G-COE GLOPE II Working Paper Series

Estimating the Causal Effect of Survey Modes

Tetsuya Matsubayashi and Yoshitaka Nishizawa

Working Paper No. 55

Estimating the Causal Effect of Survey Modes

Tetsuya Matsubayashi* Yoshitaka Nishizawa†

January 27, 2012

Abstract

This paper offers a test for an important assumption in the study of survey modes and data quality. Prior research neglects the possibility that respondents' characteristics affect their decision to participate in a survey administered by a particular mode. The potential correlation of observable and unobservable personal characteristics with survey participation may cause a bias in the estimates of survey modes in a typical cross-sectional analysis. This study addresses how serious this problem can be by comparing the estimated effects of survey modes from a cross-sectional (CS) and difference-in-difference (DD) approach. The latter approach using panel data allows us to isolate the effect of survey modes from observable and unobservable personal characteristics of respondents by holding them constant. Our analysis demonstrates that the estimates based on the CS approach with and without observable control variables and the DD approach are quite similar in their sizes.

*Department of Political Science, University of North Texas (tmatsubayashi@unt.edu).

†Department of Political Science, Doshisha University (ynishiza@mail.doshisha.ac.jp).

1 Introduction

Since the invention of scientific survey, public opinion data have been collected by a variety of methods including face-to-face interviewing, telephone interviewing, and mail surveys. The development of computer technology in the last two decades expanded the options such as computer-assisted interviews and web surveys. Groves et al. (2009, 153-8) note that alternative survey methods differ along a variety of dimensions. For example, interviewers interact more with respondents during face-to-face interviews than during telephone interviews and mail and web surveys. Higher levels of privacy are maintained when questionnaires are self-administered than when they are interviewer-administered. Technology plays a more important role in computer-assisted methods than paper-and-pencil methods.

Survey researchers have assessed whether survey modes produce any difference in the quality of responses. They typically randomize the assignment of survey modes to respondents in estimating their causal effect on responses. Half of respondents in the initial sample are randomly selected for a treatment group and interviewed by one mode, while the other half are selected for a control group and interviewed by an alternative mode. For example, a wide range of studies use this type of randomized design to examine whether self-administered questionnaires yield more reports of sensitive behavior than interviewer-administered questionnaires (See Tourangeau and Yan, 2007, for review).

Randomized procedures allow survey researchers to justify the assumption that survey modes are exogenous to respondents' personal characteristics and treatment and control groups of respondents are balanced in their baseline covariates. Accordingly, researchers claim that they successfully identify a causal impact of survey modes on the quality of responses. Yet this assumption may be unrealistic. It is always a case that not all respondents chosen for the initial sample agree to participate in an interview process. Low response rates in many surveys have been an important topic for survey researchers for years (e.g., Groves and Peytcheva, 2008). Importantly, the decision to participate in a survey is not randomly made and rather a function of respondents' characteristics (Brehm, 1993; Groves and Couper, 1998). If respondents' characteristics affect their decision to participate in a survey administered by a particular mode (Couper and Rowe, 1996), there must be a correlation between survey modes and respondents' characteristics. Some of their characteristics such as demographic attributes are observable by survey questions and can be included in a regression model as control variables, while others

such as cognitive skills and personality are unobservable and difficult to control. The potential correlation of unobservable personal characteristics with survey modes may cause a bias in the estimate of survey modes in a typical cross-sectional analysis. Previous studies have paid little attention to this possibility.

This study addresses how serious this problem can be by comparing the estimated effects of two survey modes on the quality of responses from a cross-sectional (CS) and difference-in-difference (DD) approach. We use survey data collected by a unique design between 2005 and 2007 in Japan. A national sample of respondents was initially interviewed by the Paper-and-Pencil Interviewing (PAPI) method in 2005. As a follow-up survey in 2007, half of the respondents in 2005 were randomly selected for an interview with the PAPI method, while the other half of the respondents were selected for an interview using the Computer-Assisted Self-Interviewing (CASI) method.[1] These two waves of the survey included several questions that are identical with respect to topics, question wording, and response items.

