sciendo

Journal of Official Statistics, Vol. 38, No. 4, 2022, pp. 1051–1067, http://dx.doi.org/10.2478/JOS-2022-0045

Relationship Between Past Survey Burden and Response Probability to a New Survey in a Probability-Based Online Panel

Haomiao Jin[1] *and Arie Kapteyn*[1]

We conducted an idiographic analysis to examine the effect of survey burden, measured by the length of the most recent questionnaire, or number of survey invitations (survey frequency) in a one-year period preceding a new survey, on the response probability to a new survey in a probability-based Internet panel. The individual response process was modeled by a latent Markov chain with questionnaire length and survey frequency as explanatory variables. The individual estimates were obtained using a Monte Carlo based method and then pooled to derive estimates of the overall relationships and to identify specific subgroups whose responses were more likely to be impacted by questionnaire length or survey frequency. The results show an overall positive relationship between questionnaire length and response probability, and no significant relationship between survey frequency and response probability. Further analysis showed that longer questionnaires were more likely to be associated with decreased response rates among racial/ethnic minorities and introverted participants. Frequent surveys were more likely to be associated with decreased response rates among participants with a large household. We discuss the implications for panel management and advocate targeted interventions for the small subgroups whose response probability may be negatively impacted by longer questionnaires or frequent surveys.

Key words: Survey nonresponse; survey burden; Internet panel; ideographic analysis; latent Markov chain.

1. Introduction

An online panel is a sample of persons who have agreed to complete surveys via the Internet. By tailoring key respondent burden variables like questionnaire length and survey frequency (Bradburn 1978), panel administrators can control the burden of taking surveys among panel participants. Based on common assumptions on the impacts of respondent burden (Bradburn 1978; Sharp and Frankel 1983; Baker et al. 2010), one may surmise that the experiences of long questionnaires and frequent surveys may overburden participants in panel studies and therefore decrease their propensity to complete a future survey. In this article, we aim to investigate the effects of the length of preceding surveys, or the frequency by which respondents are invited to respond to a new survey, on the probability that a respondent in a longitudinal probability-based Internet panel will respond to a new survey invitation.

[1] Center for Economic and Social Research, University of Southern California, 635 Downey Way VPD, Los Angeles, 90089 California, U.S.A. Emails: haomiaoj@usc.edu and kapteyn@usc.edu
Acknowledgments: The research in this article was supported by the U.S. National Institute on Aging under grant 5U01AG054580.

© Statistics Sweden

The context for the current article comes from the Understanding America Study (UAS), which is a U.S. nationally representative online panel of approximately 9,000 respondents aged 18 or older. The UAS was launched in 2014 and is supported by grants from several organizations, including the U.S. National Institute on Aging and the Social Security Administration. Panel members were randomly selected through address-based sampling and are compensated based on the estimated time spent on each survey, at a rate of USD 20 per 30 minutes. Interviews never exceed thirty minutes, but often may be shorter with incentive payments proportionately lower. Panel members without prior access to the Internet are provided with a tablet computer and Internet access. Since its launch, the UAS has conducted over 300 surveys of varying length on a broad number of topics including labor market behavior, income and wealth, mental and physical health, financial literacy, consumer behavior, personality, cognitive ability, health, and political views. Frequently, surveys may target specific groups. Hence, different subsamples of the UAS may receive different numbers of invites to do a survey. Details about the UAS can be found on the UAS website (UAS).

Although there exists an extensive literature on the effects of questionnaire length on response probabilities, upon closer inspection it is hard to draw firm conclusions. Most literature on online panels deals with cross-sectional surveys. Surprisingly, the way the length of surveys is defined is often ambiguous. In several cases, respondents are not told *ex ante* how long the survey may be expected to be, or at least it is not clear from the published studies whether respondents were informed about the length of the survey (Guo et al. 2016; Deutskens et al. 2004; Blumenberg et al. 2019). Koitsalu et al. (2018) find that a longer questionnaire leads to higher response rates, but in the announcement to potential respondents, everyone was told the same survey length (20 minutes). In other studies, survey length is defined after the fact, that is, by the realized length of the interview (Nicoletti and Peracchi 2005; Hill and Willis 2001). This makes it difficult to interpret the effect of interview length as a measure of burden. Potentially a longer interview may be an indicator of engagement, which would then predict higher future response rates, which indeed is often found.

The most direct evidence on the effect of survey length on response rates comes from randomized controlled trials that vary the announced length of surveys (Crawford et al. 2001; Galesic and Bosnjak 2009; Marcus et al. 2007). In these studies, there generally appears a clear negative effect of interview length on response rates. However, Crawford et al. (2001) report that although more respondents started the announced short survey, those respondents were also more likely to break off.

One might suspect that the negative effect of survey length on response rates would translate to longitudinal settings where preceding long surveys, may discourage respondents to accept a new survey invitation. Neither Hill and Willis (2001) for the Health and Retirement Study (HRS), nor (Zabel 1998) for the Survey of Income and Program Participation (SIPP) and Panel Study of Income Dynamics (PSID) find such a negative effect. In both studies, survey length is defined by the measured number of minutes the survey took, which once again may not be a measure of burden, but possibly of engagement. Zabel (1998) also takes advantage of a conscious decision by PSID staff in 1973 to reduce the length of the survey. It is found that this does reduce attrition. McCarthy et al. (2006) compare response rates to a number of agricultural surveys. Survey

durations are based on the U.S. Office of Management and Budget's estimates and hence are exogenous to the respondents. The authors report statistically marginally significant and quantitatively small negative effects of the length of preceding surveys on the likelihood that a new survey is returned.

The studies cited here sometimes provided incentives to increase response rates, but generally incentives were not tied to the length of an interview. Deutskens et al. (2004) varied the incentive in that the short questionnaire earned EUR 2 and the long questionnaire earned EUR 5. Whether that is an adequate compensation for the time spent on the survey is debatable, as the short survey already took 15–30 minutes.

