Survey Research

Are Lower Response Rates Hazardous to Your Health Survey? An Analysis of Three State Telephone Health Surveys

*Michael Davern, Donna McAlpine, Timothy J. Beebe,
Jeanette Ziegenfuss, Todd Rockwood, and Kathleen Thiede Call*

Objective. To examine the impact of response rate variation on survey estimates and costs in three health telephone surveys.

Data Source. Three telephone surveys of noninstitutionalized adults in Minnesota and Oklahoma conducted from 2003 to 2005.

Study Design. We examine differences in demographics and health measures by number of call attempts made before completion of the survey or whether the household initially refused to participate. We compare the point estimates we actually obtained with those we would have obtained with a less aggressive protocol and subsequent lower response rate. We also simulate what the effective sample sizes would have been if less aggressive protocols were followed.

Principal Findings. Unweighted bivariate analyses reveal many differences between early completers and those requiring more contacts and between those who initially refused to participate and those who did not. However, after making standard poststratification adjustments, no statistically significant differences were observed in the key health variables we examined between the early responders and the estimates derived from the full reporting sample.

Conclusions. Our findings demonstrate that for the surveys we examined, larger effective sample sizes (i.e., more statistical power) could have been achieved with the same amount of funding using less aggressive calling protocols. For some studies, money spent on aggressively pursuing high response rates could be better used to increase statistical power and/or to directly examine nonresponse bias.

Key Words. Health survey, response rates, health insurance, survey methods, drug use, health care access

Telephone surveys are commonly used in public health for surveillance, evaluation, and monitoring of important public health topics. The most widely used measure of the quality of such surveys is response rate (Atrostic et al.

2001; Biemer and Lyberg 2003). A working assumption has been that for a survey to be construed as "good," it must attain a high response rate (e.g., 70 percent) (Groves 2006). General population telephone surveys have rarely attained response rates higher than 60–70 percent (Brehm 1993; Centers for Disease Control and Prevention 2006). Moreover, participation in telephone surveys has been dropping rapidly. The median response rate for the 2008 Behavior Risk Factor Surveillance System (BRFSS) survey, a decentralized telephone survey conducted by states, is 53 percent, whereas rates for general population telephone surveys in the late 1980s were typically in the vicinity of 70 percent (Groves et al. 2004). Similar patterns have been observed in the University of Michigan's Survey of Consumer Attitudes, where response rates have declined approximately 1.5 percent every year since 1996, so that by 2003, the response rate was 48 percent (Curtin, Presser, and Singer 2005).

In order to secure the highest possible response rate, many survey organizations make multiple call attempts to a sampled telephone number and make an effort to convert initial refusals into respondents (Frey 1983; Lavrakas 1993; Groves and Lyberg 2001). Most survey vendors typically finalize the status of a telephone call after eight calls (Allison and Yoshida 1989), but if initial participation is low, the available budget allows, and the local Institutional Review Board approves, many more attempts can be made (e.g., 50 or more attempts). Going to such lengths increases costs per case and consequently reduces the total number of completed surveys one could obtain within a given budget (Groves 1989). That is, the effort invested in making many more additional attempts to reach a number or convert an initial refuser could have been put toward reaching a new number that has a higher probability of response on the next attempt (Groves 1989; Triplett 2002). Refusal conversion and making numerous call attempts to a household are also associated with respondent (or nonrespondent) burden. By recontacting those who refused to participate in the survey, researchers create a situation in which people have to refuse again, taking up more of their time and potentially

Address correspondence to Michael Davern, Ph.D., Vice President, Director of Public Health Research, NORC at the University of Chicago, 55 E. Monroe, Suite 3000, Chicago, IL 60603; e-mail: davern-michael@norc.org. Donna McAlpine, Ph.D., Associate Professor, Todd Rockwood, Ph.D., Associate Professor, and Kathleen Thiede Call, Ph.D., Associate Professor, are with the Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, MN. Timothy J. Beebe, Ph.D., Associate Professor of Health Services Research, and Jeanette Ziegenfuss, Ph.D., Assistant Professor of Health Services Research, are with the Division of Health Care Policy and Research, Mayo Clinic College of Medicine, Rochester, MN.

angering them. Similarly, many calls to a household that screens telephone calls may create a fair amount of annoyance and burden (even for those who may never answer the phone).

It is not clear whether multiple contact attempts over a protracted period of time or refusal conversion are worth the additional resources required. On the one hand, multiple contact attempts and refusal conversions do increase the response rate. On the other hand, these efforts may just be bringing in the same types of people who have already responded and may do little to reduce any "nonresponse bias." Several recent studies suggest that the respondents we work hardest to recruit may be somewhat different in sociodemographic characteristics, but quite similar in their substantive responses, compared with their more accessible and receptive counterparts (Keeter et al. 2000, 2006; Triplett 2002; Blumberg et al. 2005; Holle et al. 2006). For example, Keeter et al. (2000) found that respondents to a telephone survey with a 36 percent response rate were significantly different on characteristics such as race/ethnicity and socioeconomic status than an identical survey that deployed more rigorous contact protocols and attained a response rate of 61 percent. However, there were very few differences in measures of social and political attitudes, including attitudes toward surveys in general. This analysis was updated by Keeter and colleagues in 2006 and was found to hold at even lower response rate levels.

