

Working PAPER

BY DEBRA WRIGHT, MATT SLOAN, AND KIRSTEN BARRETT

Is There a Trade-off Between Quality and Cost? Telephone Versus Face-to-Face Interviewing of Persons with Disabilities

February 2012

ABSTRACT

Mode of data collection is a key design feature affecting both response and data quality. Particularly when interviewing persons with severe health conditions, researchers must weigh the benefits of approaches which maximize inclusion and comprehension with cost. Mathematica incorporated an experiment into the 2010 National Beneficiary Survey to compare the use of two modes, telephone and face-to-face interviewing, on the quality of data collected from a sample of persons with disabilities. We found some evidence that mode of interview has a modest impact on data quality for this population, particularly for items which are sensitive or complex. We suggest some specific steps that researchers can take to design surveys to minimize mode effects and maximize the quality of the survey data.

I. INTRODUCTION¹

While persons with disabilities are included in general population surveys through random selection alone, many surveys focus on list-frame samples in which all, or most, sample members are presumed to have a severe health condition or disability. Particularly in these cases, researchers must design the survey with accessibility in mind and weigh the cost and benefits of approaches to foster inclusion and adequate representation of this population.

Mode of data collection is a key design feature affecting response. For surveys of persons with disabilities, the choice is often between telephone and face-to-face interviewing. With appropriate accommodations, including the use of teletypewriter (TTY), Telecommunications Relay Service (TRS), amplifiers, and instant messaging technology to interview persons with speech or hearing impairments, it is possible to collect data from people with disabilities by relying exclusively on telephone interviewing (Mitchell et al. 2006). Careful attention to instrument design (including structured probes, questionnaire wording, and built-in stopping points) can offset many cognitive or stamina barriers, as can specialized interviewer training designed to sensitize interviewers to common challenges associated with telephone interviews of persons with disabilities.

Face-to-face interviewing, however, offers several advantages and may be particularly effective both in removing barriers to survey participation among people with disabilities and achieving high response rates. Contacting sample members in person increases opportunities to locate hard-to-reach sample members, offers them the convenience of being interviewed at home or in another setting, and facilitates interviewing persons whose impairments make it difficult for them to participate in telephone interviews. In-person interviewing is considerably more expensive than other data collection modes, however, and may be cost prohibitive for some surveys. To offset these costs, researchers may choose to employ a dual-mode design, in which telephone interviewing is the primary mode and face-to-face interviewing is used as a secondary mode to follow up with telephone nonrespondents.

Another key question to consider is whether the quality of data collected by telephone and face-to-face is comparable. This is important for obtaining accurate estimates, particularly when making comparisons across studies using different data collection modalities and combining data collected using multiple modes. While telephone and in-person interviews are similar because both involve an interviewer, they are different in norms of communication (de Leeuw 2005). For example, while both modes rely on verbal communication, in a face-to-face setting, interviewers may find it easier to match people's pace and communication style and pick up on their confusion or frustration through body language or other cues. This may provide an advantage in the interviewers' ability to build rapport and engage respondents and their willingness to provide information. By contrast, over the telephone, interviewers generally control the pace of the conversation. To keep up with the pace and avoid lapses, telephone respondents may not spend as much time as they would like on any given item (Schwarz 1997). Pressure to answer within a certain amount of time may interfere with

¹ This study was supported by the National Institute on Disability and Rehabilitation Research, U.S. Department of Education, through its Rehabilitation Research and Training Center on Disability Statistics and Demographics grant to Hunter College, CUNY (No. H133B080012-09A). Mathematica Policy Research is a subcontractor under this grant. The contents of this paper do not necessarily represent the policy of the Department of Education or any other federal agency (Edgar, 75.620 (b)). The authors are solely responsible for all views expressed.

cognitive processing and increase “satisficing” behavior, which is a reliance on strategies that simplify the task—for example, overselecting “Don’t Know” responses (Krosnick et al. 2002).