This panel dataset allows us to use two different approaches to estimating the effects of survey modes. First, we apply a cross-sectional approach that exploits potential differences in the quality of responses between the CASI and PAPI respondents in the second-wave survey. Second, we apply a difference-in-difference approach that exploits changes in the survey modes and responses between the first- and second-wave surveys. The first approach relies on the assumption that the survey modes are exogenous to respondents' characteristics, whereas the second approach does not because respondents' characteristics are held constant over time. If the potential correlation of unobservable personal characteristics with the interview modes causes a bias, the estimates based on these two approaches should be significantly different. We compare the estimates based on two approaches using several survey questions regarding political participation and civic engagement.

The paper proceeds as follows. The subsequent section presents two approaches for estimating the effect of survey modes on the quality of survey responses. The third section describes data for our empirical analysis. The fourth section compares the estimated effects of survey modes between two approaches. The fifth section offers concluding remarks.

[1]The CASI mode was designed as CASI using text.

2 Two Approaches

We begin by presenting a simple model to estimate the effect of interview modes on the quality of response. Let y_i be a response reported by respondent i to a survey question. We assume that y_i has a continuous scale. Suppose that the question includes topics about politics and that we are interested in comparing the effect of self-administered and interviewer-administered methods on the way respondents answer this question. The answer to the question depends on an interview mode and respondent i's characteristics. Thus, y_i can be expressed as:

$$y_i = \alpha[self]_i + \mathbf{x}_i\beta + \eta_i + \epsilon_i \tag{1}$$

where $[self]_i$ is an indicator variable that equals one if respondent i is interviewed by a self-administered mode and 0 if interviewed by an interviewer-administered mode; $\mathbf{x}_i$ is a vector of respondent's characteristics that can be captured by survey questions such as demographic attributes and political orientations; η_i denotes respondent's characteristics that cannot be captured by survey questions such as cognitive skills, motivation, and personality; and ϵ_i denotes a respondent-specific error term.

The main coefficient of interest in equation (1) is α. It measures a difference in a survey response when the questionnaire is self-administered, as compared to when it is interviewer-administered. If respondents interviewed by the self-administered method tend to choose a higher score in y_i than those interviewed by the interviewer-administered method, α is estimated to be positive. If opposite, α is estimated to be negative.

If the interview modes are randomly assigned to respondents, one can assume that $[self]_i$ is exogenous to respondent's characteristics represented by $\mathbf{x}_i$ and η_i and thus the error term ϵ_i. If this assumption is valid, no selection bias occurs and α is estimated to be unbiased.

Yet the above assumption of exogeneity is may be violated even if a randomized design is employed. This assumption is valid when all respondents selected for the sample agree to participate in surveys. However, when any large-scale surveys are conducted, some respondents refuse to participate.[2] Respondents' decision to participate is not randomly made and rather a function of their personal characteristics (Brehm, 1993; Groves and Couper, 1998). Brehm (1993, 51-68) argues that the decision to participate in surveys is affected by respondent's relationship to strangers, the interviewer, the survey, and self-image. For example, a

[2]This problem is known as noncompliance in experiments.

respondent decides to participate in a survey if she feels comfortable speaking with a stranger, if she is interested in a survey topic, or if she feels capable of answering questions. Further, respondent's characteristics may affect participation in a survey administered by a particular mode. For example, Couper and Rowe (1996) note that computerized surveys may provoke negative reactions of respondents who are unfamiliar with using a computer and as a result fail to motivate them to engage in the interview.