The research that has focused on estimates of health-related variables has reported somewhat mixed findings. For example, Heje, Vedsted, and Olesen (2006) found that the use of a reminder postcard did increase response rates, but the inclusion of individuals who responded due to the postcard did not change key estimates of patients' views of their provider. In contrast, Paganini-Hill et al. (1993) report significant differences between early and late responders to a health survey on emotional health and some measures of health services use. However, they do not assess whether weighting for differential response might minimize the potential bias. Others have suggested that persons who are hardest to reach may be different than those easier to reach in terms of some demographic variables, but accounting for these differences in analyses may negate much of the differential impact on the outcomes of interest (Mishra et al. 1993).

This paper extends this research by examining whether increased efforts to obtain higher response rates affect the estimates of population characteristics that are of interest to health services researchers—health insurance coverage, health status, utilization of care, health behaviors, and the like. Utilizing three surveys that used rigorous methods (up to 50 calls and attempts to

convert initial refusers), we compare the point estimates and effective sample sizes we obtained with those we would have obtained with a less aggressive protocol and subsequent lower response rate.

Similar to Keeter et al.'s (2000, 2006) work, this paper does not speak to the issue of nonresponse bias. The efforts put into higher response rates are often (at least implicitly) justified by the assumption that higher response rates mean lower nonresponse bias. However, like many other studies (with the exception of Groves 2006), we can only examine the impact of aggressive efforts to obtain higher survey response rates on the ultimate estimates of interest as well as both monetary and nonmonetary costs. We examine whether the efforts to increase response rates in three health surveys led to significantly different outcomes than would have been observed with lower response rates. In other words, what did we get for the extra efforts other than a higher response rate?

METHODS

Three data sources were used for this analysis. Two state surveys (Minnesota and Oklahoma) used the Coordinated State Coverage Survey, an instrument designed to measure health insurance coverage and health care access (survey information is available at http://www.shadac.org/content/coordinated-state-coverage-survey-cscs). The third data source used in this analysis is the Minnesota Treatment Needs Assessment Survey (MN Treatment Needs Survey), designed to collect information about substance use (McAlpine, Beebe, and McCoy 2005). All three surveys were fielded by the same survey research center at the University of Minnesota.

The 2004 Oklahoma Health Care Insurance and Access Survey (OK Access Survey) was fielded between March and June 2004. This survey was designed to be representative at both state and substate levels of the noninstitutionalized population of all ages in the state of Oklahoma. A stratified random sample design that disproportionately sampled geographic regions was utilized in order to obtain reliable estimates for American Indian and low-income households. The response rate (AAPOR RR4) was 45 percent (American Association for Public Opinion Research 2006) and the final sample size was 5,847. The 2004 Minnesota Health Access Survey (MN Access Survey) was fielded between July and December 2004. Like the OK Access Survey, the MN Access Survey was designed to be representative at state and regional levels; thus, it utilized a stratified random sample based on geographic regions.

The final response rate (AAPOR RR4) for this survey was 59 percent (American Association for Public Opinion Research 2006) and the final sample size was 13,802. For both of these surveys proxy interviews were allowed and the person most knowledgeable about the household's health insurance coverage answered the survey for the randomly chosen target household member.

The 2004/2005 MN Treatment Needs Survey was designed to obtain estimates of need for substance abuse treatment in the state of Minnesota. The survey instrument used for this investigation was based on the 2002 State Treatment Needs Assessment Program survey core protocol questionnaire designed by the Center for Substance Abuse Treatment (McAlpine, Beebe, and McCoy 2005). This survey utilized a stratified random sample to obtain estimates representative of noninstitutionalized adults living in Minnesota representative for geographic regions and the state as a whole. This design, combined with a Hispanic and Asian surname oversample, also allowed for reliable estimates by ethnicity. The response rate for this survey (AAPOR RR4) was 54 percent (American Association for Public Opinion Research 2006) and the final sample size was 16,891. Proxy interviews were not allowed and one random adult within the household was chosen to participate.

All three surveys were conducted by the same survey center following the same basic calling protocol. The number of sampled elements released into the field was controlled to ensure that there was a range of cases being worked at all times from never attempted to cases with multiple attempts. Attempts to contact a household occurred initially across evenings on weekdays and afternoons/evenings on the weekend. Later attempts integrated weekday afternoons and mornings into the call schedule. Numbers for which there was no response were allowed to "rest" for 3–4 weeks and then were reactivated and attempted once again. Soft refusals were routed to experienced interviewers who attempted to convert the refusal. If the attempted refusal conversion was not successful, then the case was "rested" for 3–4 weeks before the experienced interviewers recontacted the household.

The key outcomes of interest in the MN and OK Access Surveys were measures of health status, access to health care, and health insurance coverage (uninsured, privately or publicly insured). In the MN Treatment Needs Survey, the main outcomes of interest included measures of health status, substance use, and health insurance coverage (see Table 1 for a list of health variables examined from each survey).

