Panel Conditioning and Attrition in the AP-Yahoo! News Election Panel Study

Yelena Kruse[1], Mario Callegaro[2], J. Michael Dennis[3], Charles DiSogra[4]
Stefan Subias[5], Michael Lawrence[6], Trevor Tompson[7]
[1,2,3,4,5,6] Knowledge Networks, 1350 Willow Road, Ste 102, Menlo Park, CA 94025
[7]Associated Press, 1100 13th Street NW, Ste 700, Washington, DC 20005

Abstract
This paper examines the potential for attrition bias and panel conditioning in a longitudinal online election panel survey. The 2008 Associated Press-Yahoo! News Poll was conducted by Knowledge Networks with contributions from political scientists at Harvard University and Stanford University. The study involved an eleven-wave Web panel election survey of general population U.S. adults. All the interviews were conducted with KnowledgePanel® respondents. Our assessment of panel conditioning is made possible by the study's sample design, which includes both the longitudinal sample component as well as three separate fresh cross-sectional samples. We examined the potential for the impact of panel conditioning on self-reports of certain attitudes, preferences, and behaviors such as the propensity to remain undecided and to report being certain about voting or having voted early. We employ Extended Cox hazard modeling to estimate risk factors responsible for attrition from the panel study. We found some evidence of panel conditioning for one political knowledge question. Not surprisingly, longitudinal respondents were more likely to correctly name Obama's religion than cross-sectional respondents, confirming previous results from the literature for panel conditioning on knowledge questions. For the other seven items about the presidential election, only two showed some evidence of panel conditioning. Regarding panel attrition, the rare (i.e., non-whites, adults ages 18 to 29, or less than high school education) and non-rare respondent groups were attriting from the panel at the same rate during the first four waves, probably due to a specific incentive system put in place for the rare respondents. Undecided Republicans were more likely to quit the panel during the first four waves than Republicans who chose John McCain. In subsequent waves, late-participating respondents were more likely to drop from the study than early respondents in both rare and non-rare groups.

Key Words: election, Internet panel, KnowledgePanel, attrition, conditioning

1. Introduction

Online election studies are becoming more popular among researchers. There are three main reasons for it: cost, speed and experimentation (Clarke, Sanders, Stewart, & Whiteley, 2008). This means more flexibility in changing the questionnaire to adapt it to the current political situation (e.g., when a candidate drops out from the race). Online surveys allow more possibilities to run experiments and randomize subjects to conditions.

1.1 Literature review

1.1.1 Panel conditioning

Panel conditioning, also called time-in-sample bias (Kalton & Citro, 1993), time-in-survey effect, repeated measurement effect, interview effect, panel effect (Cantor, 1989) or reactivity in panel studies (Van Der Zouwen & Van Tilburg, 2001) is "observed in

repeated surveys when a sample unit's response is influenced by prior interviews or contacts" (Cantwell, 2008, p. 556). For space reasons we cannot review the entire literature on panel conditioning. We refer the reader to summaries by Nancarrow and Cartwright (2007), by Sturgis, Allum and Brunton-Smith (2009) and by Van der Zouwen and Van Tilburg (2001). Instead, we concentrate on studies about panel conditioning in election studies, attitudinal responses and online panel conditioning.

Panel conditioning in election studies on attitudinal responses

Early studies on conditioning in election surveys were focused on a single behavior and studied the following hypothesis: Does being interviewed before the election increase the respondent's likelihood to vote in comparison to somebody who is not interviewed? Although not always based on experimental data, and with some other methodological problems (Sturgis, 2002; Traugott & Katosh, 1979), many U.S.-based studies supported the above hypothesis (Clausen, 1968; Kraut & McConahay, 1973; Traugott & Katosh, 1979; Yalch, 1979). A vote validation study conducted in Sweden (a country with a historically high turnout rate) came to the same conclusion, showing that the increase in likelihood to vote due to being interviewed was strong for respondents with a low interest in politics but made almost no difference for respondents with a high interest in politics (Granberg & Holmberg, 1992). In a study comparing American National Election Study panel respondents to a fresh cross-sectional sample, Bartels (1999) concludes that panel conditioning is really minor in general if not for two variables. "The estimated relative value of panel data in my analysis of campaign interest and turnout ... implies an appropriate discount rate of about 40% by comparison with fresh cross-section data" (p.15).

Waterton & Lievesley tested six different hypotheses of panel conditioning on three waves of the British Attitudes Panel Survey compared to a fresh cross-sectional sample. Three of the hypotheses found support: respondents in the panel became less likely to answer "don't know", and became more honest but more "politicized." Effect of panel conditioning for attitudinal variables has also been found for different attitudes (Sturgis et al., 2009; Veroff, Hatchett, & Douvan, 1992; Waterton & Lievesley, 1989), but not always (Wang, Cantor, & Safir, 2000).

Panel conditioning in online panels

To test for the hypothesis of panel conditioning for political attitudes and behaviors, Clinton (2001) compares respondents of four KnowledgePanel cohorts recruited at different points in time with a fresh, just recruited cohort. The author finds little and non-systematic evidence for panel conditioning. Two other studies focused on panel conditioning effect in a cross-sectional fashion on KnowledgePanel. Dennis (2001) compared answers to the same survey questions by different groups created by panel tenure. He found no conditioning effect for awareness of alcoholic drinks, attitudes towards new products, investment and financial services. Analyzing five more sensitive questions, Dennis found some evidence of panel conditioning in one of the items: members with longer panel tenure had a lower level of comfort with a shop owner with AIDS than members with six months or less of panel tenure (51 vs. 57%). Similar results were found in a later study by Pineau, Nukulkij and Tang (2005). Panel conditioning was assessed using membership tenure (number of months in the panel) and number of same topic surveys taken in a study by Nukulkij, Hadfield, Subias, & Lewis (2007). Panel tenure was never statistically significant; the number of foreign policy surveys completed was significant in only one of the six analyses: the emphasis on military or democracy. Members with a higher number of foreign policy surveys taken were more likely

(OR=1.12) to agree that the current administration should put more emphasis on diplomacy.

