

Jason D. Stockdale¹, Vanessa E. Thornburg², and Arnie P. Aldridge³
^{1,2,3} RTI International, 3040 Cornwallis Rd, RTP NC 27709

Abstract

There are many instances in which survey research shows 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 often 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, illustrating a “mode effect.” The particular mode effect which is the focus of our paper occurs when a respondent provides more socially acceptable or “socially desirable” answers to the interviewer; most mode effect literature finds this effect more prevalent during CATI interview mode than during CAPI interview mode. Our paper examines whether the social desirability mode effect remains salient during a longitudinal survey using the same set of respondents asked the exact same set of questions first in CAPI mode and then in CATI mode. The 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 baseline CAPI interview answered that they had *not* used those drugs during their lifetime in the follow-up CATI 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 this mode effect.

Key Words: Mode effects, CAPI, CATI, Social desirability bias, p-value

1. Introduction

Since its inception as a formal science in the 1930's, survey research has evolved to take advantage of cutting edge technologies, using them to help design and improve surveys. From paper and pencil interviewing to computer assisted interviewing using in-person, telephone, IVR, and web modes of administration, survey researchers now have a variety of modes to choose from when collecting data. Even with improvements in technology, 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 alcohol and illicit drug use, or other health-related behaviors (Aquillino, 1994; Beck, 2001; Dillman, 2006).

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 time periods. Using this unique opportunity from the SBIRT survey, we test whether the mode effect demonstrated in the survey literature remains evident when our respondents are asked the same set of sensitive questions during their baseline CAPI interview and six months later during their CATI interview. 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 to CATI mode effect; and finally, we examine respondent demographic data to test for what, if any, influence those characteristics may have on the mode effect.

2. Background

Survey research literature reports that 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). Survey research has shown that the mode of interview administration influences the levels of certain behaviors reported by respondents. This social desirability bias seems to be more pronounced when questions focus on sensitive topics such as alcohol, tobacco, illicit drug use, and health-related behaviors. And, researchers seem to agree that interview mode influences the degree to which social desirability bias affects the reporting of the prevalence 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 criminal behavior 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, findings we discuss in more detail below. 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 modes (Woltman, et al., 1980; McHorney, et al 1994; Beck, et al., 2002).

Researchers postulate that face-to-face interview modes can offset the social desirability bias a number of ways. First, they 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. And, during face-to-face interviews, the respondent's attention is more likely to be focused solely on the interview, not permitting 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, 2003; Roberts, 2006).

It is important to note that the survey literature regarding social desirability mode effects we have described above is based on data derived from *different* sets of respondents from face-to-face and telephone interview modes of administration. The study design described in this paper gives 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.

3. Study Data

3.1 Study 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 United States. A significant body of literature describes the success of SBIRT in reducing substance use behaviors of individuals (Babor, 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.

3.2 Study 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 baseline¹.

Table 1 displays the demographic characteristics of the screen negative sample (our focus for this paper) and includes characteristics for the total sample for contextual purposes only.

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

Table 1
Sample Respondent Demographics

| Variable | Screen Negative/No Intervention | Total Sample |
|--------------------------|---------------------------------------|-----------------|
| Emergency Department | 26.9% | 22.5% |
| Inpatient | 41.9% | 52.7% |
| Female | 69.8% | 59.6% |
| Black | 42.3% | 49.6% |
| Other Race | 31.7% | 27.8% |
| Hispanic | 30.0% | 25.9% |
| Age | 42.8 years | 42.4 years |
| Married | 30.2% | 25.5% |
| Employed | 43.5% | 48.9% |
| High School Graduate | 62.2% | 60.7% |
| College Degree or Higher | 7.4% | 6.3% |
| N | 540 | 826 |

In addition to demographic and other study data, the study 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 were asked about more recent use for each positive response. It is important to note that all questions about lifetime use were asked before the follow-up questions about those substances they said they had used; therefore, a respondent was less likely to learn that if they answered "yes" they got more questions. So, they were less likely to change their responses in order to shorten the interview.²

4. Analysis and 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 individuals who reported at CAPI baseline that they had used a particular substance in their lifetime, but reported that they had not at CATI follow-up. The reporting of lifetime alcohol and marijuana use exhibit the greatest probability of changing responses from "have used" to "never used" at 12.5% and 8.8% respectively. We focus our attention on respondents who change from positive at baseline to negative at follow-up lifetime use for two reasons. Changing from "never used" at baseline to "have used" at follow-up, while unlikely over a six month period, is logically possible. Focusing on the purely inconsistent change presents a stronger argument for a mode effect. Furthermore, in Table 3 that follows, we see evidence that the number of respondents changing from lifetime "have used" to lifetime "never used," is significantly larger than the opposite change for alcohol, marijuana, and cocaine, the three most prevalent substances in the population at both time points.

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.

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 |

Table 3 also investigates whether the number of respondents changing from lifetime “have used” to lifetime “never used” might simply be due to random variation and, therefore, not be statistically 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 though we included 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’ levels of reporting for these substances are significantly lower during their six month follow-up CATI interview than during their baseline CAPI interview. This implies that the changes in substance use reporting do systematically vary with the mode and are not entirely due to random variation.

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 | 0.055* |
| Cocaine | 0.093* |
| Amphetamines | 0.117 |
| Inhalants | 0.817 |
| Sedatives | 0.892 |
| Hallucinogens | 0.290 |
| Opiates | 0.578 |
| Other | 0.327 |
| N | 543 |

We also 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 wanted 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' differing answers in their CAPI and CATI interviews.