Some evidence suggests that certain response errors are more likely to occur in telephone than face-to-face interviewing. For instance, respondents generally provide less information when asked an open-ended question on the telephone. In addition, telephone respondents have been found to display more acquiescence, choose more extreme categories, refuse more items, and display more evidence of recency effects than in-person respondents (Jordan et al. 1980; Locander and Burton 1976). There is also evidence of differences in respondents’ willingness to report socially undesirable behaviors, although findings across studies are not entirely consistent. For example, Jackle et al. (2010) found that face-to-face survey respondents overreported socially desirable behavior; Sykes and Collins’ (1988) review of four comparative studies also revealed that in-person respondents consistently gave more socially desirable answers. However, Holbrook et al. (2003) reported that telephone respondents are more likely to present themselves in socially desirable ways than respondents interviewed face to face.

Despite these findings, little research has examined the effect of telephone versus face-to-face interviewing for persons with disabilities as a specific population, although there is some evidence that respondents with lower cognitive abilities are more susceptible to mode effects. Chang and Krosnick’s study (2010) comparing responses to a political survey in which sample members were randomly assigned to either a computer-assisted self-administered mode or an interviewer-administered mode revealed that differences in responses were most pronounced for those with more limited cognitive skills. However, respondents to this study were not persons with disabilities, but college students with cognitive ability measured using SAT or ACT scores. Thus, while this study suggests that respondents with minor cognitive limitations may provide better quality data face to face, it is unknown whether this effect would be found for people with both physical and mental limitations and across a range of questions.

Prior research also suggests that self-reports of impairment are susceptible to mode effects. For example, Walsh and Khatuskay (2007) found that respondents’ reporting of their level of disability (number of Activities of Daily Living impairments) in a patient population varied substantially across data collection mode (mail, telephone, face-to-face, and clinical evaluations), with all of the survey-based approaches yielding lower levels of impairment than the clinical evaluations. However, while this sample was composed of individuals with functional impairments, sample members were not randomly assigned to mode of interview, making it difficult to parse out true mode effects from differences possibly due to self-selection into mode.

To fill these gaps in research, we incorporated an experiment into the 2010 National Beneficiary Survey (NBS), a dual-mode survey combining computer-assisted telephone interviewing (CATT) and computer-assisted in-person interviewing (CAPI). The purpose of the experiment was to determine whether there are differences in the quality of data collected by telephone versus face to face in a survey of persons with disabilities that are attributable to the data collection mode.

II. METHODOLOGY

A. Description of National Beneficiary Survey

The NBS, conducted by Mathematica Policy Research and sponsored by the Social Security Administration (SSA), is a multi-round, nationally representative survey of SSA disability beneficiaries and recent Ticket to Work (TTW) participants. The NBS is one of several components of an evaluation to assess the impact of TTW relative to the current system—the SSA Vocational Rehabilitation Reimbursement Program, which has been in place since 1981. A voluntary employment program for people with disabilities, TTW was authorized by the Ticket to Work and Work Incentives Improvement Act of 1999. Under the program, SSA provides beneficiaries with a “Ticket,” or coupon, that they may use to obtain employment-support services, including vocational rehabilitation, from an approved provider of their choice (called Employment Networks or ENs).²

The 45-minute, dual-mode, CATI/CAPI survey gathers information on health, insurance, employment, income, and demographic characteristics. Interviewers first attempt telephone contact and then conduct face-to-face interviews with people whose phone numbers cannot be located and those who request or require an in-person interview, are evasive to telephone attempts, or refuse to participate by telephone. The survey instrument is identical in both modes. Approximately 64 percent of all respondents report having a physical or other impairment as their main limiting condition; 32 percent report a psychiatric or cognitive impairment; and 4 percent report a sensory impairment. If a beneficiary is incapable of responding due to a severe cognitive impairment, the interviewer attempts a proxy interview with a knowledgeable informant. In 2010, the fourth round of the survey was fielded with a sample of 8,017 SSA beneficiaries. A total of 5,078 cases were completed, for a weighted response rate of 73 percent. Overall, 3,936 interviews were completed by CATI and 1,142 by CAPI.

