Survey Research Methods (2015)
Vol. 9, No. 2, pp. 125-137
doi:10.18148/srm/2015.v9i2.6128

ISSN 1864-3361
http://www.surveymethods.org

Straightlining in Web survey panels over time

Matthias Schonlau
University of Waterloo, Canada

Vera Toepoel
University of Utrecht, The Netherlands

Straightlining, an indicator of satisficing, refers to giving the same answer in a series of questions arranged on a grid. We investigated whether straightlining changes with respondents' panel experience in the LISS panel in the Netherlands. Specifically, we considered straightlining on 10 grid questions in LISS core modules (7 waves) and on a grid of evaluation questions in the LISS panel (150+ waves). For both core modules and evaluation questions we found that straightlining increases with respondents' panel experience for at least three years. Straightlining is also associated with younger age. Where straightlining corresponded to a plausible set of answers, prevalence of straightlining was much larger (15-40%) than where straightlining was implausible (<2% in wave 1).

Keywords: data quality, grid question, matrix question, satisficing, panel conditioning

1 Introduction

Data quality is a fundamental concern in survey research. With the rise of online surveys, panels are increasingly being used to collect data. In these panels, the same sets of individuals are repeatedly asked to answer questions. An important advantage of panels is that the same questions can be asked in a series of waves to the same respondents, making it possible to look at individual changes over time. However, a concern about survey quality arises when panelists that have participated in previous waves may respond differently than panelists without experience. In particular, many researchers fear that online survey panels, where respondents are interviewed at a high frequency such as once a month or more, create trained respondents (Hillygus, Jackson, & Young, 2014). Often, this concern is more a matter of speculation than of empirical investigation. There appears to be little empirical research to date how prior survey participation affects survey answers (Binswanger, Schunk, & Toepoel, 2013).

Straightlining has increasingly been used as an indicator for poor response quality (Greszki, Meyer, & Schoen, 2014; Zhang & Conrad, 2014). Straightlining refers to giving non-differentiated (identical) ratings to a series of questions with the same answer choices that are arranged in a grid format, i.e. choosing answers in the same column for the entire grid (visually indicated by the appearance of a straight line). In a survey panel the same respondents are asked to complete multiple surveys over different waves of the panel. Respondents may over time answer questions differently, for example because of survey fatigue. It is relatively unknown whether this so-called panel conditioning manifests itself in the form of increased straightlining.

The purpose of this paper is to study whether straightlining increases with respondents' experience in answering the same questions in previous waves. We use data obtained from the LISS panel, a Dutch probability-based survey panel. We investigate core modules that use the same questions every year, a common approach in longitudinal studies. In addition, we use evaluation questions that are administered in every survey as an extreme form of administering the same questions to the same individuals. In our analysis, we take socio-demographics including immigration status into account to see if particular subgroups are more prone to straightlining behavior. The study is designed to show whether data quality due to straightlining decreases with panel experience and whether this is related to personal characteristics.

An outline of the paper is as follows: in Section 2 we give some background on panel conditioning, satisficing and straightlining in particular. Section 3 presents the methods used, including a description of the LISS panel. Sections 4 and 5 investigate straightlining among grid questions in core modules and evaluation questions, respectively. Section 6 concludes with a discussion.

2 Background

Panel data, where the same individuals are repeatedly asked questions over several waves, have important advantages compared to independent cross sections. However, there are also two potential drawbacks: attrition bias and panel conditioning effects (Das, Toepoel, & van Soest, 2011;

Contact information: Matthias Schonlau (corresponding author), University of Waterloo, 200 University Avenue West, Waterloo, Bldg M3, Department of Statistics and Actuarial Science, ON N2L 3G1, Canada (Email: schonlau@uwaterloo.ca)

Toepoel, Das, & Van Soest, 2008). Attrition bias arises if dropping out of the panel is correlated with a variable of interest. Panel conditioning arises if respondents are influenced by participation in previous surveys, such that their answers differ from the answers of individuals who are interviewed for the first time. They may prepare for future surveys (increase their knowledge), or develop attitudes towards certain topics addressed in previous surveys. In addition, they may become familiar with the question-answering process, learn how to interpret questions, and make fewer errors than new respondents. Or the opposite: experienced respondents may also make more errors than unexperienced respondents – they may speed through the survey or answer strategically to avoid follow-up questions. Panel conditioning has also been called "panel fatigue" when it refers to bias introduced from respondents tiring of participation in the panel or answering the same question (Lundmark & Gilljam, 2013). Posing the same question in several waves to the same respondents can be frustrating and might likely increase the likelihood to satisfice (Scherpenzeel & Saris, 2007).

Panel conditioning may be more likely when waves of the panel occur closely together in time. The longer the interval between waves, the more likely that survey questions tend to be forgotten. Lundmark and Gilljam (2013, p. 16) reported no significant differences between seven waves of the same questions where the timing between the different waves varied between two and seven months. They conclude that "it seems as if there is no additional frustration effect that decreases data quality from posing the same question over and over, at least not when the time between the waves are two to seven months". Halpern-Manners, Warren, and Torche (2014, p. 6) suggested that when panel waves are separated by a month or less, panel conditioning effects are usually observed. When waves are separated one to 12 months apart, the results in literature are mixed, while panel conditioning is rarely observed when baseline and follow-up surveys are separated by more than a year. Das et al. (2011) found examples, however, with a significant panel conditioning effect when there were one or two years between waves.

Panel conditioning effects are strongest in the early waves of a panel (Halpern-Manners et al., 2014). Using panel datasets from the German Socioeconomic Panel (SOEP), the British Household Panel Survey (BHPS) and the Swiss Household Panel (SHP), Van Landeghem (2012, pp. 15–16) investigated a question on subjective well-being posed up to five times in the panels. He found the panel effect may accumulate over multiple survey waves rather than being fully established by the second wave. In West-Germany, for example, there is a substantial panel effect from the second to the fifth interview. This does not hold for all countries, however, since in Switzerland no panel effect is measured any longer after the third interview. Both Toepoel et al. (2008) and Struminskaya (2014) found little to no evidence of panel conditioning in the CentERpanel (administered by CentERdata, the Netherlands) and GESIS Panel (administered by GESIS, Germany). However, Toepoel et al. noted experienced respondents are more likely to take shortcuts in the response process than inexperienced respondents based on higher inter-item correlations for multiple-item-per-screen formats (a.k.a. grid or matrix questions).

