Mode Effects in American Trends Panel: A Closer Look at the Person-Level and Item-Level Characteristics

Stanislav Kolenikov[1], Kyley McGeeney[2], Scott Keeter[2], Nancy Mathiowetz[3], Andrew Mercer[2], Courtney Kennedy[2]

[1]Corresponding author: s.kolenikov@srbi.com
Abt SRBI, 8405 Colesville Road, Suite 300, Silver Spring MD 20910
[2]Pew Research Center, 1615 L Street, NW, Suite 700 Washington, DC 20036
[3]Independent Consultant, Oconomowoc, WI

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, several surveys have been conducted on different modes, including web for most panel members, and mail or phone for those who do not have access to the Internet. We analyze the results of the July 2014 wave (Wave 5) that included a comprehensive, large-scale mode-of-interview experiment that randomly assigned 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 in the 75-question instrument, we build a cross-classified 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. We discuss how the decomposition of the total survey error and explained variance helps identifying the properties of the questions that are associated with the mode effects, such as question format, topic, and the potential impact of social desirability.

Key Words: American Trends panel, mode effect, social desirability, multimode survey

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. 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 (US 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 data interview modes that rely on sufficient literacy of respondents are followed up by active modes with interviewer involvement that are more appropriate for the units that are less literate and/or more reluctant to participate in surveys.

Mode effects are differences in results for the same survey based on data collected in different modes. The existing methodological literature identifies several principal

components of mode effects. 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), and if these demographic characteristics are in turn associated with outcomes of interest, mode effects would result. These differences can be mitigated by weighting, regression modeling, multiple imputation or other forms of control for respondent characteristics (Kolenikov and Kennedy 2014). Another source of mode effects are differences in *presentation format* (Chang and Krosnick 2009; Tourangeau and Smith 1996, Tourangeau, Conrad and Couper 2013) where the questions administered in different modes 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). Presence of the interviewer in active modes such as phone and face-to-face may lead to *social desirability biases* (Presser and Stinson 1998; Kreuter, Presser and Tourangeau 2008) as respondents are more likely to select response options associated with the behaviors or outcomes that represent them in a more positive light in eyes of the person with whom the respondent communicates, i.e., the interviewer.

Multimode surveys differ in how respondents are matched with response modes. In many practical situations, including the sequential multimode designs like the above cited example of ACS, respondents are effectively left to choose the mode that is the most convenient for them (Martin and Lynn 2011; Olson, Smyth and Wood 2012). Analyzing mode effects in these studies requires joint modeling of the mode choice and outcomes. Analysis is simpler in the more rigorous, but more expensive, split-sample experiments where respondents are randomized into a specific mode and not allowed to switch.

A separate strand of literature (Elliott et. al. 2009, Kolenikov and Kennedy 2014) deals with attempts to compensate for the mode effects, and provide unified estimates that have these effects purged, when the gold standard results are available, and/or reduced to the reference mode otherwise. We do not attempt any such corrections in this work, and concentrate on quantifying the correlates of mode effects instead.

Thus our research concentrates on the following research question:

What are the item-level and/or person-level correlates of the magnitude of mode effects?

In answering this question, we will be abstracting, to the extent possible, of the question contents, and replace it by the coded characteristics of the question (e.g., format, topic, and an expert evaluation of the degree of sensitivity and social desirability.)

2. Data

This study is based on a mode experiment conducted in American Trends Panel (ATP). Panel participants were recruited from a large RDD telephone survey conducted in early 2014 on the subject of political polarization. The study had a total sample size of about 10,000, providing a large base for the panel recruitment. Out of 5,338 participants from the base study who agreed to join the panel, approximately 3,200 completed each subsequent wave of data collection. All respondents in the original telephone survey received a common core of questions about their political values and engagement, along with a comprehensive set of demographic questions, ensuring a good baseline of information about respondents who agreed to join the panel as well about those who refused. 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. Individuals who do not have

access to the Internet or did not want to use the Internet for ATP surveys (about 10% of the panel respondents) complete them by mail with a paper questionnaire. We provide 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, and are being conducted approximately every two-three months in 2015. 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 (2015a).

