Non-response		Recent Weighing
Conditioning	0.02	0.00	0.05	-0.02	-0.19
	(0.20)	(0.21)	(0.61)	(0.62)	(0.16)
Recent weighing		-1.13***		-1.41*	---
		(0.19)		(0.76)	---
N	678	672	685	679	781

Women	Rounding		Weight Non-response		Recent Weighing
Conditioning	-0.08	-0.05	-0.78	-0.65	0.21
	(0.17)	(0.17)	(0.60)	(0.61)	(0.14)
Recent weighing		-0.57***		-1.10**	---
		(0.17)		(0.52)	---
N	769	766	792	787	918

$* p < 0.10, ** p < 0.05, *** p < 0.01$

Notes. Shown are coefficients. Age and education are controlled in the models but are not shown in the table. Standard errors, which are adjusted for clustering in sample design, are shown in parentheses.

Hypotheses 7 and 8 focus on the extent to which conditioning reduces the propensity to provide a round number for bodyweight and non-response at self-reported bodyweight questions. Table 5 contains coefficients from a set of logistic regressions of conditioning on rounding, bodyweight non-response and recent weighing. Hypothesis 7 suggests that the coefficient for conditioning when predicting item non-response at the bodyweight question should be negative. For women, this is in fact the case though the effects are not statistically significant. The conditioning effect for men, though also non-significant, changes from positive to negative once recent weighing is controlled. Hypothesis 8 suggests that conditioning should reduced the probability of providing a rounded number as a response strategy to avoid socially undesirable disclosure of weight. Here we would expect to find a negative effect for conditioning. The coefficient for men is positive, though very close to zero and not statistically significant. The coefficient for women is in fact negative, though not statistically significant. Although there is no clear hypothesis about recent weighing, conditioning seems to have a negative effect on recent weighing for men and positive effect for women, though the coefficients are not significant.

Table 6. Estimates from logistic regression of employment on categorisation of overweight or obese, and body-mass index

Men	Unconditioned		Conditioned	
	I	II	III	IV
Overweight	0.35		0.43	
	(0.45)		(0.38)	
Obese	-0.13		-0.21	
	(0.46)		(0.50)	
BMI		0.01		-0.02
		(0.05)		(0.04)
F	1.13	1.18	1.59	1.56
p-value	0.35	0.32	0.15	0.17
N	239	239	247	247

Women	Unconditioned		Conditioned	
	I	II	III	IV
Overweight	-0.44		0.15	
	(0.35)		(0.40)	
Obese	-0.75		-0.15	
	(0.49)		(0.49)	
BMI		-0.04		0.01
		(0.03)		(0.04)
F	3.02	3.83	4.46	4.57
p-value	0.00	0.00	0.00	0.00
N	237	237	240	240

Notes: Shown are coefficients. Respondent age, education and motherhood (women only), are included in all models but not shown in these tables. Standard errors, which are corrected for clustering in the sample design, are shown in parentheses.

Table 6 presents results from logistic regression of being obese and overweight, and BMI, on the probability of being employed for both conditioned and unconditioned men and women. Hypothesis 9 suggests that the effect sizes for the conditioned sample should be greater than for the unconditioned sample, particularly for women. The literature on obesity and employment generally finds that men do not typically experience a penalty for being overweight or obese and the results in Table 6 conform to this general finding, regardless of conditioning treatment.   Note that the coefficients for being overweight and obese are of a similar sign for both conditioned and unconditioned men, and that the size of the coefficients is generally larger for conditioned as compared to unconditioned men, as is hypothesised. However, none of these coefficients are statistically significant, so it would be inappropriate to test for significant differences between them.

The results in Table 6 for women are exactly opposite to hypothesised. Amongst unconditioned women, the coefficients for being overweight and obese are negative – as one might expect from the literature – with the effect for obesity being larger than the effect for being overweight. The results also show a negative association between BMI and the odds of employment for unconditioned women. The results for conditioned women suggest a lessening of the association rather than a strengthening. The coefficient for being overweight is positive, whilst the coefficient for being obese is negative. Moreover, there is a slight positive effect of BMI for conditioned women on employment. However, none of these coefficients for women are statistically significant in either the conditioned or unconditioned samples.