Table 1: Dependent Variables and Data Source

Concept	Question	OK Access (2004)	MN Access (2004)	MN Treatment Needs (2004/2005)
Insurance status	Created variable:	X	X	X
Health status	Would you say your health, in general, is excellent, very good, fair, or poor?	X	X	X
Chronic condition	(Not including pregnancy) do you now have any medical conditions that have lasted for at least 3 months?	X		
Depressive symptoms	Created variable: Depression screening score (Kroenke 2003)			X
Serious mental illness	Created variable: Meet criteria for serious mental illness in the past month (Kessler 2003)			X
Confident will get care	How confident are you that you can get the health care you need?		X	
Usual source of care	Is there a regular place that you go for medical care?	X	X	
Dental visit	During the past year did you go to the dentist?		X	
ER/UC in past year	During the past 12 months have you been to a hospital emergency room or urgent care center?	X		
No doctor visit in past 6 months	In the past 6 months how many visits did you make to a doctor's office, outpatient clinic, or any other place for medical care? Do not include overnight hospital stays, emergency room, or urgent care visits.	X		
Alcohol disorder	Created variable: DSM-IV criteria for alcohol abuse or dependence			X
Drug disorder	Created variable: DSM-IV criteria for drug abuse or dependence			X
Binge drink—year	Created variable: Drank 4 (women)/5 (men) drinks on same occasion in past 12 months			X
Binge drink—month	Created variable: Drank 4 (women)/5 (men) drinks on same occasion in past 30 days			X

Table 1. *Continued*

Concept	Question	OK Access (2004)	MN Access (2004)	MN Treatment Needs (2004/2005)
Nondrinker	Created variable: Drank <12 drinks in lifetime or no drinks in past year			X
Prescription drug use (ever)	Created variable: Nonmedical use of prescription drugs ever?			X
Prescription drug use (in last year)	Created variable: Nonmedical use of prescription drugs in last year?			X
Illegal drug use (ever)	Created variable: Ever used illegal drugs in lifetime?			X
Illegal drug use (in last year)	Created variable: Ever used illegal drugs in past year?			X
Smoking status	Created variable: BRFSS smoker—have smoked >100 cigarettes in lifetime and currently smoke some days or every day			X

Note. X indicates that an item is asked on the survey; a blank indicates that it was not asked.

ANALYSIS

We examine the extent to which multiple contact attempts and refusal conversions impact the health survey estimates obtained from three state telephone surveys. Contact disposition was operationalized to include those respondents who completed the survey within 1–4 contact attempts, 5–8 contact attempts, and 9 or more contacts.[1] The refusal conversion disposition was operationalized as those whose household at one time refused to participate. "Hard refusals" are refusals made by people who adamantly refuse to participate when contacted, people who use profane language, and/or who use threats. Hard refusals were not recontacted. Initial "soft refusals" needed at least two contact attempts (the first initial refusal and at least one more) to eventually be converted and typically there was an extended interval between contact attempts (a soft refusal the case is allowed to "rest" for up to 3 weeks).

We ran independent sample t-tests (Table 3) in order to determine whether respondents who initially refused are different than those who did not refuse. We also compare the demographic characteristics of early responders (1–4 contacts) with later responders (5–8 contact and 9 or more contact attempts). This analysis is done unweighted. As a result, the estimates do not control for the fact that some groups of people within the states were oversampled (e.g., areas with high concentrations of minority group members were oversampled in both of the Minnesota surveys). However, this analysis does tell us if there are demographic differences between those respondents who we went to greater lengths to obtain responses from and those we did not. We only highlight as significant those differences showing a p value of $<.01$.[2]

To answer the question of whether our final substantive survey estimates of interest would have been different if we would have followed a less aggressive protocol, we constructed survey weights for three subgroups of respondents in each of the three surveys: (1) persons who took 1–4 attempts to become a complete without any refusals, (2) those who took 1–8 attempts to become a complete without any refusals, and (3) those who took 1–8 attempts including those with at least one refusal. We construct survey weights for the three samples following a similar weighting scheme originally used to weight each dataset. The survey weight takes into account the differential probability of selecting any one respondent and the weight is poststratified to state population control totals by age, sex, race, geography, and ethnicity (Hispanic versus non-Hispanic). We then compare the weighted estimates from each sample with the overall weighted estimate from the survey using t-tests adjusted for nonindependence of the estimates (as each

subsequent estimate includes the new cases plus everyone in the less aggressive estimates).

Our final analysis is a counterfactual analysis that answers the question of how many "effective sample size" cases we would have had in our study if we used a less aggressive telephone attempt protocol within the same overall study budget. Effective sample size is a commonly used survey research analytic tool that expresses the impact of complex survey design effect in terms of how many "simple random sample with replacement" cases would have been needed to produce the same size standard error. The design effect of survey data collected using a complex sampling method is often greater than one for most telephone survey estimates (the design effect is the ratio of the variance of an estimate calculated taking the complex sample design into account divided by the variance of the same estimate had the same number of cases been collected through a simple random sample) (Kish 1965). The effective sample size will typically be lower than the actual sample size[3] and allows us to compare different estimates with different design effects on the same scale (the effective sample size).

