Assessment of Mode-Effects in a Web-enabled Study of Civic Attitudes

Glen Laird, Elizabeth Wiebe, Paul Pulliam, Lisa Thalji, RTI International,
and Vicki Huggins, Knowledge Networks

Keywords: Weighting, Web-based sampling, mode effect, nonresponse follow-up survey

Abstract: Using an RDD-recruited Web-enabled panel to collect survey data combines the convenience of self-administered Web data collection with the statistical rigor of probability sampling. To reduce levels of nonresponse and attempt to adjust survey outcomes to reduce nonresponse bias, a nonresponse follow-up survey (NRFUS) with Web nonrespondents may be conducted via CATI. Although the purpose is to reduce nonresponse bias, this approach potentially introduces bias due to the change in data collection mode from the main study. This study examines the potential for mode bias in NRFUS using a split sample design with CATI and Web-enabled data collection.

Mode Effects in Web-enabled Surveys

The use of Web surveys has grown considerably in recent years. Speed, flexibility, increase in population coverage, audio/video capabilities and cost factors make the use of the Internet appealing as a new survey tool. Application of probability-based sampling methods and survey research methods via the Internet is of great interest and in infant stages of investigation. Among the issues warranting study is the existence and level of mode effects for surveying using the Internet, which to this point have largely been studied in the context of mail, phone, and personal interview surveys. In order to increase coverage and response rates, Web and phone administration can be combined into a mixed mode survey approach. For questions that exhibit a mode effect, one can consider making an adjustment to the weights to compensate for the change in mode. In this paper, we investigate one simple adjustment of this type.

The aural nature of a phone questionnaire tends to manifest itself in terms of recency effects. That is, respondents are more likely to report the last item they heard read to them.

Web Methodology

The sample for this study was selected from subjects recruited for Knowledge Networks' nationally representative Web-enabled panel. The Web-enabled panel is an on-going research panel based on a nationally representative, list-assisted, random-digit-dial (RDD) sample drawn from all 10-digit telephone numbers in the U.S. These telephone numbers are grouped into banks of 100 numbers each, and identified by the first 8 of the 10 digits. The Knowledge Networks panel recruits from the "1+ banks", which are banks with at least one known working residential telephone number. Only those banks of telephone numbers that have zero directory-listed phone numbers are excluded. Sampling is implemented without replacement to ensure that numbers already fielded do not get fielded twice. Additionally, as of October 2001, Knowledge Networks began over-sampling telephone exchanges of Hispanic and African Americans.

The selected phone numbers are first screened for confirmed disconnected numbers, and for businesses. The numbers are also screened to identify numbers that are not in the WebTV Internet Service Provider network. A subsample of numbers that are outside the WebTV Internet Service Provider network are included in the sample to represent these areas, yet minimize the cost of paying higher Internet costs in those areas. Because this is an RDD-based sample, it should also be noted that households that do not have a telephone are not covered in the sample (approximately 4% of the U.S. households).

Households that agree to participate in the panel are then shipped identical WebTV hardware to install. Every participating household receives free hardware, free Web access, free e-mail accounts for each resident over the age of 13, ongoing technical support, and an incentive program to encourage continued participation. These households then complete an initial demographic profile survey via the Web.

Overview of the Survey of Civic Attitudes and Behavior

Research Triangle Institute and the Odum Institute conducted the Survey of Civic Attitudes and Behavior from January 11, 2002, through February 28, 2002 using the Web-enabled panel described above. The survey examined attitudes about politics, bioterrorism, volunteerism, and community after the terrorist attacks of September 11, 2001. A total of 3,879 interviews were completed across all sample groups, and 900 of these interviews were conducted over the telephone to facilitate investigation of mode effects in the nonresponse follow-up. Of the 2,979 Web respondents, 429 were from North Carolina. North Carolina residents were over-sampled since some of the questions applied only to North Carolina residents. Six groups comprise the sample for the Survey of Civic Attitudes and Behavior:

- Group 1: the Web-enabled panel sample surveyed using the Internet;
- Group 2: the Web-enabled panel sample surveyed using the telephone;
- Groups 3 – 6: four groups of Web-enabled panel nonrespondents were included in the formal nonresponse follow-up study.

We describe the sample groups in the nonresponse follow-up study below. Additionally, one sample group that was not included in the Survey of Civic Attitudes and Behaviors are those Web panel members who attrited from the panel.

Designing the Nonresponse Follow-up

Nonresponse in the survey was measured at one of four levels of participation:

(1) RDD recruitment to the panel,
(2) actual installation of the Web TV device,
(3) completion of the first household level survey, and
(4) response to the e-mail request to participate in the survey.