B. Sample and Response Rates for Experiment

Sample members for the NBS are randomly selected from administrative records provided by SSA. The NBS uses a multistage sampling design with a supplemental single-stage sample for some TTW participant populations. For the multistage design, data from SSA on the counts of eligible beneficiaries in each county are used to form the primary sampling units (PSUs), consisting of one or more counties. We selected PSUs with probability proportional to size of their beneficiary population. A sample of SSA beneficiaries then is selected from among beneficiaries residing in these PSUs, using age-defined sampling strata. To ensure a sufficient number of persons seeking work for key analyses, the beneficiary sample is classified into sampling strata based on age, with persons in the younger age categories selected at higher rates than those in the oldest category. The target population for both the national sample of SSA beneficiaries and the TTW participant sample consists of Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI) beneficiaries between ages 18 and 64.

For this experiment, a subset of 645 beneficiaries was randomly selected from the same sample frame used to select the nationally representative beneficiary sample for the NBS. This sample

² For more information on the Ticket to Work program, see Thornton et al. (2004).

comprised the experimental face-to-face-only group. The sample was restricted to beneficiaries ages 18 to 49 to reduce design effects, since 50- to 64-year-olds were undersampled for the overall survey. The comparison group comprised individuals selected from the regular representative beneficiary sample, restricted to 18- to 49-year-olds who completed interviews by telephone. Beneficiaries in the regular sample who did not complete in CATI and were forwarded to face-to-face interviewers (per the study design) were considered nonrespondents for the purposes of this experiment. Interviews completed by proxy were excluded from the analysis. This process yielded a data set with 1,574 observations comprising 1,229 cases completed in CATI and 345 cases completed in CAPI.

The response rate for the experimental sample was 54 percent, compared with 73 percent for the comparison group.³ Although lower than expected, the experimental group response rate was comparable to that of the nonexperimental NBS sample sent to the field for in-person locating and interviewing (54 percent)⁴. While the experimental group had a significantly lower response rate than the nonexperimental group, comparisons between key characteristics revealed few significant differences between respondents in the two groups. There was a somewhat larger number of Hispanics in the experimental group; fewer SSI, SSDI, and more concurrent beneficiaries; and more individuals with early age of disability onset than those in the telephone comparison sample (see Table 1).

Table 1. Characteristics of Face- to- Face and Telephone Respondents (Percentages, Unweighted)

	Face-to-Face (n=345)	Telephone (n=1,229)	All Interviews (n=1,574)
Age			
18-29	32.2	31.3	31.5
30-39	30.7	33.7	33.0
40-49	37.1	35.0	35.5
Sex			
Male	50.7	49.7	50.4
Female	49.3	49.3	49.6
Education			
< High school	35.9	31.0	32.1
High school grad	45.6	46.5	46.3
> High school	18.5	22.5	18.9
Ethnicity*			
Hispanic	17.1	11.9	13.1
Non-Hispanic	82.9	88.1	86.9
Race			
White	62.9	63.6	63.5
Black	26.1	22.9	23.6
Other	11.0	13.5	13.0
Disabling Condition			
Psychiatric	28.7	30.1	29.8
Cognitive	11.8	7.6	8.5
Musculoskeletal	11.8	11.7	11.8
Sensory	4.1	4.7	4.6
Other	43.5	45.9	45.4

³ The higher response rate for telephone is partially explained by the fact that these cases had more contact attempts than those sent directly to the field for face-to-face interviewing (due to the NBS study design).

⁴ This includes cases sent to the field but who ultimately completed the interview by telephone.