9. Discussion

The theoretical approach outlined in Section 3 implies that more extremely sensitive questions might be less affected by panel conditioning than more moderate questions. Self-reported height and bodyweight may cause embarrassment when posed, motivating respondents to mis-report in cross-sectional studies, but this motivation dissipates at a subsequent administration. The test of this approach controlled for survey context and all other conditions of the interview, varying only

whether the questions were asked or not to respondents in the "conditioned" group.   In this way, the design should evaluate the extent to which the questions themselves, as distinct from the survey experience, fosters more accurate reporting. Admittedly, validation data would be beneficial in evaluating the extent to which accurate reporting is obtained.   Nevertheless, the nature of bias is established in the literature so one can surmise that, were conditioning to be ameliorative, then reporting in ways opposite to those biases would be observed for the conditioned treatment.

The results presented here are somewhat disappointing as very few findings reach statistical significance. Nevertheless, many of the findings – particularly for women – are in the directions hypothesised. Conditioned women tend to report being heavier and taller than unconditioned women – both of which are in directions contrary to the expected direction of bias.   This suggests that conditioning does induce some degree of more accurate reporting, particularly for weight, amongst women. Consequently, heavier conditioned women tend to have a calculated body-mass that is about one point higher than unconditioned women. In terms of data quality, conditioned women are less likely to provide rounded bodyweights, (i.e. to the whole or half stone), they are less likely to non-respond when asked for their body weight, and they are more likely to have weighed themselves recently.   Taken together, this all suggests that conditioning enhances the reporting of accurate data as compared to data reported by unconditioned women.   Although all but the findings for self-reported bodyweight fail to reach statistical significance at normal levels, the findings align with the proffered theory.   Therefore, the theory may have some relevance for women.

The set of effects for men are less easily interpreted than for women. While the results indicate that conditioned lightweight men tend to report lower weights than unconditioned men, conditioning seems to induce men across the distribution to report lower weights. As for height, conditioning leads to reporting of taller heights, not shorter as hypothesised.   Consequently, conditioned men tend to have body-mass index values lower than unconditioned men across all points in the distribution of body-mass. If we consider that the nature of validation bias in self-

reported height and bodyweight for men is against being small or slight, the results for calculated body-mass would conform to the theory. That is, conditioned men report being more slight at all points in the distribution of body-mass than unconditioned men. Yet given that all but the results for tall men's height are not statistically significant and the results for height and bodyweight do not otherwise conform to prediction, the theory proffered in Section 3 would not seem to hold much salience for men.

The effects of obesity on employment should be affected by biases in self-reported height and bodyweight. If reports are biased towards cultural ideals, we might expect that effects would be attenuated as compared to those based on anthropometric measures of height and bodyweight. If conditioning reduces these biases, the effects should strengthen. The results presented here do not clearly support this idea. The results for men, though non-significant, do seem to align in the expected ways however. The coefficients for conditioned men are generally larger than for unconditioned men, though non-significant for either group. Even though panel conditioning does not seem to ameliorate the reporting bias we might expect for men given validation, the results for the effects on the relationship between obesity and employment seem to support the approach proffered in Section 3. On the other hand, the results for women's employment do not support these ideas. The coefficients for being overweight are of different sign across conditioning treatments, and the coefficient for being obese, though of the same sign, is of a lower magnitude amongst the conditioned sample as compared to the unconditioned sample. The fact that these results are not statistically significant does not help with interpretation. It would seem that, although there is some evidence that women's reporting of height and bodyweight may be more accurate, this increased accuracy does not lead to a strengthening of the association between obesity and employment as observed using anthropometric measures.

10. Conclusion

This paper examines whether panel conditioning ameliorates the effects of social desirable responding in panel surveys. The approach rests on the degree of threat posed by survey questions rather than the context in which they are asked. Sensitive questions can be more or less sensitive or sensitive in qualitatively different ways. Self-reported height and bodyweight are argued to be less threatening than most topics covered in the literature on social desirability. Panel conditioning is argued to lessen the threat of these questions and therefore reduce social desirability bias irrespective of continued participation and familiarity with the survey as others have argued. One strength of this research over previous examinations of panel conditioning is that it uses an experimental design to test these ideas. Though coefficients were generally found to not reach statistical significance, the theory seems to account for conditioning effects on women's self-reports better than men's.