The cumulative response rate for panel surveys (including Web surveys) can be low due to the variety of ways in which a subject can become a non-respondent. To improve response rates and analyze error due to nonresponse, RTI designed a special nonresponse follow-up study (NRFUS) targeting nonrespondents at the four levels of participation described above. The NRFUS approach collects data from a random sample of the persons who have not completed the survey. These data are combined with those from the initial respondents to generate population estimates (Lessler and Kalsbeek 1992). The NRFUS allows us to compare Web panel respondents with those persons who were nonrespondents to either the panel recruitment effort or to the Web panel survey. The NRFUS was conducted using computer-assisted telephone interviewing (CATI). This procedure is therefore a mixed-mode survey. Because the data collection mode was different for these NRFUS samples than for the Web panel survey, RTI also conducted interviews with Web panel members over the telephone in order to facilitate investigation of possible mode effects in these nonresponse follow-up data.

For the households who refused to participate in the Web-enabled panel, hereafter referred to as "RDD Nonrespondents," a total of 300 (or 28.4% using AAPOR Response Rate 3) of these cases were interviewed in the NRFUS. For the households that agreed to participate in the Web-enabled panel, but had not yet hooked up the WebTV device, a total of 100 (or 50.6% using AAPOR Response Rate 3) were converted. Similarly, 100 non-responders to the profile survey and 100 non-responders to the SCAB were interviewed via CATI.

To allow estimation of mode effects in comparing the Web panel responders to the nonrespondents, 300 interviews (or 68.8% using AAPOR Response Rate 3) were completed over the telephone. Upon completion, the data were weighted to account for several factors, including probability of recruitment, probability of selection for the survey, sub-sampling, nonresponse adjustment, and post-stratification.

Primacy and Recency in Scaled Attitudinal Items

Table 1 shows the weighted distribution of response to survey question 4: "How much do you agree or disagree with the following statement: Bioterrorism is one of the most serious problems facing the country today." Response options were listed on screen (Web survey) and read to the respondent (telephone survey) in the same order they are listed in Table I, beginning with "Strongly Disagree" and ending with "Strongly Agree." These results

provide some initial support for the mode effects hypothesis. For this scaled attitudinal item, we expected to see a recency effect among the telephone respondents and a primacy effect among the Web respondents.

Primacy and recency effects are apparent in the results. Web respondents were more likely to select the first answer categories, and telephone respondents were more likely to select the last answer categories. The evidence in Table 1 suggests that mode effects appear to be the driving force in establishing differences between the main Web respondents and the NRFUS respondents. Chi-square tests comparing the web respondents to the phone respondents were highly significant ($p<.01$), but results among the various phone groups were not significant.

Social Desirability Effects in Items about Socially Favored Behaviors

Questions about sensitive behaviors such as drug use and sexual activity tend to generate the strongest social desirability effects. Among items on the survey was question 27: "Over the past year, how often would you say that you have volunteered for community service?" for which results are displayed in Table 2. In general, the mode effects literature led us to expect people to over-report this kind of good deed in a telephone interview. True to our expectations, the phone respondents reported volunteering at a higher rate than Web respondents.

Modeling of Mode Effects

We will describe the change in the estimates over different modes via proportional odds models (McCullagh, 1980). The proportional odds model is appropriate to use for ordinal dependent variables and categorical or interval independent variables. Suppose we wish to model the dependent variable, which we call Y, and a single binary independent variable, which we call X. In our models, Y will be the subject's response to either question 4 or question 27, and X will be the mode by which the subject was contacted. This model can be expanded to include other independent variables, such as demographic factors. Let X=0 indicate a phone case and X=1 indicate a Web case. Also, let Y be coded in ordinal fashion from 1 to K so that a "1" indicates the lowest-ranked response (*e.g.* "strongly disagree") and "K" indicates the highest-ranked response (*e.g.* "strongly agree"). Then the relationship between the dependent variable and independent variable is described as:

$$\text{Log}\,(P(Y\leq k)/P(Y>k)) = \alpha + \beta x \quad k=1,2,\ldots K \; ; \; (1)$$

It can be seen from (1) that the odds a respondent gives a response no greater than k is $\exp\{\beta\}$ times larger when the Web mode of survey is used. Note this model assumes the value of β does not depend on k. Therefore, the effect of mode is assumed to be uniform over the range of ordinal responses. A positive β indicates that Web respondents tend to give lower-ranked responses while a negative β indicates the Web respondents tend to give higher-ranked responses.