Das, Toepel and van Soest (2007) compared several questions answered in a two-wave longitudinal design, to answers from a refreshment sample in the Dutch probability-based online panel, CenterData. Taking attrition into account, the authors found evidence for conditioning for knowledge questions but not for questions about attitudes, actual behaviors or expectations concerning the future. Similar results were found in another study by the same authors (Toepoel, Das, & van Soest, 2009).

Lastly, we found only one study on panel conditioning using an opt-in panel. The authors, Nancarrow & Cartwright (2007), used a longitudinal design to test panel conditioning effects; three experimental groups were compared among each other: a high frequency group where the same interview was administered five times four weeks apart, a medium frequency group (three interviews with an eight week interval) and a low frequency group with two interviews only, 16 weeks apart. A control group was also part of the design and was interviewed at the same time of the last wave of the other three groups. In order to qualify for the survey, the respondent had to claim to use or buy toothpaste regularly. Using weighted data to control for attrition the authors found conditioning effects mostly for the high frequency (HF) group. The effects were going into two directions. In some cases, e.g. using toothpaste regularly and recall of a brand (ad shown at the end of each interview), the actual percentages were lower in the HF group. The authors explain this phenomenon with boredom or fatigue in taking the same interview five times. In other cases, e.g., when a dummy brand was placed in a list to measure its awareness, the percentage of respondents recognizing it increased as the number of surveys increased: a clear example of panel conditioning due to learning effect.

1.1.2 Panel attrition

Definition of attrition in online panels

Attrition in online panels depends on three factors (Callegaro & DiSogra, 2008):

- Voluntary attrition by the panel member. The panel member decides not to participate anymore in the online panel.
- Involuntary attrition. The panel, depending on some rules, decides to drop some members.
- Mortality

When a probability-based online panel is used in a cross-sectional fashion, attrition is a problem in the sense that it affects recruitment cost (Tortora, 2009). The more people attrit, the more recruitment efforts need to be done in order to maintain the panel size at the same level that is estimated to support multiple studies at the same time. If we are talking about a probability based panel, attrition has a higher cost, because recruitment is much more expensive than for opt-in panels (e.g. recruitment via RDD vs. having volunteers enrolling online in the panel).

Previous research on attrition for online panels

Pioneering research on online panels, specifically on the precursor of probability-based online panels, the Telepanel, has been conducted in the Netherlands. The Telepanel, originated in the 1980s, was a nationally representative panel of Dutch households that were given a computer and a modem to participate in weekly online surveys. After 100 weeks from the initial start of Telepanel, only 50% of the original households were still

members. Three variables were identified as correlated with attrition: type of research, number of questions per study, and frequency of data collection (Saris, 1998). Felix and Sikkel (1996) analyzed four classes of psychological variables to investigate possible attrition bias: general (e.g. need for cognition), values, personality and political involvement. The analysis was done correlating duration of panel membership with the aforementioned scales. Their conclusion was that there was hardly any relationship between panel tenure and these psychological variables: for example, members who stayed longer were feeling no lonelier than members with lower tenure. The maximum possible stay in the panel was seven years (Sikkel & Hoogendoorn, 2008). The authors also analyzed demographic variables, finding age as a strong predictor of attrition: younger people dropped off sooner than older people. After controlling for age, other predictors were income and political voting behaviors. The authors found lower associations with variables like living with a partner, urbanization, region, and degree of rurality. In another study of the Telepanel, response burden – measured in number of times a budget survey was administered to panel members – was correlated with attrition (Hoogendoorn & Sikkel, 1998, p. 199). Lastly, an analysis of the CentERpanel, successor of the Telepanel and now via the Internet, shows that education was one of the strongest predictors of panel tenure (Sikkel & Hoogendoorn, 2008).

Two attrition studies were conducted on the Gallup Poll Panel. The GPP is not fully an Internet panel, because non-Internet households or low Internet users are surveyed via mail or occasionally by phone and IVR. It is however a probability-based panel recruited via RDD (Tortora, 2009). Approximately 48% of panel members are surveyed online (Rookey, Hanway, & Dillman, 2008). In the first study Sayles and Arens (2007) analyzed attrition in the GPP from July 2004 to March 2006. In univariate analyses, attrition in the GPP looked very similar to attrition in traditional household panels with some exceptions: older respondents, white and female were less likely to drop out. Non-attritors were also more likely to be home owners, had lived longer in their current location, were married and had fewer children. Lastly, non-attritors had higher education and were more likely to be retired. Urbanicity did not have an effect. The next set of analyses focused on investigating the association between survey burden and attrition. The authors found that a higher burden (number of surveys assigned per month) was associated with higher attrition, controlling for everything else. They also found that the more surveys completed, the less likely that panel members were to drop out.