The relationship between survey frequency and survey response has been studied much less, which is not surprising as most surveys are cross sectional and most longitudinal studies have a fixed periodicity, so that there is no variation in frequency that can be exploited. Zabel (1998) compares PSID and SIPP and finds that attrition is higher in the latter panel. During the observation period for his study, PSID respondents were interviewed once a year and SIPP respondents three times a year. He finds a higher attrition rate in SIPP. Although conceivably that might reflect the effect of the more frequent interview requests in SIPP, there are of course many other differences between the two panels, which might explain differences in attrition rates. McCarthy et al. (2006) exploit the fact that different agricultural operations have different probabilities of being contacted for a survey. They find only a negligible relation between frequency of survey invitations and response rates. A few studies have evaluated the relationship between survey frequency and survey response in ecological momentary assessments, where respondents may be prompted for brief surveys up to six times a day. The studies mostly suggest no relationship between survey frequency and completion rates (Eisele et al. 2020; Walsh and Brinker 2015; Stone et al. 2003; Ono et al. 2019).

Response rates are not just a function of questionnaire length and number of invites. In particular saliency of a survey topic is found to be important (Marcus et al. 2007). However, what is salient to one respondent may not be salient to another. Leverage-saliency theory proposed by Groves et al. (2000) speaks to this. Essentially, the likelihood that a survey invitation will be accepted depends on many factors, such as survey sponsorship, survey topic, survey length, incentive level, and unobserved personal characteristics. Although the survey topic is of course observable, how salient that topic is to a respondent is generally unknown *ex ante*.

In panel studies, respondents may receive larger incentives from taking longer or more frequent surveys if they are paid based on questionnaire length or time spent on taking surveys (Baker et al. 2010). Depending on how the incentives are evaluated against the other features of a survey, panel participants may exhibit diverse response behaviors to new surveys. Some participants may be motivated by the larger incentives from taking longer or more frequent surveys as a means to earn extra income. Other participants, though, may be less motivated by financial incentives to take more surveys. Such individual differences call for an idiographic (bottom up) analysis focusing on individual-specific relationships between questionnaire length and survey frequency and response propensity to a new survey.

To the best of our knowledge, there have been no published studies looking at the relationships between the length and frequency of past surveys and the response

propensity to a new survey in online panels. The present study examined the relationships between survey burden, measured by the length of the last questionnaire (defined exogenously) and the survey frequency (number of surveys in a one-year period preceding a new survey), and response to a new survey in the UAS. Specifically, the study conducted an idiographic analysis that examines individual-specific relationships between survey burden and probability of response to a new survey. Next, the different response patterns were related to observable demographic and socio-economic characteristics. Given the exploratory nature of the study, no specific hypotheses were postulated about what subgroups would be identified and which characteristics would differ between subgroups. The present study filled the knowledge gap about the impacts of experiencing longer questionnaires and frequent surveys on the response propensity to a new survey in online panels. The study also contributed to the methodology of analyzing individual survey response processes in panel studies by implementing a new dynamic model based on latent Markov Chain and Monte Carlo Markov Chain methods.

2. Methods

2.1. Overview of the Methods

The present study conducted an idiographic analysis for a total of 1,333 individuals randomly selected from all currently enrolled or former (attrited) UAS panel members. The analysis was based on the latent Markov chain method (Bartolucci et al. 2019), which is a flexible and powerful analytic approach to modelling the individual survey response processes, by incorporating unobserved variables like survey commitment, in addition to questionnaire length and survey frequency as covariates. Due to the computational burden of fitting the latent Markov chain model, we did not carry out the analysis for all current and past panel members. A Monte Carlo based method was used to estimate the associations between past survey burden and response to a new survey for each individual. The individual estimates were pooled to derive estimates of the overall relationships and to identify specific groups whose responses were mostly impacted by questionnaire length or survey frequency. Finally, an ordered logistic regression analysis was conducted to examine the relationships between the group memberships and sociodemographic characteristics and personality traits.

2.2. Measures

The survey response of each panel participant was modeled as the outcome of a latent Markov chain model including questionnaire length, survey frequency, and panel commitment as explanatory variables. Questionnaire length was measured by the number of screens viewed by the survey taker in the last survey they responded to. This is not a perfect proxy for questionnaire length. The UAS typically shows one question per screen, but it also uses grids, where respondents have to choose a category on each line. On the other hand, grid use is relatively rare and there is no reason to assume a systematic relation between the number of screens shown and the percent of those that would use grid questions (For example, the UAS does not try to reduce the number of screens by packing

more questions in a grid). Survey frequency was measured by the number of surveys to which a respondent was invited in the year preceding a survey.

Panel commitment was defined as an unobserved continuous variable indicating the feelings of obligation by panel participants to remain with the panel through taking surveys. It was defined as an unobserved variable because it is impossible to measure panel commitment at the time when there was a nonresponse. Survey response was coded as a binary variable with one indicating a survey had been completed and zero indicating the survey had not been completed.

Participants were grouped based on the relationships between questionnaire length, survey frequency and survey response. Group membership was modeled as the outcome of an ordered logistic regression model including socio-demographic characteristics and personality traits as predictors. Selection of the predictors was based on a systematic review of potential factors affecting response behavior in online surveys (Fan and Yan 2010). Socio-demographic variables included gender (coded as female (Yes/No)), age, racial/ethnical minority (not non-Hispanic White (Yes/No)), marital status (married (Yes/No)), education level (high school or less (Yes/No)), household income, household size, and the big five personality traits (McCrae 2004; McCrae and Costa 2008). We included personality traits in addition to commonly examined demographic variables since prior research has suggested certain personality traits like extroversion, conscientiousness, and openness to new experience are related to the probability of response to a subsequent wave in online panels (Cheng et al. 2020; Lugtig 2014). Data on the personality measures were obtained from the first survey every UAS participant takes when joining the panel.