The tendency for respondents to seek cognitive shortcuts that involve less-than-optimal responding is called "Satisficing" (Krosnick, 1991). Indicators of satisficing include straightlining or non-differentiation in grid or matrix questions (Bethlehem & Biffignandi, 2011, Sec. 4.2.1.5) as well as speeding. Evidence of straightlining has been used to compare the quality of data from different survey modes (Fricker, Galesic, Tourangeau, & Yan, 2005). Because straightlining is easily evaluated automatically, it is a popular indicator of data quality for marketers. For example, the market research company GfK has developed a program called "ALBERTA" which identifies, among other things, items sets for which straight-lining occurred (Menold, Winker, Storfinger, & Bredl, 2014). Some argue straightlining the middle response option can be seen as endorsing the status quo (Bethlehem & Biffignandi, 2011).

Evidence of a relationship between speeding and survey experience is mixed. Analysing several online panels of market research agencies in the Netherlands, Van Ossenbruggen (2008) noted a strong correlation between panel experience and speeding through the survey. On the other hand, Coen, Lorch, and Piekarski (2005) found no evidence of speeding in their panel, while Hillygus et al. (2014) found that experienced respondents took slightly longer to complete the survey in the 2010 Cooperative Congressional Election Study (CCES). More research is needed to find out when – for which populations, topics, and panel tenure- speeding is most likely to affect data quality.

Straightlining is associated with undesirable speeding in answering questions (Revilla & Ochoa, 2015; Zhang, 2013). Greszki et al. (2014) found several indicators of satisficing to be negatively correlated with response time based on their investigation of the German Longitudinal Election Study (GLES) and American National Election Study (ANES). Straightlining was the strongest indicator for speeding: a one-unit increase in straightlining was accompanied by a decrease of 27 seconds on the respective survey sites. However, excluding speeders did not affect substantive conclusions.

Prevalence estimates of straightlining are sparse. In the 2010 National Survey of Student Engagement (NSSE), 8.5% of respondents straightlining in 3 grids or more (Cole, McCormick, & Gonyea, 2012)). In another study, Couper, Tourangeau, Conrad, and Zhang (2013) noted that only 1.1% of respondents gave the same response to all 13 items, and only 2.3% gave the same response to 12 or more items. Cole, Bowers, McCormick, and Brummett-Carter (2012, p. 7) in-

vestigated several sets of items and concluded that some sets of items displayed very low rates of straightlining (under 2% of respondents) while as many as one in five respondents straightlined in other sets. This suggests that straightlining is dependent on question or topic type. Cole et al. noted that the vast majority of straightliners chose positive response options which were presented first. This is in accordance with satisficing behavior.

Zhang and Conrad (2014) demonstrate that speeding is mostly associated with low-educated respondents. This is also found by (Cole, McCormick, & Gonyea, 2012) who add that male respondents straightline significantly more often than female respondents in nine of eleven items under investigation. Straightlining is also associated with survey reluctance as measured by the number of contact attempts (Kaminska, McCutcheon, & Billiet, 2010)). Kaminska et al. (2010) argue that reluctant respondents are important as their responses may more closely resemble those of non-respondents and that a straightlining respondent may be preferable than a non-responder. The literature on the relationship between straightlining and survey experience is scarce. Smith and Brown (2006) showed experienced and inexperienced panelists were equally likely to straightline. Other than this study, very little is known about straightlining in relation to experience.

It is important to note straightlining in itself does not necessarily imply a loss in data quality or satisficing behavior. Respondents may have thoughtfully considered each response category which resulted in identical responses. For example, a non-religious person may answer "no" to all items related to religion. Some item batteries contain similar items which can be asked in different directions. For example, some items are positively and some are negatively formulated. In that case straightlining – in particular straightlining on the endpoints - becomes increasingly implausible and likely represents poor data quality.

3 Method

Matrix or grid questions are a type of survey question in which multiple items with the same response scales are displayed in a table-like format in which the items – usually statements – are the rows and the response categories the columns (see, for example Zhang & Conrad, 2014). Here we investigate straightlining in grid questions both in core modules and from evaluation questions in the LISS Panel. Core modules are (nearly) identical questionnaires that are asked annually or bi-annually. Evaluations questions are asked as part of most surveys regardless of survey topic. We next describe the LISS panel followed by GEE logistic regression.

The LISS panel is an open-access Internet panel based on a probability sample of households drawn from the Dutch population register in 2007. Households that could not otherwise participate are provided a computer and Internet connection. In 2009 and again in 2010/2011 refreshment samples were drawn. Respondents are paid an incentive of 15 Euro per hour (and proportionally less for shorter surveys). The LISS panel has roughly between 6,000 and 10,000 respondents who participate in monthly Internet surveys. The number fluctuates a bit with attrition and replenishment samples. The socio-demographic composition of respondents to wave 7 of the Leisure Core module is given in Table 1.

Table 1
Socio-demographic composition of respondents to wave 7 of the core module "Leisure"

Variable	N	Percent
Age		
18–35	1608	24.2
35–49.9	1612	24.3
50–64.9	1876	28.3
≥65	1544	23.3
Gender		
male	3578	53.9
female	3062	46.1
Marital Status		
married	3524	53.1
single	2072	31.2
divorced/separated/widowed	1044	15.7
Education		
primary	576	8.7
secondary	2323	35.1
vocational	3078	46.5
university	639	9.7
Immigration status		
Dutch	5298	84.8
1st generation, western	227	3.6
1st generation, non-western	258	4.1
2nd generation, western	317	5.1
2nd generation, non-western	149	2.4

(N=6,616)

At the beginning of each month, LISS panel respondents are invited to participate in one or several surveys. At least one survey each month is conducted with the whole panel rather than with a subset of panel members.