The mode experiment analyzed in this paper is based on one of the waves of the 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. A set of 60 questions like those commonly asked by the Center's research programs was administered to each respondent in their assigned mode. Respondents in 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 were about 5%, with a range from 0% to 18%. The biggest differences were observed for questions regarding the quality of respondents' family and social life, as well as some of the questions about societal discriminations, where the mode effects differed between the groups pointed out by the discrimination questions. Also, there were strong effects in ratings of the various 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 (2015b).

In addition to the survey data, the data on items were coded by survey methodologists at Pew Research Center to describe the following item-level characteristics:

- 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
- Social desirability (SD) scale: not subject to SD, possible SD, subject to SD
- Number of response options

According to the typical findings in the existing methodological literature on mode effects, the following effects can be expected.

- Topic area: relatively neutral topics like media and journalism or Internet and technology use are likely to have lower or no mode effects compared to topics like social and demographic trends, politics, and religion.
- Type of question: factual questions (e.g., demographics and nonsensitive behaviors) are likely to have lower or no mode effects compared to attitude questions. Knowledge items may have mode effects to the extent that deeper thought can be given on the web.
- Question format: longer questions with more response categories may be more difficult to perceive on the phone, thus leading to larger mode effects. Open ended questions were omitted from the subsequent analysis.
- Social desirability: mode effects are expected to be larger for the more sensitive questions subject to SD.

3. Analysis

Typically, mode effect analyses found in the literature consist of cross-tabulating the survey items across mode, and possibly across demographic groups. The unique feature of the current analysis is that the existing data set that combines survey data across modes, demographic characteristics of survey respondents, and item characteristics allows joint modeling of the questions and person characteristics, and their interactions. Our analysis is thus aimed at isolating the magnitudes of mode effects, and relating them to the available covariates.

3.1 Data sets

The following data sets will be utilized in the analysis:

1. *Survey data*: the data set containing the responses of the 3,513 participants of the mode experiment, with their demographics and answers to the survey questions.
2. *Item data*: the data set with the question characteristics, as described in the end of the previous section.
3. *Long data*: the combined data set of survey responses, demographics and item characteristics, with one line per person per item.
4. *Regression data*: the intermediate data set of mode effect estimates for each item and each demographic category.

The next section describes in detail how the latter two data sets are produced.

3.2 Main procedure

To achieve the analysis goal, the following procedure is adopted.

1. All items in the *item data* were transformed to the 0/1 format.
    a. 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.
    b. Multinomial questions were recoded to 0/1 category-specific dummy variables (party affiliation: party == Republican, party == Democrat; religion: Religion == Protestant, Religion == Catholic, Religion == unaffiliated)

The remaining data set had 57 binary items.

2. *Survey data* on the 3,513 respondents were recoded into a *long data* format, with one line per person per item. Accounting for the item missing data, there were 170,259 lines in this data set.
3. *Item data* characteristics were merged into the resulting *long data* set.
4. Mode effects were estimated within each item-by-demographic group cell using the appropriate final weights. Each line in the resulting intermediate data set is identified by the item-by-demographic group interaction, with ~100–1000 aggregated lines.
5. The absolute value of thus estimated mode effect was taken.
    a. Since this intermediate result is a strictly positive quantity, mostly clustered near zero, with some occasional high values, it is heavily skewed. A square root transformation was applied to reduce skewness.

This produced the *regression data* set.

6. A regression model was fit to the resulting *regression data*, with transformed magnitude of the mode effect as the dependent variable, and the demographic variables over which the mode effect was estimated, along with the item characteristics, as explanatory variables. The within-demographic-cell contrasts

were weighted by the inverse of the estimated variance (i.e., the more accurate estimates were given greater weight).

Steps 4–6 were repeated for the various demographic groups and interaction of demographic variables.

3.3 Standard errors

To quantify uncertainty in the regression coefficients, we used two approaches to compute the standard errors.

1. Standard errors that were clustered on the items used (Rogers 1994);
2. Survey bootstrap standard errors (Rao, Wu and Yue 1992).