2.3. *Modeling the Individual Survey Response Process*

The individual survey response process was modeled by a latent Markov model. Let panel participants be indexed by $i = 1, 2, ..., I$, and assume that in the past twelve months participant i was invited for a total of J_i surveys indexed by $j = 1, 2, ..., J_i$ in chronological order. Participant $i's$ completion of the jth invited survey, that is, $Y_{i,j}$, was related to survey length, survey frequency, and the unobserved panel commitment level through a logit link function as below.

$$Y_{i,j} \sim Bernoulli\left(\frac{e^{\eta_{i,j}}}{1+e^{\eta_{i,j}}}\right) \tag{1}$$

$$\eta_{i,j} = \theta_i + \lambda_i x_{i,j} + \beta_{i,1} S_{i,j} + \beta_{i,2} S_{i,j}^2 + \delta_{i,1} N_{i,j} + \delta_{i,2} N_{i,j}^2 \tag{2}$$

$\eta_{i,j}$ is the linear predictor. $x_{i,j}$ is the unobserved panel commitment level. $S_{i,j}$ is the number of screens viewed by the survey taker in the last survey. $N_{i,j}$ is the number of survey invitations in the preceding twelve months. Second-order terms $S_{i,j}^2$ and $N_{i,j}^2$ are included to model the potential non-linear relationships between questionnaire length, survey frequency, and survey response. The intercept θ_i, the impact of panel commitment on response λ_i, the impacts of questionnaire length on response $\beta_{i,1}$ and $\beta_{i,2}$, the impacts of survey frequency on response $\delta_{i,1}$ and $\delta_{i,2}$ are unknown coefficients to be estimated from data.

Dynamics of the unobserved panel commitment level $x_{i,j}$ was modeled by a latent Markov chain as below.

$$x_{i,j} = \left(\phi_{i,0} + \phi_{i,1}T_{i,j} + \phi_{i,2}T_{i,j}^2\right)x_{i,j-1} + \varepsilon_i \tag{3}$$

$T_{i,j}$ are the days between the $(j-1)$th and the jth survey invitation. The second-order term $T_{i,j}^2$ are included to model the potential nonlinear impact of the survey interval on the transition of panel commitment level. ε_i are an error term assumed to be normally distributed with mean zero and unknown variance. The impacts of survey interval on the transition of panel commitment $\phi_{i,1}$ and $\phi_{i,2}$ are unknown coefficients.

The unknown coefficients in Equations (2) and (3) were estimated in an idiographic way, that is, for each individual separately. Since the present article focuses on the relationships between questionnaire length, survey frequency, and survey response, we are most interested in the estimates of $\beta_{i,1}$, $\beta_{i,2}$, $\delta_{i,1}$ and $\delta_{i,2}$. The unknown coefficients were estimated by using the Particle Markov Chain Monte Carlo method (Andrieu et al. 2010). This method improves the standard MCMC method by incorporating particle filers to facilitate efficient Bayesian inference on latent Markov chain models (Andrieu et al. 2010). The prior of θ_i was set to be normally distributed with mean 2 and standard deviation 1, reflecting that most UAS participants have a high albeit varying initial response propensity. The prior distribution of λ_i was set to a normal distribution with mean 0.2 and standard deviation 0.1 to reflect the potential positive association between panel commitment level and survey response. The priors of $\beta_{i,1}$, $\beta_{i,2}$, $\delta_{i,1}$ and $\delta_{i,2}$ were set to be normally distributed with mean 0 and standard deviation 0.2. This setting was informed by the prevalence of no-association findings in prior relevant studies (Yetter and Capaccioli 2010; Blumenberg et al. 2019; Eisele et al. 2020; Walsh and Brinker 2015; Stone et al. 2003; Ono et al. 2019). Finally, the priors of $\phi_{i,1}$ and $\phi_{i,2}$ were set to be normally distributed with mean 0 and standard deviation 0.2, reflecting our experience that the survey interval has little impact on the change of panel commitment level. The estimation was implemented using the R package "nimble" (De Valpine et al. 2017).

2.4. *Pooling Individual Estimates, Identifying Subgroups Based on the Direction of Relationship, and Ordered Logistic Regression Analysis*

We used the means of the posterior distributions as individual estimates for $\beta_{i,1}$, $\beta_{i,2}$, $\delta_{i,1}$ and $\delta_{i,2}$ and then pooled those individual estimates to derive the overall estimates of the associations between questionnaire length, survey frequency on survey response.

We used a multi-step procedure to group the individual estimates based on whether the associations are likely to be positive, negative, or indicate no relationship. For the association between questionnaire length and survey response, we first grouped individuals into three subgroups based on the estimates of $\beta_{i,1}$ and $\beta_{i,2}$, respectively. Those subgroups are the positive association group (i.e., the estimate is larger than zero plus two times the standard error of the individual estimates), the negative association group (i.e., the estimate is smaller than zero minus two times the standard deviation), and the no-association group (i.e., the estimate is between the two thresholds). This subgrouping procedure is often used in the area of statistical quality control to identify

subjects that are most impacted by a factor (Oakland and Oakland 2018). After this calculation, each individual had two group memberships for $\beta_{i,1}$ and $\beta_{i,2}$, respectively. Our analysis showed that only two individuals were assigned to groups of opposite directions, that is, the positive association group for $\beta_{i,1}$ and negative association group for $\beta_{i,2}$, or the negative association group for $\beta_{i,1}$ and positive association group for $\beta_{i,2}$. We thereafter removed those cases from the subsequent analysis. For removed cases, the effect of more surveys was in one direction for small number of screens, and the effect of more surveys was in the opposite direction for large number of screens.

The remaining individuals were grouped into the ultimate positive association group if they were in the positive association group for either $\beta_{i,1}$ or $\beta_{i,2}$ or both. They were grouped into the ultimate negative association group if they were in the negative association group for either $\beta_{i,1}$ or $\beta_{i,2}$ or both. Finally, they were grouped into the ultimate no-association group if they were in the no-association groups for both $\beta_{i,1}$ and $\beta_{i,2}$ Similar procedures were applied to $\delta_{i,1}$ and $\delta_{i,2}$ to derive the group membership based on the association between survey frequency and survey response.