When outcomes are binary (such as straightlining or not straightlining) usually logistic regression is employed. Logistic regression is a member of the family of distributions called generalized linear models. Logistic regression has two potentially serious shortcomings: 1) The variance is a function of the mean. Often real life data sets exhibit overdispersion, that is, the observed variance is greater than would be predicted by the model. 2) Logistic regression assumes that observations are independent which means it cannot account for clustering. Because some respondents may have a greater

tendency to straightline than others, we want to account for possible clustering within respondent.

Generalized Estimating Equations (GEE) (Liang & Zeger, 1986) is a technique that accomplishes both these goals: 1) the variance can be specified independently of the mean; and in particular can accommodate a multiplicative overdispersion parameter. Here we specify a logit link function and a Binomial-like variance structure

$$E(y) = \mu$$

and

$$\phi \mathrm{Var}(\mu) = \phi\mu(1-\mu)$$

where ϕ is the overdispersion parameter. 2) a "working" correlation matrix can be specified for the within cluster correlation. Observations from different clusters are still assumed to be independent. This GEE logistic regression will estimate the coefficients and the standard errors consistently, however the standard errors are only estimated well if the "working" correlation matrix is specified correctly. We therefore take the standard next step and instead estimate the standard errors via the robust of sandwich estimate (Huber, 1967; White, 1982). Unlike the "na ive" model-based estimate of standard erorrs, the robust estimate does not require correct specification of the model and the working correlation matrix to give consistent estimates. Without the burden of having to specify the correct "working" correlation matrix, we specify an exchangeable working correlation matrix here. "Exchangeable" means the correlation between any two observations of the same respondent is assumed to be the same.

Because both use the logit link, estimates from the GEE logistic and logistic regressions are interpreted in the same way. Logistic regression output typically contains a likelihood-based pseudo R^2 value to gauge the strength of the relationship between covariates and the response on a 0-1 scale. GEE output does not contain pseudo R^2 values because GEE is based on quasi-likelihood rather than likelihood.

4 Results

4.1 Straightlining in Core modules

In addition to researcher supplied questionnaires, the LISS panel contains several longitudinal core studies in such topics as health, religion, and personality. Core studies were conducted once a year for six or seven years, depending on the module, with identical questionnaires. All panel members participate in core studies.

The number of grid questions varies considerably by module. Some modules contained no grid questions with at least five questions posed to all respondents: family and household, work and schooling, assets, and housing. Other modules such as politics and personality contained multiple grid questions.

We investigate a variety of grid questions from core modules. First, we report the percentage of straightlining over time. We then repeat the analyses on the same grid questions using regression to adjust for socio-demographic covariates and to establish statistically significant differences.

Straightlining over time: Descriptive analyses. Table 2 shows the percentage of straightlining in core modules for each wave for a variety of grid questions. The question texts and answer choices are given in Appendix A. Note the percentage of straightlining refers to straightlining on *any* answer choice.

The grids can be divided in two groups: those where straightlining is plausible and those where it is not plausible. For example, the "Religion 2" question asks whether the respondent believes in Adam and Eve, the bible, the Koran, the Torah, and praying (see appendix A for the exact wording). These are all religious terms; a non-religious respondent might plausibly say "no" to all of these answers. The leisure question asks whether the respondent had attended any of 11 types of performances ranging from theatre performances to movies in the last 12 months. Again, it is possible that respondents did not attend any of such events making straightlining on "no" a plausible answer. An example of a question where straightlining is not plausible is "Politics 2". This question asks respondents how much confidence they have in each of 16 institutions including the Dutch parliament, United Nations and democracy. It is hard to believe that a respondent upon reflection would rate all 16 institutions identically on an 11-point scale. Another example of a question where straightlining is not plausible is "Personality 3". Here some items are asked in opposite directions, for example, "I would prefer complex to simple problems" and "I prefer to think about small, daily projects to long-term ones".

Table 2 shows the percentage of straightlining across a range of grid questions in different modules. The difference in the prevalence of straightlining among questions in the group "plausible" (14.5%-46.2% in wave 1) and questions in the group "not plausible" (<2% in wave 1) is striking.

Among questions where straightlining is not plausible, the percentage of straightlining doubles for all questions except health where there is a mild non-significant increase. This is further illustrated in Figure 1. Possibly, the health question is more salient and concrete to respondents. Among questions where straightlining is plausible, the percentage of straightlining increases from wave 1 percentages for all questions except income.

Straightlining for the income question increases until wave 4 and then falls in wave 5 and 6. However, for income the percentage of missing values jumps abruptly at wave 3. The percentage of missing values for income in order from wave 1 to 6 is: 6.08%, 3.74%, 17.02%, 17.73%, 16.39%, 17.77%. Increased satisficing does occur in later waves; but not in the form of increased straightlining but rather in the

Table 2
Percent straightlining across a range of grid questions in core modules by wave.

Question	Straight lining plausible	Wave 1	2	3	4	5	6	7
Personality 2	–	0.5	0.9	1.1	1.5	–[a]	1.1	–[b]
Personality 1	–	0.5	0.9	1.1	1.3	–[a]	1.3	–[b]
Politics 2	–	1.0	2.3	3.0	3.3	2.7	2.8	3.1
Politics 1	–	1.1	2.3	3.1	3.7	3.5	3.4	4.0
Health	–	1.8	1.7	1.5	1.9	1.9	2.0	2.1
Personality 3	–	1.9	2.7	3.1	3.9	–[a]	3.5	–[b]
Leisure	yes	14.5	15.3	17.2	19.5	21.6	21.4	20.9
Religion 2	yes	26.5	26.5	27.2	28.1	31.0	31.0	31.5
Religion 1	yes	31.3	31.8	33.4	35.6	38.0	38.4	39.3
Income	yes	46.2	47.9	48.9	49.8	46.9	45.2	–[b]

Rows are sorted by wave 1 percentages.
[a] Grid questions of Wave 5 of module personality are denoted missing as they were only administered to a small subset of respondents. [b] Modules with missing grid questions for wave 7 were only administered over 6 years.