	Face-to-Face (n=345)	Telephone (n=1,229)	All Interviews (n=1,574)
Benefit Type*			
SSI	31.9	39.6	37.9
SSDI	25.8	36.2	33.9
SSI and SSDI	42.3	24.2	28.2
Age at Onset*			
< 18	53.9	44.6	46.6
18 and Older	46.1	55.5	53.4

* p ≤ .05

We also used SSA administrative data to compare nonrespondents to responders within each group as an additional check for differential nonresponse. This analysis revealed some differences in benefit status, condition, and age within the experimental group, and for race within the telephone-only group. Although the differences were not large, we used all of these variables as controls in our analyses to minimize the potential impact of differential attrition on our results.

C. Procedures

Data collection procedures for the experimental cases (face-to-face) mirrored those for the regular NBS sample. All sample members received an advance letter on SSA letterhead explaining the study prior to being contacted (the only difference was that letters sent to individuals selected for the experimental group omitted references to a telephone option). The incentive amount, a post-paid check for \$10, was the same for both groups. The questionnaire administered to the experimental group was the same as that administered to the regular NBS sample. Under normal study operations, field interviewers who find a telephone number for an individual assigned to CAPI are instructed to pass this information along to telephone operations if appropriate so that the individual can be interviewed by phone. Because individuals assigned to the experimental group could not be interviewed by telephone even if a phone number was located, field interviewers were told which cases were “face to face” and could not be forwarded to telephone operations. They did not, however, know the purpose of the experiment and were instructed to apply the same effort and methods to locate and interview the experimental as they did for the nonexperimental cases.

D. Measures

To evaluate the data collected by telephone and face to face, we identified four general domains of quality, including (1) item nonresponse, (2) proportion of socially desirable responses, (3) amount of non-differentiation in a series of questions, and (4) acquiescence. We purposively selected items to represent each domain we thought would be the most sensitive to mode effects, focusing on questions that were more subjective, sensitive, or vague, or could be construed as cognitively demanding. Finally, we examined only items administered to all respondents.⁵

⁵ An item that is not asked of all respondents cannot usually be statistically evaluated for mode effects. The routing creates a subset of the matched data set, and this subsetting can lead to incomplete data for some random selected cases (Pierzchala et al. 2005). The implication is that only variables asked of all beneficiaries are used for this analysis.

1. Item Nonresponse

Item nonresponse occurs when a respondent is unable or unwilling to answer a question. To estimate the impact of data collection mode on item nonresponse, we created dummy variables for key items that had more than two percent nonresponse overall to indicate whether the response was valid or missing (“don’t know” or “refused”). These items included age at which the beneficiary first became limited; work goals (“You see yourself {continuing to work/working} for pay in the next year,” and “You see yourself {continuing to work/working} for pay in the next five years”); household income in 2004; and four demographic questions (race, father’s education, mother’s education, and weight). In addition, we created a dummy variable for two series of similar items to indicate whether a response to any question in the series was missing. The first series included five items asking about awareness of particular SSA programs (ever heard of Blind Work Expenses, Expedited Reinstatement, Benefits Specialists, and the Ticket to Work Program). The second series was a number of items asking whether the respondent had received income from any of eight sources in the preceding month.

2. Social Desirability

Individuals may provide socially desirable responses when they are unwilling to admit holding or reporting what may be perceived as an undesirable opinion or behavior (Holbrook et al. 2003; Kreuter et al. 2008). To represent this domain, we selected several items that could be perceived as sensitive for this survey and compared the estimates between modes. The selected items included three work goal items (“Your personal goals include {getting a job/moving up in a job}, or learning new job skills?”; “You see yourself {continuing to work/working} for pay in the next year”; and “You see yourself {continuing to work/working} for pay in the next five years”); alcohol use (“In the past 12 months, have you ever felt you ought to cut down on your drinking?”); drug use (“During the 12 months, have you used drugs on your own more than 5 times?”); and household income in 2004. We considered the work goal items to be potentially sensitive for this survey since a “yes” response to these items could be perceived as providing the answer that SSA desired.