The estimates of cost were provided by the survey center that fielded the surveys. A completed survey that was attempted up to 4 times with no refusal conversion costs U.S.$42.40, which is 43 percent of the cost per complete of the full aggressive protocol. The full aggressive protocol was estimated to cost U.S.$98.90 per complete. Increasing the number of maximum attempts to 8, once again without refusal conversion, costs U.S.$60.96, which is 62 percent of the cost per call of the aggressive protocol. Finally, if refusal conversion is added to the maximum of 8 attempts, the costs increase slightly to U.S.$61.96, or 63 percent of the full protocol. Utilizing this information, we simulate how many effective sample size cases we could have obtained if we had spent the entire project budget for data collection under each of the three simulated contact protocols.[4] Although cost structures differ across survey centers, it seems plausible that the ratio of costs for different protocols would be similar, making the analysis somewhat robust to exact costs.

RESULTS

Table 2 shows the differences in response rates under the three hypothetical follow-up protocols and the protocol as implemented in the original study for each of the three surveys. The highest overall response rate was for the 2004 MN Access Survey with 59 percent and the lowest was for the 2004 OK Access

Table 2:   Response Rates under Variously Aggressive Follow-up Protocols in Three State Health Surveys

	OK Access (%)	MN Access (%)	MN Treatment Needs (%)
Total sample (AAPOR RR4)	45	59	54
Removing cases with at least one refusal	39	51	47
Removing cases with 9 or more contacts to complete	36	48	41
Removing cases with 5 or more contacts to complete	26	37	30

Note. All responses are unweighted: 5,847 (OK Access), 13,802 (MN Access), and 16,891 (MN Treatment Needs).

Source: 2004 Oklahoma Health Care Insurance and Access Survey (OK Access), 2004 Minnesota Health Access Survey (MN Access), 2004/2005 Minnesota Treatment Needs Assessment Survey (MN Treatment Needs).

Survey at 45 percent. If we did not include either those who were initial refusals or those people who took 5 or more contact attempts, the response rates would have ranged from a low of 26 percent in the 2004 OK Access Survey to 37 percent in the 2004 MN Access Survey.[5]

Table 3 shows differences in demographic characteristics between respondents who initially refused to participate and those who never refused to complete the survey. It also compares characteristics of early responders and late responders. Overall, similar patterns were observed across surveys with certain exceptions. For example, households that took 5–8 contact attempts to complete or 9 or more contact attempts were more likely to have selected a child target (0–17 years old) than those that took 4 or fewer contact attempts in the two surveys that allowed proxy interviews for this age group (MN and OK Access Surveys). Respondents in the higher contact groups were more likely to be 18–30 years old and less likely to be 65 years or older in all three surveys. Households with at least one initial refusal were less likely to have an interview conducted with someone 30–64 years old and more likely to have an interview conducted with someone 65 years or older. Across all three surveys, households that took 9 or more contact attempts to complete were more likely to be in a metropolitan statistical area (MSA) and conversely less likely to be in non-MSA than early responders (1–4 attempts). Those who took longer to complete were more likely to be Hispanic/Latino and less likely to be white.[6]

Next we examined key estimates and effective sample sizes under simulated contact protocol that are less aggressive than the protocol we used: 1–4 attempts, no refusal conversion; 1–8 attempts, no refusal conversion; and 1–8 with refusal conversion (see Tables 4–6). For the OK Access Survey (Table 4) and the MN Access Survey (Table 5), none of the estimates projected under

Table 3: Selected Demographic Characteristics of Those with an Initial Refusal, and by Number of Call Attempts

	OK Access Survey					MN Access Survey					MN Treatment Needs Survey				
	At Least One Refusal (%)	No Refusal (%)	1–4 Attempts (%)	5–8 Attempts (%)	9 or More Attempts (%)	At Least One Refusal (%)	No Refusal (%)	1–4 Attempts (%)	5–8 Attempts (%)	9 or More Attempts (%)	At Least One Refusal	No Refusal	1–4 Attempts	5–8 Attempts	9 or More Attempts
Age															
0–17 years old	15.5	17.0	14.6	19.0**	20.8**	17.8	16.3	15.0	18.2**	19.9%**	N/A	N/A	N/A	N/A	N/A
18–30 years old	11.2	13.0	11.4	13.0	16.4**	15.4	19.4**	16.3	21.2**	25.5**	14.4	17.8**	15.1	18.8**	21.4**
31–64 years old	42.6	50.0**	47.3	51.9*	51.2	40.7	45.3**	43.6	46.6*	46.2	58.5	65.1**	61.0	69.7**	67.0**
65+ years old	30.8	20.1**	26.8	16.2**	11.6**	26.1	19.0**	25.2	14.0*	8.5**	27.1	17.2**	24.0	11.6**	11.6**
Gender															
Male	44.4	45.0	43.7	45.0	48.3*	45.5	45.9	44.4	47.5*	48.9**	38.4	40.8	38.2	42.4**	44.3**
Female	55.6	55.0	56.3	55.0	51.7*	54.5	54.1	55.6	52.6*	51.1**	61.6	59.2	61.9	57.6**	55.7**
Geographic area															
MSA	50.6	46.3	45.3	46.7	51.3**	49.2	53.7**	50.6	53.5*	61.0**	42.8	51.8**	49.3	52.4**	52.1*
Non-MSA	49.4	53.8	54.7	53.3	48.7**	50.8	46.4**	49.4	46.5*	39.0**	57.2	48.3**	50.8	47.6**	47.9*
Race/ethnicity															
White	83.6	81.3	83.9	79.7*	77.1**	87.2	83.9**	86.4	82.4**	79.3**	82.9	84.0	86.6	82.2**	78.9**
Black	6.1	4.8	4.2	5.1	7.3**	3.2	4.5*	3.4	5.0**	7.0**	2.9	3.5	2.9	3.7	4.3**
American Indian	5.7	5.6	5.5	6.1	5.6	2.6	3.8*	3.7	3.3	3.9	4.2	3.4	3.2	3.4	4.1
Hispanic	3.0	6.2**	4.9	7.3**	6.8	3.7	3.4	3.6	4.8*	5.7**	5.0	5.7	4.4	6.7**	7.4**
Other/multiple	1.7	2.1	1.6	1.9	3.3*	3.3	4.4*	2.9	4.5*	4.0	5.1	3.4*	2.8	4.0*	5.2**