Next, we look at the role that respondent demographic characteristics may play in influencing the mode effects. Table 4 displays the results of an ordinary least squares (OLS) regression³ of an indicator of changing from reported lifetime use of alcohol to no lifetime use of alcohol on multiple respondent characteristics. Again, the models are estimated using the sample of respondents who received no intervention and also 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 8 percentage points less likely to switch from lifetime "have used" to lifetime "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, although the estimate for Married is suggestive. For the subsample of individuals who still reported at least one substance at follow-up, having graduated from High School was negatively associated with the illogical change.

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 |
|--------------------------|-------------------------------------|--|
| Emergency Department | 0.076 [0.048] | 0.0287 [0.03965] |
| Inpatient | 0.087 [0.048] | 0.008 [0.042] |
| Female | 0.072 [0.061] | 0.068 [0.048] |
| Black | 0.069 [0.063] | 0.073 [0.050] |
| Other Race Category | -0.027 [0.121] | 0.088 [0.099] |
| Black and Female | -0.096 [0.076] | -0.058 [0.061] |
| Other Race and Female | 0.055 [0.088] | -0.011 [0.071] |
| Hispanic | 0.098 [0.100] | -0.04 [0.084] |
| Age | -0.003 [0.007] | -0.007 [0.005] |
| Married | -0.071 [0.034]* | -0.01 [0.028] |
| Employed | -0.082 [0.030]** | -0.047 [0.025] |
| High School Graduate | -0.052 [0.034] | -0.08 [0.029]** |
| College Degree or Higher | -0.062 [0.056] | -0.087 [0.046] |
| N | 537 | 434 |

Standard errors in brackets

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

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.

The same analysis was calculated using an indicator variable for a switch from positive to negative on *any* substance as the dependent variable. These results are shown in Table 5. Again, Employed stands out more strongly in these models and its impact is statistically significant at the .01 level in all models. Married is negatively correlated with changing the response. For the subsample, the level of education (College Degree or Higher) again plays a role with a statistically significant ($p < .05$) decrease in the probability changing a response for. It should be noted that Age and claiming Hispanic ethnicity are positively associated with changing the response.

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 |
|--------------------------|-------------------------------------|--|
| Emergency Department | -0.039 [0.068] | -0.092 [0.074] |
| Inpatient | -0.018 [0.068] | -0.088 [0.078] |
| Female | -0.042 [0.086] | -0.054 [0.090] |
| Black | -0.019 [0.090] | -0.039 [0.094] |
| Other Race Category | -0.308 [0.172] | -0.298 [0.186] |
| Black and Female | -0.063 [0.108] | -0.004 [0.114] |
| Other Race and Female | 0.01 [0.125] | 0.013 [0.133] |
| Hispanic | 0.343 [0.143]* | 0.292 [0.157] |
| Age | 0.021 [0.009]* | 0.022 [0.010]* |
| Married | -0.118 [0.048]* | -0.052 [0.054] |
| Employed | -0.13 [0.043]** | -0.12 [0.048]* |
| High School Graduate | -0.044 [0.049] | -0.1 [0.055] |
| College Degree or Higher | -0.109 [0.080] | -0.174 [0.086]* |
| N | 537 | 434 |

Standard errors in brackets

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

We also estimated count models (negative binomial regressions) to investigate how different respondent characteristics may be associated with the number of substances that are switched conditional on at least one switch. We find no significant relationships; however, this is likely due to the smaller samples and the low prevalence of poly-drug users.

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 |
|--------------------------|-------------------------------------|--|
| Emergency Department | -0.005 [0.222] | 0.124 [0.269] |
| Inpatient | 0.01254 [0.20676] | 0.0334 [0.26432] |
| Female | -0.168 [0.257] | -0.014 [0.287] |
| Black | -0.138 [0.256] | -0.119 [0.290] |
| Other Race Category | 0.398 [0.804] | 0.249 [0.830] |
| Black and Female | 0.172 [0.324] | 0.036 [0.370] |
| Other Race and Female | 0.073 [0.369] | 0.064 [0.414] |
| Hispanic | -0.591 [0.753] | -0.314 [0.783] |
| Age | 0.009 [0.031] | 0.006 [0.035] |
| Married | 0.054 [0.160] | -0.048 [0.186] |
| Employed | -0.034 [0.144] | -0.091 [0.175] |
| High School Graduate | 0.155 [0.149] | 0.126 [0.173] |
| College Degree or Higher | -0.096 [0.301] | -0.023 [0.339] |
| N | 187 | 143 |

Standard errors in brackets

* significant at $p < .05$; ** significant at $p < .01$ **5. Discussion**

Between CAPI baseline and CATI six month follow-up we found strong evidence that individuals switched their reports of lifetime use of certain illicit substances from “have used” to “never used.” Taking into account the survey literature on mode effects, we believe that our findings illustrate the possibility of a social desirability bias in action that is related to the interview mode effect. 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 does not fully explain the large number of response switches between the CAPI and CATI interview modes.

Furthermore, our data show that being employed was the main characteristic that diminished the social desirability 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. This is likewise the case for the other characteristics associated with consistent responses, namely, married and more highly educated.

As we review our respondent data, we’d like to 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. At baseline our respondents were interviewed in a hospital or health clinic setting, and this more intimate setting may have contributed to our 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.

Finally, we believe that our results are consistent with social desirability bias found when comparing prevalence data for sensitive survey topics between CAPI and CATI interview administration modes. This type of mode effect appears to occur even when the same respondents are asked the same questions in both interview 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.