3. Non-Differentiation

Non-differentiation occurs when respondents fail to distinguish between different questions and select the same answer choice on a scale for all, or almost all, similar questions. For analysis of the amount of non-differentiation, we examined response patterns to the series of four health and functional status items (“How much does your physical health limit your usual physical activities such as walking or climbing stairs?”; “During the past 4 weeks, how much difficulty have you had doing daily work, both at home and away from home because of your physical health?”; “During the past 4 weeks, how much did your physical health or emotional problems limit your usual social activities with family and friends?”; and “During the past 4 weeks, how much did personal or emotional problems keep you from doing your usual work, school, or other daily activities?”). The response scale for all items was “Not at all,” “Very little,” “Somewhat,” “Quite a lot,” or “Could not do.” We created a dummy variable indicating whether a respondent chose the same response option for all four items or not.

4. Acquiescence

To measure the amount of acquiescence, we counted the number of “yes” responses to the four items in a series of questions about awareness of SSA programs (heard of impairment-related/individual work expenses, heard of expedited reinstatement, heard of benefits specialist, and

heard of Ticket to Work). These items were chosen because they were asked in the same series and because it was reasonable to assume that respondents inclined to acquiesce might say they had heard of these programs without making the effort to determine whether, in fact, they had.

III. RESULTS

We present the analyses below in four sections, related to the domains outlined above. We used cross-tabulations to compare differences in reporting by mode (using the Chi-Square statistic to test for significance) and t-tests to test the hypothesis of no difference between modes for continuous variables. For these analyses, we excluded proxy respondents and limited the sample to those sampled persons able to respond for themselves.

In addition to the Chi-Square analyses, we conducted sensitivity tests by fitting logistic regression models for all categorical outcome variables showing a significant difference by mode in the cross-tabulations and ordinary least squares regressions for variables with significant t-test statistics. This allowed us to test for mode differences while controlling for sample characteristics that could influence the association of mode and the outcome variable. Control variables included sex; whether the sample person first became limited during childhood or in adulthood (under versus over 18 years of age); race (white, black, other); ethnicity (Hispanic, not Hispanic); age (18–29, 30–39, 40–49); benefit type (SSI, SSDI, or both SSI and SSDI); and condition (mental versus physical/other). We included mode as a dummy variable in all models. For all outcomes with a significant Chi-Square statistic, mode remained significant when these variables were included in the model. Therefore, we present only the results of the Chi-Square analyses below.

A. Item Nonresponse

Among the seven individual items analyzed, only one item—household income—had significantly different rates of missing data by mode. For this item, data were missing for 30.7 percent of telephone respondents and 43.5 percent of face-to-face respondents ($\chi^2 = 19.83$, $p < .0001$; see Table 2). Among the two series of items we analyzed (awareness of SSA work incentive programs and health insurance), item nonresponse for the series of health insurance items was significantly higher for telephone respondents (with 6.8 percent omitting at least one response in the series, compared to 1.5 percent of face-to-face respondents; $\chi^2 = 14.65$, $p = .000$).

Table 2. Item Nonresponse, by Mode

Item	Telephone Nonresponse (% and number missing)	Face-to-Face Nonresponse (% and number missing)	Total Nonresponse (% and number missing)	Chi-Square Statistic, p-value
Expect to work for pay next year	2.44 (n=30)	1.45 (n=5)	2.22 (n=35)	$\chi^2 = 1.22$, $p = .270$
Expect to work for pay next five years	3.58 (n=44)	4.64 (n=16)	3.81 (n=60)	$\chi^2 = .82$, $p = .265$
Race	5.37 (n=66)	3.77 (n=13)	5.02 (n=79)	$\chi^2 = 1.45$, $p = .229$
Father's education	35.39 (n=435)	32.17 (n=111)	34.69 (n=546)	$\chi^2 = 1.23$, $p = .267$
Mother's education	22.13 (n=272)	19.42 (n=67)	21.54 (n=339)	$\chi^2 = 1.72$, $p = .279$
Respondent's weight	2.85 (n=35)	2.32 (n=8)	2.73 (n=43)	$\chi^2 = .284$, $p = .594$

Item	Telephone Nonresponse (% and number missing)	Face-to-Face Nonresponse (% and number missing)	Total Nonresponse (% and number missing)	Chi-Square Statistic, p-value
Household income	30.68 (n=377)	43.48 (n=150)	33.48 (n=527)	$\chi^2 = 19.83$, $p < .0001$

Although the effects were not significant, there was a pattern of higher nonresponse in telephone mode for four other items (expect to work for pay next year, race, father’s education, and mother’s education).