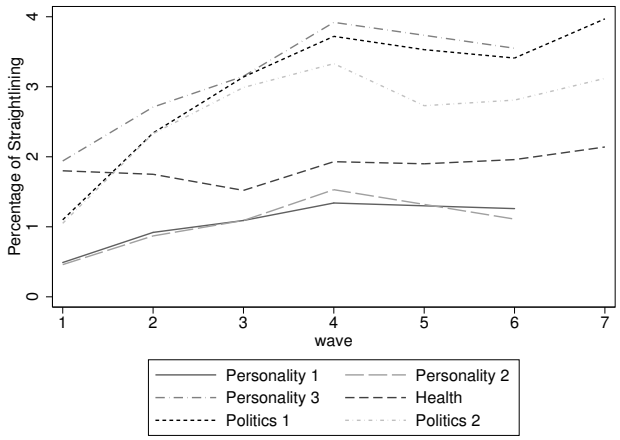

Figure 1. Percentage of straightlining over time for the questions in the group where straightlining was not plausible.

form of not providing an answer. In all other grid questions the percentage of missing values was negligible and is not further discussed here.

Next, we repeat these analyses on the same grid questions but use regression analysis to adjust for covariates and to establish statistical analyses.

Straightlining over time: Regression analyses. Table 3 and Table 4 contain GEE logistic regressions of an indicator of straightlining on study wave and other covariates. Table 3 contain the regressions related to implausible straightlining and Table 4 those to plausible straightlining. GEE regressions account for within-person correlation. We specified an exchangeable correlation structure and used robust standard errors. The regression includes indicator variables for study wave and socio-demographic covariates including immigration status. Because study wave is highly correlated with respondent experience, we did not include additional measures for respondent experience such as "number of surveys taken". We did not find meaningful interactions.

Adjusted for covariates, the regressions confirm the increase in straightlining over waves.[1] Straightlining significantly decreased with age. Male gender was associated with significantly increased straightlining in 6 out of 10 grid questions and a significant decrease in one grid question. Marital status "Divorced/Separated/Widowed" relative to "married" was associated with significant increased straightlining in out of 10 grid questions and a significant decrease in one question. Marital status "single" relative to "married" had no significant effect on straightlining in 8 of 10 grid questions. Relative to secondary school education, vocational and university education was associated with significantly decreased straightlining in the same 8 out of 10 grid questions. University educated respondents showed a significant increase in straightlining in the two remaining questions about religion.

4.2 Straightlining among evaluation questions

The prevalence of straightlining among evaluation questions is 6.6%. To investigate whether straightlining increases with panel tenure, we consider so-called bubble plots (Everitt, 1994) in Figure 2. A circle represents all respondents with the same value on the horizontal axis: number

[1] Straightlining also increases when respondent experience is measured as "number of survey taken" instead of "number of waves". Also see Section 4.3.

Table 3
Log odds ratios and standard errors from GEE logistic regression of indicators of implausible straightlining on study wave and other covariates.

	Personality 1		Personality 2		Personality 3		Politics 2		Politics 1		Health	
	b	S.E.	b	S.E.	b	S.E.	b	S.E.	b	S.E.	b	S.E.
Waves (*Ref.:* Wave 1)												
2	0.847***	0.24	0.862**	0.27	0.409***	0.12	0.861***	0.17	0.968***	0.17	0.017	0.15
3	1.013**	0.34	1.128**	0.36	0.549**	0.19	1.215***	0.17	1.322***	0.18	−0.059	0.15
4	1.311***	0.25	1.634***	0.27	0.938***	0.12	1.409***	0.17	1.544***	0.19	0.13	0.15
5	–	–	–	–	–	–	1.294***	0.18	1.652***	0.19	0.131	0.16
6	1.386***	0.26	1.484***	0.29	0.887***	0.12	1.353***	0.18	1.639***	0.18	0.24	0.14
7	_[a]	_[a]	_[a]	_[a]	_[a]	_[a]	1.500***	0.18	1.812***	0.18	0.317*	0.15
Age (*Ref.:* 18–35)												
35 to <50	−0.975***	0.24	−0.958***	0.24	−0.514***	0.13	−0.333*	0.13	−0.535***	0.12	−0.036	0.15
50 to <65	−1.638***	0.27	−1.826***	0.33	−1.069***	0.16	−1.146***	0.15	−0.958***	0.15	−0.252	0.16
65 and older	−2.227***	0.37	−2.385***	0.37	−1.600***	0.20	−1.829***	0.20	−1.509***	0.19	−0.441*	0.19
Gender (*Ref.:* Female)												
Male	0.461**	0.17	0.467**	0.18	0.246*	0.10	−0.332***	0.10	−0.013	0.10	−0.084	0.10
Marital Status (*Ref.:* Married)												
Single	−0.065	0.21	0.065	0.22	−0.137	0.13	0.209	0.12	0.187	0.12	0.053	0.13
Divorced/Sep/Wid	0.761**	0.25	0.569*	0.28	−0.13	0.17	0.424**	0.15	0.342*	0.15	0.420**	0.14
Immigration (*Ref.:* Dutch)												
1st gen (western)	0.639	0.46	0.355	0.54	0.189	0.31	0.157	0.27	−0.291	0.36	−0.035	0.28
1st gen (non-western)	1.141***	0.26	0.923**	0.32	0.647***	0.19	0.563**	0.19	0.955***	0.17	0.612**	0.19
2nd gen (western)	0.218	0.38	0.169	0.37	0.087	0.25	0.163	0.22	−0.321	0.29	0.157	0.23
2nd gen(non-western)	0.595	0.40	0.434	0.42	0.123	0.32	0.38	0.26	0.731***	0.20	0.631*	0.28
Education (*Ref.:* Secondary education)												
Primary	−0.519*	0.24	−0.329	0.24	0.157	0.13	0.064	0.14	0.328**	0.12	0.127	0.15
Vocational[c]	−0.920***	0.18	−0.912***	0.20	−0.534***	0.11	−0.572***	0.11	−0.642***	0.10	−0.382***	0.11
University	−3.240***	0.83	−2.058***	0.53	−1.559***	0.25	−1.423***	0.22	−1.404***	0.25	−0.822***	0.22
Constant	−4.653***	0.32	−4.808***	0.35	−3.278***	0.17	−3.983***	0.21	−4.181***	0.21	−3.842***	0.18
Observations	23,287		23,276		23,102		38,229		38,099		37,139	
Respondents	8,382		8,382		8,364		8,873		8,865		8,471	

