Mode Effects in American Trends Panel: Bayesian Analysis of a Cross-Classified Item-Person Mixed Model

Jeff Gill* Stas Kolenikov† Kyley McGeeney‡ Nancy Mathiowetz§

Abstract
American Trends Panel is a probability panel with RDD recruitment developed by Pew Research Center and Abt SRBI. Over the life of the panel, surveys have been conducted primarily via web mode, with mail mode for those who do not have access to the Internet or do not provide an email address. We analyze the results of the July 2014 wave (Wave 5) that included a comprehensive, large-scale mode-of-interview experiment that randomly assigned web respondents to telephone and web modes, with approximately 1,500 respondents in each mode. To quantify the contributions to the mode effects of the different question characteristics, we build a cross-classified random effects model with effects of person and question characteristics to identify the properties of survey questions that make them susceptible to mode effects, as well as the demographic groups that tend to exhibit mode effects. The model was estimated by Markov chain Monte Carlo computational Bayesian methods using a combination of R and JAGS packages.

Key Words: Bayesian, mixed model, mode effect, survey methodology

1. Mode effects

From the turn of the century, an important trend in survey data collection that affects both operations and statistical aspects of survey data analysis is proliferation of multimode surveys, in which the survey data are collected in more than one of web, phone, mail, face-to-face, and sometimes other modes of data collection (see de Leeuw (2005) for an outline of various approaches to multi-mode surveys). For instance, the American Community Survey (ACS) first requests that sample units complete the survey online. Then after two weeks, the web non-respondents are mailed a paper questionnaire. The non-respondents to the Web and mail phase are followed-up via computer-assisted telephone interviewing (CATI), and a subsample of persistent non-respondents is ultimately followed-up in-person (U.S. Census Bureau 2014). This sequence demonstrates the typical trade-offs in multimode survey design: the least expensive Internet mode with least coverage and lowest response rates is followed by the modes that are better suited for the balance of nonresponding sample, at the expense of increasing costs. Also, passive, self-administered interview modes that require sufficient literacy on the part of respondents are followed by active modes with interviewer involvement that are more appropriate for units that are less literate or more reluctant to participate in surveys.

Mode effects are differences in results for the same survey based on data collected in different modes. Groves (1989) first highlighted the challenges of studying mode effects by noting that one could look at the marginal effect of the mode (related solely to the medium for communication) or the overall effect of the mode, that is, the joint effect of the medium of communication and the operational differences arising from administering surveys in different modes (e.g., the joint effects of coverage, nonresponse, and measurement). More recent methodological literature (e.g., (Tourangeau et al. 2013)) stresses the importance of

*Washington University, One Brookings Drive, St. Louis, MO 63108
†Abt SRBI, 8405 Colesville Road, Suite 300, Silver Spring, MD 20910
‡Pew Research Center, 1615 L Street, NW, Suite 800 Washington, DC 20036
§Independent Consultant, Oconomowoc, WI

separating the effects of who responds across different modes from the effects of how the different modes effects how a respondent answers. In comparisons of the fielding of studies across two different modes, the differences in respondent mix (Elliott et al. 2009) may cause differences in the marginal results between modes to the extent that different modes are populated with respondents of different demographic characteristics (e.g., Internet users may be younger and more educated than non-users). If these demographic characteristics are in turn associated with outcomes of interest, mode selection effects would result. These differences can be mitigated by weighting, regression modeling, multiple imputation, or other methods of controlling for respondent characteristics (Kolenikov & Kennedy 2014).

In contrast, mode measurement effects are those effects specifically related to the question and answer process. Such effects may arise from differences in the medium of communication due to the presence of an interviewer (e.g., any comparison of self-administered questionnaire to an aurally delivered questionnaire) or the presentation format (Chang & Krosnick 2009, Tourangeau & Smith 1996, Tourangeau et al. 2013) Differences in the medium of communication may lead to primacy effects (in visual modes like web and mail, respondents may be more likely to choose the first option they read) and recency effects (in the aural mode like phone, respondents may be more likely to choose the last option that the interviewer read). Presence of the interviewer in active modes such as phone and face-to-face may lead to social desirability biases (Presser & Stinson 1998, Kreuter et al. 2008) as respondents are more likely to select response options associated with the behaviors or outcomes that present them in a more positive light in eyes of interviewer. Questions administered in different modes may lead to different cognitive processes in formatting the response (e.g., “other” option may or may not be offered in the Internet mode whereas it is a response option that can be volunteered by the respondent in the phone version of the survey). In addition, aural communication may lead to differences with respect to time taken for memory searches or referencing other records to assist in formulating responses.