B. Social Desirability

There was a significant association between interview mode and providing a “yes” response for two of five items tested: work goals include moving up in a job or learning new skills, and work goals include working in the next five years (see Table 3). Neither of the two questions about use of alcohol or drugs yielded significant associations with respondent mode. While CATI respondents reported an average household income of \$24,082 and CAPI respondents reported an average of \$20,468 ($t=1.72$, $p=.085$), this difference was not significantly different after controlling for sample characteristics.

Table 3. Social Desirability, by Mode

Item	Telephone S.D. Response (% and number)	Face-to-Face S.D. Response (% and number)	Chi-Square Statistic, p-value
Goals include moving up (Yes)	51.43 (n=611)	43.32 (n=146)	$\chi^2 = 6.90$, $p=.009$
See working for pay next year (Strongly Agree or Agree)	30.61 (n=367)	25.59 (n=87)	$\chi^2 = 3.21$, $p=.073$
See working for pay next 5 years (Strongly Agree or Agree)	48.35 (n=573)	37.99 (n=125)	$\chi^2 = 11.12$, $p=.001$
Felt need to cut down on drinking in last 12 months (No)	84.60 (n=390)	79.35 (n=73)	$\chi^2 = 1.55$, $p=.213$
Used drugs in last 12 months (No)	94.20 (n=1,153)	95.94 (n=331)	$\chi^2 = 1.60$, $p=.207$

C. Non-Differentiation

There were no significant differences by mode for the series of four health-related items we tested. Only about one-quarter of respondents in either mode gave the same answer to all four of these items (21.48 percent in CATI and 20.58 percent in CAPI; $\chi^2 = .131$, $p=.718$).

D. Acquiescence

As shown in Table 4, we found a significant association for mode and providing an affirmative response on two of the four awareness items tested. The third item, “Heard of Expedited Reinstatement,” was marginally significant. In all cases, telephone respondents were more likely to give a “yes” response to these items than were face-to-face respondents.

Table 4. Acquiescence, by Mode

Item	Telephone "Yes" Response (% and number)	Face-to-Face "Yes" Response (% and number)	Chi-Square Statistic, p-value
Heard of Impairment-Related Work Expenses/Blind Work Expenses	7.62 (n=92)	5.56 (n=19)	$\chi^2 = 1.71$, p=.191
Heard of Expedited Reinstatement	12.53 (n=152)	8.50 (n=29)	$\chi^2 = 4.19$, p=.041
Heard of Work Incentive and Planning Assistance Programs	20.12 (n=243)	15.84 (n=54)	$\chi^2 = 3.14$, p=.076
Heard of Ticket to Work	34.21 (n=414)	26.61 (n=91)	$\chi^2 = 7.03$, p=.008

IV. DISCUSSION

In general, we found evidence that mode of interview has a modest impact on data quality for this population. Our findings are generally consistent with prior research on other populations and suggest that a similar pattern of mode effects can be expected for this population.

Item Nonresponse. While we saw a general trend towards more item nonresponse in the telephone than face-to-face mode for several of the items tested, we found few statistically significant differences (only for household income and health insurance). It is unclear why telephone respondents would provide less information about the health insurance items, although it is possible that face-to-face respondents were more likely to look up such information while someone was physically present in their homes. The finding for household income was in the opposite direction than the trend for the other items, suggesting that respondents were less willing to reveal this information face to face than by telephone, a mode offering greater perceived anonymity.