Finally, to examine the relationships between participant characteristics and group membership, we conducted an ordered logistic regression analysis with the group membership as an ordinal outcome variable (i.e., negative association $<$ no association $<$ positive association) and sociodemographic characteristics and personality traits as predictors. The personality traits were standardized before entering into the regression analysis. The model was fitted using the "polr" function from the R package "MASS" (Venables and Ripley 2013).

3. Results

A total of 1,333 individuals were randomly selected from all currently enrolled or attrited UAS panel members for the analysis. Individuals in the analysis sample were 58.2% female (n = 774), 35.5% racial/ethnic minority (n = 472), 53.9% married (n = 718), 24.3% with high school or lower education (n = 324), and 21.5% attrited (n = 286). The age at enrollment into the panel was on average 46.8 years old with a standard deviation of 16.2 years. Individuals in the analysis sample were invited for an average of 74.2 surveys with a standard deviation of 45.6. They had been in the UAS panel for an average of 35.5 months with a standard deviation of 21.4 months.

Pooling individual estimates from the latent Markov chain model showed an overall positive relationship between questionnaire length of the last survey, measured by the number of screens, and response to a new survey (estimate for the linear term = 0.043, 95% CI: 0.039 ~ 0.047, $P < 0.001$; estimate for the quadratic term = 0.066, 95% CI: 0.061 ~ 0.070, $P < 0.001$). There was no significant association between survey frequency, measured by the number of survey invitations in the preceding year, and response to a new survey (estimate of the linear term = 0.003, 95% CI: -0.001–0.006, P = 0.22; estimate of the quadratic term = -0.004; 95% CI: -0.009–0.001, P = 0.06).

As shown in Table 1, the multi-step grouping procedure identified a total of 414 individuals for whom seeing more screens in the last survey was associated with increased response propensity to a new survey. In contrast, seeing more screens in the last survey was associated with decreased response propensity to a new survey for 42 individuals. The

Table 1. Descriptive statistics and multiple-group comparisons based on the association between the number of screens in the last survey and responding to a new survey.

	More screens associated with decreased response (N = 42)	No association (N = 871)	More screens associated with increased response (N = 414)	P-value of test for equality across columns
Female	28 (66.7%)	495 (56.9%)	247 (59.8%)	0.32
Age	43.69 (16.88)	46.46 (16.03)	47.85 (16.35)	0.16
Racial/ethnic minority	20 (48.8%)	328 (37.7%)	121 (29.3%)	0.003
Married	22 (52.4%)	470 (54.0%)	222 (53.6%)	0.98
High school or less	10 (23.8%)	209 (24.0%)	105 (25.4%)	0.86
Household income				
• Less than USD 35,000	13 (31.0%)	219 (25.1%)	120 (29.0%)	0.28
• USD 35,000 – USD 59,999	11 (26.2%)	260 (29.9%)	106 (25.6%)	0.27
• USD 60,000 – USD 99,999	9 (21.4%)	184 (21.1%)	92 (22.2%)	0.90
• More than USD 100,000	9 (21.4%)	202 (23.2%)	93 (22.5%)	0.93
Three or more people in household	29 (69.0%)	392 (45.0%)	202 (48.8%)	0.006
Big five personality scores				
• Extroversion	26.53 (6.52)	25.99 (6.22)	27.17 (6.29)	0.007
• Agreeable	37.27 (5.04)	36.14 (5.50)	36.59 (5.65)	0.21
• Conscientiousness	36.17 (5.54)	36.12 (5.69)	36.17 (5.62)	0.99
• Neuroticism	22.41 (6.82)	21.97 (6.58)	21.33 (6.84)	0.23
• Openness	36.05 (6.05)	36.59 (6.06)	36.75 (6.45)	0.76

remaining 871 individuals were classified into the "no-association" group. Multi-group comparisons of sociodemographic variables and personality traits showed that the three groups were significantly different in racial/ethnic minority status, household size, and extroversion. The ordered logistic regression analysis (Table 2) revealed two significant predictors for group membership. For respondents who belong to a racial/ethnic minority, the odds of their response rates being negatively associated with more screens in the last survey was 1.56 times (i.e., 1/0.64) that of respondents who are non-Hispanic Whites, holding constant all other variables. For every one unit increase in respondents' standardized extroversion score, the odds of their response rates being not associated or positively associated with more screens in the last survey increases by 1.19, holding constant all other variables.

As shown in Table 3, having received more survey invitations in the preceding year was associated with increased response propensity to a new survey for a total of 126 individuals. Receiving more survey invitations in the preceding year was associated with decreased response propensity to a new survey for 152 individuals. The remaining 1,052 individuals were classified into the "no-association" group. Multi-group comparisons revealed significant differences by household size and by extroversion. The ordered logistic regression analysis (Table 4) revealed only one significant predictor of group membership. Respondents with three or more people in their household were 1.56 times (i.e., 1/0.64) more likely than respondents with smaller households to lower their response in reaction to being invited to more surveys, holding constant all other variables.

Table 2. Ordered logistic regression results for group membership based on the association between the number of screens in the last survey and responding to a new survey.

	Odds ratio	95% CI – lower bound	95% CI – upper bound	P-value
Female	1.11	0.86	1.43	0.41
Age group				
• 18–29		Reference		
• 30–44	1.04	0.72	1.51	0.84
• 45–64	1.17	0.81	1.70	0.42
• 65 and older	1.21	0.78	1.89	0.40
Racial/ethnic minority	0.64	0.49	0.82	< 0.001
Married	0.99	0.76	1.30	0.95
High school or less	1.12	0.84	1.50	0.43
Household income				
• Less than USD 35,000		Reference		
• USD 35,000 – USD 59,999	0.78	0.56	1.08	0.13
• USD 60,000 – USD 99,999	0.87	0.61	1.25	0.46
• More than USD 100,000	0.81	0.56	1.18	0.27
Three or more people in household	1.13	0.87	1.46	0.35
Big five personality scores				
• Extroversion	1.19	1.05	1.35	0.008
• Agreeable	1.00	0.87	1.15	0.99
• Conscientiousness	0.93	0.81	1.07	0.33
• Neuroticism	0.92	0.79	1.06	0.23
• Openness	1.00	0.89	1.14	0.95

Note: The ordered logistic regressions explain the probability of being in one of the three groups. A significant odds ratio below one means that one is more likely to belong to a group where more screens decrease the response probability; if the odds ratio is significantly above one, the demographic category is more likely to belong to the group for whom more screens imply a higher likelihood of responding.