We can model the effect of mode by using the data from the samples of panel members. Note that both of these samples consist of cooperative panel members randomly assigned for contact by the internet or phone. The demographic makeup of samples 1 and 2 is therefore very similar. Models were fit for both question 4 and question 27 using SUDAAN® software procedure MULTILOG (Research Triangle Institute, 2001). The fit of the model was assessed by also fitting a generalized logit model. The fit was found to be good for both question 4 and question 27. Results for question 4 are shown in Table 3.

We see from Table 3 that subjects who responded via the Web were significantly more likely to give low-valued responses. In addition, an intercept-only model was fit to the NRFUS data for both questions 4 and 27. Results for question 4 are shown in Table 4. Using the coefficients from Table 4 in (1) will simply yield the observed sample probabilities from the weighted data.

The basis of our mode effect adjustment is to combine these two models by assuming the mode effect observed in the panel members is appropriate for use with non-respondents. That is, we assume the coefficient for the mode effect can be used in conjunction with the intercepts for the nonrespondents. The change in intercepts between the two models is assumed to be due to other factors than mode, most notably the level of cooperation exhibited by the subject. The logits in model (1) are computed using the coefficients from the mode effect model and the intercepts from the NRFUS sample.

The mode effect model for question 27 is complicated slightly by the presence of an interaction effect involving gender. However, the basic form of the model is unchanged. The results of the mode effect model and the intercept model are given in tables 5 and 6.

As can be seen from Table 5, the mode effect is considerably larger for men than it is for women. Perhaps the social desirability of volunteering for community service in the wake of September 11th had a greater impact upon men. Analogously to Table 4, Table 6 presents the intercept model for question 27.

The predicted probabilities of the NRFUS groups are combined in Table 7.

The predicted probabilities are considerably more in line with the Web response values than the initial values. Note that it is not necessarily appropriate for them to match the Web mode values precisely since our goal was to compensate for mode effects, not nonresponse.

Re-post-stratification

We can now view the issue of weight adjustment from a post-stratification standpoint. The salient difference between this post-stratification and the usual application is that here we are poststratifying to random values that are the output of a statistical model. Using the total weight from the NRFUS participants, we can re-poststratify their values to the probabilities from Table 7 via proportional adjustment so that the data conforms to the predicted values.

When more than one post-stratification total is used, this is often achieved via a process of iterative proportional fitting known as raking. In this method, proportional adjustments are made according to each of the post-stratification totals iteratively until each of the marginal distributions converges. If post-stratification to standard demographic totals was already planned, the predicted totals for the mode effect questions could simply be added to the list of post-stratification variables. In that case, all of the demographic totals would still be satisfied, as well as the predicted probabilities of the mode questions. The resulting weight would be designated for use with any of the questions used in the mode adjustment.

For illustration, we post-stratified to the predicted values of questions 4 and 27 after the data had already been post-stratified to demographic totals from the Current Population Study (CPS) in order to observe what changes would occur in the estimates of demographics. A difference of 4% was the largest difference observed for any demographic category, but most of the differences were around 1%. The values for gender are unchanged since post-stratification was conducted separately by gender for question 27 due to the presence of an interaction effect.

Discussion

The issue of mode effects involving split phone/Web surveys requires further study. Since a considerable proportion of the U.S. population is Internet ready, Web panels present an opportunity to have a nationally representative population to use for Internet surveys. However, since many households do not wish to join such a panel, nonresponse follow-up surveys are often necessary to conduct via phone. As the proportion of Internet-ready households increases in the U.S., it may be possible to conduct NRFUS by an alternate e-mail address instead of by phone. In the meantime, the mode effect of conducting a study via Web or phone remains a topic of interest.

We have adjusted data for the Survey of Civic Attitudes and Behavior using a simple proportional odds model and a special mode effect sample taken for this purpose. This method is not the only which should be studied, but it does have the advantage of being simple and not requiring any special software or advanced programming beyond what one would normally use for post-stratification. The standard errors of the new estimates have not been given because, ideally, they should be adjusted to account for the fact that the post-stratification totals used are actually random and not known (Singh & Folsom, 2001).

References

Bishop, G. and Smith, A. 2001. *Response-order Effects and the Early Gallup Split-ballots*. Public Opinion Quarterly. 65: 479-505.

Green, M.C., Krosnick, J.A. and Holbrook, A.L. 2001. The Survey Response Process in Telephone and Face-to-Face Surveys: Differences in Respondent Satisfaction and Social Desirability Response Bias. Report published at http://www.psy.ohio-state.edu/social/krosnick.htm

Lessler, J.T. and Kalsbeek, W.D. 1992. *Nonsampling Error in Surveys.* New York: John Wiley and Sons.

McCullagh, P. 1980 "Regression models for ordinal data.", JRRS B, 42, 109-142.