Most of these and other studies have been looking at one aspect of mode effects at a time. There has been some limited work grounded in structural equation modeling approach to evaluation of survey instruments that attempted to quantify the relative impact of social desirability, topic, question form, as well as respondents’ demographic characteristics on the magnitude of measurement mode effects (Saris & Andrews 1991).

In addition to our interest in examining the effects of question characteristics on the magnitude of mode effects, we are also interested in determining the extent to which demographic characteristics impact the magnitude of measurement mode effects. Previous empirical research has found significant mode or measurement error effects related to levels of education (Chang & Krosnick 2010), gender (Hewitt 2002), race and ethnicity (Nelson et al. 2003), and age (Holbrook et al. 2006). Cognitive theory would suggest that comprehension and retrieval strategies may differ across modes (due to communication medium as well as length of time to search memory and formulate responses), suggesting that we would expect to see differences as a function of education and age.

Using data collected from a rigorous random assignment of respondents to mode treatments, we attempt to address the void in the literature by examining the item-level and person-level correlates of measurement mode effects. While previous work in this area has mostly examined individual questions for mode effects, we examine mode effects by analyzing the expert-coded characteristics of the question: the degree of sensitivity or social desirability, the question topic, the type of question (factual vs. attitude), and the question format.

Our research hypotheses are informed by both the empirical literature with respect to mode effects as well as the broader cognitive and social psychological literature addressing measurement errors in general. The four question traits of interest to our research and the hypotheses related to those characteristics are as follows:

- Social Desirability: larger mode effects are expected in questions identified as more sensitive or related to socially desirable or undesirable attitudes or behaviors.
- Topic area: we include topic area as an exploratory factor and offer no specific hypothesis with respect to differences by topic area.
- Type of question: there has been little research with respect to mode effects and question type (attitude, behavior, knowledge, or demographic).
- Question format: the more cognitively complex the response format, the more likely it is to be subject to mode effects. In particular, we expect question format to interact with personal characteristics, namely education, in that less educated individuals have larger mode effects when faced with the complex format questions.
- Number of response categories: measurement mode effects will increase as the number of response options increases.

With respect to the person-level variables, we hypothesize that the following individual characteristics will impact mode effects:

- Education: people with lower level of education are more susceptible to mode effects, especially in situations with higher cognitive processing demands, such as with difficult question formats, or with questions that are subject to social desirability that require extra mental processing to evaluate how pleasing the answer will be to the interviewer.
- Age: as cognitive function decreases with age, we expect some mode effects of age, at least in the older individuals.

2. Study Design

This study is based on a mode experiment conducted in the Pew Research Center's American Trends Panel (ATP). Panel participants were recruited from a large dual frame random-digit-dial (RDD) telephone survey conducted in early 2014 on the subject of political polarization. The donor study had a total sample size of 10,013, providing a large base for the panel recruitment. A total of 5,338 participants from the telephone study agreed to join the panel, and approximately 3,200 completed each subsequent wave of data collection . The telephone survey and panel recruitment was funded in part by grants from the William and Flora Hewlett Foundation and the John D. and Catherine T. MacArthur Foundation and the generosity of Don C. and Jeane M. Bertsch.

The standard mode of interview for panelists with access to the Internet is self-administration on a desktop, laptop, tablet or smartphone. At the time the data were collected, individuals who did not have access to the Internet or did not want to use the Internet for ATP surveys (about 10% of the panel respondents) completed them by mail with a paper questionnaire. Beginning in 2015, all of ATP respondents use Internet as the response mode. Panelists were given a small incentive for joining the panel ($10 in cash) and for completing each panel survey ($5 or $10). During 2014, surveys were conducted approximately once per month. The American Trends Panel was designed by Pew Research Center staff in collaboration with staff at Abt SRBI. Overall direction of the panel is the responsibility of Pew Research Center. Ongoing data collection is conducted and managed by

Abt SRBI. Additional information about the ATP can be found in Pew Research Center (2015*a*).

From the base wave of data collection, the following person-level characteristics are available:

- Sex (male, female).
- Race and ethnicity (non-Hispanic White, non-Hispanic black, non-Hispanic other, Hispanic).
- Education (high school or below, some college, college degree and above).
- Age (18-29, 30-49, 50-64, 65+).
- Ideological consistency: consistent liberal, lean liberal, neutral, lean conservative, consistent conservative.