Table 4: Weighted Estimates from Selected Subsets of Respondents Compared with Estimates Using All Cases with Average Effective Sample and Simulated Effective Sample: 2004 Oklahoma Health Care Insurance and Access Survey

Selected Variables	1–4 Attempts to Complete with No Refusal Conversion		1–8 Attempts to Complete with No Refusal Conversion		1–8 Attempts to Complete with Refusal Conversion		Observed Sample with up to 50 Attempts and Refusal Conversion	
	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)
Any public coverage	33.1	1.2	32.2	1.0	32.4	1.0	32.1	0.9
Any private coverage	53.2	1.3	54.0	1.1	53.9	1.0	54.7	0.9
Uninsured	17.6	1.0	17.6	0.9	17.6	0.8	17.3	0.7
At least good health	84.5	1.0	84.9	0.8	84.8	0.8	85.5	0.7
Chronic condition	38.3	1.2	38.0	1.1	37.5	1.0	36.8	0.9
Usual source of care	85.4	1.0	85.2	0.8	84.8	0.8	84.2	0.7
User of ER/UC care in past year	24.8	1.2	25.0	1.0	24.5	1.0	24.5	0.8
No doctor visit past 6 months	26.7	1.2	26.6	1.0	26.6	0.9	26.9	0.8
Average effective sample size								
Simulated effective sample size	6,225		4,506		4,418		2,868	

Notes. Some of the dependent variables are missing values so the number of cases will be slightly lower than the total *n*. All rates within the four subsets are weighted to be representative of the population using poststratification adjustments (by age, race, sex, and geography) of survey design-based weights.

*p<.01, **p<.001 significant difference between rate in subpopulation and the full sample rate with up to 50 attempts and refusal conversion.

Table 5: Weighted Estimates from Selected Subsets of Respondents Compared with Estimates Using All Cases with Average Effective Sample and Simulated Effective Sample: 2004 Minnesota Health Access Survey

Selected Variables	1–4 Attempts to Complete with No Refusal Conversion		1–8 Attempts to Complete with No Refusal Conversion		1–8 Attempts to Complete with Refusal Conversion		Observed Sample with up to 50 Attempts and Refusal Conversion	
	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)
Any public coverage	25.3	0.7	25.0	0.6	24.9	0.5	24.9	0.5
Any private coverage	78.5	0.7	78.7	0.6	79.0	0.5	78.8	0.5
Uninsured	7.2	0.4	7.0	0.4	6.9	0.3	6.9	0.3
At least good health	89.9	0.4	90.1	0.4	90.1	0.4	90.1	0.3
Confident in ability to get care	92.2	0.4	92.5	0.4	92.6	0.3	92.6	0.3
Dental visit in past year	75.6	0.7	76.0	0.6	76.0	0.6	76.3	0.5
Usual source of care	90.4	0.5	90.2	0.4	90.1	0.4	90.1	0.4
Average effective sample size								
Simulated effective sample size	16,767		11,991		11,840		7,672	

Notes. Some of the dependent variables are missing values so the number of cases will be slightly lower than the total *n*. All rates within the four subsets are weighted to be representative of the population using poststratification adjustments (by age, race, sex, and geography) of survey design-based weights.

*$p<.01$, **$p<.001$ significant difference between rate in subpopulation and the full sample rate with up to 50 attempts and refusal conversion.

Table 6:   Weighted Estimates from Selected Subsets of Respondents Compared with Estimates Using All Cases with Average Effective Sample and Simulated Effective Sample: Minnesota Treatment Needs Assessment Survey