Singh, A.C, and Folsom, R. 2001 "Calibration Estimation and Calibration-adjusted Variance Estimation", Research Triangle Institute.

Research Triangle Institute. 2001. SUDAAN User's Manual, Release 8.0. Research Triangle Park, NC: Research Triangle Institute.

Table 1: Bioterrorism is one of the most serious problems facing the country today.

	Web Panel Members (Web)	Web Panel Members (Phone)	RDD Non-respondents	Panel Non-connectors	Profile Non-respondents	Survey Non-respondents
Strongly Disagree	13.6 (1.25)	5.0 (2.58)	8.9 (2.33)	9.3 (4.5)	3.0 (1.67)	2.0 (1.54)
Somewhat Disagree	19.9 (1.39)	13.0 (3.38)	12.3 (2.13)	10.2 (3.37)	14.2 (4.91)	5.4 (1.99)
Neither Agree nor Disagree	18.9 (1.24)	12.7 (4.11)	11.9 (2.39)	4.7 (2.27)	7.4 (2.98)	36.4 (11.11)
Somewhat Agree	29.0 (1.44)	36.6 (5.67)	35.6 (3.09)	42.3 (6.64)	40.6 (8.17)	32.1 (8.01)
Strongly Agree	18.6 (1.49)	32.7 (5.30)	31.4 (3.13)	33.4 (6.44)	34.9 (6.83)	24.1 (6.81)

Table 2: Over the past year, how often would you say that you have volunteered for community service?

	Web Panel Members (Web)	Web Panel Members (Phone)	RDD Non-respondents	Panel Non-connectors	Profile Non-respondents	Survey Non-respondents
Never	46.4 (1.71)	28.7 (4.63)	30.5 (3.17)	31.3 (6.76)	18.5 (5.14)	41.1 (9.68)
1-3 times	28.5 (1.53)	33.8 (5.92)	28.2 (3.11)	23.8 (5.56)	26.9 (5.73)	26.5 (7.16)
4-6 times	7.7 (0.74)	13.1 (3.71)	14.0 (2.27)	8.9 (2.99)	22.3 (9.05)	7.2 (3.42)
> 6 times	17.4 (1.23)	24.5 (4.90)	27.3 (2.77)	36.1 (6.35)	32.4 (6.58)	25.2 (10.02)

Table 3: Model of Mode Effect for Question 4

Parameter	**Coefficient**	**Standard Error**	**T-value[1]**	**P-value**
Intercept 1: Strongly Disagree	-2.85	.16	---	---
Intercept 2: Disagree	-1.61	.15	---	---
Intercept 3: Neither	-.86	.14	---	---
Intercept 4: Agree	.66	.14	---	---
Mode[2]	.90	.15	5.99	<.0001

[1] 3005 degrees of freedom

[2] Phone contact is the reference group

Table 4: Intercept-only Model for Question 4

Parameter	Coefficient	Standard Error
Intercept 1: Strongly Disagree	-2.36	.25
Intercept 2: Disagree	-1.39	.15
Intercept 3: Neither	-.75	.13
Intercept 4: Agree	.78	.13

Table 5: Mode Effect Model for Question 27

Parameter	Coefficient	Standard Error	T-value[1]	P-value
Intercept 1: Never	-.66	.18	---	---
Intercept 2: 1 to 3 times	.53	.18	---	---
Intercept 3: 4 to 6 times	1.04	.18	---	---
Gender[3]	-.15	.25	-.60	.55
Mode[2]	.15	.19	.79	.43
Mode*Gender	.54	.27	2.00	.05

Table 6: Intercept-only Model for Question 27

Parameter	Coefficient	Standard Error
Intercept 1: Never	- .79	.13
Intercept 2: 1 to 3 times	- .35	.11
Intercept 3: 4 to 6 times	.90	.12

Table 7: Predicted Probabilities for Mode-Adjusted NRFUS Respondents

Response	Web Completes	RDD NR	Non-connect NR	Profile NR	Survey NR
Question 4					
Strongly Disagree	.14	.18	.26	.08	.05
Disagree	.20	.21	.17	.27	.10
Neither	.19	.16	.05	.10	.45
Agree	.29	.30	.36	.33	.27
Strongly Agree	.19	.16	.16	.21	.13
Question 27					
Men					
Never	.53	.44	.53	.33	.68
1 to 3 times	.27	.27	.25	.35	.16
4 to 6 times	.07	.11	.03	.24	.04
>6 times	.13	.17	.19	.08	.11
Women					
Never	.41	.36	.32	.17	.23
1 to 3 times	.29	.28	.22	.26	.36
4 to 6 times	.09	.11	.17	.09	.04
>6 times	.21	.24	.29	.48	.36

[3] Female is the reference group