Tortora (2009) focused his analysis on the period from September 2004 to November 2005. The results from a demographics analysis mimicked Sayles and Arens (2007) work with the additional finding that teens (13-17) showed the lowest attrition rates than any adult group. The author explains this with the low frequency of surveys that teen received, therefore being in a group with lower burden than any other group. Interesting findings were given when focusing on the relationship between attrition rate and number of survey requests. The relationship is definitely non-linear: it starts with a very high rate of attrition (70%) among respondents who receive two survey requests and then declines sharply until it reaches 11 requests. It reaches 63% among those with 14 requests to decline sharply again. Mode of data collection also makes a difference. In an experiment where panel members were randomly assigned to receive the first survey of each month by telephone, and the following self-administered (or self-administered only), the attrition rate for the first group was 29.9% in comparison to 24.1% in the second group. Lastly, Tortora found a relationship between survey topic and attrition. By giving a score on survey topics as being poll-like or market research, the author found that attrition rate

was lower for panel members who completed relatively more poll-like or social surveys than market research surveys.

In two attrition analyses done when KnowledgePanel was just established (1999), attrition effects were found to be very modest in nature. Clinton (2001) concentrates on six cohorts recruited from January 2000 to June 2000 showing a slightly higher attrition for respondents age 55 and older and less than a high school education. On other demographic variables the differences were very minor. Dennis (2001) uses panel tenure from 0 to 12 months showing very minor changes in demographics as tenure increases for variables such as gender; age (18 to 34 years, 35 to 54 years); education (less than high school, have a BA or more); and having an income less or above $40,000. A possible explanation is the novelty effect. At that time respondents in the cohort analyzed were all given an MSN WebTV and provided with free Internet access for the duration of panel membership. The motivation to stay and the free use of Internet as an incentive are very likely to have played a big role in retention during the year 2000, when Internet penetration was very low at a household level or 41.5%. The higher attrition of respondents 55 and older noted by Clinton (2001) can be explained by resistance to new technology of older people. In a more recent attrition analysis DiSogra and colleagues (2007) analyzed 15 months of panel tenure using survival analysis. Their findings show higher attrition rates for younger people, lower income, non-whites, people with children in their household, and less educated respondents.

2. Data and Methods

2.1 Study description

Knowledge Networks conducted an eleven-wave longitudinal election survey on behalf of the Associated Press (AP) from November 2007 to December 2008. The wave 1 survey (baseline) was fielded to a sample of 3548 panel members 18 years of age or older who represented a general population sample. In addition to the longitudinal samples, three cross-sectional samples were fielded at wave 3 (681 completed), wave 6 (548 completed), and wave 9 (453 completed). 2279 baseline respondents completed wave 3, 1748 responded in wave 6, and 1674 completed wave 9. 1086 baseline respondents participated in all eleven waves of data collection.

Table 1. Research design

Waves / Samples	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11
Longitudinal	X	X	X	X	X	X	X	X	X	X	X
Cross-sectional			X			X			X		

Whether an incentive was given to the longitudinal sample respondents, and the amount of incentive they received, varied by wave and respondent group. In waves 1–4, only those in rare population groups (younger than 30 years old, less than high school education, or non-white) were given an incentive. Starting at wave 5 the incentives were changed to also include number of days before response, based on previous waves, and a date to complete the survey by in order to receive the incentive. This was used to improve the initial sample sizes for the AP-Yahoo! news stories. Early respondents were defined as respondents from the longitudinal sample who completed all of the first four waves in seven days or less after fielding; those who did not were categorized as late respondents.

In this paper we will analyze the first nine waves of data collection. Waves 1–4 will be examined separately from wave 5–9 due to the difference in incentives: in the first four waves two groups (rare respondents vs. non-rare respondents) received different amounts; starting in wave 5 there were three different incentive amounts depending on the respondent group.

2.2. Panel conditioning analysis

2.2.1 Comparing longitudinal and cross-sectional samples

To compare longitudinal and cross-sectional samples, we used an approach in which logistic regression analyses in SAS LOGISTIC were performed on respondents' answers as dependent variables, demographics as covariates, and number of surveys taken and the type of sample as the independent variables of interest. While controlling for potential covariates we compared responses in three waves of the cross-sectional samples with responses of the longitudinal samples on a number of attitudinal and behavioral questions and a knowledge question. Each time, the analysis was performed on the weighted longitudinal sample and the weighted cross-sectional sample using post-stratification weights that incorporated the probabilities of selection to account for panel attrition.

The following independent variables were used in the models:

- Gender (Male as a reference group)
- Age (In years)
- Education (Less than high school as a reference group, high school diploma, some college, bachelor's degree or higher)
- Race/Ethnicity (White as a reference group, African-American, Hispanic, other racial descent)
- Income
- Party affiliation (Republican as a reference group, Independent/Other/None, Democrat)
- Evangelical beliefs (No/Yes)
- Number of previous KN surveys completed
- Sample (Longitudinal/Cross-sectional)

The following dependent variables were used in the models:

- Certainty to vote in the upcoming presidential election (waves 3 and 6)
- Voting early in the presidential election (wave 9)
- Correct answer about Obama's religion (wave 6)
- Feeling excited about the upcoming presidential election (wave 3 and 9)
- Feeling interested in the upcoming presidential election (wave 3 and 9)
- Feeling hopeful about the upcoming presidential election (wave 3 and 9)
- Feeling bored with the upcoming presidential election (wave 3 and 9)
- Feeling frustrated with the upcoming presidential election (wave 3 and 9)

Table 2 reports only the odds ratios and their associated p values for the dichotomous variable cross-sectional/longitudinal, where longitudinal is the reference group. The complete tables are available from the authors.

