Same Respondents, Same Questions; Different Modes, Different Responses

Jason D. Stockdale, Vanessa E. Thornburg, and Arnie P. Aldridge
RTI International

May 2008

Abstract

Survey research shows in many instances that respondents report higher levels of sensitive behaviors during an in-person interview than during a telephone interview (Beck, et al., 2002; Woltman, et al., 1980). Specifically, researchers have found that respondents provide higher levels of alcohol, tobacco, and illicit drug use during in-person interviews, and report lower levels of those behaviors when a telephone interviewer asks the same set of questions. This "mode effect" is found when comparing different sets of CAPI and CATI respondents, and our paper examines whether this mode effect remains salient during a longitudinal survey using the same set of respondents. This paper addresses this question by analyzing data collected from the Screening, Brief Intervention, Referral and Treatment (SBIRT) study, funded by the Center for Substance Abuse Treatment in the Substance Abuse and Mental Health Services Administration. Data were analyzed from a CAPI baseline interview and a CATI six month follow up interview with the same respondents and the same questions. Findings show that a significant number of respondents who answered that they had used alcohol or illicit drugs in their lifetime during the CAPI baseline interview answered that they had not used those drugs during their lifetime in the CATI follow-up interview. This paper further investigates the established mode effect between CATI and CAPI interviewing using a within subjects design. Based on our findings, we compare the accuracy of prevalence data collected from a CAPI and CATI study. We also take a look at other mediating factors that may play a role in diminishing this mode effect.

Introduction

Since its inception as a formal science in the 1930's, survey research has evolved to take advantage of cutting edge technologies designed to improve survey outcomes. From paper and pencil interviewing to computer assisted interviewing using in-person, telephone, IVR, and web modes of administration, survey researchers now have numerous modes to choose from when collecting data. Even with these improvements, however, research has shown that the mode used to collect data can have a direct impact on the types of responses a survey participant provides. In fact, the impact of the mode used is often more pronounced when survey questions involve sensitive topics, such as or alcohol and illicit drug use, or other health-related behaviors.

This paper examines preliminary data from a national, cross-site evaluation of a program designed to address health related behaviors. These data are gathered from the same respondent sample group across two survey collection periods, and we draw on this unique opportunity to look for detectable interview administration mode effects in this longitudinal format. Specifically, we look to see if the mode effect demonstrated in the survey literature remains evident when our respondents are asked the same set of sensitive questions during their CAPI baseline and CATI sixth month follow-up interviews. First, we review the literature on mode effects to provide the context for our research. Next, we describe our sample group and offer evidence of the CAPI-CATI mode effect, and we examine respondent demographic data to see what intervening influence those characteristics may have on the mode effect.

Background

Survey research literature reports there is a tendency for research participants to portray themselves in the most positive light possible (Dovidio and Fazio, 1992). This tendency, referred to as the "social desirability bias," is defined as the response bias introduced when a participant provides responses in a way that favors social approval (Dillman, 2006). In addition, survey research has shown that the mode

of interview administration influences the levels of certain behaviors reported by respondents. Moreover, this social desirability bias seems to be more prevalent when questions focus on sensitive topics such as alcohol, tobacco, illicit drug use, and health-related behaviors. Researchers agree that interview mode influences the degree to which social desirability bias affects reporting out of certain respondent behaviors (Holbrook, et al, 2003; Roberts, et al 2006).

Survey research literature provides a reasonable hierarchy of interview administration modes that mitigate social desirability bias. Prior to the use of computerized interviewing techniques, Woltman, et al. (1980) found that respondents reported higher levels of crime when they were interviewed in-person versus via telephone. Holbrook, et al. found that telephone respondents were more likely to be influenced by social desirability issues than with other modes. Roberts, et al. (2006) noted that face-to-face interviewing, in general, provides some advantages over telephone interviewing that diminish the social desirability bias. Eicheldinger, et al. (2003) and Roberts, et al. (2006) concluded that telephone respondents may be more influenced by social desirability than respondents interviewed with other mode (Woltman, et al., 1980; McHorney, et al 1994; Beck, et al., 2002).