This research is limited in that it does not address the full range of potentially sensitive questions. One extension would be to look at the effects of conditioning on a set of questions that vary in degree of sensitivity. If the theory holds, then highly sensitive questions should be more resilient to ameliorative panel conditioning.

Finally, researchers may interpret these data to mean that conditioning does not influence self-reported height and bodyweight at all. Moreover, one might conclude that the effects of conditioning on social desirable reporting do not influence the association between obesity and employment. Were this analysis replicated on the full *Understanding Society* sample, we may discover that these types of measurement problems are reassuringly small.

11. References

Abrevaya J. (2001) The Effects of Demographics and Maternal Behavior on the Distribution of Birth Outcomes. *Empirical Economics* 26, 247-257.

Alwin D. (2007) *Margins of Error*. Wiley and Sons, Inc, Hoboken, NJ.

Arulampalam W, Booth AL and Bryan ML. (2007) Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wages Distribution. *Industrial and Labor Relations Review* 60,163-186.

Bailar BA. (1989) Information Needs, Surveys and Measurement Errors. In D. Kasprzyk, G Duncan, G Kalton and MP Singh. eds. *Panel Surveys*. Pp. 1-24. John Wiley, New York.

Bailar BA. (1975) The effects of rotation group bias in estimates from panel surveys. *Journal of the American Statistical Association,* 70**,** 23-30.

Bartels LM. (1999) Panel effects in the American National Election Studies. *Political Analysis,* 8**,** 1-20.

Battaglia MP, Zell ER and Ching PLYH. (1996) Can participating in a panel sample introduce bias into trend estimates? *1996 Proceedings of Survey Research Methods Section of the American Statistical Association***,** 1010-1013.

Baum CL and Ford W. (2004) The wage effects of obesity: a longitudinal study. *Health Economics,* 13**,** 885-899.

Borkan GA, Hults DE and Glynn RJ. (1983) Role of longitudinal change and secular trend in age differences in male body dimensions. *Human Biology,* 55**,** 629-64.

Bostöm G and Diderichsen F. (1997) Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *International Journal of Epidemiology,* 25**,** 860-866.

Bradburn NM and Sudman S. (1979) *Improving interview method and questionnaire design: response effects to threatening questions in survey research,* San Francisco, Jossey-Bass.

Brannen J. (1993) The effects of research on participants: findings from a study of mothers and employment. *The Sociological Review,* 41**,** 328-346.

Bridge RG, Reeder LG, Kanouse D, Kinder DR, Nagy VT and Judd CM. (1977) Interviewing changes attitudes - sometimes. *Public Opinion Quarterly,* 41**,** 56-74.

Buchinsky M. (1994) Changes in the U.S. wage structure 1963-1987: an application of quantile regression. *Econometrica,* 62**,** 405-458.

Buchinsky M. (1995) Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study. *Journal of Econometrics,* 68**,** 303-338.

Buchinsky M. (1998) Recent advances in quantile regression models: a practical guideline for empirical research. *Journal of Human Resources,* 33, 88-126.

Cantor D. (1989) Substantive implications of longitudinal design features: the national crime survey as a case study. In D Kasprzyk, G Duncan, G Kalton and MP Singh. eds. *Panel Surveys*, Pp 25-51. John Wiley, New York.

Cawley J. (2004) The impact of obesity on wages. *Journal of Human Resources,* 39, 451.

Cawley J, Grabka M and Lillard DR. (2005) A comparison of the relationship between obesity and earnings in the US and Germany. *Journal of Applied Social Science Studies,* 125, 119-129.

Clausen A. (1968) Response validity: vote report. *Public Opinion Quarterly,* 32, 588-606.

Cohen SB and Burt VL. (1985) Data collection frequency effects in the National Medical Care Utilization and Expenditure Survey. *Journal of Economic and Social Measurement,* 13, 125-151.

Corder, LS and Horvitz DG. (1989) Panel effects in the National Medical Care Utilization and Expenditure Survey. In D Kasprzyk, G Duncan, G Kalton and MP Singh. eds. *Panel Surveys*, Pp 304-318. John Wiley, New York.