4. Discussion

This study examined the relationships between past survey burden and response propensity to a new survey in the UAS. The results show an overall positive relationship between questionnaire length of the last survey, measured by the number of screens, and the probability of response to a new survey. This finding may be explained by the relatively generous incentives in UAS, which compensate participants at a rate of USD 20 per thirty minutes for estimated completion time. This rate is about 5.5 times the U.S. minimum wage and about twice the average hourly earnings of employees on private nonfarm payrolls (U.S. Bureau of Labor Statistics n.d.). Further analysis in this article identified subgroups of participants whose response to a new survey was not or negatively related to the length of the last survey. Specifically, the results suggest that racial/ethnical minorities were less likely to complete a new survey following a prior long questionnaire. Our finding adds to the existing evidence showing that minorities are less likely to respond to re-interviews (Broman et al. 1995; Chen and Kandel 1995; Griffin 2002) and less likely to complete surveys in panels (Harris-Kojetin and Tucker 1998). Strategies to identify and address the challenges of conducting surveys in hard-to-survey populations including racial/ethnic

Table 3. Descriptive statistics and multiple-group comparisons based on the association between the number of surveys in the one-year period preceding a new panel survey and responding to a new survey.

	Group membership based on the association between the number of surveys in the one-year period preceding a new survey and responding to a new survey			
	More preceding surveys associated with decreased response (N = 152)	No association (N = 1052)	More preceding surveys associated with increased response (N = 126)	P-value of test for equality across columns
Female	83 (55.0%)	619 (55.0%)	71 (56.3%)	0.60
Age	46.12 (15.29)	46.68 (16.39)	48.51 (15.54)	0.42
Racial/ethnic minority	49 (32.2%)	380 (36.2%)	40 (31.7%)	0.42
Married	87 (57.2%)	561 (53.3%)	68 (54.0%)	0.66
High school or less education	41 (27.0%)	250 (23.8%)	31 (24.6%)	0.68
Household income				
• Less than USD 35,000	38 (25.0%)	282 (26.8%)	32 (25.4%)	0.86
• USD 35,000 – USD 59,999	38 (25.0%	302 (28.7%)	39 (31.0%)	0.52
• USD 60,000 – USD 99,999	31 (20.4%)	227 (21.5%)	28 (22.2%)	0.93
• More than USD 100,000	44 (28.9%)	234 (22.2%)	26 (20.6%)	0.15
Three or more people in household	86 (56.6%)	491 (46.7%)	48 (38.1%)	0.008
Big five personality scores				
• Extroversion	27.59 (6.10)	26.18 (6.26)	26.67 (6.23)	0.03
• Agreeable	36.80 (5.66)	36.19 (5.49)	36.70 (5.72)	0.33
• Conscientiousness	36.87 (5.79)	35.98 (5.65)	36.51 (5.63)	0.15
• Neuroticism	20.87 (6.96)	21.98 (6.55)	21.23 (7.08)	0.10
• Openness	37.01 (6.05)	36.59 (6.25)	36.64 (5.78)	0.74

minorities were discussed in depth by Harkness et al. (2010) and Tourangeau et al. (2014). Specific solutions to address the challenges of surveying racial/ethnic minorities included specially designed sampling frames (Duffy et al. 2006; Stueve et al. 2001; Heckathorn 2007), adaptation and translation (Harkness et al. 2010; Schroepfer et al. 2009; Lee and Schwarz 2014), culture-sensitive survey design (Sun 2012; Pennell et al. 2004; Harkness et al. 2003), and pretesting (Canales et al. 1995). Our analysis found that extroverted participants were more likely to complete a new survey following a prior long questionnaire compared with introverted participants. Prior research has suggested that extraversion was positively associated with participation in probability-based online panels (Saßenroth 2013). The finding may be explained by the relationships between extraversion and reward processing, which suggests that extroverts are more sensitive to rewards, often show greater engagement with rewarding stimuli and situations, and are more motivated to gain a promised reward (Smillie 2013; Lucas et al. 2000; Fu 2013; Furnham et al. 1999).

Table 4. Ordered logistic regression analysis result for the group membership based on the association between the number of surveys in the one-year period preceding a new survey and responding to a new survey.

	Odds ratio	95% CI – lower bound	95% CI – upper bound	P-value
	1.08	0.80	1.44	0.62
Female				
Age group				
• 18–29		Reference		
• 30–44	1.27	0.83	1.95	0.27
• 45–64	1.28	0.83	1.96	0.26
• 65 and older	1.38	0.83	2.32	0.22
Racial/ethnic minority	1.05	0.78	1.42	0.76
Married	0.96	0.70	1.32	0.81
High school or less education	0.87	0.62	1.22	0.42
Household income				
• Less than USD 35,000		Reference		
• USD 35,000 – USD 59,999	1.08	0.74	1.57	0.70
• USD 60,000 – USD 99,999	0.99	0.65	1.51	0.96
• More than USD 100,000	0.80	0.51	1.23	0.30
Three or more people in household	0.64	0.47	0.86	0.004
Big five personality scores				
• Extroversion	0.92	0.79	1.07	0.29
• Agreeable	0.98	0.83	1.14	0.76
• Conscientiousness	0.98	0.83	1.14	0.77
• Neuroticism	1.03	0.87	1.22	0.77
• Openness	0.98	0.85	1.14	0.82

Note: The ordered logistic regressions explain the probability of being in one of the three groups. A significant odds ratio below one means that one is more likely to belong to a group where more survey invitations decrease the response probability; if the odds ratio is significantly above one, the demographic category is more likely to belong to the group for whom more survey invitations imply a higher likelihood of responding.