If participation in a survey administered by a particular mode is conditional on respondents' characteristics, there must be a relationship between $[self]_i$ and respondent's characteristics represented by $\mathbf{x}_i$ and η_i in equation (1). Further, their characteristics may also have an impact on the way they answer survey questions. Importantly, $\mathbf{x}_i$ are observable and can be included as control variables, while η_i cannot be measured by survey questions and will be captured by ϵ_i. This violates the assumption that $E(\epsilon_i|self_i) = 0$ and as a consequence α is estimated to be biased. For example, suppose that η_i captures respondents' motivation. Less motivated respondents may be more likely to refuse to participate in the interviewer-administered survey than in self-administered survey. In addition, their motivation may also increase the likelihood that they choose a higher score as a response to the question. In this case, the relationship between motivation and participation in a self-administered survey is negative, while the relationship between motivation and a response to a survey question is positive. This would generate a negative bias in the estimate of α in equation (1). In short, the potential correlation of η_i with the interview mode is likely to cause problems in a typical cross-sectional analysis.

We address this problem by exploiting panel data. Suppose that respondent i answers the identical question in two waves of the survey. We can then explicitly rewrite equation (1) as:

$$y_{i1} = \alpha[self]_{i1} + \mathbf{x}_{i1}\beta + \eta_{i1} + \epsilon_{i1} \tag{2}$$

$$y_{i2} = \alpha[self]_{i2} + \mathbf{x}_{i2}\beta + \eta_{i2} + \epsilon_{i2} \tag{3}$$

where 1 and 2 denote the first- and second-wave of the survey. Suppose that the two waves of the survey are conducted between a short time period. In addition, suppose that all respondents in the first-wave survey are interviewed by the interviewer-administered method, while half of the second-wave survey are interviewed by the interviewer-administered method and the other half were interviewed by the self-administered method. Thus, $[self]_{i1}$ is equal to zero

in equation (2), whereas $[self]_{i2}$ can take either 0 or 1 in equation (3).

In order to eliminate the effect of respondent's personal characteristics, we take a difference between (2) and (3):

$$\Delta y_i = \alpha \Delta [self]_i + (\mathbf{x}_{i1} - \mathbf{x}_{i2})\beta + (\eta_{i1} - \eta_{i2}) + \nu_i \tag{4}$$

where $\Delta y_i = y_{i1} - y_{i2}$, $\Delta [self]_i = [self]_{i1} - [self]_{i2}$, and $\nu_i = \epsilon_{i1} - \epsilon_{i2}$. Since respondent's characteristics are unlikely to vary within a short time period, we assume that $(\mathbf{x}_{i1} - \mathbf{x}_{i2}) = 0$ and $(\eta_{i1} - \eta_{i2}) = 0$. As a consequence, equation (4) can be expressed as:

$$\Delta y_i = \alpha \Delta [self]_i + \nu_i \tag{5}$$

where $\Delta [self]_i$ is equal to zero if the respondent was interviewed by the interviewer-administered method in the first- and second-wave of the survey and one if the respondent was interviewed by the interviewer-administered method in the first-wave and the self-administered method in the second-wave of the survey. The identification assumption is that there is no correlation between $\Delta [self]_i$ and ν_i. This is likely to be met because the interview modes are randomly assigned and all respondent-specific characteristics are differenced out.

Equation (5) suggests that we are now comparing respondent i's answers to the identical question between the first- and second-wave. If the respondent changes her answer depending on the interview mode, the size of Δy_i must be larger for the respondent interviewed by the two different methods than that for respondents interviewed by the same method. Note that α can be either negative or positive, but it should be statistically significant if the interview mode has any impact.

We compare the estimate of survey modes based on equation (1) and equation (5). Equation (1) applies a cross-sectional approach, while equation (5) applies a difference-in-difference approach. Almost all previous studies relied on the cross-sectional approach. If the the potential correlation of unobservable personal characteristics with the interview modes causes no bias, the estimates based on these two approaches should be similar to each other.