[a] Some wave 5 and wave7 coefficients are missing because corresponding grid questions did not have a wave 5 or wave 7. questions did not have a wave 7.
[b] Secondary education is comprised of the Dutch school acronyms VMBO, HAVO, and VWO. [c] Vocational education is comprised of the Dutch school acronyms MBO and HBO.
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 4
Log odds ratios and standard errors from GEE logistic regression of indicators of implausible straightlining on study wave and other covariates.

	Leisure		Religion 1		Religion 2		Income	
	b	S.E.	b	S.E.	b	S.E.	b	S.E.
Waves (*Ref.:* Wave 1)								
2	0.074	0.04	0.043	0.03	0.024	0.03	0.026	0.04
3	0.191***	0.04	0.099***	0.03	0.089**	0.03	−0.01	0.04
4	0.358***	0.04	0.198***	0.03	0.147***	0.03	−0.006	0.04
5	0.501***	0.04	0.304***	0.03	0.264***	0.03	−0.122**	0.04
6	0.497***	0.05	0.337***	0.03	0.271***	0.03	−0.229***	0.04
7	0.534***	0.05	0.386***	0.03	0.312***	0.03		
Age (*Ref.:* 18–35)								
35 to < 50	0.312***	0.08	0.192***	0.05	−0.008	0.05	0.285***	0.05
50 to <65	0.585***	0.08	0.307***	0.05	−0.103	0.05	0.589***	0.05
65 and older	0.718***	0.09	0.301***	0.06	−0.350***	0.06	1.243***	0.06
Gender (*Ref.:* Female)								
Male	0.262***	0.04	0.807***	0.04	0.557***	0.04	0.068*	0.03
Marital Status (*Ref.:* Married)								
Single	−0.425***	0.07	0.056	0.05	0.231***	0.05	0.072	0.05
Divorced/Sep/Wid	0.054	0.06	−0.105*	0.05	0.009	0.05	0.217***	0.05
Immigration (*Ref.:* Dutch)								
1st gen (western)	0.234*	0.12	0.159	0.1	−0.184	0.11	−0.152	0.1
1st gen (non-western)	0.784***	0.1	0.350***	0.1	0.075	0.1	−0.184	0.1
2nd gen (western)	0.001	0.1	−0.022	0.09	−0.116	0.09	−0.215**	0.08
2nd gen(non-western)	0.017	0.18	0.589***	0.14	0.141	0.14	0.126	0.16
Education (*Ref.:* Secondary education)								
Primary	0.317***	0.06	−0.043	0.06	−0.062	0.06	−0.039	0.06
Vocational[b]	−0.523***	0.05	−0.071	0.04	−0.003	0.04	−0.161***	0.04
University	−1.512***	0.11	0.309***	0.07	0.249***	0.07	−0.300***	0.06
Constant	−2.080***	0.09	−1.429***	0.06	−1.329***	0.06	−0.543***	0.06
Observations	39,914		39,769		39,766		26,962	
Respondents	9,750		9,343		9,343		7,506	

[a] Secondary education is comprised of the Dutch school acronyms VMBO, HAVO, and VWO. [b] Vocational education is comprised of the Dutch school acronyms MBO and HBO.
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

of months in panel or number of previous surveys, respectively. The size of the circle is proportional to the number of respondents. We plot the percentage of straightlining over both the number of months in the panel (left) and the number of previous surveys (right). The percentage of straightlining increases for about 4 years until it reached almost one in ten respondents (about 8%) and then dips somewhat.

To see whether this trend is statistically significant even when adjusted for covariates, we regress an indicator of straightlining on respondent experience (number of waves) and socio-demographic variables. To improve interpretation of the model, we parameterized the number of surveys using indicator variables of groups (e.g. 5-9 previous waves) rather than as log(number of waves). Each observation represents a survey of an individual respondent. Responses for the same individual in different surveys may be correlated. Again, we account for this correlation by estimating the logistic regression using generalized estimating equations specifying an exchangeable correlation structure and using robust standard errors. The interpretation of coefficients, standard errors and p-values remains the same as in logistic regression.

The LISS panel sends out at least one survey to all panel respondents each month and sometimes sends additional surveys to the panel or subsets of the panel. We use data from the one monthly survey that is sent out to all respondent from April 2007, through September 2013. Evaluation questions were asked in 62 of the 78 months.

Table 5 shows the regression of the indicator variable of straightlining on indicator variables of number of previous waves in the panel and other covariates. The number of waves refers to the number of panel surveys a respondent has participated in. Because the relationship of number of previous waves and straightlining does not appear linear (Figure 2), we chose indicator variables for the number of previous waves rather than a single linear term. We did not find meaningful interactions. As previously seen in the graphs, straightlining increases over time. The odds of straightlining decreases with older age, and is lower for divorced /separated/ widowed respondents than for married respondents. Education and Gender are not related to straightlining with the exception of vocational education.

4.3 Robustness of Results

We found straightlining increases on a wide range of grid questions for at least three annual waves in the LISS panel. Straightlining among evaluation questions which are asked in most surveys continues to increase in the LISS panel even after respondents have answered dozens of surveys. The finding is robust with respect to the measurement of panel experience. We have also considered measuring panel experience as the number of months elapsed since a respondents' first survey for the evaluation questions (instead of number of previous surveys). The findings are consistent with those presented in the paper.