Dekkers JC, van Wier MF, Hendriksen IJM, Twisk JWR and van Mechelen W. (2008) Accuracy of self-reported body weight, height and waist circumference in a Dutch overweight working population. *BMC Medical Research Methodology,* 8, 69.

DeMaio, TJ. (1984) Social desirability and survey measurement: a review. In CF Turner and E Martin. eds. *Surveying Subjective Phenomena (Vol. 2)*. Pp. 257-281. Russell Sage Foundation, New York.

DePaulo BM, Kashy DA, Kirkendol SE, Wyer MM and Epstein JA. (1996) Lying in everyday life. *Journal of Personality and Social Psychology,* 70, 979-995.

Fendrich M and Vaughn CM. (1994) Diminished lifetime substance use over time: an inquiry into differential underreporting. *Public Opinion Quarterly,* 58, 96-123.

Foddy W. (1993) *Constructing questions for interviews and questionnaires: theory and practice in social research*. Cambridge University Press, Cambridge.

Ghangurde PD. (1982) Rotation group bias in the LFS estimates. *Survey Methodology,* 8, 86-101.

Goffman E. (1963) *Stigma: notes on the management of spoiled identity*. Penguin, Harmondsworth.

Hanuchek EA and Jackson JE. (1977) *Statistical methods for social scientists*. Academic Press, Orlando, FL.

Hao L and Naiman DQ. (2007) *Quantile regression*. Sage Publications, Thousand Oaks, CA.

Holt D. (1989) Panel conditioning: discussion. In D Kasprzyk, G Duncan, G Kalton and MP Singh. eds . *Panel Surveys,* Pp 340-347. John Wiley, New York.

Kalton G and Citro CF. (2000) Panel surveys: adding the fourth dimension. In D Rose. ed. *Researching social and economic change*. Pp. 36-51. Routledge, London.

Kennickell AB. (1996) Using range techniques with CAPI in the 1995 Survey of Consumer Finances. *Paper presented at the Joint Statistical Meetings, August 1995, Chicago, IL.*

Koenker R and Bassett G. (1978) Regression quantiles. *Econometrica,* 46, 33-50.

Koenker R and Hallock KF. (2001) Quantile regression: an introduction. *Journal of Economic Perspectives,* 15, 143-156.

Kraut RE and McConahay JG. (1973) How being interviewed affects voting: an experiment. *Public Opinion Quarterly,* 37, 398-406.

Kreuter F and Valiant R. (2007) A survey on survey statistics: what is done and can be done in Stata. *The Stata Journal,* 7, 1-21.

McCormick MK, Buterl DM and Singh RP. (1992) Investigating time in sample effects for the Survey of Income and Program Participation. *Proceedings of the Survey Research Section of the American Statistical Association*, 554-559.

McLean RA and Moon M. (1980) Health, obesity and earnings. *American Journal of Public Health,* 70, 1006-1009.

Menard S and Elliott DS. (1993) Data set comparability and short-term trends in crime and delinquency. *Journal of Criminal Justice,* 21, 433-445.

Mensch BS and Kandel DB. (1988) Underreporting of substance use in a National Longitudinal Youth Cohort: individual and interviewer effects. *Public Opinion Quarterly,* 52, 100-124.

Moore JC, Stinson LL and Welniak EJ. (1999) Income reporting in surveys: cognitive issues and measurement error. In MG Sirken, DJ Herrmann, S Schechter, N Scharz, JM Tanur and R Tourangeau . eds. *Cognition and survey research*. Pp 155-176. John Wiley, New York.

Morris S. (2007) The impact of obesity on employment. *Labour Economics,* 14, 413-433.

Morris S. (2006) Body mass index and occupational attainment. *Journal of Health Economics,* 25, 347-364.

Osgood DW, O'Malley PM, Bachman JG and Johnston LD. (1989) Time trends and age trends in arrests and self-reported illegal behavior. *Criminology,* 27, 389-417.

Pagán JA and Dávila A. (1997) Obesity, occupational attainment, and earnings. *Social Science Quarterly,* 78, 756-770.

Palta M, Prineas RJ, Berman R and Hannan P. (1982) Comparison of self-reported and measured height and weight. *American Journal of Epidemiology,* 115, 223-230.