Bootstrap samples were taken independently within each mode. Shortcut bootstrap weights were produced as the product of the final weight with the bootstrap frequencies. Another, more methodologically rigorous, version of the bootstrap weights would be the weights that were calibrated (raked) to the same margins as the final weights. These were produced as well, with results virtually identical to the shortcut bootstrap weights. Steps 4–6 of the main procedure were repeated with the bootstrap weights for each demographic group analyzed to produce alternative standard errors.

We found that the bootstrap standard errors were consistently much smaller than the clustered standard errors by a factor of between 2 and 3. The bootstrap standard errors appear to only account for the sampling error in the estimates, while the clustered standard errors account for model uncertainty and misspecification, which may be a greater source of variability for these data. In the results below, we only report the clustered standard errors that are more conservative.

3.4 Item main effects

Given the above main procedure outlined, the basic model is the one that contains no demographics and the question effects only. Table 1 reports the basic statistics for that model. The interpretation of the coefficients is that the positive coefficients indicate a larger mode effect, while negative coefficients, a smaller mode effect; and a greater coefficient represents a larger effect of the variable. There are clear effects of social desirability (in the expected direction), as well as effects of the topics, where questions on religion and media exhibited lower mode effects than politics and policy as well as social and demographic trends.

Additional summary statistics reported in the table are as follows.

- R^2: variance explained by the regression model.
- # items: total number of items that were used in the regression model (c.f. 57 binary items in the recoded item-level data)
- Total contrast cells: number of item-by-demographic group cells.
- Singleton demos: number of item-by-demographic group cells that had all the respondents concentrated in a single mode. (While this did not happen in the item-only main effect regression, this was becoming an issue with higher order demographic interactions that lead to small demographic-group-by-mode cells.)
- Usable demos: number of item-by-demographic group cells that had estimable mode effects (i.e., at least one respondent in each mode).
- # usable respondents: ratio of the number of lines in the “long” data set corresponding to the “usable demographics” divided by the number of items.
- # nonusable respondents: ratio of the number of lines in the “long” data set corresponding to the “singleton demographics” divided by the number of items.

Table 1: Item characteristics only (baseline) model of mode effects.

	b/se		b/se
Topics: Politics and Policy	(base)	Question Type: Attitude	(base)
Social and Demographic	-0.0275	Behavior	-0.0512
Trends	(0.0401)		(0.0324)
Religion	-0.0834*	Demographic	-0.0196
	(0.0390)		(0.0331)
Internet	0.0567	Knowledge	0.0266
	(0.0595)		(0.0393)
Journalism	-0.1296***	# of original response options	0.0063
	(0.0320)		(0.0051)
SD: No social desirability	(base)	R^2	0.3550
Possible SD	0.0854*	# items	57
	(0.0346)	Total contrast cells	57
Subject to SD	0.1004*	Singleton demos	0
	(0.0397)	Usable demos	57
		# usable respondents	2923.70

3.5 Person main effects

Table 2 (see the end of the paper) reports the results of applying the main procedure to the variety of demographic groups with standard errors clustered by items (to account for the likely dependence of the mode effects for a given item).

The table is sorted across the models (given in columns) by the regression R^2 from smallest to largest. The bold column represents the base model with no demographics (same as Table 1). Some of the models with demographic variables had *lower* reported R^2 than the baseline model. This somewhat counterintuitive finding is due to the fact that the estimates for the different demographic models use different aggregate samples, and have lower precision due to lower sample sizes than the whole survey. As precision of the mode effect estimates is used to weight the regression data set cases for the final regression, the dilution of the explanatory power due to insignificant variables may result in lower R^2. These models still had lower standard errors than the baseline model, though, which points to improved prediction of the mode effects.

Across the topics, questions on journalism and media had consistently lower mode effects than others. Questions with potential social desirability produced higher mode effects, although in some specifications, the magnitudes were moderated by the demographic variables in regression specifications. Behavior questions tended to have lower mode effects than attitude questions. Demographic questions were found to have significantly lower mode effects than the base category of attitude only in regression with ideology measurement. The magnitudes of the item effects are generally reduced somewhat compared to the baseline model, which serves as an indirect evidence of interactions between item-level and person-level characteristics.