Researchers postulate that face-to-face interview modes seem to offer the interviewer and respondent a variety of visual cues to guide their progress through the interview that are lacking during telephone interviews. During face-to-face interviews, the respondent's attention is more likely to be focused solely on the interview, not permitting the distractions that could occur during a telephone interview. In addition, the face-to-face interview itself typically runs at a "slower pace" than a telephone interview. This slower pace offers the respondent more time to fully understand questions and thus provide more accurate responses than may be provided in a telephone format. In addition, the literature suggests that computer-assisted modes foster a greater sense of privacy and increase the willingness of respondents to provide higher levels of reporting sensitive behaviors, including higher levels of alcohol, tobacco, and illicit drug use (Tourangeau and Smith 1996; De Leeuw, 1998; Wright, 2001; Brenner, et al, 2006). Taken together, these factors result in a more intimate interview setting that helps to create rapport between the interviewer and respondent and thus diminish the effects of social desirability bias (Holbrook; Roberts).

The survey research described thus far are based on mode effect data derived from *different* sets of respondents for face-to-face and telephone interview modes of administration. The study data that we examine give us the unique opportunity to examine whether these mode effects hold true for the *same* set of respondents and the *same* questions across CAPI and CATI interview administration modes.

Study Data

Sample

Our sample comes from data collected to support the Cross-Site Evaluation of the National Screening, Brief Intervention, Referral and Treatment (SBIRT) initiative. Funded by the Center for Substance Abuse Treatment (CSAT) in the Substance Abuse and Mental Health Services Administration (SAMHSA), SBIRT is designed to demonstrate evidence-based screening and interventions for at risk substance use in variety of health care systems in the US. A significant body of literature describes the success of SBIRT in reducing substance use behaviors of individuals (Babor, TF, et al 2007). The philosophy of SBIRT is to screen individuals presenting for general health care to determine if they are at-risk substance users and provide them with a brief intervention, brief treatment or refer them to specialist treatment. Screening consists of questions concerning an individual's substance use behaviors. Only individuals who screen positive for risky levels of behavior receive an intervention.

Since receiving an intervention may change attitudes about substance use, and, thus, decisions to admit substance use, this paper focuses on examination of data from individuals who screen negative and, therefore, receive no intervention. Individuals were recruited into the study between March 2007 and December 2007 from SBIRT programs in New Mexico, Pennsylvania, Texas, and Illinois.

Data Collection Protocol

Individuals are screened by SBIRT practitioners in primary care and public health clinics, emergency departments, and as hospital inpatients. After screening them, SBIRT practitioners introduced interested potential respondents to an RTI field interviewer (FI) who was located in the health care facility. The FIs used Tablet PCs (handheld computers) to conduct CAPI interviews with the respondents. (We conducted CAPI rather than ACASI interviews because the required Tablet PC did not contain the memory capabilities required for auditory files.) The study followed up with respondents six months after baseline using CATI to conduct a survey identical to the baseline[1]. Table 1 displays the demographic characteristics of the screen negative sample and includes characteristics for the total sample for contextual purposes only.

Table 1

Sample Respondent Demographics

Variable	Screen Negative/No Intervention	Total Sample
Emergency Department	29.7%	25.9%
Inpatient	35.8%	45.5%
Female	70.1%	60.9%
Black	36.8%	43.15
Other Race	22.7%	21.7%
Hispanic	32.1%	28.6%
Age	42.8 years	42.3 years
Married	30.3%	26%
Employed	44.9%	41.2%
High School Graduate	61.1%	60%
College Degree or Higher	9.0%	7.7%
N	489	718

In addition to demographic and other study data, the survey collected information on substance use via the Alcohol, Smoking, Substance Involvement Severity Test (ASSIST). Developed by the World Health Organization (WHO), the ASSIST has been successfully tested for validity and reliability as a screening instrument for risky substance use (World Health Organization [WHO] *Addiction* 97:1183-1194). The structure of the ASSIST is relevant to this study because it asks a series of questions about ten drugs or drug categories including alcohol. The first series of questions are gate questions that ask whether a respondent has *ever* used each substance, during their lifetime. Alcohol precedes all other substances except tobacco. After lifetime use has been determined for each substance, subsequent questions are asked about more recent use for each positive response. It's

[1] The follow-up survey did include several new questions, but they were asked after the substance use sections.

important to note that since all of the gate questions are asked together, a respondent is less likely to learn that there are gates and thus change their responses to shorten the interview[2].