Table 2. Probability values and odds ratios for predicting attitudes and behaviors for cross-sectional sample (longitudinal is a reference group)

Variable of interest	Wave 3		Wave 6		Wave 9	
	p	OR	p	OR	p	OR
Certainty to vote	0.150	0.854	0.694	0.951		
Voting early					0.575	0.930
Correct answer for Obama's religion			**<0.01**	**0.719**		
Feeling excited about the election	0.841	1.024			0.169	0.823
Feeling interested in the election	0.241	0.898			0.238	1.147
Feeling hopeful about the election	**<0.05**	**0.834**			0.879	1.018
Feeling bored with the election	0.238	0.851			**<0.001**	**0.444**
Feeling frustrated with the election	0.206	0.880			0.188	0.855

Only three of the 24 odds ratios were statistically significant; more specifically:

Providing correct answer about Obama's religion in wave 6
In waves 4 and 6, longitudinal respondents were asked about the religion of presidential candidates. Their responses were compared to the wave 6 responses of the cross-sectional sample. The answer "Protestant" was coded as 1 (correct), and any other answers were coded as 0. Based on the results of the direct logistic regression (a test of the full model with all predictors against a constant-only model was statistically significant with χ^2 = 285.72, n=2296, p<.001, but Nagelkerke R^2 =.16), sample origin was a significant predictor (p<.01) of answering the question about Obama's religion correctly. Controlling for other variables, cross-sectional sample respondents were 28% less likely to answer the question correctly, than the longitudinal sample respondents (Table 2). In addition, those with higher number of previous KN surveys completed were more likely to provide the correct response (with 10 additional completed KN surveys, the likelihood of providing the correct answer increased by 1%).

Feeling of hope in wave 3
According to the wave 3 model (χ^2 = 186.74, n=2960, p<.001, Nagelkerke R^2 = .08), cross-sectional respondents in wave 3 were 17% less likely to report being hopeful about the presidential election than the longitudinal sample respondents, controlling for other variables. In wave 9 this effect disappeared.

Feeling of boredom in wave 9
In wave 3, there was no significant effect of the sample origin on reporting being bored with the presidential election; however, in wave 9 the effect of sample origin became significant (p<.001, Table 2). Controlling for other variables, cross-sectional sample respondents were 56% less likely to report being bored with the presidential election than longitudinal sample respondents.

2.2.2 Comparing responses to the same questions in the consecutive waves for the longitudinal sample
The methods of comparing the responses of longitudinal and cross-sectional samples have been criticized in some literature (Sturgis et al., 2009). Instead it has been proposed to examine the answers for the same respondents over time. In particular, Sturgis et al. tested four different measures to examine panel conditioning. First, they studied the reliability of attitudinal scales between the first and later waves using a confirmatory factor model; second, they examined the stability of attitude items by comparing

correlations between the summed scales in adjacent waves; third, they assessed the degree of opinionation by comparing proportions of "don't know" and "not sure" responses across waves; and fourth, Sturgis et al. compared self-reported interest in politics across waves. Because we had neither validated scales nor "interest in politics" questions in our longitudinal study, and we offered explicit "don't know" responses on a very small number of items, we could only test one hypothesis: *The stability of attitude items will increase from the first to subsequent waves* (Sturgis et al., 2009). We selected four items for this analysis that we believed to be less sensitive to change over time. Pearson's correlations were run on the responses to each of these items in subsequent waves, and the Haan *t*-test for significant trend was performed. It is worth noting that none of these items were asked in wave 6 and wave 8; therefore the time lag between the responses was longer toward the end of the data collection than in the beginning.

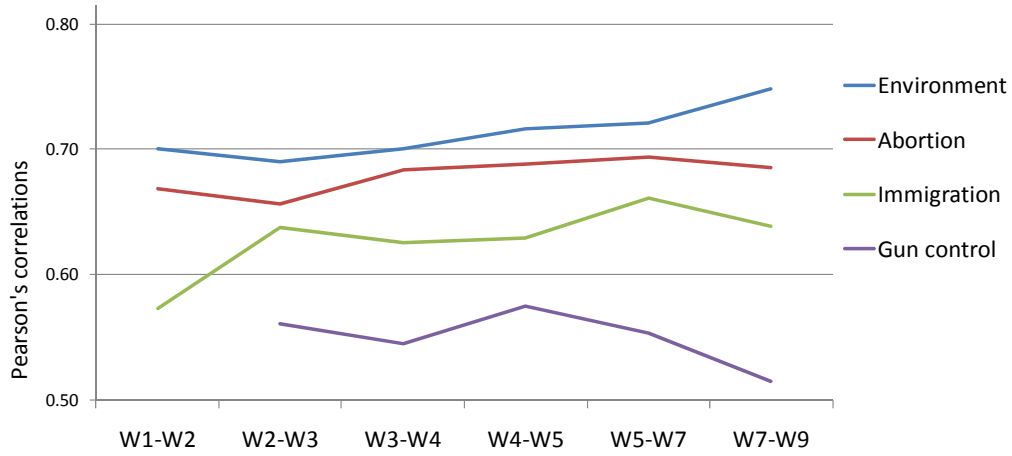

Figure 1. Interwave Stabilities (Wave 1 – Wave 9).