3 Data

Our data come from the Waseda-GLOPE opinion surveys in Japan. The Waseda-GLOPE group has conducted several surveys mostly before and after the national elections between 2003 and 2010. We use the panel data collected in November 2005 and February 2007 because they offer an ideal setting for our purpose. For the first-wave survey in 2005, the sample of 3000 respondents were chosen from the list of eligible voters in Japan. The respondents were drawn by a cluster sampling method. About 1400 respondents agreed to participate in the 2005 survey. The response rate was 46.6 percent. The 2005 survey aimed to assess people's voting behavior in the 2005 Lower House election, party support, trust in political and non-political institutions, views on a variety of social and political issues, psychological involvement in politics, and engagement in civic activities.

The second-wave survey in February 2007 was a follow-up of the first-wave survey. Of 1398 respondents of the first-wave survey, 899 respondents participated in the second-wave survey. The response rate is 64.3%.[3] The second-wave survey included questions about people's voting behavior in the 2007 Upper House election, views on social and political issues, psychological involvement in politics, and political participation.

The 2005-2007 panel study employed a unique design with respect to interview modes. All of the respondents in the first-wave survey in 2005 were interviewed by an interviewer-administered Paper-and-Pencil Interviewing (PAPI) method. In contrast, the second-wave survey used two distinct interview modes. Half of the respondents in the first-wave survey were chosen for an interview by the PAPI method, while the other half of the respondents were chosen for an interview by a Computer-Assisted Self-Interviewing (CASI) method. In the PAPI mode, the interviewers asked questions and recorded answers from the respondents on a paper questionnaire. In contrast, in the CASI mode, the respondents read questions and recorded responses on a computer screen by themselves.[4] The interview modes in the second-wave survey were randomly assigned to the respondents. In the second-wave survey, the PAPI mode includes includes 468 respondents, while the CASI mode includes 423 respondents. The structure of the survey was summarized in Figure 1. Note that the first- and second-wave of the survey included several questions that are identical with respect to topics, question

[3]See the code book for more details of the sampling design.

[4]See Kohno et al. (2008) for more information on the survey design.

wording, and response items.

[Figure 1 Here]

A major difference between the CASI and PAPI mode is the degree of privacy. The CASI mode is self-administered and thus ensures privacy to the greater extent, in comparison to the PAPI mode. Since the respondents implemented an interview by themselves using a computer, their responses were not shared by the interviewer or other family members at home.

4 Results

Using the panel data described above, we estimate equations (1) and (5). Equation (1) is estimated using data from the second-wave survey, while equation (5) is estimated using data from the first- and second-wave of the survey.

In both equations, $[self]_i$ equals one if the second-wave respondent was interviewed by the CASI mode and 0 if interviewed by the PAPI mode. As a control variable, equation (1) includes major demographic variables such as gender, age, education, and income. The first variable, female, equals one if the respondent is female and zero otherwise. Age is measured in years. Education is measured by a scale including four categories of educational attainment such as high school and college graduates. Income is measured by five categories. One of the categories denotes respondents who did not report their income status. The regression model includes four income categories as dummy variables.[5] We estimate equation (1) with and without these control variables to examine whether the presence of these variables improve the performance of the model.

Using these demographic variables, we compare demographic characteristics of the respondents in the second-wave survey by the interview modes. Table 1 reports the percentages of PAPI and CASI respondents who belong to particular types of demographic groups. Except for the level of household income, we find no statistically significant difference in the demographic compositions of the respondents between the survey modes. About one third of the PAPI respondents refused to report their household income. The CASI respondents were more likely to report their income partly because higher privacy is ensured by self-administration. In short, we find that the respondents of two survey modes are comparable with respect to observable demographic features.

[5]The baseline category is the one for respondents who did not report their income status.