Pennell S and Lepkowski JM. (1992) Panel conditioning effects in the survey of income and program participation. *Proceedings of the Survey Research Section of the American Statistical Association*, 566-571.

Pevalin DJ. (2000) Multiple applications of the GHQ-12 in a general population sample: an investigation of long-term retest effects. *Social Psychiatry and Psychiatric Epidemiology,* 35, 508-512.

Porst R and Zeifang K. (1987) A description of the German General Social Survey Test-Retest Study and a report on the stabilities of the sociodemographic variables. *Sociological Methods and Research,* 15, 177-218.

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

Roberts JM and Brewer DD. (2001) Measures and tests of heaping in discrete quantitative distributions. *Journal of Applied Statistics,* 28, 887-896.

Rowland ML. (1990) Self-reported weight and height. *American Journal of Clinical Nutrition,* 52, 1125-1133.

Sarlio-Lähteenkorva S and Lahelma E. (1999) The association of body mass index with social and economic disadvantage in women and men. *International Journal of Epidemiology,* 28, 445-449.

Schaeffer NC. (2000) Asking questions about threatening topics: a selective overview. In AA Stone, JS Turkkan, CA Bachrach, JB Jobe, HS Kurtzman and VS Cain. eds. *The science of self-report: implications for research and practice*, Pp. 105-121. Lawrence Erlbaum Associates Inc, Mahwah, NJ.

Silberstein AR and Jacobs CA. (1989) Symptoms of repeated interview effects in the Consumer Expenditure Interview Survey. In D Kasprzyk, G Duncan, G Kalton and MP Singh. eds. *Panel Surveys,* Pp. 289-303. John Wiley, New York.

Spencer EA, Appleby PN, Davey G and Key TJ. (2002) Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutrition,* 5, 561-565.

StataCorp. (2007) Stata statistical software: release 10. StataCorp LP, College Station, TX.

Stewart AW, Jackson RT, Ford MA and Beaglehole R. (1987) Underestimation of relative weight by use of self-reported height and weight. *American Journal of Epidemiology.* 125, 122-126.

Sturgis P, Allum N and Brunton-Smith I. (2009) Attitudes over time: the psychology of panel conditioning. In P Lynn. ed. *Methodology of longitudinal surveys*, Pp. 113-126. John Wiley & Sons Ltd, New York.

Tangney JP and Fischer K. (1995) *Self-conscious emotions: shame, guilt, embarrassment, and pride*. Guildford Press, New York.

Tangney JP, Miller RW, Flicker L and Barlow DH. (1996) Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology,* 70, 1256-1269.

Tourangeau R, Raskinski K, Jobe JB, Smith TW and Pratt W. (1997) Source of error in a survey of sexual behavior. *Journal of Official Statistics,* 13, 341-365.

Tourangeau R, Rips LJ and Raskinski KA. (2000) *The psychology of survey response*. Cambridge University Press, Cambridge.

Traugott MW and Katosh JP. (1979) Response validity in surveys of voting behavior. *Public Opinion Quarterly,* 43, 359-377.

Viner R and Cole TJ. (2005) Adult socioeconomic, educational, social and psychological outcomes of childhood obesity: a national birth cohort study. *British Medical Journal,* 330 [doi:10.1136/bmj.38453.422049.E0 (published 17 May 2005)].

Wagner HM. (1959) Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, 54, 206-207.

Wagstaff DA, Kulis S and Elek E. (2009) A six-wave study of the consistency of Mexican/Mexican-American preadolescents' lifetime substance use reports. *Journal of Drug Education,* 39, 361-384.

Waterton J and Lievesley D. (1989) Evidence of conditioning effects in the British Social Attitudes Panel. . In D Kasprzyk, G Duncan, G Kalton and MP Singh. eds. *Panel Surveys,* Pp. 319-339. John Wiley & Sons, New York.

Wilson SE and Howell BL. (2005) Do panel surveys make people sick? US arthritis trends in the Health and Retirement Survey. *Social Science and Medicine,* 60, 2623-2627.

Yalch RF. (1976) Pre-election interview effects on voter turnout. *Public Opinion Quarterly,* 40, 331-336.