By looking at Figure 1 it is difficult to assess if there is a linear upward trend (i.e. that the correlation among pairs of waves is increasing over time). For this reason we applied the Haan *t*-test for trend (Önöz & Bayazit, 2003) in a time series statistical analysis given by: $t = r\sqrt{n-2}/\sqrt{1-r^2}$ The analysis follows Student's t distribution with degrees of freedom equal to $n - 2$, where r is the correlation between the data points and the time points and n is the number of time points. Only the environment variable showed a significant increase over time.

2.3. Panel attrition analysis

As mentioned earlier, occasional attrition was allowed in the design of the study (respondents could miss some waves of the data collection and come back to later ones). For the purpose of the analysis that follows, once a respondent skipped a wave of data collection for the first time, he/she was considered to be a permanent attritor. Group differences of key demographics, party ID, and public opinion poll exposure in overall survival on the AP-Yahoo! News panel were tested using a log-rank test in the SAS LIFETEST Kaplan-Meier method. It is important to note that this is a univariate test, which does not control for any covariates.

Survival on the panel by respondent group in waves 1 through 11(rare vs. non-rare)

A test of group differences performed for rare and non-rare respondents showed significant differences in the survival plot for the two groups in waves 1 through 11

(χ^2=4.91, 1df, n=2735, p<.05). The non-rare respondent group had higher survival probabilities than the rare respondent group, but the curves did not differ substantially.

Survival on the panel by party identification in waves 1 through 11
Of the respondents who completed the first wave of the study, 44.5 percent were Democrats (n=1217), 40.3% were Republican (n=1101), and 15.2% were Independent or belonged to other parties or to no party (n=417). Forty percent of longitudinal sample respondents (n=1086) completed all 11 waves of data collection. Out of these, 45.6% were Democrats (n=495), 41.7% were Republicans (n=453), and 12.7% were Independent or from other parties or those who said they did not belong to any party (n=138) as indicated by respondents' answers in wave 1. Test of group differences in survival on the panel by party ID showed that Republicans and Democrats had very similar survival rates; however, Independents and respondents from other parties had lower survival probability on the panel than Democrats and Republicans (χ^2=9.60, 1df, n=2735, p<.01).

2.3.1 Waves 1through 4 analysis
In wave 1, non-rare respondents received a $5 incentive, and rare respondents received a $10 incentive; in waves 2, 3, and 4, however, only the rare respondent group (younger than 30 years old, non-white, or less than high school education) received a $5 incentive, while non-rare respondents did not receive any money. Our first research objective was to look at the survival rates for rare vs. non-rare respondents in waves 1–4 and determine what variables had a relationship with attrition in the AP-Yahoo! panel. Univariate analysis was performed using SAS PROC LIFETEST; there were no significant differences in survival plot for the two groups in waves 2–4 (χ^2=1.931, 1df, n=2735, p=.165).

When examining attrition of rare vs. non-rare respondent groups, the following demographic, behavioral, and attitudinal covariates at the time of baseline were included in the models:

- Age (age 30 or older as a reference group)
- Ethnicity/race (white as a reference group)
- Education (high school or higher education as a reference group)
- Employment status: (non-working as a reference group)
- Gender (male as a reference group)
- Marital status (married as a reference group)
- Income
- Children in household (no children as a reference group)
- Internet status (no Internet as a reference group)
- Media exposure

  The survey respondent received a media exposure score (range 0–150) by answering five questions about media use in the baseline survey (see Online Appendix, MED1 for question wording). The frequency of media use in each category was given a number from 0 for those who answered "never" to 30 for those who answered "Every day." Then a summary score was calculated for each person.
- Party ID (Republican as a reference group, Independent/Other/None, Democrat)

  The variable was based on the responses to the baseline survey (see Online Appendix, PID1); for ten respondents who refused to answer the party ID question in the baseline survey, their party identification values were received from the Knowledge Networks 2007 profile data that was on file.

- Political ideology (conservative as a reference group, moderate, liberal)
  The variable was based on the responses to the baseline survey (see Online Appendix, ID1); for 57 respondents who refused to answer the ideology question in the baseline survey, their ideology values were received from the Knowledge Networks 2007 profile data that was on file.
- Candidate preference for Democratic party (Obama as a reference group, Clinton, other candidate, don't know/Refused)
  The variable was based on the responses to the baseline survey of those who indicated voting in the Democratic Primary (see Online Appendix, VOT1).
- Candidate preference for Republican party (McCain as a reference group)
  The variable was based on the responses to the baseline survey of those who indicated voting in the Republican Primary (see Online Appendix, VOT2).
- Religion/Evangelism ("no" as reference group)
  The variable was based on the responses to the baseline survey (see Online Appendix, REL1). For 51 respondents who refused to answer this question in the baseline survey, their evangelical belief values were received from the Knowledge Networks 2007 profile data that was on file.
- Self-rated general health status ("excellent" as a reference group)
  This variable was based on Knowledge Networks Profile data. 67 respondents had missing values; Solas Hot Deck Imputation method was used to impute these cases ("Solas Imputation User Reference," 2008).

In order to have a better model, multicollinearity among covariates needs to be assessed. We conducted the test of multicollinearity using SPSS Factor Principal Axis Factoring to obtain squared multiple correlations. The highest initial communality was 0.32, and because none exceeded 0.90, we concluded that there was no issue of multicollinearity in the set of predictor variables (Tabachnick & Fidell, 2007).