This study detected no significant relationships between survey frequency in the one-year period preceding a new survey and the response to the new survey. This finding is consistent with evidence from ecological momentary assessment studies (Eisele et al. 2020; Walsh and Brinker 2015; Stone et al. 2003; Ono et al. 2019). Further analysis identified small subgroups of participants whose response was negatively associated with survey frequency. Specifically, the response of participants with three or more members in their households was more likely to be negatively associated with survey frequency. One possible explanation is that participants with large households can experience higher respondent burden especially if surveys contain questions about their household members (Walsh and Fields 2012). As a result, participants with a large household may be more likely to be overburdened when the frequency of surveys is high.

The results from this study have several implications for the management of an Internet panel. First, the results suggest that longer surveys may not be a problem when compensation is high enough and positively related to questionnaire length. However, the results should not be interpreted as supporting the use of extraordinarily long questionnaires. Surveys in the UAS are limited to a maximum of thirty minutes, so that

the results cannot be generalized beyond the thirty-minute time frame. Second, the results suggest that targeted interventions are needed to improve response of participants from minority groups. Panel managers may consider specially designed sampling frames for minorities and introduce culture-sensitive customizations into survey and website design. Third, although our findings suggest that incentives may work well for extroverts, they may work less well for introverts. Research has suggested that while the extroverts are more easily motivated to seek more rewards, the introverts are motivated to avoid punishment (Furnham et al. 1999). Fourth, our analysis suggests that it may not be a problem to engage participants in frequent surveys. In the UAS, participants respond to an average of about two surveys per month. Fifth, our analysis suggests that targeted interventions may be needed to manage the burden of participants with large households. Panel managers may consider higher compensation or breaking down a survey into shorter pieces for participants with a large household when the survey contains a relatively large number of questions about other household members. Lastly, the estimated associations between questionnaire length/survey frequency and response may be informative for identifying panel members at high risk of attrition. Future research may use machine learning techniques to examine whether the estimated impacts are strong predictors of attrition.

It is worth noticing that this study has several limitations. First, results from the present study may not be generalizable to online panels that are not probability-based. Second, the analysis is based on observational data only. The above suggested panel management strategies would need to be tested in randomized controlled trials. Third, although we have tried to choose explanatory variables that can be considered exogenous (e.g., the number of survey invitations, rather than the number of surveys one has answered, or the number of screens shown in previous surveys rather that the total time taken to respond to a survey), these variables can be refined in several ways. For instance, there is a certain seasonality in survey burden in the UAS (the Fall tends to generate the largest number of surveys), so that rather than taking the total number of surveys in the previous year as an explanatory variable, we may also consider shorter time periods to detect these possible peak affects. Another aspect that has not been analyzed yet is the nature of the surveys. We know from the UAS "end-of-year" surveys that respondents vary in the type of surveys they like most (or least). We can in principle group surveys by topic and use these as extra explanatory variables, possibly interacted with respondent characteristics. Fourth, all racial/ethnic minority groups were combined together into one group in the present ordinal regression analysis due to the limited sample sizes of each minority group. This limited our ability to examine the differences between specific minority groups. Finally, we did not carry out the analysis for all current and past panel members due to the computational burden of fitting the latent Markov chain model. Future research can expand the analysis to all current and past panel members.

In conclusion, this study suggests that experiences of longer questionnaires and frequent surveys may not lead to a decreased response propensity to a new survey for the majority of participants in a large probability-based panel. The study advocates targeted interventions for the small subgroups of participants whose response may be negatively impacted by longer questionnaires and frequent surveys.

5. References

Andrieu, C., A. Doucet, and R. Holenstein. 2010. "Particle Markov Chain Monte Carlo Methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72 (3): 269–342. DOI: https://doi.org/10.1111/j.1467-9868.2009.00736.x.

Baker, R., S.J. Blumberg, J.M. Brick, M.P. Couper, M. Courtright, J.M. Dennis, D. Dillman, M.R. Frankel, and P.Garland. 2010. "AAPOR Report on Online Panels." *Public Opinion Quarterly* 74 (4): 711–781. DOI: https://doi.org/10.1093/poq/nfq048.

Bartolucci, F., A. Farcomeni, and F. Pennoni. 2019. *Latent Markov Models for Longitudinal Data*. Boca Raton, FL: Chapman and Hall/CRC.

Blumenberg, C., A.M.B. Menezes, H. Gonçalves, M.C. Formoso Assunção, F.C. Wehrmeister, F.C. Barros, and A.J.D. Barros. 2019. "The Role of Questionnaire Length and Reminders Frequency on Response Rates to a Web-Based Epidemiologic Study: A Randomised Trial." *International Journal of Social Research Methodology* 22 (6): 625–35. DOI: https://doi.org/10.1080/13645579.2019.1629755.

Bradburn, N. 1978. "Respondent Burden." In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 35: 35–40. Alexandria, VA: American Statistical Association. Available at: http://www.asasrms.org/Proceedings/y1978f.html (accessed August 2021).

Broman, C.L., V.L. Hamilton, W.S. Hoffman, and R. Mavaddat. 1995. "Race, Gender, and the Response to Stress: Autoworkers' Vulnerability to Long-Term Unemployment." *American Journal of Community Psychology* 23 (6): 813–842. DOI: https://doi.org/10.1007/BF02507017.

Canales, S., P.A. Ganz, and C.A. Coscarelli. 1995. "Translation and Validation of a Quality of Life Instrument for Hispanic American Cancer Patients: Methodological Considerations." *Quality of Life Research* 4 (1): 3–11. DOI: https://doi.org/10.1007/BF02507017.

Chen, K., and D.B. Kandel. 1995. "The Natural History of Drug Use from Adolescence to the Mid-Thirties in a General Population Sample." *American Journal of Public Health* 85 (1): 41–47. DOI: https://doi.org/10.2105/ajph.85.1.41.