Selected Variables	1–4 Attempts to Complete with No Refusal Conversion		1–8 Attempts to Complete with No Refusal Conversion		1–8 Attempts to Complete with Refusal Conversion		Observed Sample with up to 50 Attempts and Refusal Conversion	
	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)	Rate (%)	Standard Error (%)
Any public coverage	27.9	0.7	27.3	0.6	27.2	0.5	26.8	0.5
Any private coverage	79.9	0.6	80.3	0.5	80.4	0.5	80.6	0.4
Uninsured	6.6	0.4	6.6	0.4	6.7	0.3	6.8	0.3
At least good health	83.4	0.6	83.5	0.5	83.6	0.5	84.3	0.4
Depressive symptoms	8.2	0.4	7.8	0.4	7.6	0.3	7.7	0.3
Serious mental illness	2.6	0.3	2.5	0.2	2.4	0.2	2.3	0.2
Alcohol disorder	7.5	0.5	8.0	0.4	7.7	0.4	8.0	0.3
Drug disorder	2.5	0.3	2.4	0.3	2.3	0.3	2.2	0.2
Binge drink in past year	33.8	0.8	34.4	0.7	34.4	0.6	35.0	0.5
Binge drink in past month	17.9	0.7	18.0	0.6	18.2	0.5	18.8	0.5
Nondrinker	25.7	0.7	25.4	0.6	25.4	0.6	24.6	0.5
Any prescription drug use—lifetime	8.8	0.5	9.0	0.4	8.6	0.4	8.5	0.3
Any prescription drug use—year	3.0	0.3	3.2	0.3	3.0	0.3	3.0	0.2
Any illegal drug use—lifetime	42.1	0.8	42.2	0.7	42.0	0.6	42.1	0.6
Any illegal drug use—year	9.1	0.5	9.2	0.5	8.9	0.4	8.8	0.4
Current smoker	21.2	0.7	20.9	0.6	21.1	0.6	20.8	0.5
Former smoker	25.9	0.7	26.0	0.6	26.3	0.5	26.4	0.5
Never smoker	52.9	0.8	53.1	0.7	52.7	0.6	52.7	0.6
Average effective sample size								
Simulated effective sample size	16,483		11,847		11,370		7,474	

Notes. All rates within the four subsets are weighted to be representative of the population using poststratification adjustments (by age, race, sex, and geography) of survey design-based weights.

*$p < .01$, **$p < .001$ significant difference between rate in subpopulation and the full sample rate with up to 50 attempts and refusal conversion.

the varying contact strategies were statistically different from the observed estimates using *t*-tests for comparing two correlated estimates. In other words, although the response rates are much lower under the less aggressive contact scenarios, key estimates are not different. For example, while the response rate in the OK Access Survey was 45 percent, we would have obtained estimates that did not significantly differ for key variables with a simulated response rate of 23 percent under the least aggressive protocol (1–4 attempts with no refusal conversion).

The last two rows of Tables 4–6 also present the average effective sample size across all the estimates that were observed as well as the average effective sample size that would have been generated if we had used the varying contact protocol with the same overall survey budget. For each survey, we would have obtained dramatically larger effective sample sizes had we followed the various less aggressive calling protocols. For example, we would have had an average effective sample size of 6,225 in the OK Access Survey, if we had stopped calling numbers after four attempts. This is more than double the actual average effective sample size (2,868) we obtained. Similarly, in the MN Access Survey, the effective sample size in the least aggressive contact protocol (16,767) was more than twice that we obtained using the most aggressive protocol (7,474).

Table 6 presents parallel estimates from the MN Treatment Needs Survey. In this table, we also find no estimates in the simulated subset of call protocols that are significantly different from the overall sample estimate using *t*-tests. Again this table shows that dramatically larger effective sample sizes could have been achieved with the same costs using less aggressive fielding operations. In the MN Needs Assessment study, the simulated average effective sample size is 16,483 for the least aggressive protocol compared with the observed effective sample size of 7,474. Note that the response rate would have been 27 percent for the less aggressive protocol and 54 percent for the more aggressive protocol, but there were no significant differences in the point estimates.

DISCUSSION

A long-held maxim in telephone survey research is that we should maximize response rates (Frey 1983; Lavrakas 1993; Groves and Lyberg 2001). Our results, however, temper this long-standing belief and demonstrate why Groves (2006) recommended that "[b]lind pursuit of high response rates in

probability samples is unwise; informed pursuit of high response rates is wise" (Groves 2006, p. 668). As Groves points out, surveys with high response rates can have as much nonresponse bias as surveys with lower response rates, and estimates within the same survey (with the same response rate) can have highly variable levels of nonresponse bias. Moreover, as our results show, there are cost and statistical power considerations when deciding to blindly pursue a high response rate.

We found that if we had accepted a lower response rate, our estimates would not have varied significantly from those that we did obtain after aggressive contact attempts. This finding is consistent with earlier work (Keeter et al. 2000, 2006; Triplett 2002; Blumberg et al. 2005; Holle et al. 2006). After reweighting the subset of data elements (adjusting for basic demographic characteristics used to poststratify survey weights), we found no statistically significant differences between the estimates of key variables between less and more aggressive contact protocol. However, the effective sample sizes could have been much larger had we stopped contacts after the fourth attempt. A larger effective sample size has the advantage of improving the precision of the estimates.

There are two additional cost considerations that are often overlooked in the pursuit of high response rates. First, going to greater lengths to achieve higher response rates could pose the risk of measurement error introduced by aggressive follow-up protocols (Groves 1989; Lavrakas 1993; Olson 2006). Second, survey researchers who use aggressive call protocols increase both respondent burden as well as nonrespondent burden. Converting refusals creates extra burden on those who already declined to participate at one point and calling people up to 50 times can create substantial nonrespondent burden for those people who screen their calls and do not pick up unrecognized numbers.