Results

We first investigated how often our respondents changed their answers to the lifetime use questions between their two interviews. Table 2 shows the percent of screen negative individuals who responded positively at baseline that they had used a particular substance in their lifetime, but provided a negative response at follow-up. Negative screened lifetime alcohol and marijuana use respondents exhibit the greatest probability of changing their responses from "have used" to "never used" at 12.5% and 8.8% respectively. The number of respondents changing from "have used" to "never used" is significantly larger than the opposite change for alcohol, marijuana, amphetamines, and hallucinogens, and for the rest of the substances, there does not appear to be much difference. Therefore, we focus our attention on respondents who change from positive to negative lifetime use.

Table 2

Percent Changing From Lifetime Use to No Lifetime Use

Substance	Screen Negative/ No Intervention
Alcohol	12.50%
Marijuana	8.80%
Cocaine	6.10%
Amphetamines	3.50%
Inhalants	1.00%
Sedatives	6.10%
Hallucinogens	2.00%
Opiates	6.10%
Other	0.40%
N	489

[2] The mean and median times to complete the ASSIST portions of the interview were similar between CATI and CAPI and were less than three minutes.

Next we investigated whether the number of respondents changing from positive to negative use might simply be due to random variation and, therefore, not be statistically significantly different for the different modes. In Table 3 we show the results of t-tests of net changes in lifetime use responses between the two modes. The change to "never used" is statistically significantly different from the option of no change for alcohol ($\alpha = .01$), marijuana, and cocaine ($\alpha = .1$). This is true even after including the new "takeups", that is, those respondents who reported having lifetime use during the follow-up who had reported no lifetime use at baseline. In other words, respondents' level of reporting for these substances is significantly lower during their six month follow-up CATI interview than during their baseline CAPI interview.

Table 3

P Values from One-Tailed T-tests of the Net* Proportion of Individuals Changing Lifetime Use Status

Substance	Screen Negative/ No Intervention
Alcohol	0.001
Marijuana	0.077
Cocaine	0.070
Amphetamines	0.214
Inhalants	0.873
Sedatives	0.883
Hallucinogens	0.322
Opiates	0.339
Other	0.327
N	489

Furthermore, we examined the data to determine if the decrease in respondents' reports of alcohol, marijuana, and cocaine use was due to them learning about the gate questions during their baseline CAPI interview. If they want a shorter interview then they would simply say no to the lifetime use questions and avoid any follow up questions. However, 621 (86%) of the respondents reported lifetime use of at least one substance during their CATI follow-up interview, while only 270 respondents switched from yes to no for lifetime use. Of those 270 who switched their lifetime use answers, 230 (85%) changed their lifetime use status on at least one substance, making it less likely that it was due simply to learning the gate questions. Although our data cannot rule out the existence of gate learner effect, we believe our data still show that the social desirability mode effect remains a significant explanation for our respondents' switched answers in their CATI interviews.

Next, we took a look at the role that respondent demographic characteristics may play in mitigating our mode effects. Table 4 displays the results of an ordinary least squares (OLS) regression[3] of an indicator of changing from lifetime use of alcohol to no lifetime use of alcohol on multiple respondent

[3] * We chose not to use a logit or probit model to estimate the binary outcomes for several reasons. Under most conditions, they produce close approximates of the OLS linear probability model. The OLS model produces coefficients that are marginal effects without any retransformation. Finally, the nonlinear models may yield biased coefficients for categorical regressors.

characteristics. The models are estimated using the sample of respondents who received no intervention, and a subsample of that group who reported use of at least one substance at follow-up. Employment status stands out as the respondent characteristic that influences the probability of respondents switching from lifetime use to no lifetime use. Specifically, individuals who are currently employed are 7 percentage points less likely to switch from lifetime "have used" to "never used" responses. This calculation is statistically significant at the .01 level. Overall, the remainder of these characteristics is not strongly associated with higher probabilities of switching from positive to negative lifetime alcohol use.