[Table 1 Here]

We use a series of questions on political participation as a dependent variable. They are chosen for our analysis because the identical questions were included in the two waves of the surveys. Further, Nishizawa and Kuriyama (2010) report that the PAPI and CASI modes make a significant difference in the reported frequency and willingness of participating in political activities. They argue that social desirability bias explains the difference in the responses between the survey modes. Accordingly, we also expect that the two survey modes produce a difference in the reported level of political participation and civic engagement. Note that our focus is not on the estimated difference between the survey modes. Our ultimate goal is to detect whether the two estimation approaches generate a difference in the estimated effects of survey modes.

The first- and second-wave of the survey asked respondents about whether they have ever engaged in a variety of political activities. Those activities include involvement in (A) a neighborhood organization, (B) volunteer activities, (C) political demonstrations and rallies, (D) campaign activities, (E) vote solicitation, (F) candidate support group, (G) party acclivities such as contribution, and (H) contact with national or local representatives. Respondents chose "(1) never," "(2) once or twice," or "(3) several times."[6] For difference-in-difference estimation, we take a difference in the reported level of participation between the second-wave and the first-wave of the survey. Thus, higher values in the difference denote that respondents chose higher values in the second-wave than in the first-wave.

Figure 2 presents the estimated results based on three estimation approaches. Three horizontal lines for each of the activities denote the 95% confidence intervals of the estimates associated with $[self]_i$ using equation (1) without control (solid lines with black circles), equation (1) with controls (dashed lines with triangles), and equation (5) (dotted lines with squares). Black circles, triangles, and squares denote the sizes of actual estimates. If the confidence intervals do not overlap the vertical gray line that denotes no difference between the survey modes (=0), the difference in the responses between the modes is statistically significant. We find only minor differences in the estimates across three approaches. Notably, the estimates based on equation (1) with control variables and equation (5) are very similar to each other. These results indicate that the potential correlation between unobservable personal character-

[6] In the original questionnaire, (1) denotes "several times," whereas (3) denotes "never." We recoded the items so that higher values denote more active involvement.

istics and survey modes plays little role in estimating the effect of survey modes on the quality of responses when a cross-sectional approach is used.

[Figure 2 Here]

Next, we use a series of questions about whether respondents are willing to participate in the activities noted above in the future. Respondents chose either "(1) would like to do it" or "(2) would rather not to be involved with it." As similar to the definition above, we take a difference in the reported level of willingness of participation between the second-wave and the first-wave for difference-in-difference estimation. The results are reported in Figure 3. As in Figure 2, Figure 3 shows that the estimates across the different estimation approaches show no major difference.

[Figure 3 Here]

Finally, we use a series of questions about how actively respondents engage in various types of civic activities such as (A) neighborhood associations; (B) consumer groups; (C) volunteer groups; (D) local residents' campaigns; (E) citizens actions; and (F) religious organizations. The estimated results by three estimation approaches are summarized in Figure 4. Again, regardless of the estimation approaches, the estimated effects of the survey modes are are almost identical.

[Figure 4 Here]

5 Conclusion

This paper offers a test for the important assumption that the potential correlation of observable and unobservable personal characteristics with survey participation has no impact on the estimates of survey modes in a typical cross-sectional analysis. We estimate the effects of two different survey modes on survey responses regarding political participation and civic engagement using the cross-sectional (CS) and difference-in-difference (DD) approach. The latter approach using the panel data allows us to isolate the effect of survey modes from observable and unobservable personal characteristics of respondents by holding them constant. Our analysis demonstrates that the estimates based on the CS approach with and without observable control variables and the DD approach are quite similar in their sizes. Our findings

imply that the random assignment of the survey modes was successful in the survey data we have used for our analysis.

References

Brehm, John. 1993. *The Phantom Respondents : Opinion Surveys and Political Representation*. Michigan Studies in Political Analysis. Ann Arbor: University of Michigan Press.

Couper, Mick P and Benjamin Rowe. 1996. "Evaluation of a Computer-Assisted Self-Interview Component in a Computer-Assisted Personal Interview Survey." *Public Opinion Quarterly* 60:89.