While Groves (2006) argued against "blind pursuit" of response rates, he also argues that the "informed pursuit of high response rates is wise." Informed pursuit requires considering the respondents most likely to be reached by extra efforts to increase response rate—if they are not different after controlling for basic demographic characteristics used in weighting, pursuing a higher response rate may not be the best use of money. He further elaborates on ways to think about nonresponse theoretically to improve surveys (Groves and Peytcheva 2008). Health survey researchers are urged to be thoughtful in their decision to include or exclude excessive methods to reach a higher response rate. At the very least, analyses should be undertaken to assess the merits of choosing higher response rates at the added cost of diminished

statistical power resultant of a fixed survey budget being spent on increasing response rate instead of achieving more completed surveys.

Our analysis speaks to the sensitivity of survey estimates to response rate. In our study, analyzing only the data from easy-to-get respondents compared with data from all respondents yielded essentially the same empirical findings. Thus, there was no evidence that a lower response rate would have been associated with more response bias than the higher response rate. Survey researchers should be cautioned against aggressively chasing high response rates, and thus costing the survey statistical power and increasing respondent burden, when there is not a good reason to believe it will impact overall substantive estimates.

Acknowledging that response rate should not be the summary gold standard data quality measure for judging the success of a survey does not negate the importance of pursing probability samples and following rigorous survey procedures. Probability sampling typically reduces the bias of estimates (Elliott and Haviland 2007; Malhotra and Krosnick 2007) and should be favored over convenience samples. Also, survey researchers should invest more effort to reaching out to sampled elements using mixed modes as there has been evidence that such efforts lead to less-biased survey estimates, whereas pursuing respondents through only the telephone mode does not seem to help reduce bias (Baines et al. 2007). However, when combining survey data collected from different modes, there is the additional risk of introducing mode effects into the survey (Dillman, Smyth, and Christin 2009).

Our analysis does not directly address the issue of nonresponse bias. If people who did not respond to these surveys at all are systematically different than those who did respond, bias in the estimates will remain a serious problem (regardless of whether we use an aggressive protocol). Given that we do not know the characteristics of nonrespondents, we could not assess the degree of nonresponse bias in this study. However, we are not alone in this constraint. In the influential work published by Groves (2006) showing a very slight relationship between absolute nonresponse bias and survey response rates, he located only 30 articles that presented estimates of nonresponse bias.

Clearly, nonresponse bias is one of the most pressing concerns facing survey researchers. The best way to address this issue is to have comparative information from both respondents and nonrespondents from administrative data, medical records, or some other secondary data source. Many of the datasets health services researchers use are derived from sampling frames that are rich with auxiliary information such as demographic characteristics, ser-

vices use, and health status of both respondents and nonrespondents (e.g., health plan data, hospital discharge data, Medicare or Veterans Administration data, etc.). This auxiliary information should be used to examine nonresponse bias explicitly to inform our understanding of what types of survey items are more likely to be biased and what types of people cause the bias to occur by not responding. This type of work is critically important to survey researchers in general (see Groves 2006; Groves and Peytcheva 2008 for theoretical development on what causes nonresponse bias) and health services researchers in particular. Instead of allowing large proportions of research budgets to be aimed at measures that increase response rates, we need to consider models that allow us to better explicitly control for nonresponse bias in our survey estimates and potentially use multimode strategies. We believe that survey funds that go to aggressive contact follow-up attempts can be used for these efforts as well as to increase the overall sample size with a fixed survey budget. Reliance on response rates as being the critical summary measure of survey quality has distracted survey researchers from conducting this type of vital work.

ACKNOWLEDGMENTS

Notes

1. We choose 1–4 contacts because approximately 50 percent of the completed surveys from all three surveys came within this range (see Table 2 for a breakdown). We chose 9 or more because of evidence that most survey vendors finalize the status of a telephone call after eight calls (Allison and Yoshida 1989). As such, we chose 8 contact attempts as the threshold demarcating the transition from standard to extra effort. We have also performed this analysis using number of days a piece of sample was in the field with similar results (results of that analysis are available from the corresponding author).
2. We use .01 as our basic significance level throughout the paper because we have large sample sizes and do not make multiple comparison adjustments as this is an exploratory exercise to uncover patterns (not a theory confirming exercise).
3. For this analysis, we use "effective sample" size numbers as opposed to raw sample size numbers because, in general, the scenarios where we call people fewer times will have a slightly higher design effect due to the larger impact of poststratification adjustments in the survey weights.
4. When examining the tradeoffs with sample size, it is instructive to consider how much bias there is in the estimates by constructing mean squared errors (variance+bias squared) to compare various protocols. However, for reasons we point out in the discussion section, we do not know what the "unbiased" estimate is and assuming the estimates made with the data using the highest response rate are unbiased is not supported by the current evidence (Groves 2006).
5. We do not have the full call disposition history, so the simulated response rates are calculated by adjusting only the numerator number of completes and assuming the denominator remains the same.
6. There were differences in unweighted health-related characteristics as well. Health status and health insurance coverage status are the only two outcomes that are observed in all three surveys. For those measures, we see that those requiring 5–8 contact attempts or 9 or more contact attempts to complete are less likely to have public insurance and in the two Minnesota surveys, they were more likely to be uninsured compared with those who responded early. In all three surveys, those who took longer to become a complete were more likely to be in good health. However, we underscore the fact that our analysis in Tables 3–5 shows that these differences between groups are mitigated after making basic weighting adjustments. This analysis is available from the corresponding author upon request.