The mode experiment analyzed in this paper is based on one wave of ATP data collection. Panelists who normally take their surveys on the Web were randomly assigned to either the phone mode (n=1,494 completed by phone) or the Web mode (n=1,509 completed on the Web), and interviewed July 7-Aug. 4, 2014. Panel participants who did not have access to the Internet, n=348, were interviewed on the phone. Their responses were used for cross-sectional reporting for the survey, but they are excluded from the mode experiment. A set of 54 questions like those commonly asked by the Centers research programs was administered to each respondent in their assigned mode. Respondents in the responding samples from each mode were independently weighted to be representative of the U.S. adult population in an effort to ensure that any differences observed between the groups were a result only of mode-of-interview effects. The differences between responses by mode ranged from 0 percentage points to 18 percentage points. The largest differences were observed on questions regarding the quality of respondents family and social life, as well as some of the questions about views on discrimination, where the mode effects differed for members of the group facing discrimination. There were also strong effects in ratings of political figures, where the members of the opposite party of each figure rated were more likely to give a “very unfavorable” rating on the web than on the phone. Additional information about the mode experiment, including further methodological details and descriptive analysis, can be found in Pew Research Center (2015*b*).

In addition to the survey data, each survey question was coded by survey methodologists at the Pew Research Center to describe the following question-level characteristics:

- Social desirability (SD) scale: not subject to SD, possible SD, subject to SD.
- Topic area: social and demographic trends, politics, religion, media and journalism, Internet and technology use.
- Type of question: attitude, behavior, knowledge, demographic.
- Question format: unipolar, bipolar, frequency, yes/no, forced choice, open, closed nominal categories.
- Number of response options (top-coded at 5).

2.1 Data preparation

All questions in the item survey data were transformed to the 0/1 format. Unipolar, bipolar, and frequency questions were transformed to 0/1 variables using a split that was as close as possible to 50/50 within the 20/80 to 80/20 range of possible splits; items that could not be split closer to 50/50 than 20/80 were not used. For instance, item "Q18. In general, how safe would you say you are from crime when walking in your neighborhood?" had the following distribution of responses: "Very safe", n=1760 (59.0%); "Somewhat safe", n=1021(34.2%); "Not too safe", n=157 (5.26%); "Not at all safe", n=47 (1.57%). For the purposes of this analysis, the variable was dichotomized into "1 Very safe" vs. "0 Other than very safe" (n=1225, 41.0%). Multinomial questions were recoded to 0/1 category-specific dummy variables (party affiliation: indicator for Republican, indicator for Democrat; religion: indicator for Protestant, indicator for Catholic, indicator for unaffiliated) The resulting data set had 57 binary items.

3. Mixed modeling of mode effects

Given the research questions above, and the available data, the analysis model must have the following properties.

1. The response is a binary 0/1 variable.
2. Probability of a "positive" response (i.e., the value of 1) varies between items and individuals, and the effect of individual-level variables such as education is item-specific.
3. One mode is selected as the reference mode. (Given that most data collection in the American Trends Panel is done on the web, it is chosen as the reference mode.) By definition, there are no mode effects in the reference mode.
4. Mode effects affect the probabilities of a "positive" response. Direction and magnitude of the mode effects depend on both the item-level and person-level covariates.

Let us denote the binary response given by person i on item j as y_{ij}, and the probability of probability of positive response, as $\mathrm{Prob}[y_i j = 1] = p_i j$. Let x_i be characteristics of person i; z_j be the characteristics of item j, and the mode in which the person i is randomized to, m_i (=0 for web, =1 for phone). Then the following model formalizes the above requirements:

$$\log \frac{p_{ij}}{1-p_{ij}} = \alpha_j + x_i'\beta_j + m_i v_{ij}, \tag{1}$$

$$v_{ij} \sim N(0, \sigma_{ij}^2), \quad \sigma_{ij}^2 = \sigma^2 \exp(\phi' x_i + \psi' z_j) \tag{2}$$

where α_j are item-specific intercepts, β_j are item-specific slopes for the person-level covariates in the linear predictor. Furthermore, v_{ij} is item-person-specific random effect whose variance is described by the second equation (2), where σ^2 is the typical magnitude of mode effects; ϕ is the vector of regression coefficients for person-level characteristics in the equation for the magnitude of the mode effects; and ψ is the vector of regression coefficients for the item-level characteristics in that equation.

Note that by the matter of notation, the mode effect $m_i v_i j$ is zero for the reference mode, and is "turned on" for the phone mode. This term is specific to the person-item combination. Note also that requirement 2 rules out the use of item characteristics in the probability/fixed effect equation (1). In the presence of item-specific intercepts, item-level covariates are not identifiable in this equation.