Groves, Robert M and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72(2):167–189.

Groves, Robert M, Floyd J Fowler Jr., Mick P Couper, James M Lepkowski, Eleanor Singer and Roger Tourangeau. 2009. *Survey Methodology*. Wiley Series in Survey Methodology. 2nd ed. Hoboken, N.J.: Wiley-Interscience.

Groves, Robert M and Mick Couper. 1998. *Nonresponse in Household Interview Surveys*. Wiley Series in Probability and Statistics. Survey Methodology Section. New York: Wiley.

Kohno, Masaru, Koichi Kuriyama, Yuko Morimoto, Aiji Tanaka and Motoki Watabe. 2008. "The Making of Making CASI Possible in Japan.".

Nishizawa, Yoshitaka and Koichi Kuriyama. 2010. "Social Desirability Bias in Survey Interviews." *Leviathan* 46(1):51–74.

Tourangeau, Roger and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5):859.

Table 1: Demographic Characteristics of PAPI and CASI Respondents

	PAPI	CASI
Gender		
Male	47.57	48.83
Female	52.43	51.17
χ^2 (df; p-value)	0.10 (1; 0.757)	
Age (in years)		
20–34	11.23	10.80
35–49	23.73	21.83
50–64	32.63	34.98
65–	32.42	32.39
χ^2 (df; p-value)	0.76 (3; 0.860)	
Education		
Less Than high school	19.02	24.11
High school degree	47.86	44.21
Some college	12.82	14.66
College degree and higher	20.30	17.02
χ^2 (df; p-value)	5.16 (3; 0.161)	
Income (in 1000 yen)		
No Report	33.62	3.87
0–2000	5.71	11.86
2000–6000	33.83	50.36
6000–10000	17.55	24.46
10000–	9.30	9.44
χ^2 (df; p-value)	128.07 (4; 0.000)	

Note: Table entries are percentages and χ^2 statistics with corresponding degrees of freedom and p-values. Data are unweighted. In total, the second-wave survey includes 891 respondents (PAPI=468; CASI=423).

Figure 1: Structure of the Survey

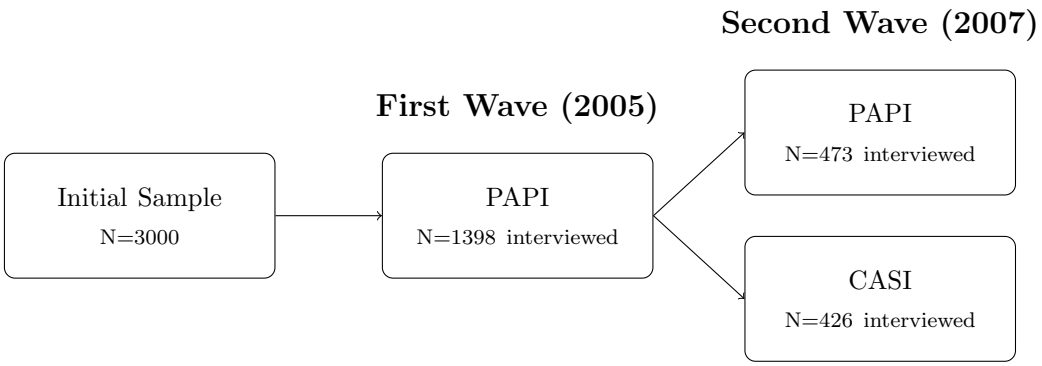

Figure 2: Comparing the Estimated Effect of Survey Modes on the Reported Level of Political Participation in the Past