References

Allison, K. R., and K. K. Yoshida. 1989. "Increasing Response Rates in Community Health Surveys Administered by Telephone." *Canadian Journal of Public Health* 80 (1): 67–70.

American Association for Public Opinion Research. 2006. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* Lenexa, KS: AAPOR.

Atrostic, B. K., N. Bates, G. Burt, and A. Silberstein. 2001. "Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights." *Journal of Official Statistics* 117 (2): 209–26.

Baines, A., M. R. Partin, M. Davern, and T. Rockwood. 2007. "Mixed Mode Administration Reduced Bias and Enhanced Post-Stratification Adjustments in a Health Behavior Survey." *Journal of Clinical Epidemiology* 60 (12): 1246–55.

Biemer, P., and L. Lyberg. 2003. *Introduction to Survey Quality.* New York: Wiley.

Blumberg, S., K. Davis, M. Khare, and M. Martinez. 2005. *The Effect of Survey Follow-up on Nonresponse Bias: Joint Canada/United States Survey of Health, 2002–03.* Paper presented at the annual meeting of the American Association for Public Opinion Research, Miami, FL, May 12–15.

Brehm, J. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation.* Ann Arbor, MI: University of Michigan Press.

Centers for Disease Control and Prevention. 2006. *2005 Behavioral Risk Factor Surveillance System Data Quality Report Handbook.* Atlanta, GA: Centers for Disease Control and Prevention.

Curtin, R., S. Presser, and E. Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69 (1): 87–98.

Dillman, D. A., J. D. Smyth, and L. M. Christin. 2009. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* Hoboken, NJ: Wiley & Sons.

Elliott, M. N., and A. Haviland. 2007. "Use of a Web-Based Convenience Sample to Supplement a Probability Sample." *Survey Methodology* 33 (2): 211–5.

Frey, J. H. 1983. *Survey Research by Telephone.* Beverly Hills, CA: Sage Publications.

Groves, R. M. 1989. *Survey Errors and Survey Costs.* New York: Wiley.

———. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70 (4): 646–75.

Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology.* New York: Wiley.

Groves, R. M., and L. E. Lyberg. 2001. *An Overview of Nonresponse Issues in Telephone Surveys.* New York: Wiley.

Groves, R. M., and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72: 167–89.

Heje, H. N., P. Vedsted, and F. Olesen. 2006. "A Cluster-Randomized Trial of the Significance of a Reminder Procedure in a Patient Evaluation Survey in General Practice." *International Journal for Quality in Health Care* 18: 232–7.

Holle, R., M. Hochadel, P. Reitmeir, C. Meisinger, and H. E. Wichman. 2006. "Prolonged Recruitment Efforts in Health Surveys." *Epidemiology* 17 (6): 639–43.

Keeter, S., C. Kennedy, M. Dimock, J. Best, and P. Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70 (4): 125–48.

Keeter, S., A. Kohut, A. Miller, R. Groves, and S. Presser. 2000. "Consequences of Reducing Non-Response in a Large National Telephone Survey." *Public Opinion Quarterly* 64 (2): 125–48.

Kessler, R. C., P. R. Barker, L. J. Colpe, J. F. Epstein, J. C. Gfroerer, E. Hiripi, M. J. Howes, S. L. Normand, R. W. Manderscheid, E. E. Walters, and A. M. Zaslavsky. (2003). "Screening for Serious Mental Illness in the General Population." *Archives of General Psychiatry* 60 (2): 184–9.

Kish, L. 1965. *Survey Sampling.* New York: John Wiley & Sons, Inc.

Kroenke, K., R. L. Spitzer, and J. B. Williams. (2003). "The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener." *Medical Care* 41 (11): 1284–92.

Lavrakas, P. J. 1993. *Telephone Survey Methods: Sampling, Selection, and Supervision.* Thousand Oaks, CA: Sage Publications.

Malhotra, N., and J. A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis* 15: 286–323.

McAlpine, D. D., T. J. Beebe, and K. McCoy. 2005. *Estimating the Need for Treatment for Substance Abuse among Adults in Minnesota: Results from the 2004/2005 Minnesota Treatment Needs Assessment Survey.* Minneapolis, MN: University of Minnesota.

Mishra, S. I., D. Dooley, R. Catalono, and S. Serxner. 1993. "Telephone Health Surveys: Potential Bias from Noncompletion." *American Journal of Public Health* 83: 94–9.

Olson, K. 2006. "Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias." *Public Opinion Quarterly* 70 (5): 737–58.

Paganini-Hill, A., G. Hsu, A. Chao, and R. K. Ross. 1993. "Comparison of Early and Late Respondents to a Postal Health Survey Questionnaire." *Epidemiology* 4: 375–9.

Triplett, T. 2002. *What Is Gained from Additional Call Attempts and Refusal Conversion and What Are the Cost Implications?* Washington, DC: Urban Institute.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.