The main assumption of the Cox PH model is that the hazard ratios of predictor categories are constant over time (i.e., they do not cross). If this assumption is violated, the Cox PH model cannot be used; instead, stratified or extended Cox models have to be employed. One of the ways to assess whether the assumption is met for continuous variables is to find the "correlation between the Schoenfeld residuals for a particular covariate and the ranking of individual failure times. If the proportional hazard assumption is met, then the correlation should be near zero" (Kleinbaum & Klein, 2005). Predictor variables *income* and *media* were examined using the above method, and the categorical variables were examined using a graphical approach of comparing log-log survival curves for variable groups. The assumption of the proportionality of hazards was not met for *Internet* (Internet status at the time of baseline: No/Yes) and *employment* (working status No/Yes). Therefore, instead of using the PH Cox Regression, Extended Cox regression including interaction of Internet with time and stratifying by employment was deemed more appropriate.

First we wanted to examine what variables had an effect on the overall panel attrition in waves 1–4. Non-whites and respondents with children were more likely to drop out than whites and respondents without children in the household (HR=1.22 and HR=1.33, respectively), and higher media exposure seemed to have a positive effect on panel survival.

2.3.2 Wave 5 through 9 analysis

In waves 5–9, we studied the rare group independently from the non-rare group, due to the difference in the incentive amounts for early responders and late responders in rare and non-rare groups. As mentioned earlier, in order to increase response rates, in wave 5 additional incentives were introduced for late respondents (those who did not answer all of waves 1–4 in the first seven days of data collection). This addition added a potential confounding factor to studying rare and non-rare respondents together, because all rare group respondents received an incentive in waves 5–9 (it only differed in the amount), but non-rare group respondents only received an incentive if they were also late respondents. Therefore, we examined rare and non-rare groups separately. For the purpose of these analyses we treated wave 5 as a baseline, and therefore everyone who completed wave 5, even if they had dropped out in the previous waves, was included in the analyses (n=2055).

To test for group differences, a univariate analysis (not controlling for any covariates) was performed on the variable for Late/Early Respondents using SAS PROC LIFETEST. We found significant differences in survival plots for the two groups in waves 5–9 (χ^2=151.10, 1df, p<.001). Early respondents consistently have higher survival probability on the panel than late respondents, and the difference increases in later waves. The survival curves are shown in Figure 2.

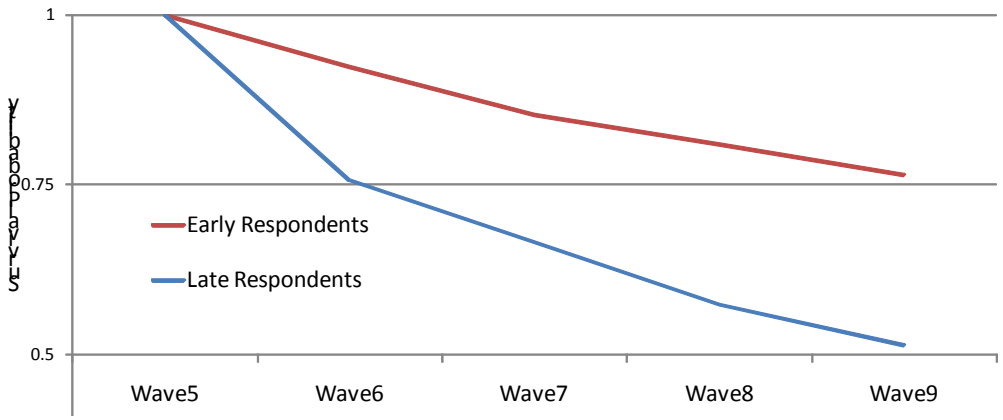

Figure 2: Survival Curves of Late vs. Early respondents (Waves 5–9).

We reexamined the assumption of proportionality of hazards for wave 5–9 data and concluded that the assumption was met for all of the variables but the media score. Therefore an Extended Cox model was fitted to wave 5–9 data with a *media* x *time* interaction term included.

Analysis of survival rates for non-rare respondents (early vs. late respondent groups)

Because the non-rare group included only white respondents and those with at least a high school education, we excluded the ethnicity variable from the model, and redefined the reference group for education to be "High School." In the full multivariate model, comparing early and late respondents in the non-rare group (n=1270), there was a significant effect (p<.001) of being a late vs. early respondent on panel survival after adjusting for other covariates. In the non-rare group (white, older than 29 years old, and with at least a high school diploma), the probability of dropping out for late respondents (who received a $5 incentive in waves 5–9) was almost 2.5 times more than the probability of dropping out for early respondents (who did not receive any incentive in

waves 5–9). In addition, females were more likely to drop out than males, and those without Internet at home were 50% more likely to drop out than those who had an Internet connection. Candidate preference, or lack thereof, was not a significant predictor of the survival of the non-rare sample on the AP-Yahoo! panel in waves 5–9 (p=.60).

3. Conclusion

Compared to most longitudinal studies, the AP-Yahoo! election survey presents an optimal laboratory for testing the hypothesis that respondents will change as a result of study participation. If panel conditioning were to be a factor in any longitudinal survey, it should manifest itself in the AP-Yahoo! survey data. This is because respondents in most cases participated in an election survey each month during the entire campaign, many times being asked similar questions. In other words, this study can be seen as a worst case scenario for panel conditioning if it were to occur.

In comparison to traditional household panels, where the period between waves is more expanded (e.g., yearly or quarterly data collection), the AP-Yahoo! design had a much more compressed schedule with the time between waves being approximately four weeks. This feature makes it ideal to study panel conditioning because respondents are asked the same or similar questions at a rapid pace. Another feature of the panel is being an online election study, where these respondents were more likely constantly exposed to Internet political campaign messages. Given these characteristics, our limited findings of panel conditioning effects are reassuring.