Cheng, A., G. Zamarro, and B. Orriens. 2020. "Personality as a Predictor of Unit Nonresponse in an Internet Panel." *Sociological Methods & Research* 49 (3): 672–698. DOI: 10.1177/0049124117747305.

Crawford, S.D., M.P. Couper, and M.J. Lamias. 2001. "Web Surveys: Perceptions of Burden." *Social Science Computer Review* 19 (2): 146–162. DOI: https://doi.org/10.1177/089443930101900202.

Deutskens, E., K. de Ruyter, M. Wetzels, and P. Oosterveld. 2004. "Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study." *Marketing Letters* 15 (1): 21–36. DOI: https://doi.org/10.1023/B:MARK.0000021968.86465.00.

De Valpine, P., Daniel Turek, Christopher J. Paciorek, Clifford Anderson-Bergman, Duncan Temple Lang, and Rastislav Bodik. 2017. "Programming with Models: Writing Statistical Algorithms for General Model Structures with NIMBLE." *Journal of Computational and Graphical Statistics* 26 (2): 403–413. DOI: https://doi.org/10.1080/10618600.2016.1172487.

Duffy, D., J. Goldberg, and D. Buchwald. 2006. "Using Mail to Reach Patients Seen at an Urban Health Care Facility." *Journal of Health Care for the Poor and Underserved* 17 (3): 522–531. DOI: https://doi.org/10.1353/hpu.2006.0104.

Eisele, G., H. Vachon, G. Lafit, P. Kuppens, M. Houben, I. Myin-Germeys, and W. Viechtbauer. 2020. "The Effects of Sampling Frequency and Questionnaire Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population." *Assessment*. DOI: https://doi.org/10.1177/1073191120957102.

Fan, W., and Z. Yan. 2010. "Factors Affecting Response Rates of the Web Survey: A Systematic Review." *Computers in Human Behavior* 26 (2): 132–139. DOI: https://doi.org/10.1016/j.chb.2009.10.015.

Fu, Y. 2013. "On the Nature of Extraversion: Variation in Conditioned Contextual Activation of Dopamine-Facilitated Affective, Cognitive, and Motor Processes." *Frontiers in Human Neuroscience* 7: 288. DOI: https://doi.org/10.3389/fnhum.2013.00288.

Furnham, A., L. Forde, and K. Ferrari. 1999. "Personality and Work Motivation." *Personality and Individual Differences* 26 (6): 1035–1043. DOI: https://doi.org/10.1016/S0191-8869(98)00202-5.

Galesic, M., and M. Bosnjak. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73 (2): 349–360. DOI: https://doi.org/10.1093/poq/nfp031.

Griffin, D.H. 2002. "Measuring Survey Nonresponse by Race and Ethnicity." In Proceedings of the Annual Meetings of the American Statistical Association, August 2002: 11–15. Alexandria, VAL: American Statistical Association. Available at: https://nces.ed.gov/FCSM/pdf/IHSNG_asa02_DG.pdf (assessed October 2021)

Groves, R.M., E. Singer, and A. Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." *The Public Opinion Quarterly* 64 (3): 299–308. DOI: https://doi.org/10.1086/317990.

Guo, Y., J.A. Kopec, J. Cibere, L.C. Li, and C.H. Goldsmith. 2016. "Population Survey Features and Response Rates: A Randomized Experiment." *American Journal of Public Health* 106 (8): 1422–1426. DOI: https://doi.org/10.2105/AJPH.2016.303198.

Harkness, J.A., M.E. Braun, B.E. Edwards, T.P. Johnson, L.E. Lyberg, P.Ph. Mohler, B.-E.E. Pennell, and T.W. Smith. 2010. *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken, NJ: Wiley.

Harkness, J.A., F.J.R. van de Vijver, P.Ph. Mohler, and J. Wiley. 2003. *Cross-Cultural Survey Methods*. 325. Hoboken, NJ: Wiley.

Harris-Kojetin, B.A., and C. Tucker. 1998. "Longitudinal Nonresponse in the Current Population Survey (CPS)." In International Workshop on Household Survey Nonresponse, September 24, 1997: 263–272. Mannheim, Germany. Available at: https://www.ssoar.info/ssoar/bitstream/handle/document/49724/ssoar-1998-harris-kojetin_et_al-Longitudinal_nonresponse_in_the_Current.pdf?sequence = 1 = y = ssoar-1998-harris-kojetin_et_al-Longitudinal_nonresponse_in_the_Current.pdf (assessed October 2021).

Heckathorn, D.D. 2007. "6. Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment." *Sociological Methodology* 37 (1): 151–208. DOI: https://doi.org/10.1111/j.1467-9531.2007.00188.x.

Hill, D.H., and R.J. Willis. 2001. "Reducing Panel Attrition: A Search for Effective Policy Instruments." *Journal of Human Resources*: 416–38. DOI: https://doi.org/10.2307/3069625.

Koitsalu, M., M. Eklund, J. Adolfsson, H. Grönberg, and Y. Brandberg. 2018. "Effects of Pre-Notification, Invitation Length, Questionnaire Length and Reminder on Participation Rate: A Quasi-Randomised Controlled Trial." *BMC Medical Research Methodology* 18 (1): 1–5. DOI: https://doi.org/10.1186/s12874-017-0467-5.

Lee, S, and N. Schwarz. 2014. "Question Context and Priming Meaning of Health: Effect on Differences in Self-Rated Health between Hispanics and Non-Hispanic Whites." *American Journal of Public Health* 104 (1): 179–185. DOI: https://doi.org/10.2105/AJPH.2012.301055.

Lucas, R.E., E. Diener, A. Grob, E.M. Suh, and L. Shao. 2000. "Cross-Cultural Evidence for the Fundamental Features of Extraversion." *Journal of Personality and Social Psychology* 79 (3): 452. DOI: https://doi.org/10.1037//0022-3514.79.3.452.

Lugtig, P. 2014. "Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers." *Sociological Methods & Research* 43 (4): 699–723. DOI: https://doi.org/10.1177/0049124113520305.