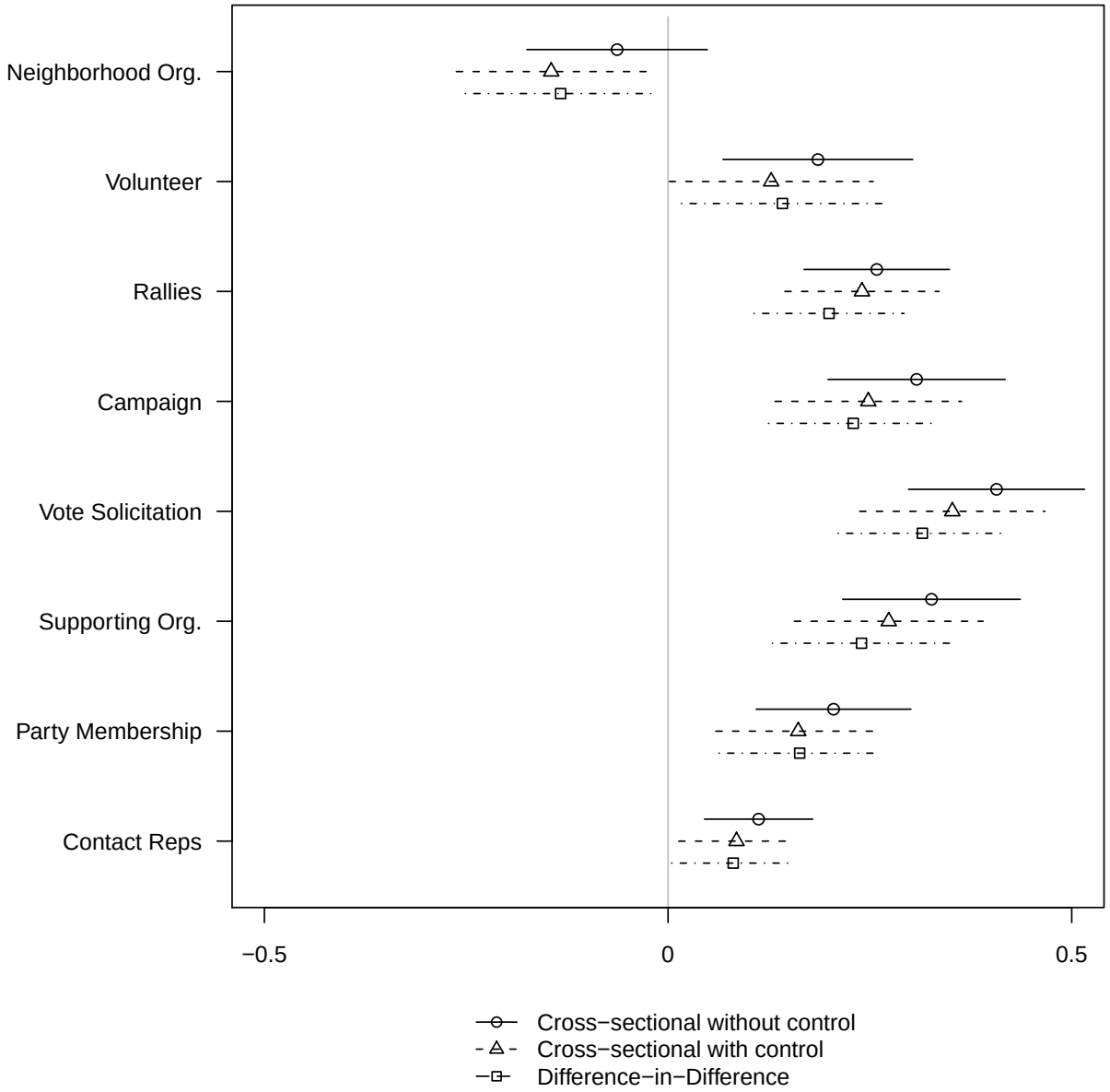

Figure 3: Comparing the Estimated Effect of Survey Modes on the Anticipated Level of Political Participation in the Future