3.1 Bayesian estimation of the mixed model

The variables are labeled in the following fashion. First, `Y[i,j]` in the code is a version of how the respondents answered on question j of interest and the corresponding explanatory variables are describers of this question. Then we have the explanatory variables that are constant across items but varying over respondents ($n = 2072$):

- `age`: age in years
- `educ3`: 1 = College graduate+, 2 = Some college, 3 = H.S. graduate or less
- `mode`: 1 = phone, 2 = web
- `sex`: 1 = male, 2 = female
- `income3`: 1 = $75,000+ 2 = $30-$74,999 3 = less than $30,000 4 = Don't know/Refused
- `racethn4` 1 = White non-Hispanic 2 = Black non-Hispanic 3 = Hispanic 4 = Other
- `has_smartphone`: 0 = no, 1 = yes
- `has_tablet`: 0 = no, 1 = yes

Next we specify the explanatory variables that are constant across respondents but varying over items ($k = 57$):

- `metaqsd`: 1 = not subject to SD, 2 = possible SD, 3 = subject to SD
- `metaqtype`: 1 = Attitude, 2 = Behavior, 3 = Demographic, 4 = Knowledge
- `metaqtopic`: 1 = Politics and Policy, 2 = Social and Demographics Trends, 3 = Religion, 4 = Internet, 5 = Media and Journalism
- `metaqformat`: 1 = Forced Choice, 2 = Yes-No Binary, 3 = recode

The model (1)–(2) is a location-scale model, where both the location (1) of the linear predictor and the scale of random effects (2) are being modeled (a two-level, non-nested logit multilevel specification with a single dichotomous response). Hierarchical specification are especially well-suited to social science data since it often comes in different levels of aggregation. See Cohen et al. (1998), Gill & Witko (2013), Gill & Casella (2009), Park et al. (2004), Gelman & Rubin (1995), Shor et al. (2007), Gelman & Rubin (1995), Grimmer (2011).

To our knowledge, no off-the shelf software can fit the above model. For example, Hedeker & Nordgren (2013) needed to develop Fortran code to implement a variation of the location-scale mixed model for repeated measurements of a normal outcome. Given the above computational limitations, we chose to implement Bayesian Markov chain Monte Carlo estimation of the model (1)–(2) (Gill 2014) (Geyer 1992, Tierney 1994, Robert & Casella 2011) using `JAGS` ("Just Another Gibbs Sampler") software, which can be obtained from `http://www-fis.iarc.fr/~martyn/software/jags/`. The code is structured as follows inside the `model` { } statement. First loop through the respondents and then for each respondent loop through the items constructing the logit of the response probability, `p[i,j]`, which then gets related to the observed outcome variable, `Y[i,j]`, with a Bernoulli distribution:

```
    for (i in 1:N.RESPONDENTS) {
        for (j in 2:N.ITEMS) {
            Y[i,j] ~ dbin(p[i,j],1)
            logit(p[i,j]) <- alpha
            + beta[1,j]*mode[i] + beta[2,j]*educ3[i]
            + beta[3,j]*age[i] + beta[4,j]*sex[i]
            + beta[5,j]*inc.30.75.v.75[i] + beta[6,j]*inc.30.v.75[i]
            + beta[7,j]*inc.DK.v.75[i] + beta[8,j]*Black.v.White[i]
            + beta[9,j]*Hispanic.v.White[i]
            + beta[10,j]*Other.v.White[i]
            + beta[11,j]*Rural.v.Suburban[i]
            + beta[12,j]*Urban.v.Suburban[i]
            + beta[13,j]*has_smartphone[i] + beta[14,j]*has_tablet[i]
        }
    }
```

Since this is a Bayesian specification we need to stipulate the priors on each of these estimated coefficients:

```
    alpha   ~ dnorm(0,tau.a)
    tau.a   ~ dgamma(1,1)
    sigma.a <- 1/sqrt(tau.a)
    for (k in 1:N.ITEMS) {
        beta[1,k]    ~ dnorm(0,tau.b[1,k])
        tau.b[1,k]   ~ dgamma(1,magnitude[k])
        sigma.b[1,k] <- 1/sqrt(tau.b[1,k])
        for (m in 2:14) {
            beta[m,k]    ~ dnorm(0,tau.b[m,k])
            tau.b[m,k]   ~ dgamma(1,magnitude[k])
            sigma.b[m,k] <- 1/sqrt(tau.b[m,k])
        }
        magnitude[k] <- exp(-0.5*(gamma[1]
            + gamma[2]*metaqsd[k]
            + gamma[3]*attitude.v.knowledge[k]
            + gamma[4]*behavior.v.knowledge[k]
            + gamma[5]*demographic.v.knowlede[k]
            + gamma[6]*forced.v.recode[k]
            + gamma[7]*binary.v.recode[k]
            + gamma[8]*pp.v.journo[k] + gamma[9]*sd.v.journo[k]
            + gamma[10]*religion.v.journo[k]
            + gamma[11]*internet.v.journo[k] ))
    }
    for (l in 1:N.MAGS) {
        gamma[l]   ~ dnorm(0,tau.g[l])
        tau.g[l]   ~ dgamma(1,1)
        sigma.g[l] <- 1/sqrt(tau.g[l])
    }
}
```