Marcus, B., M. Bosnjak, S. Lindner, S. Pilischenko, and A. Schütz. 2007. "Compensating for Low Topic Interest and Long Surveys: A Field Experiment on Nonresponse in Web Surveys." *Social Science Computer Review* 25 (3): 372–383. DOI: https://doi.org/10.1177/0894439307297606.

McCarthy, J.S., D.G. Beckler, and S.M. Qualey. 2006. "An Analysis of the Relationship between Survey Burden and Nonresponse: If We Bother Them More, Are They Less Cooperative?" *Journal of Official Statistics* 22 (1): 97–112. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/an-analysis-of-the-relationship-between-survey-burden-and-nonresponse-if-we-bother-them-more-are-they-less-cooperative.pdf.

McCrae, R.R. 2004. "Human Nature and Culture: A Trait Perspective." *Journal of Research in Personality* 38 (1): 3–14. DOI: https://doi.org/10.1016/j.jrp. 2003.09.009.

McCrae, R.R., and P.T. Costa Jr. 2008. "The Five-Factor Theory of Personality." Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions: 91-104. Hoboken, NJ: Wiley.

Nicoletti, C., and F. Peracchi. 2005. "Survey Response and Survey Characteristics: Microlevel Evidence from the European Community Household Panel." *Journal of the Royal Statistical Society:* 168 (4): 763–781. DOI: https://doi.org/10.1111/j.1467-985X.2005.00369.x.

Oakland, R.J., and J.S. Oakland. 2018. *Statistical Process Control*. London, UK: Routledge.

Ono, M., S. Schneider, D.U. Junghaenel, and A.A. Stone. 2019. "What Affects the Completion of Ecological Momentary Assessments in Chronic Pain Research? An Individual Patient Data Meta-Analysis." *Journal of Medical Internet Research* 21 (2): e11398. DOI: https://doi.org/10.2196/11398.

Pennell, B.-E., A. Bowers, D. Carr, S. Chardoul, G.-Q. Cheung, K. Dinkelmann, N. Gebler, S.E. Hansen, S. Pennell, and M. Torres. 2004. "The Development and Implementation of the National Comorbidity Survey Replication, the National Survey of American Life, and the National Latino and Asian American Survey." *International Journal of Methods in Psychiatric Research* 13 (4): 241–269. DOI: https://doi.org/10.1002/mpr.180.

Saßenroth, D. 2013. "Personality Effects on Participation in the LISS Panel." In *The Impact of Personality on Participation Decisions in Surveys*, 129–160. Berlin, Germany: Springer.

Schroepfer, T.A., J. Matloub, P. Creswell, R. Strickland, and D.M. Anderson. 2009. "A Community-Specific Approach to Cancer Research in Indian Country." *Progress in Community Health Partnerships: Research, Education, and Action* 3 (4): 317. DOI: https://doi.org/10.1353/cpr.0.0096.

Sharp, L.M., and J. Frankel. 1983. "Respondent Burden: A Test of Some Common Assumptions." *Public Opinion Quarterly* 47 (1): 36–53. DOI: https://doi.org/10.1086/268765.

Smillie, L.D. 2013. "Extraversion and Reward Processing." *Current Directions in Psychological Science* 22 (3): 167–172. DOI: https://doi.org/10.1177/0963721412470133.

Stone, A.A., J.E. Broderick, J.E. Schwartz, S. Shiffman, L. Litcher-Kelly, and P. Calvanese. 2003. "Intensive Momentary Reporting of Pain with an Electronic Diary: Reactivity, Compliance, and Patient Satisfaction." *Pain* 104 (1–2): 343–351. DOI: https://doi.org/10.1016/s0304-3959(03)00040-x.

Stueve, A., L.N. O'Donnell, R. Duran, A. San Doval, and J. Blome. 2001. "Time-Space Sampling in Minority Communities: Results with Young Latino Men Who Have Sex with Men." *American Journal of Public Health* 91 (6): 922. DOI: https://doi.org/10.2105/ajph.91.6.922.

Sun, H. 2012. *Cross-Cultural Technology Design: Creating Culture-Sensitive Technology for Local Users.* Oxford, UK: Oxford University Press

Tourangeau, R., B. Edwards, T.P. Johnson, K.M. Wolter, and N. Bates. 2014. *Hard-to-Survey Populations*. Cambridge, UK: Cambridge University Press.

U.S. Bureau of Labor Statistics. n.d. *Average Hourly and Weekly Earnings of All Employees on Private Nonfarm Payrolls by Industry Sector, Seasonally Adjusted*. Available at: https://www.bls.gov/news.release/empsit.t19.htm (Assessed May 2021)

UAS (Understanding American Study). Available at: https://uasdata.usc.edu/index.php (accessed October 2022).

Venables, W.N., and B.D. Ripley. 2013. *Modern Applied Statistics with S-PLUS*. Berlin, Germany: Springer Science & Business Media.

Walsh, E., and J.K. Brinker. 2015. "Temporal Considerations for Self-Report Research Using Short Message Service." *Journal of Media Psychology* 28 (4): 200–206. DOI: https://doi.org/10.1027/1864-1105/a000161.

Walsh, R., and J.M. Fields. 2012. "Reducing Respondent Burden: Evaluating the Progress of the SIPP-EHC." In *Federal Committee on Statistical Methodology 2012 Biennial Conference. FCSM, Washington, DC*. Available at: https://nces.ed.gov/FCSM/pdf/-Walsh_2012FCSM_IX-B.pdf (Assessed October 2021).

Yetter, G., and K. Capaccioli. 2010. "Differences in Responses to Web and Paper Surveys among School Professionals." *Behavior Research Methods* 42 (1): 266–272. DOI: https://doi.org/10.3758/BRM.42.1.266.

Zabel, J.E. 1998. "An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior." *Journal of Human Resources*: 479–506. DOI: https://doi.org/10.2307/146438.

Received January 2021
Revised June 2021
Accepted October 2021