Specifically, for the evaluation questions straightlining also increases over "number of months in panel, the odds of straightlining decreases with older age, and are lower for divorced/separated/widowed respondents than for married respondents. Gender and education are not related to straightlining with the exception of a possible slight increase for primary education relative to secondary education. Unlike in the model presented in Table 5, straightlining does not increase as much for 1st generation non-western migrants (a mild interaction with time-in-panel reduces the main effect somewhat). There are too many regressions for the core questions to present in detail, but the overall findings when specifying "number of months" instead of "number of waves" are the same.

We have also considered "near straightlining" meaning straightlining with a single non-straightlining value. While straightlining increased with respondent experience, "near straightlining" remained roughly unchanged.

5 Discussion

Why does straightlining increase with panel experience? Possibly, as the novelty of being a panel member wears off the respondents gradually take more liberties with burdensome questions such as grid questions. Satisficing behavior might a more benign form of reacting to burden than breakoff and grid questions have been associated with higher breakoff rates (Peytchev, 2009).

We have distinguished between grids with plausible and implausible straightlining. Plausible straightlining occurs in grids where straightlining may be a reasonable answer for some respondents. As expected, far more straightlining occurred in grids with plausible straightlining. Couper et al. (2013) suggested implausible straightlining typically occurs with behavioral items whereas attitudinal items typically lead to plausible straightlining. All this suggests merely counting the number of grids in which straightlining occurred is too coarse a measure of data quality. Plausible straightlining is not evidence of poor data quality.

Zhang (2013) found that persistent speeding decreases as respondent get older and attributed this to older respondents being slower. We found straightlining among the evaluation questions also decreases with older age. Because persistent speeding and straightlining are correlated (Zhang, 2013), this suggests being slower may not be the only explanation for reduced persistent speeding in older age.

While it is important to diagnose increasing straight lining, are there any implications for panel management? One option is to try to reduce straightlining through immediate feedback giving the respondent the choice to change their answer (Zhang, 2013). This may be particularly appealing if the grid of question is of high importance. Straightlining might also be reduced by including, where possible, posi-

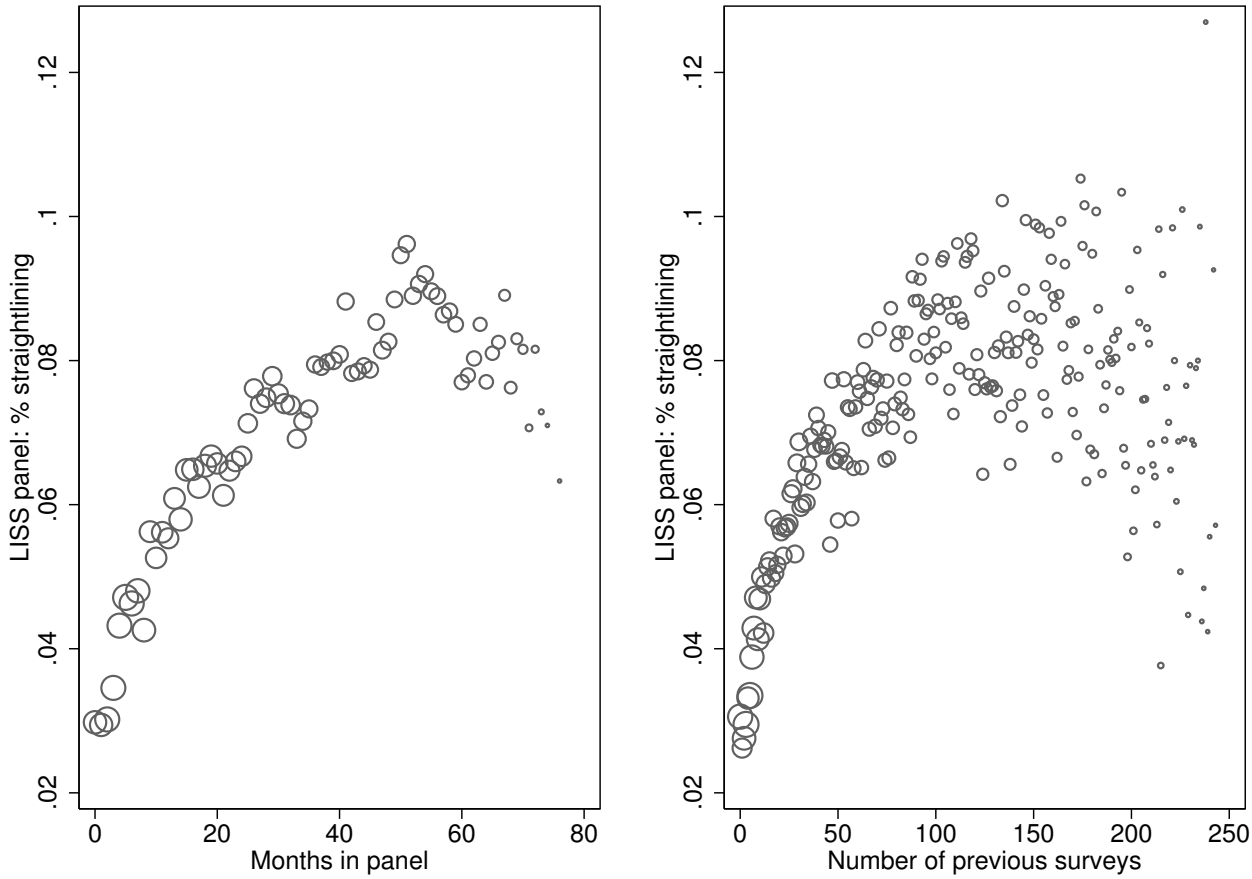

Figure 2. The percentage of straightlining as a function of respondent experience measured as months in panel (left) and number of previous surveys (right) in the LISS panel. The size of the circle is proportional to the number of respondents with the same value on the horizontal axis.

tively and negatively keyed questions in the grid that should not be answered in the same way. A second choice is to exclude surveys of respondents who straightline, possibly in conjunction with persistent speeding. Excluding straightlining respondents may reduce measurement error but increase non-response bias, especially given that straightlining is associated with reluctant respondents. A third choice is to report the percentage of implausible straightlining as one indicator of data quality among an array of quality indicators including the response rate. Of course, questionnaires without grid questions cannot report on straightlining. Unlike reporting a single value for the response rate, a report on straightlining would need to include, at a minimum the size of the grid. In general, however, reporting both response rates and straightlining as indicators of data quality would be a very welcome change, in part because it would reduce the overreliance on response rates and in part because they measure different aspects (potential for non-response bias vs non-differentiation) of data quality.