All of these are normal and gamma forms with large variance to be conservative and to have semi-conjugate distributions for numerical stability in the running of the chain. Notice the use of `k` to loop through the items again. Inside this loop we parameterize the magnitude term to specify variability in the precision (1/variance) of the δ terms. The Gibbs sampler (Gelfand & Smith 1990) serially updates each parameter as the Markov chain progresses by drawing from their full conditional distribution, where the conditionality is on the most recent production of the other sampled parameters. As an illustration Figure 1 shows the orthogonal generation of three parameters in two graphical frames.

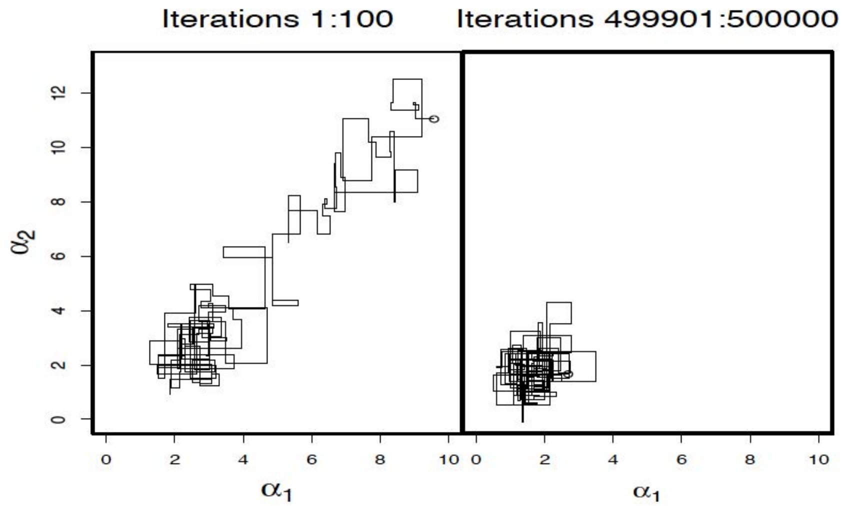

Figure 1: Gibbs Sampling of Three Parameters

The model was estimated by Markov chain Monte Carlo computational Bayesian methods using the `JAGS` package with 500,000 iterations keeping the second half. There was no evidence of nonconvergence using the standard diagnostics as provided in the `R` package `superdiag`. Specifically the Geweke, Gelman & Rubin, and Heidelberger & Welsch tests were applied without failures that indicate nonconvergence. See Hobert & Jones (2004) or Gill (2008) for discussions of convergence issues in this context.

Since this model produces over 1,000 parameters (e.g., the set of β parameters (1) is a full cross-classification of items and demographic variables, which leads to $14 \times 57 = 798$ coefficients), it is necessary to focus on a subset of interest. Table 1 presents a selected set of "significant" posterior parameter summaries for the β coefficients of equation (1) where the highest posterior density regions do not cover zero.

The direction and magnitude of the mode effects is given by β_1 coefficients. Table 2 demonstrates that none of the items demonstrated significant mode effects.

The mode effect specifically is described by the γ terms of equation (2). The posterior distributions are summarized in Table 3. Given that mode effects were found to be insignificant, as evidenced above by Table 2, the overall constants indicates low magnitudes of mode effects, and the coefficients of the predictors of the mode effect are empirically underidentified.

References

Chang, L. & Krosnick, J. A. (2009), 'National surveys via RDD telephone interviewing versus the internet', *Public Opinion Quarterly* **73**(4), 641–678.

Chang, L. & Krosnick, J. A. (2010), 'Comparing oral interviewing with self-administered computerized questionnaires: An experiment', *Public Opinion Quarterly* **74**(1), 154–167.

Cohen, J., Nagin, D., Wallstrom, G. & Wasserman, L. (1998), 'Hierarchical Bayesian analysis of arrest rates', *Journal of the American Statistical Association* **93**(444), 1260–1270.

de Leeuw, E. D. (2005), 'To mix or not to mix data collection modes in surveys', *Journal of Official Statistics* **21**(2), 233.

Elliott, M. N., Zaslavsky, A. M., Goldstein, E., Lehrman, W., Hambarsoomians, K., Beckett, M. K. & Giordano, L. (2009), 'Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores.', *Health services research* **44**(2 Pt 1), 501–518.

Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409.

Gelman, A. & Rubin, D. B. (1995), 'Avoiding model selection in Bayesian social research', *Sociological methodology* **25**, 165–173.