Predictors that produced models with better explanatory power than the baseline model were the ones based on race/ethnicity, education and phone usage. Minority respondents had consistently higher mode effects. More educated respondents had lower mode effects. Among the models with diminished R^2 compared to the baseline, middle age adults had somewhat lower mode effects; ideology and gender did not contribute to explanation of mode effects, while political activism by the respondent (defined as the respondent having conducted at least one of the three actions in the political activism

block of questions) produced a mildly lower mode effect, possibly indicating that politically active respondents have better defined attitudes that are less affected by the mode of data collection and/or presence of interviewer.

3.6 Demographic interactions

Table 3 reports the results of applying the main procedure to interactions of demographic person-level characteristics. Since the strongest predictors in the main effect analysis of the previous section were race, education, and phone usage, we concentrate on their interactions. The detailed coefficients are only reported for the race+age+education (three main effects specification; R+A+E) in Table 4; other results are cumbersome to present, but are available from the corresponding author upon request.

Table 3: Item-level results for models with interactions of person-level characteristics.

Person level characteristics and interactions	**None (baseline)**	Race + age + education	Race by education	Phone by education by race	Phone by education
	b/se	b/se	b/se	b/se	b/se
Topics: Politics and Policy	***(base)***	*(base)*	*(base)*	*(base)*	*(base)*
Social and Demographic	**-0.0275**	-0.0156	-0.0199	0.0021	-0.0072
Trends	**(0.0401)**	(0.0197)	(0.0269)	(0.0191)	(0.0218)
Religion	**-0.0834***	-0.0351	-0.0403	-0.0065	-0.0449
	(0.0390)	(0.0211)	(0.0269)	(0.0198)	(0.0232)
Internet	**0.0567**	0.0137	0.0021	-0.0110	0.0222
	(0.0595)	(0.0270)	(0.0419)	(0.0301)	(0.0334)
Journalism	**-0.1296*** **	-0.0762***	-0.0557*	-0.0305	-0.0696***
	(0.0320)	(0.0182)	(0.0257)	(0.0162)	(0.0186)
# of original response options	**0.0063**	0.0089**	0.0028	0.0034	0.0038
	(0.0051)	(0.0028)	(0.0029)	(0.0027)	(0.0038)
Social desirability: none	***(base)***	*(base)*	*(base)*	*(base)*	
Possible SD	**0.0854***	0.0324	0.0475*	0.0243	0.0642***
	(0.0346)	(0.0171)	(0.0224)	(0.0163)	(0.0167)
Subject to SD	**0.1004***	0.0253	0.0537*	0.0318*	0.0579**
	(0.0397)	(0.0156)	(0.0241)	(0.0157)	(0.0189)
Question type: Attitude	***(base)***	*(base)*	*(base)*	*(base)*	*(base)*
Behavior	**-0.0512**	-0.0363*	-0.0370	-0.0199	-0.0503*
	(0.0324)	(0.0154)	(0.0197)	(0.0157)	(0.0189)
Demographic	**-0.0196**	-0.0865***	-0.0186	-0.0595**	-0.0408
	(0.0331)	(0.0227)	(0.0244)	(0.0173)	(0.0206)
Knowledge	**0.0266**	-0.0222	-0.0262	-0.0219	-0.0187
	(0.0393)	(0.0182)	(0.0221)	(0.0254)	(0.0301)
R^2	**0.3550**	0.3883	0.4113	0.4810	0.5171
# items	**57**	57	57	57	57
Total contrast cells	**57**	2736	684	2052	513
Singleton cells	**0**	575	9	512	9
Usable cells	**57**	2161	675	1531	504
# non-usable respondents	**0.00**	103.74	8.33	52.19	4.63
# usable respondents	**2923.70**	2819.96	2915.37	2871.51	2919.07

Notes: standard errors are clustered on items. Detailed coefficient estimates for demographic variables are omitted; available from the corresponding author upon request.