Finally, and perhaps most importantly, what should analysts do? We are not aware of any text book recommendations to cope with straightlining as part of the analysis, but we offer some thoughts on possible analysis strategies: First, conduct a sensitivity analysis with and without straightlined observations. Second, include an indicator variable for straightlining as an independent variable in the analysis. This enables a quantitative statement about the effect of straightlining on the outcome variable. If appropriate, include additional indicator variables for straightlining on specific values. Third, impute straight-lined values. Because straightlining represents zero variation, the primary goal of the imputation is to increase variation. The imputation model should include the straight-lined value as an independent (x-) variable because straightlining on a mid-level value is probably different than straightlining on an end-point. The imputation model might also include answers to the same grid in

past waves.

These analysis strategies are primarily designed to address implausible straightlining. Plausible straightlining is challenging as it is difficult to distinguish between best-effort responses and satisficing behavior. For plausible straightlining imputing straight-lined values seems conceptually less desirable. In this case the first two strategies seem more appropriate, however, if straightlining has a significant effect interpretation should proceed with caution.

6 Acknowledgement

This research was supported by the Social Sciences and Humanities Research Council (SSHRC, grant number 430-2013-0301). In this paper we make use of data of the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). We wish to thank Marcel Das and Joris Mulder in particular for their support with the data. We wish to thank the anonymous referees for their review.

References

Bethlehem, J. & Biffignandi, S. (2011). *Handbook of web surveys (Vol. 567)*. New York: John Wiley & Sons.

Binswanger, J., Schunk, D., & Toepoel, V. (2013). Panel conditioning in difficult attitudinal questions. *Public Opinion Quarterly*, *77*(3), 783–797. doi:10.1093/poq/nft030

Coen, T., Lorch, J., & Piekarski, L. (2005). *The effects of survey frequency on panelists' responses*. Paper presented at the ESOMAR Conference on Panel Research, Budapest, Hungary.

Cole, J., Bowers, A., McCormick, A., & Brummett-Carter, H. (2012). *Straight-lining and survey reluctance: prevalence and implications*. Paper presented at the American Association for Public Opinion Research. Retrieved from http://cpr.iub.edu/uploads/2012_AAPOR.pdf

Cole, J., McCormick, A. C., & Gonyea, R. M. (2012). *Respondent use of straight-lining as a response strategy in education survey research: prevalence and implications*. Paper presented at the nnual meeting of the American Educational Research Association, Vancouver.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The design of grids in web surveys. *Social Science Computer Review*, *31*(3), 322–345.

Das, M., Toepoel, V., & van Soest, A. (2011). Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociological Methods and Research*, *40*(1), 32–56.

Everitt, B. (1994). Exploring multivariate data graphically: a brief review with examples. *Journal of Applied Statistics*, *21*(3), 63–94.

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, *69*(3), 370–392.

Greszki, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: a data quality perspective* (pp. 238–262). Chichester, UK: Wiley.

Halpern-Manners, A., Warren, J. R., & Torche, F. (2014). Panel conditioning in a longitudinal study of illicit behaviors. *Public Opinion Quarterly*, *78*(3), 565–590. doi:10.1093/poq/nfu029

Hillygus, S., Jackson, N., & Young, M. (2014). Professional respondents in nonprobability online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: a data quality perspective* (pp. 219–237). Chichester, UK: Wiley.

Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions.* Paper presented at the Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, *74*(5), 956–984.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236.

Liang, K.-Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Lundmark, S. & Gilljam, M. (2013). *Participation effects in panel surveys: evidence from a seven-wave randomized experiment*. Paper presented at the American Association of Public Opinion Research (AAPOR), Boston.

Menold, N., Winker, P., Storfinger, N., & Bredl, S. (2014). Interviewers' falsifications in face-to-face surveys. In U. Engel, B. Jann, P. Lynn, S. A., & P. Sturgis (Eds.), *Improving survey methods: lessons from recent research* (pp. 86–97). New York: Routledge.

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, *73*(1), 74–97.

Revilla, M. & Ochoa, C. (2015). What are the links in a web survey among response time, quality, and auto-evaluation of the efforts done? *Social Science Computer Review*, *33*(1), 97–114. doi:10.1177/0894439314531214

Scherpenzeel, A. & Saris, W. (2007). Multitrait-multimethod models for longitudinal research. In J. Montfort K. Oud & A. Satorra (Eds.), *Longitidunal models in*

the behavioral and related sciences. Mahwah, NJ: Lawrence Erlbaum.

Smith, R. & Brown, H. H. (2006). *Panel and data quality: comparing metrics and assessing claims*. Paper presented at the ESOMAR World Research Conference: Panel Research, Barcelona, Spain.

Struminskaya, B. (2014). *Data quality in probability-based online panels.* Amsterdam: Off-Page.

Toepoel, V., Das, M., & Van Soest, A. (2008). Effects of design in web surveys comparing trained and fresh respondents. *Public Opinion Quarterly*, *72*(5), 985–1007.

Van Landeghem, B. (2012). *Panel conditioning and self-reported satisfaction: evidence from international panel data and repeated cross-sections*. SOEP Papers on Multidisciplinary Panel Data Research (Vol. 484), Berlin.