Panel conditioning, in the definition of "real change," was assessed using two different methods. Taking advantage of three independent cross-sectional samples, we were able to compare longitudinal respondents with cross-sectional respondents on a variety of questions about political attitudes and knowledge. In order to minimize the confounding factor of panel attrition, when comparing a longitudinal with a cross-sectional sample, we used a weighted logistic regression, controlling for demographic characteristics and survey experience. In only 3 out of 14 comparisons of longitudinal versus cross-sectional did we find statistical support that there is a panel conditioning effect (p<.05). We found some evidence of panel conditioning for a political knowledge question. Not surprisingly, longitudinal respondents were 28% more likely to correctly name Obama's religion than cross-sectional respondents, confirming previous results from the literature for panel conditioning on knowledge questions (Das, Toepoel, & van Soest, 2007; Toepoel, Das, & van Soest, 2009). We also found some evidence of panel conditioning for two attitudinal questions. Longitudinal respondents were 17% more likely to report feeling hopeful about the presidential election than cross-sectional respondents in wave 3; however, this effect disappeared in wave 9. Longitudinal sample respondents were also 56% more likely to be bored with the election than the cross-sectional respondents in wave 9, but not in previous waves. For the other six key items, we did not find any evidence of panel conditioning.

Using a new approach to study panel conditioning, based on the hypothesis that the stability of attitudes will increase in the subsequent waves in the panel (Sturgis, Allum, & Brunton-Smith, 2009) – the administration of the same questions at different points in time should strengthen or crystallize opinions – we found small evidence of conditioning for 1 out of 4 items. The inter-wave stabilities (correlation between the answer in a previous wave with the answer in the next wave) of attitudes about the environment increased from 70% to 75% after eight waves. Taking into account the political climate

at the time of the presidential race, this small increase in stability of attitudes about the environment is not really surprising, and it can also be attributed to external factors.

In summary, we found limited and weak evidence of panel conditioning in this study.

Regarding panel attrition, because the incentive system changed at wave 5, we ran two separate analyses for panel attrition: one for waves 1–4, and one for waves 5–9. During waves 1–4, non-white respondents were 1.2 times more likely to leave the panel than white respondents, and those with children in the household were 1.3 times more likely to attrit from the panel than respondents without children, adjusting for other covariates. These results are in line with previous findings for online panel attrition (DiSogra et al., 2007; Sayles & Arens, 2007). Although the incentive system was not set up with a control group, overall, it appears that providing rare (younger than 30 years old, non-white, or less than high school education) panel members with an incentive of $5 per survey kept them in the panel at the same rate as non-rare panel members.

When examining the candidate preference in waves 1–4, Democrats were no more or less likely to drop out from the study regardless of their candidate preference (Obama, Clinton, others, or undecided). This was not the case for Republicans: undecided Republicans were 1.7 times more likely to leave the panel than Republicans who indicated that they would vote for McCain. This was not the case for the subsequent waves, keeping also in mind that the number of Republican candidates was smaller and that the closer it is to an election, the number of undecided voters decreases.

In waves 5–9, there was a similar pattern of attrition as in waves 1–4: Internet households were less likely to attrite; candidate preference, however, was not a significant predictor of attrition in waves 5–9.

In waves 5–9, late-responding panel members (those who did not complete waves 1–4 in the first seven days after having received the email invitation), in both non-rare and rare groups, were more likely to drop out from the panel than early-responding panel members when adjusting for other covariates. In particular, non-rare late respondents were 2.4 times as likely to leave the panel as non-rare early respondents, and rare late respondents were 2.7 times as likely to attrit from the panel as rare early respondents. Although we did not set up an experiment on this issue, we changed the incentive system from wave 5 forward, giving extra money to late respondents to be more "punctual" in answering the questionnaire each wave. Without this incentive, it is plausible that the attrition rate for late respondents would have been even higher than what we experienced. In the literature on attrition, this appears to be a new finding worth investigating further. It might be the case that we identified a possible new predictor variable in panel attrition.

Acknowledgements

We seek to acknowledge and thank our colleagues at Stanford University, in particular Norman Nie, and at Harvard University, especially Sunshine Hillygus and Mike Henderson, for their contributions to the study. In addition, we would like to express our gratitude to Tom Wells, Xiaolu Yan, and Jing Yan Quek for their research support in providing KnowledgePanel profile data.

References

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

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly, 72*, 1008-1032.

Cantor, D. (1989). Substantial implications of longitudinal design features: the National Crime Survey as a case study. In G. Duncan, G. Kalton, D. Kasprzyk & M. P. Singh (Eds.), *Panel surveys* (pp. 25-51). New York: Wiley.

Cantwell, P. J. (2008). Panel conditioning. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (Vol. 2, pp. 566-567). Los Angeles, CA: Sage.

Clarke, H. D., Sanders, D., Stewart, M. C., & Whiteley, P. (2008). Internet surveys and national election studies: A symposium. *Journal of Elections, Public Opinion & Parties, 18*, 327-330.

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

Clinton, J. (2001). Panel bias from attrition and conditioning: A case study of the Knowledge Networks panel. Paper presented at the AAPOR 55th Annual Conference.

Das, M., Toepoel, V., & van Soest, A. (2007). Can I use a panel? Panel conditioning and attrition bias in panel surveys. Unpublished working paper. CentERdata Tilburg University.