Table 4: Comparison of results for models with age, race/ethnicity and education: separate vs. joint modeling.

	Age	Race / ethnicity	Education	Race+age+education
	b/se	b/se	b/se	b/se
Age 18-29	*(base)*			*(base)*
Age 30-49	-0.0399**			-0.0074
	(0.0117)			(0.0133)
Age 50-64	-0.0324*			-0.0140
	(0.0132)			(0.0136)
Age 65+	-0.0256			0.0217
	(0.0181)			(0.0200)
NH White		*(base)*		*(base)*
NH Black		0.0999***		0.1751***
		(0.0180)		(0.0153)
Hispanic		0.1110***		0.1772***
		(0.0125)		(0.0114)
Other		0.1208***		0.1976***
		(0.0192)		(0.0138)
College graduate			-0.0644***	-0.1166***
			(0.0123)	(0.0097)
Some college			-0.0232*	-0.0615***
			(0.0099)	(0.0093)
High school or less			*(base)*	*(base)*
R2	0.2643	0.3540	0.3739	0.3883
# items	57	57	57	57
Total contrast cells	228	228	171	2736
Singleton cells	0	0	0	575
Usable cells	228	228	171	2161
# non-usable respondents	0.00	0.00	0.00	103.74
# usable respondents	2923.70	2923.70	2923.70	2819.96

One general problem with using more demographic variables is that the main procedure works by forming smaller cell with full interactions of these demographic variables (regardless of whether the interactions are fully modeled later on in the regression, or just used as main effects, as in the R+A+E specification). With more variables, a much larger number of cells is being produced (compare 228 cells for the age or race/ethnicity 4-category variables vs. 2736 cells for the R+A+E specification in Table 4), and some cells are becoming so small that the mode effect is not estimable for them (575 singleton cells in R+A+E), because they only have respondents in one mode. As Table 3 shows, this leads to losing anywhere from 5 to more than 100 respondents from the analysis.

While the impact of the item-level characteristics is reduced, as was the case of the main effect models in Table 2, the patterns of changes of the person-level characteristics estimates is more complicated in Table 4. On one hand, the magnitude of age effects is diminished, so these effects are no longer significant in the R+A+E specification. On the other hand, the magnitudes of both race/ethnicity and education effects increase, indicating potentially complicated interplay of these effects in a comprehensive mode effect model.

4 Discussion

We analyzed a rich data set from a rigorous mode experiment implemented on the American Trends Panel, an online and mail probability panel developed and maintained by Pew Research Center and Abt SRBI. We proposed a joint model of item-level and

person-level contributions to mode effects that required sophisticated estimation techniques. Our approach pools all of the questions together (recoding ordinal and frequency questions into binary ones, trying to approximate 50/50 split between the recoded categories) and assumes that either demographic variables or item characteristics uniformly increase or decrease mode effects. The heterogeneity of mode effects in specific questions may be further revealed by analysis of item-person interactions, which is a possible direction for future research.

The necessity to compute the mode effect as the difference between two cell means for each mode implies that cases with unique demographics are removed from this analysis. Some higher order demographic cells may be small or empty. The effects of this happening (beyond the obvious loss of representation) are unclear.

An attempt was made to add the technology to the model, namely, completion on a tablet or on a mobile phone, as identified from user agent strings. However, these variables were found to be multicollinear (with the demographic characteristics), and were dropped out of the model by statistical package.

The explanatory power of most models, as expressed by R^2, is somewhat limited: mode effects are elusive to modeling.

As a side note on an unsuccessful modeling attempt, at earlier stages of statistical modeling of mode effects, we fit a logistic regression model with all two-way item-characteristics vs. person-characteristics interactions. Except for a great number of interaction cells that were empty, the model converged. However, interpretation of the several hundred estimated coefficients was impossible, unless another model would be built to fit the coefficient estimates themselves. Difficulties in dealing with the results of that model led us to formulate the procedure outlined in Section 3.2.