Table 4

Estimates from a Linear Probability Model (OLS) of Changing from Lifetime Use of Alcohol to No Lifetime Use of Alcohol

Variable	Screen Negative/ No Intervention	No Intervention and One or More Substances Reported at Follow-up
Received Intervention		
Emergency Department	0.051 [0.047]	0.006 [0.039]
Inpatient	0.029 [0.041]	0.007 [0.034]
Female	0.049 [0.056]	0.031 [0.045]
Black	0.035 [0.063]	0.034 [0.051]
Other Race Category	-0.062 [0.085]	-0.042 [0.068]
Black and Female	-0.034 [0.076]	-0.001 [0.062]
Other Race and Female	0.074 [0.091]	0.095 [0.076]
Hispanic	0.097 [0.049]*	0.021 [0.041]
Age	0.046 [0.079]	0.028 [0.060]
Married	-0.053 [0.035]	-0.006 [0.030]
Employed	-0.084 [0.032]**	-0.052 [0.027]
High School Graduate	-0.056 [0.036]	-0.075 [0.031]*
College Degree or Higher	-0.061 [0.058]	-0.082 [0.047]
N	485	393

Standard errors in brackets

* significant at p<.05; ** significant at p<.01

We also ran the same models for the same samples using an indicator variable for a switch from positive to negative on *any* substance as the dependent variable. Employment stands out more strongly in these models and its impact is statistically significant at the .05 level in all models. These results are shown in Table 5.

Table 5

Estimates from a Linear Probability Model (OLS) of Changing from Lifetime Use to No Lifetime Use of One or More Substances

Variable	Screen Negative/ No Intervention	No Intervention and One or More Substances Reported at Follow-up
Received Intervention		
Emergency Department	-0.056 [0.067]	-0.109 [0.073]
Inpatient	0.071 [0.059]	0.064 [0.063]
Female	-0.07 [0.080]	-0.065 [0.083]
Black	-0.034 [0.091]	-0.04 [0.094]
Other Race Category	-0.17 [0.121]	-0.18 [0.126]
Black and Female	-0.034 [0.108]	-0.01 [0.114]
Other Race and Female	0.024 [0.131]	0.016 [0.140]
Hispanic	0.113 [0.070]	0.082 [0.076]
Age	-0.156 [0.113]	-0.147 [0.112]
Married	-0.101 [0.050]*	-0.041 [0.055]
Employed	-0.15 [0.046]**	-0.153 [0.050]**
High School Graduate	-0.07 [0.051]	-0.138 [0.057]*
College Degree or Higher	-0.136 [0.083]	-0.218 [0.088]*
N	485	393

Standard errors in brackets

* significant at p<.05; ** significant at p<.01

Finally, we estimate count models (negative binomial regressions) to investigate how different respondent characteristics may be associated with the number of substances that are switched. As in the above models, the impact of employment shown in Table 6 is significant and robust regardless of the sample being analyzed; we can conclude that employed respondents are less likely to switch their lifetime use responses.

Table 6

Estimates from a Negative Binomial Regression of the Number of Changes from Lifetime Use to No Lifetime Use

Variable	Screen Negative/ No Intervention	No Intervention and One or More Substances Reported at Follow-up
Received Intervention		
Emergency Department	-0.155 [0.239]	-0.193 [0.264]
Inpatient	0.209 [0.201]	0.085 [0.223]
Female	-0.343 [0.260]	-0.19 [0.277]
Black	-0.176 [0.286]	-0.191 [0.313]
Other Race Category	-0.547 [0.412]	-0.566 [0.453]
Black and Female	0.01 [0.351]	-0.053 [0.386]
Other Race and Female	0.283 [0.447]	0.296 [0.499]
Hispanic	0.157 [0.244]	0.136 [0.269]
Age	-0.172 [0.323]	-0.188 [0.316]
Married	-0.235 [0.176]	-0.137 [0.196]
Employed	-0.489 [0.162]**	-0.573 [0.185]**
High School Graduate	-0.087 [0.164]	-0.304 [0.182]
College Degree or Higher	-0.579 [0.331]	-0.765 [0.356]*
N	485	393

Standard errors in brackets

* significant at p<.05; ** significant at p<.01

Discussion

Between baseline CAPI and six month follow-up CATI we found strong evidence that individuals switched their reports of lifetime use of certain illicit substances from “have used” to “not used." Taking into account the survey literature on mode effects, we believe that our findings illustrate the possibility of the social desirability mode effect in action. Our sample respondents’ baseline interviews

were conducted in the more intimate setting of a face-to-face interview in a healthcare facility (discussed below). Six months later our respondents were interviewed over the telephone and asked the very same questions they were asked during their face-to-face interview. Our data show that the gate question learning effect cannot fully explain the large number of response switches between CAPI and CATI.