Dennis, J. M. (2001). Are internet panels creating professional respondents? A Study of panel effects. *Marketing Research*, 34-38.

Dennis, J. M., Chatt, C., Li, R., Motta-Stanko, A., & Pulliam, P. (2005, May 12-15). *Data collection mode effects controlling for sample origins in a panel survey: Telephone versus internet.* Paper presented at the 60th Annual Conference of the American Association for Public Opinion Research, Miami Beach, FL.

DiSogra, C., Slotwiner, D., Clinton, S., Chan, E., Handarwan, E., & Zeng, W. (2007, July 29 - August 2). *Nonresponse bias in two methods of panel recruitment.* Paper presented at the Joint Statistical Meeting, Salt Lake City, UT.

Felix, J., & Sikkel, D. (1996). Attrition bias in a Telepanel. *Kwantitatieve Methoden, 61*, 101-110.

Granberg, D., & Holmberg, S. (1992). The Hawthorne effect in election studies: The impact of survey participation on voting. *British Journal of Political Science, 22*, 240-247.

Hansen, J. (2008). Panel surveys. In W. Donsbach & M. W. Traugott (Eds.), *The Sage handbook of public opinion research* (pp. 330-339). Thousands Oaks, CA: Sage.

Hoogendoorn, A., & Sikkel, D. (1998). Response burden and panel attrition. *Journal of Official Statistics, 14*, 189-205.

Kalton, G., & Citro, C. F. (1993). Panel surveys: Adding the fourth dimension. *Survey Methodology, 19*, 205-215.

Kleinbaum, D., & Klein, M. (2005). *Survival Analysis. A Self-Learning Text* (Vol. 2): Springer.

Kraut, R. E., & McConahay, J. B. (1973). How being interviewed affects voting: An experiment. *Public Opinion Quarterly, 37*, 398-406.

Nancarrow, C., & Cartwright, T. (2007). Online access panels and tracking research. The conditioning issue. *International Journal of Market Research, 49*, 573-594.

Nukulkij, P., Hadfield, J., Subias, S., & Lewis, E. (2007, May 17-20). *An investigation of panel conditioning with attitudes toward U.S. foreign policy.* Paper presented at the

62nd Annual Conference of the American Association for Public Opinion Research, Anaheim, CA.

Önöz, B., & Bayazit, M. (2003). The power of statistical tests for trend detection. *Turkish Journal of Engineering and Environmental Sciences, 27*, 247-251.

Pineau, V., Nukulkij, P., & Tang, X. (2005). Assessing panel bias in the Knowledge Networks panel: Updated results from 2005 research. In American Statistical Association (Ed.), *JSM Proceedings [CD-ROM]* (pp. 3480-3486). Alexandria, VA: AMSTAT.

Rookey, B. D., Hanway, S., & Dillman, D. A. (2008). Does a probability-based household panel benefit from assignment to postal response as an alternative to Internet only? *Public Opinion Quarterly, 72*, 962-984.

Saris, W. E. (1998). Ten years of interviewing without interviewers: The Telepanel. In M. Couper, R. P. Baker, J. Bethleem, C. Z. F. Clark, J. Martin, W. L. Nicholls II & J. M. O'Reilly (Eds.), *Computer assisted survey information collection* (pp. 409-429). New York: Wiley.

Sayles, H., & Arens, Z. (2007). *A study of panel member attrition in the Gallup Panel*. Paper presented at the AAPOR 62nd Annual Conference.

Sikkel, D., & Hoogendoorn, A. (2008). Panel surveys. In E. De Leeuw, J. Hox & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 479-499). New York: Lawrence Erlbaum.

Solas Imputation User Reference. (2008).

Sturgis, P. (2002). Attitudes and measurement error revisited: A reply to Johnston and Pattie. *British Journal of Political Science, 32*, 691-698.

Sturgis, P., Allum, N., & Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 113-126). Hoboken, NJ: Wiley.

Tabachnick, B., & Fidell, L. (2007). Survival/Failure analysis. In *Using Multivariate Statistics* (Fifth ed.): Pearson Education, Inc.

Toepoel, V., Das, M., & van Soest, A. (2009). Relating question type to panel conditioning: A comparison between trained and fresh respondents. *Survey Research Methods, 3,* 72-80.

Tortora, R. (2009). Attrition in consumer panels. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 235-249). Hoboken, NJ: Wiley.

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly, 43*, 359-377.

Van Der Zouwen, J., & Van Tilburg, T. (2001). Reactivity in panel studies and its consequences for testing causal hypotheses. *Sociological Methods & Research, 30*, 35-56.

Veroff, J., Hatchett, S., & Douvan, E. (1992). Consequences of participating in a longitudinal study of marriage. *Public Opinion Quarterly, 56*, 315-327.

Wang, K., Cantor, D., & Safir, A. (2000). Panel conditioning in a random digit dial survey. In American Statistical Association (Ed.), *Proceedings of the Survey Research Section of the American Statistical Association* (pp. 822-827). Washington D.C.: AMSTAT.

Waterton, J., & Lievesley, D. (1989). Evidence of conditioning effects in the British Attitudes Panel Survey. In G. Duncan, G. Kalton, D. Kasprzyk & M. P. Singh (Eds.), *Panel surveys* (pp. 319-339). New York: Wiley.

Yalch, R. F. (1979). Pre-election interview effects on voter turnout. *Public Opinion Quarterly, 40*, 331-336.