References

Chang, L. and J. A. Krosnick (2009). "National Surveys via RDD Telephone Interviewing Versus the Internet." *Public Opinion Quarterly*, **73** (4): 641–678. http://poq.oxfordjournals.org/content/73/4/641.abstract.

Elliott, M. N., Zaslavsky, A. M., Goldstein, E., Lehrman, W., Hambarsoomians, K., Beckett, M. K. and Giordano, L. (2009), Effects of Survey Mode, Patient Mix, and Nonresponse on CAHPS Hospital Survey Scores. *Health Services Research*, **44**, 501–518. doi: 10.1111/j.1475-6773.2008.00914.x

Kreuter, F., S. Presser and R. Tourangeau (2008). "Social Desirability Bias in CATI, IVR, and Web Surveys the Effects of Mode and Question Sensitivity." *Public Opinion Quarterly*. http://poq.oxfordjournals.org/content/72/5/847.abstract.

Kolenikov, S. and C. Kennedy (2014). "Evaluating Three Approaches to Statistically Adjust For Mode Effects." *Journal of Survey Statistics and Methodology*, **2** (2), 126–158, http://jssam.oxfordjournals.org/content/2/2/126.abstract.

Martin, P., and P. Lynn (2011), "The Effects of Mixed Mode Survey Designs on Simple and Complex Analyses," *Working Paper 2011–28*, Institute for Social and Economic Research, University of Essex, UK.

Olson, K., J. D. Smyth, and H. M. Wood (2012), "Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination," *Public Opinion Quarterly*, **76** (5), 611–635.

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

Pew Research Center (2015b). From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls. Technical report, available from http://pewrsr.ch/1cT0q6r.

Presser, S. and L. Stinson (1998). “Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance.” *American Sociological Review*, **63**, 137–45

Rao, J. N. K., C. F. J. Wu, and K. Yue (1992), “Some Recent Work on Resampling Methods for Complex Surveys,” *Survey Methodology*, **18**(3), 209–217.

Rogers, W. (1994). Regression standard errors in clustered samples, *Stata Technical Bulletin*, **3** (13), 19–24.

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

Tourangeau, R. and T. W. Smith (1996). “Asking Sensitive Questions: The Impact of Data-Collection Mode, Question Format, and Question Context.” Public Opinion Quarterly. http://poq.oxfordjournals.org/content/60/2/275.abstract, **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 2: Joint regression model for the magnitude of mode effects vs. main effects of demographic variables.