Van Ossenbruggen, R. (2008). Dutch online panel comparison study (NOPVO). In Market Research Association (Ed.), *Yearbook market research association* (pp. 63–78). Haarlem: Spaar en Hout.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1).

Zhang, C. (2013). *Satisficing in web surveys: implications for data quality and strategies for reduction (Ph.D.)* Ann Arbor, MI: University of Michigan. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/97990

Zhang, C. & Conrad, F. (2014). Speeding in web surveys: the tendency to answer very fast and its association with straightlining. *Survey Research Methods*, *8*(2), 127–135.

Appendix
Question text

This appendix contains the question text and the answer choices of the grid questions asked. The questions are presented as a bulleted list followed by the answer choices that apply to each of the questions. The questions were originally asked in Dutch. The English translation shown is available from the LISS webpage (http://www.lissdata.nl/). The numbering of the question (e.g. "Politics 1") is ours and is not part of the question text.

From the module on personality

Personality 1: Please use the rating scale below to describe how accurately each statement describes *you.*

- Am the life of the party.
- Feel little concern for others.
- Am always prepared.
- Get stressed out easily.
- Have a rich vocabulary.
- Don't talk a lot.
- Am interested in people.
- Leave my belongings around.
- Am relaxed most of the time.
- Have difficulty understanding abstract ideas.

1 very inaccurate; 2 moderately inaccurate; 3 neither inaccurate nor accurate; 4 moderately accurate; 5 very accurate

Personality 2: Please use the rating scale below to describe how accurately each statement describes *you*.

- Feel comfortable around people.
- Insult people.
- Pay attention to details.
- Worry about things.
- Have a vivid imagination.
- Keep in the background.
- Sympathize with others' feelings.
- Make a mess of things.
- Seldom feel blue.
- Am not interested in abstract ideas.

1 very inaccurate; 2 moderately inaccurate; 3 neither inaccurate nor accurate; 4 moderately accurate; 5 very accurate

Personality 3: For each of the following statements please indicate to what extent they apply to you. Please note that there are no right or wrong answers. (1 = strongly disagree, . . . , 7 = strongly agree)

- I would prefer complex to simple problems
- I like to have the responsibility of handling a situation that requires a lot of thinking
- Thinking is not my idea of fun
- I would rather do something that requires little thought than something that is sure to challenge my thinking abilities
- I try to anticipate and avoid situations where there is likely chance that I will have to think in depth about something
- I find satisfaction in deliberating hard and for long hours
- I only think as hard as I have to
- I prefer to think about small, daily projects to long-term ones
- I like tasks that require little thought once I've learned them

From the module on Politics and Values

Politics 1: What is your opinion on the following statements?

- It is good if society consists of people from different cultures.
- It is difficult for a foreigner to be accepted in the Netherlands while retaining his/her own culture.
- It should be made easier to obtain asylum in the Netherlands.
- Legally residing foreigners should be entitled to the same social security as Dutch citizens.
- There are too many people of foreign origin or descent in the Netherlands.
- People of foreign origin or descent are not accepted in the Netherlands.
- Some sectors of the economy can only continue to function because people of foreign origin or descent work there.
- It does not help a neighborhood if many people of foreign origin or descent move in.

1 fully disagree; 2 disagree; 3 neither agree nor disagree; 4 agree; 5 fully agree

Politics 2: Can you indicate, on a scale from 0 to 10, how much confidence you personally have in each of the following institutions?

- Dutch government
- Dutch parliament
- the legal system
- the police
- politicians
- political parties
- European Parliament
- United Nations
- the media
- the military
- the education system
- healthcare
- science
- the economy
- democracy
- shops/firms that you deal with personally (that you visit in person)
- shops/firms on the Internet

0 no confidence at all, . . . , 10 full confidence; 99 I don' t know

From the module on health

The following questions are about how you felt over the past month. For every question, please choose the answer that best describes how you felt during this past month. This past month . . .

- I felt very anxious
- I felt so down that nothing could cheer me up
- I felt calm and peaceful
- I felt depressed and gloomy
- I felt happy

1 never; 2 seldom; 3 sometimes; 4 often; 5 mostly; 6 continuously

From the module on social integration and leisure

Did you visit any one of the following performances or facilities over the past 12 months?

- a theatre performance
- a cabaret performance
- a concert of classical music
- an opera or operetta
- a concert of popular music, pop, jazz, musical or pop opera
- a 'dance' event, houseparty
- a ballet performance
- the cinema
- a film house
- an art gallery
- a museum

1 yes; 2 no

From the module on Religion and Ethnicity

Religion 1: Do you believe in . . .

- a life after death?
- the existence of heaven?
- purgatory?
- hell?
- a devil?

1 yes; 2 no; 3 maybe; 4 I don't know

Religion 2 Do you . . .

- believe Adam and Eve existed?
- see the Bible as the word of God?
- see the Koran as the word of God?
- see the Torah as the word of God?
- think that it makes sense of you to pray?

1 yes; 2 no; 3 maybe; 4 I don't know

From the module on "Economic Situation: Income"

We now wish to ask you whether certain events may possibly change your income over the next 12 months. Please indicate whether the event listed will happen (answer: yes), will not happen (answer: no), may possibly happen (answer: possibly) or whether the event is not applicable (answer: not applicable). Think about the next 12 months. Do you expect to see any changes to the total net income per month of your household due to the fact that a member of your household who is now working will stop working.

- A member of your household who is currently not working will start working.
- A member of your household will change to another job.
- A member of your household will gain career advancement.
- Social benefits received by your household will increase considerably.
- Social benefits received by your household will decrease considerably.

1 yes; 2 no; 3 possibly; 4 not applicable

Evaluation question:

- Was it difficult to answer the questions?
- Were the questions sufficiently clear?
- Did the questionnaire get you thinking about things?
- Was it an interesting subject?
- Did you enjoy answering the questions?

1 certainly not; 2; 3; 4; 5 certainly yes