Geyer, C. J. (1992), 'Practical Markov chain Monte Carlo', *Statistical Science* **7**(4), 473–483.

Gill, J. (2008), 'Is partial-dimension convergence a problem for inferences from MCMC algorithms?', *Political Analysis* **16**(2), 153–178.

Gill, J. (2014), *Bayesian Methods: A Social and Behavioral Sciences Approach*, Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences, 3 edn, Chapman and Hall/CRC.

Gill, J. & Casella, G. (2009), 'Nonparametric priors for ordinal Bayesian social science models: Specification and estimation', *Journal of the American Statistical Association* **104**(486), 453–454.

Gill, J. & Witko, C. (2013), 'Bayesian analytical methods: A methodological prescription for public administration', *Journal of Public Administration Research and Theory* **23**(2), 457–494.

Grimmer, J. (2011), 'An introduction to Bayesian inference via variational approximations', *Political Analysis* **19**(1), 32–47.

Groves, R. M. (1989), *Survey Errors and Survey Costs*, John Wiley & Sons, New York.

Hedeker, D. & Nordgren, R. (2013), 'MIXREGLS: A program for mixed-effects location scale analysis', *Journal of Statistical Software* **52**(12).

Hewitt, M. (2002), 'Attitudes toward interview mode and comparability of reporting sexual behavior by personal interview and audio computer-assisted self-interviewing analyses of the 1995 National Survey of Family Growth', *Sociological Methods & Research* **31**(1), 3–26.

Hobert, J. P. & Jones, G. L. (2004), 'Sufficient burn-in for Gibbs samplers for a hierarchical random effects model', *The Annals of Statistics* **32**(2), 784–817.

Holbrook, A., Cho, Y. I. K. & Johnson, T. (2006), 'The impact of question and respondent characteristics on comprehension and mapping difficulties', *Public Opinion Quarterly* **70**(4), 565–595.

Kolenikov, S. & Kennedy, C. (2014), 'Evaluating three approaches to statistically adjust for mode effects', *The Journal of Survey Statistics and Methodology* **2**(2), 126–158.

Kreuter, F., Presser, S. & Tourangeau, R. (2008), 'Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity', *Public Opinion Quarterly* **72**(5), 847–865.

Nelson, D. E., Powell-Griner, E., Town, M. & Kovar, M. G. (2003), 'A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System', *American journal of public health* **93**(8), 1335–1341.

Park, D. K., Gelman, A. & Bafumi, J. (2004), 'Bayesian multilevel estimation with post-stratification: State-level estimates from national polls', *Political Analysis* **12**(4), 375–385.

Pew Research Center (2015*a*), Building Pew Research Centers American Trends Panel, Technical report, Washington, DC. available from http://pewrsr.ch/1Jo4nKE.

Pew Research Center (2015*b*), From telephone to the web: The challenge of mode of interview effects in public opinion polls, Technical report, Washington, DC. available from http://pewrsr.ch/1cT0q6r.

Presser, S. & Stinson, L. (1998), 'Data collection mode and social desirability bias in self-reported religious attendance', *American Sociological Review* **63**(1), 137–145.

Robert, C. & Casella, G. (2011), 'A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data', *Statistical Science* **26**(1), 102–115.

Saris, W. E. & Andrews, F. M. (1991), Evaluation of measurement instruments using a structural modeling approach, *in* P. P. Biemer, R. M. Groves, L. Lyberg, N. Mathiowetz & S. Sudman, eds, 'Measurement errors in surveys', John Wiley and Sons, New York, pp. 575–597.

Shor, B., Bafumi, J., Keele, L. & Park, D. (2007), 'A Bayesian multilevel modeling approach to time-series cross-sectional data', *Political Analysis* **15**(2), 165–181.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions (with discussion)', *Annals of Statistics* **22**, 1701–1762.

Tourangeau, R., Conrad, F. G. & Couper, M. P. (2013), *The Science of Web Surveys*, Oxford University Press.

Tourangeau, R. & Smith, T. W. (1996), 'Asking sensitive questions: The impact of Data-Collection mode, question format, and question context', *Public Opinion Quarterly* **60**(2), 275–304.

U.S. Census Bureau (2014), *American Community Survey: Design and Methodology*, Washington, DC. Available at: http://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_report_2014.pdf.