	Ideology	Age	Political activism	Gender	Race / ethnicity	**None**	Education	Phone usage
	b/se	b/se	b/se	b/se	b/se	**b/se**	b/se	b/se
Politics and Policy	*(base)*	*(base)*	*(base)*	*(base)*	*(base)*	***(base)***	*(base)*	*(base)*
Social and Demographic	0.0148	-0.0168	-0.0090	-0.0164	-0.0123	**-0.0275**	-0.0139	-0.0158
Trends	(0.0239)	(0.0271)	(0.0368)	(0.0298)	(0.0310)	**(0.0401)**	(0.0307)	(0.0317)
Religion	-0.0347	-0.0454	-0.0404	-0.0340	-0.0372	**-0.0834***	-0.0565	-0.0426
	(0.0316)	(0.0308)	(0.0390)	(0.0319)	(0.0361)	**(0.0390)**	(0.0289)	(0.0377)
Internet	0.0567	0.0600	0.0554	0.0857*	0.0412	**0.0567**	0.0340	0.0395
	(0.0389)	(0.0366)	(0.0578)	(0.0408)	(0.0444)	**(0.0595)**	(0.0462)	(0.0546)
Journalism	-0.1098***	-0.1089***	-0.0884**	-0.1135***	-0.1040**	**-0.1296*** **	-0.0657*	-0.1386***
	(0.0198)	(0.0227)	(0.0307)	(0.0270)	(0.0308)	**(0.0320)**	(0.0260)	(0.0245)
# of original response options	0.0085*	0.0044	0.0037	0.0041	0.0041	**0.0063**	0.0043	0.0041
	(0.0038)	(0.0045)	(0.0054)	(0.0047)	(0.0041)	**(0.0051)**	(0.0039)	(0.0052)
No social desirability	*(base)*	*(base)*	*(base)*	*(base)*	*(base)*	***(base)***	*(base)*	
Possible SD	0.0496*	0.0725**	0.0888**	0.0924**	0.0539	**0.0854***	0.0716**	0.0873**
	(0.0189)	(0.0254)	(0.0299)	(0.0267)	(0.0275)	**(0.0346)**	(0.0243)	(0.0285)
Subject to SD	0.0475*	0.0765**	0.0828*	0.0866**	0.0779**	**0.1004***	0.0769**	0.0876**
	(0.0219)	(0.0284)	(0.0324)	(0.0295)	(0.0279)	**(0.0397)**	(0.0261)	(0.0305)
Type="Attitude"	*(base)*	*(base)*	*(base)*	*(base)*	*(base)*	***(base)***	*(base)*	*(base)*
Type="Behavior"	-0.0334	-0.0592**	-0.0567*	-0.0724**	-0.0296	**-0.0512**	-0.0594*	-0.0575*
	(0.0200)	(0.0204)	(0.0278)	(0.0246)	(0.0201)	**(0.0324)**	(0.0232)	(0.0278)
Type="Demographic"	-0.0484*	-0.0247	-0.0259	-0.0331	-0.0270	**-0.0196**	-0.0477*	-0.0300
	(0.0195)	(0.0255)	(0.0308)	(0.0235)	(0.0241)	**(0.0331)**	(0.0221)	(0.0295)
Type="Knowledge"	0.0335*	-0.0299	-0.0011	0.0017	-0.0276	**0.0266**	-0.0237	0.0007
	(0.0159)	(0.0401)	(0.0534)	(0.0470)	(0.0378)	**(0.0393)**	(0.0444)	(0.0499)
R^2	0.1647	0.2643	0.2790	0.2957	0.3540	**0.3550**	0.3739	0.4435
# items	57	57	57	57	57	**57**	57	57
Total contrast cells	285	228	114	114	228	**57**	171	171
Singleton cells	1	0	0	0	0	**0**	0	0
Usable cells	284	228	114	114	228	**57**	171	171
# non-usable respondents	5.58	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00
# usable respondents	2918.12	2923.70	2923.70	2923.70	2923.70	**2923.70**	2923.70	2923.70

Table 2, continued.

	Ideology	Age	Political activism	Gender	Race / ethnicity	**None**	Education	Phone usage
	b/se	b/se	b/se	b/se	b/se	**b/se**	b/se	b/se
Demographic characteristics	Consistent Liberal: -0.0036 (0.0187) Leaning liberal: 0.0238 (0.0132) *Inconsistent: (base)* Leaning conservative: 0.0123 (0.0142) Consistent conservative: 0.0309 (0.0184)	*Age 18-29 (base)* Age 30-49: -0.0399** (0.0117) Age 50-64: -0.0324* (0.0132) Age 65+: -0.0256 (0.0181)	Politically active: -0.0245* (0.0114)	Female: 0.0055 (0.0149)	*NH White (base)* NH Black: 0.0999*** (0.0180) Hispanic: 0.1110*** (0.0125) Other: 0.1208*** (0.0192)		College graduate: -0.0644*** (0.0123) Some college: -0.0232* (0.0099) *High school or less (base)*	Landline only: 0.1658*** (0.0238) Cell phone only: 0.0167 (0.0094) *Dual user (base)*
R^2	0.1647	0.2643	0.2790	0.2957	0.3540	**0.3550**	0.3739	0.4435
# items	57	57	57	57	57	**57**	57	57
Total contrast cells	285	228	114	114	228	**57**	171	171
Singleton cells	1	0	0	0	0	**0**	0	0
Usable cells	284	228	114	114	228	**57**	171	171
# non-usable respondents	5.58	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00
# usable respondents	2918.12	2923.70	2923.70	2923.70	2923.70	**2923.70**	2923.70	2923.70

Standard errors corrected for clustering on items.