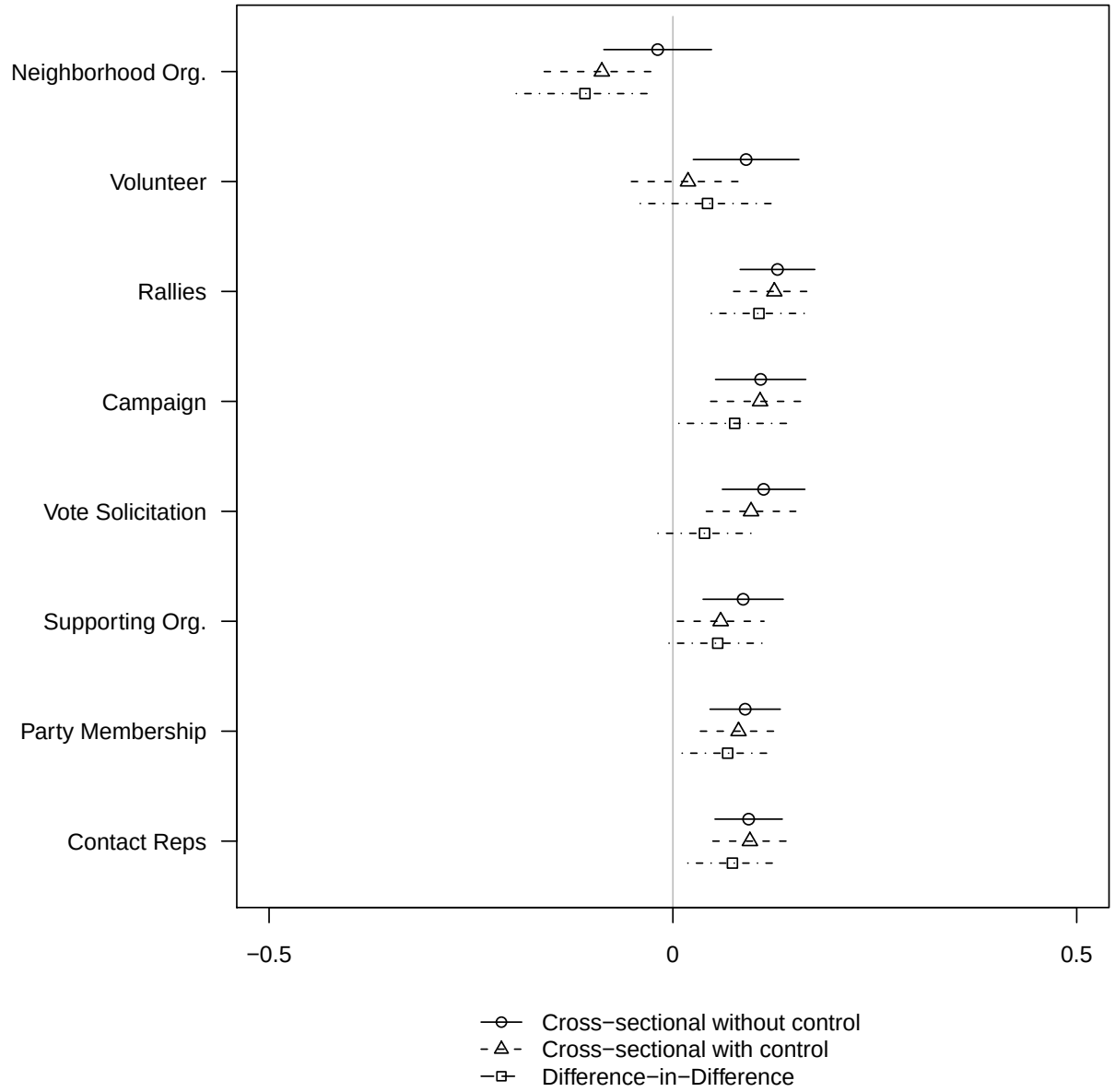

Figure 4: Comparing the Estimated Effect of Survey Modes on the Reported Level of Civic Engagement