Table 1: Significant predictors of outcomes

Parameter	Item	Covariate	Estimate	95% HPD
$\beta_{14,39}$	Favorite politician Hillary Clinton	tablet	.1538	(.0019, .3058)
$\beta_{12,40}$	Favorite politician Michelle Obama	Urban vs Suburban	-.1660	(-.3272, -.0047)
$\beta_{7,42}$	Favorite politician Mitch McConnell	inc DK	-.3954	(-.7636, -.0272)
α_{44}	Favorite politician Sarah Palin	Intercept	-.5216	(-.9377, -.1054)
$\beta_{3,44}$	Favorite politician Sarah Palin	age	.0047	(.0002, .0091)
$\beta_{8,29}$	Life satisfaction	Black vs White	.4187	(.1313, .7061)
$\beta_{11,30}$	Social life satisfaction	Rural vs Suburban	-.2291	(-.4543, -.0039)
$\beta_{12,30}$	Social life satisfaction	Urban vs Suburban	-.2308	(-.3928, -.0689)
$\beta_{5,30}$	Social life satisfaction	inc 30 - 75	.1971	(.0239, .3703)
$\beta_{11,33}$	Community as a place to live	Rural vs Suburban	.2427	(.0226, .4628)
$\beta_{9,33}$	Community as a place to live	Hisp vs White	.3651	(.0782, .652)
$\beta_{12,46}$	Conflict Israel vs. Palestine	Urban vs Suburban	.1720	(.0147, .3293)
$\beta_{9,46}$	Conflict Israel vs. Palestine	Hisp vs White	-.3235	(-.6078, -.0391)
$\beta_{14,49}$	Social class self-identification	tablet	-.1703	(-.3307, -.0099)
$\beta_{9,34}$	Frequency talking to neighbors	Hisp vs White	-.2988	(-.5843, -.0133)
$\beta_{8,50}$	Standard of living vs. parents	Black vs White	-.3055	(-.5958, -.0151)
$\beta_{6,55}$	Religion == Protestant	inc $<$ 30	-.2863	(-.5125, -.0601)
$\beta_{13,57}$	Religion == unaffiliated	smartphone	-.1345	(-.2447, -.0242)
$\beta_{14,57}$	Religion == unaffiliated	tablet	.1616	(.0012, .3221)
$\beta_{9,57}$	Religion == unaffiliated	Hisp vs White	.3234	(.0465, .6002)
$\beta_{5,28}$	Have a passport	inc 30 - 75	-.1874	(-.3617, -.0131)
$\beta_{11,12}$	Worked with neighbors to fix problem	Rural vs Suburban	.2522	(.0341, .4703)
α_{13}	General trust	Intercept	-.4114	(-.8155, -.0073)
$\beta_{12,14}$	Could not afford doctor	Urban vs Suburban	-.1797	(-.3414, -.0179)
$\beta_{6,14}$	Could not afford doctor	inc $<$ 30	.2557	(.0369, .4745)
$\beta_{4,15}$	Could not afford food in past 12 mo	sex	-.1458	(-.2886, -.003)
$\beta_{5,16}$	Smoked 100+ cigarettes	inc 30 - 75	-.1771	(-.3521, -.0022)
$\beta_{9,16}$	Smoked 100+ cigarettes	Hisp vs White	.4357	(.1534, .7181)
$\beta_{6,18}$	Discrimination against LGBT	inc $<$ 30	.2202	(.0002, .4402)
$\beta_{10,20}$	Discrimination against women	Other vs White	.3519	(.0706, .6332)
$\beta_{12,21}$	Moral, values and religion	Urban vs Suburban	-.1970	(-.3589, -.0351)
$\beta_{5,21}$	Moral, values and religion	inc 30 - 75	.2022	(.0291, .3753)
$\beta_{12,22}$	Anti-terrorism policies	Urban vs Suburban	.1958	(.0298, .3618)
$\beta_{3,23}$	U.S. involvement in global economy	age	.0053	(.0009, .0097)
$\beta_{8,1}$	Volunteered in past 12 mo	Black vs White	.2885	(.0043, .5727)
$\beta_{3,3}$	Played game on PC or mobile yesterday	age	-.0053	(-.0097, -.0009)
$\beta_{4,6}$	Visited family/friends yesterday	sex	.1944	(.0536, .3352)
α_{7}	Wrote/receive personal letter yesterday	Intercept	-.7324	(-1.141, -.3237)
$\beta_{14,7}$	Wrote/receive personal letter yesterday	tablet	.2122	(.0533, .3712)
$\beta_{9,11}$	Listened to news on the radio yesterday	Hisp vs White	.2824	(.0057, .5592)