Furthermore, our data show that being employed was the main characteristic that diminished the mode effect we found with our respondents. That is, our respondents who were employed at the follow-up interview were less likely to change their lifetime illicit substance use responses from "have used" to "not used." One hypothesis is that being employed is associated with respondents being more confident and forthcoming about substance use regardless of mode, implying that the mode effect appears to be less of a concern with "higher functioning" population, including those who are employed.

As we review our respondent data, we point out a few issues worthy of further discussion. First, we wonder if the mode effects that we found for our respondents were to some extent more pronounced due to the health care setting for the baseline face-to-face interview. Our respondents were interviewed in a hospital or health clinic setting, and this more intimate setting may have contributed to respondents' willingness to report higher levels of illicit substance use. Second, we would like to see the gate learner effect explored further in other studies. If a CATI follow-up interview were conducted closer in time to the baseline interview we would look to see if the gate learner effect might diminish the social desirability mode effect.

Finally, we believe that our data results are consistent with social desirability bias found when comparing sensitive survey responses between CAPI and CATI interview administration modes, even when the same respondents are asked the same questions in both modes.

References

Aquillino. “Interview Mode Effects in Surveys of Drug and Alcohol Abuse”. 1994.

Babor TF, McRee BG, Kassebaum PA, Grimaldi PL, Ahmed K, Bray J Subst Abus. 2007;28(3):7-30. “Screening, Brief Intervention, and Referral to Treatment (SBIRT): toward a public health approach to the management of substance abuse.”

Beck "Impact of Data Collection Methodology on Reporting of Illicit Drug Use by Adolescents” 2001.

Brenner. "The Association of Survey Setting and Mode with Self-reported health risk behaviors among high school students” 2006.

DeLeeuw. “The Effect of Computer-Assisted Interviewing on Data Quality: A Review” 1998.

Dillman, Don A. “Why Choice of Survey Mode makes a Difference” *Public Health Reports.* Volume 121. January 2006.

Eicheldinger CR, Iannacchione VG., Carpenter LC. “Rating Medicare: does mode matter?” *Proceedings of the American Statistical Association, Survey Research Methods Section.* Alexandria, VA: American Statistical Association; 2003. pp. 1318–1324.

Holbrook, A. L., Green, M. C. and Krosnick, J. A. (2003) 'Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias', Public Opinion Quarterly, 67(1): 79-125.

McHorney CA, Kosinski M, Ware JE “Comparisons of the Costs and Quality of Norms for the SF-36 Health Survey Collected by Mail Versus Telephone Interviews: Results from a National Survey” *Medical Care*. 32:6:551-567 1994.

Roberts, CA, Annette Jäckle, A and Peter Lynn, P “Causes of Mode Effects: Separating out Interviewer and Stimulus Effects in Comparisons of Face-to-Face and Telephone Surveys” AAPOR: ASA Section on Survey Research Methods, 2006.

Tourangeau, R, Smith, TW “Asking Sensitive Questions” *Public Opinion Quarterly.* Vol 60, No. 2, p. 275.

Woltman. “A Comparison of Three Mixed-Mode Interviewing Procedures in the National Crime Survey” 1980.

World Health Organization (WHO) ASSIST Working Group. “The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Development, Reliability and Feasibility.” *Addiction* 97:1183-1194. 2002.

Wright, Aquilino and Supp “Comparison of Computer-Assisted and Paper and Pencil Self Administered Questionnaires in a Survey on Smoking, Alcohol and Drug Use” *Public Opinion Quarterly.* Vol 62, No. 3, Pp331-353.