Table 2: Estimated mode effects

Parameter	Estimate	95% HPD	Parameter	Estimate	95% HPD
$\beta_{1,1}$	.0001	(-.0168, .0169)	$\beta_{1,30}$	-.0044	(-.0415, .0326)
$\beta_{1,2}$	-.0003	(-.0178, .0171)	$\beta_{1,31}$	.0022	(-.0254, .0299)
$\beta_{1,3}$	.0006	(-.0232, .0244)	$\beta_{1,32}$	.0057	(-.0358, .0472)
$\beta_{1,4}$	-.0014	(-.0248, .022)	$\beta_{1,33}$	.0017	(-.026, .0293)
$\beta_{1,5}$	.0010	(-.0161, .0182)	$\beta_{1,34}$	-.0023	(-.03, .0254)
$\beta_{1,6}$	.0006	(-.0149, .0160)	$\beta_{1,35}$	-.0036	(-.0412, .034)
$\beta_{1,7}$	.0002	(-.0153, .0157)	$\beta_{1,36}$	-.0015	(-.0272, .0242)
$\beta_{1,8}$	-.0017	(-.0326, .0293)	$\beta_{1,37}$	.0009	(-.0256, .0275)
$\beta_{1,9}$	.0012	(-.0192, .0216)	$\beta_{1,38}$	.0021	(-.0238, .0279)
$\beta_{1,10}$	.0008	(-.0153, .0169)	$\beta_{1,39}$	.0006	(-.0251, .0263)
$\beta_{1,11}$	-.0014	(-.0212, .0183)	$\beta_{1,40}$	-.0003	(-.0228, .0222)
$\beta_{1,12}$	-.0004	(-.0191, .0183)	$\beta_{1,41}$	-.0032	(-.0346, .0283)
$\beta_{1,13}$	.002	(-.0276, .0316)	$\beta_{1,42}$	-.0026	(-.0326, .0275)
$\beta_{1,14}$	-.0003	(-.016, .0154)	$\beta_{1,43}$	-.0022	(-.0334, .029)
$\beta_{1,15}$	-.0012	(-.0192, .0168)	$\beta_{1,44}$	.0021	(-.0246, .0287)
$\beta_{1,16}$	-.0015	(-.0222, .0191)	$\beta_{1,45}$	-.0008	(-.0225, .0209)
$\beta_{1,17}$	.0011	(-.0175, .0197)	$\beta_{1,46}$	.0011	(-.0249, .027)
$\beta_{1,18}$	.0001	(-.0164, .0166)	$\beta_{1,47}$	.0026	(-.0306, .0359)
$\beta_{1,19}$	.0021	(-.0228, .0271)	$\beta_{1,48}$	.0010	(-.0290, .03100)
$\beta_{1,20}$	-.0015	(-.0243, .0213)	$\beta_{1,49}$	.0045	(-.0356, .0446)
$\beta_{1,21}$	-.0046	(-.0449, .0356)	$\beta_{1,50}$	.0036	(-.0329, .0400)
$\beta_{1,22}$	-.0013	(-.0312, .0287)	$\beta_{1,51}$	-.0015	(-.0218, .0188)
$\beta_{1,23}$	.0056	(-.0394, .0506)	$\beta_{1,52}$	-.0006	(-.0256, .0245)
$\beta_{1,24}$	-.0007	(-.0232, .0218)	$\beta_{1,53}$	.0022	(-.0288, .0332)
$\beta_{1,25}$	-.0007	(-.0219, .0205)	$\beta_{1,54}$	-.0024	(-.032, .0271)
$\beta_{1,26}$	.0012	(-.026, .0285)	$\beta_{1,55}$	.0025	(-.0304, .0354)
$\beta_{1,27}$	.0002	(-.0213, .0217)	$\beta_{1,56}$	-.003	(-.034, .028
$\beta_{1,28}$	.0003	(-.0246, .0253)	$\beta_{1,57}$	-.0011	(-.0251, .023)
$\beta_{1,29}$	-.0003	(-.0214, .0207)			

Table 3: Significant predictors of outcomes

Parameter	Covariate	Estimate	95% HPD
γ_1	Overall magnitude	-10.68	(-17.62, -3.747)
γ_2	Social desirability	-.7347	(-3.694, 2.225)
γ_3	Attitude vs. knowledge	-.2757	(-3.552, 3.000)
γ_4	Behavior vs. knowledge	-.1539	(-2.179, 1.871)
γ_5	Demographic vs. knowledge	-.5197	(-2.244, 1.204)
γ_6	Forced response vs. other	-.2673	(-4.081, 3.547)
γ_7	Binary vs. other	-.454	(-3.619, 2.711)
γ_8	Public policy vs. media	.1779	(-2.411, 2.767)
γ_9	Social and demographic trends vs. media	.0072	(-2.542, 2.556)
γ_{10}	Religion vs. media	-.0935	(-3.183, 2.996)
γ_{11}	Internet vs. media	-.1877	(-3.155, 2.779)