Social Desirability. For three of the five items tested, we found that respondents were significantly more likely to report having work goals and higher expectations of future work activity when interviewed by telephone than face to face. This is consistent with the Holbrook et al. (2003) finding that respondents are more likely to report socially desirable behavior in the CATI mode than in the CAPI mode. However, this contrasts with the finding reported above regarding missing data for household income, which suggests that respondents were less comfortable reporting sensitive information face to face. It may be that these items are not particularly sensitive and, rather than providing a socially desirable response, respondents simply are acquiescing (saying “yes”) and are more likely to do so by telephone than face to face. We were surprised to find no effect for questions related to drug use and drinking, since these items generally are considered quite sensitive. However, the number of “yes” responses to these items is small, which may have contributed to the lack of effect.

Non-Differentiation. We found no significant association between the amount of non-differentiation displayed between modes for the items we tested. The questions were part of a series of like items using similar but slightly different response scales. Questions were interspersed throughout the interview. To ensure clarity, interviewers read the response options for each item as it appeared on the screen. This may have had the effect of heightening attention to the items and could have minimized satisficing in the telephone mode.

Acquiescence. For all items we tested, telephone respondents were more likely to agree that they had heard of various SSA work incentives than those interviewed face to face. This is consistent with prior research and generally supports the hypothesis that telephone respondents are more likely to exhibit satisficing behaviors than face-to-face respondents. As is true for all of our findings however, it is possible that these findings reflect real differences in the two populations.

V. CONCLUSION

In sum, we found some evidence that mode of data collection may affect data quality for this population. While telephone and face-to-face respondents exhibited about the same levels of item nonresponse and non-differentiation in this survey, telephone respondents were more likely to provide “yes” responses to questions about SSA work programs and likelihood of work. This suggests that persons with disabilities interviewed by telephone may be more likely to take shortcuts when responding, perhaps because telephone interviewing places higher cognitive demands on them. Due to the constraints on our sample sizes (particularly after excluding proxy interviews for those with cognitive impairments), we were not able to conduct subgroup analyses comparing individuals with physical and cognitive impairments, and thus cannot determine if this effect is greater for those with cognitive limitations.

These results should not be interpreted to mean that persons with disabilities cannot be interviewed effectively by telephone. For these analyses, we selected items likely to be the most sensitive to a possible mode effect. With appropriate accommodations made to include persons with disabilities, we anticipate that, for most data comparisons, there would be few differences between estimates obtained by telephone and those obtained face to face.

Our findings do suggest that researchers should consider carefully the interplay of question content or complexity and mode of data collection in the design phase, and which mode effects are most likely given the population. Specifically, in studies of people with mental and physical disabilities, the design should take into account the need to minimize item nonresponse and acquiescence for telephone respondents. For example, researchers may need to include structured follow-ups when interviewees respond “don’t know” or “refused.” Such follow-ups also can help to avoid asking items as part of a series, which can encourage providing quick—or similar—responses to subsequent items; alternatively, researchers could include structured probes reminding respondents of the response options. Our findings suggest this may be most important for items that are sensitive or vague, or otherwise demand more attention and cognitive processing. In our study, factual questions about behavior were less likely to differ by mode. Attention to these design elements is extendable to nondisabled populations and can be considered good practice for surveys of the general population.

Finally, despite the experimental design of this study, it is not entirely clear whether the variations in quality we found are due to differences in the mode of interview or real differences in respondent characteristics in each mode. Although we attempted to control for characteristics that could explain some differences, some of our conclusions could be explained by differences for which we could not control. An experiment not tied to an ongoing study could provide a more rigorous design (for example, maintaining greater equivalence in level of effort for the two groups). A larger sample size that allowed for subgroup analyses also would provide additional information about which groups of individuals with disabilities are most susceptible to providing different responses according to interview mode.

REFERENCES

- Chang, L., and J. A. Krosnick. "Comparing Oral Interviewing with Self-Administered Computerized Questionnaires." *Public Opinion Quarterly*, vol. 74, 2010, pp. 154–167.
- de Leeuw, E. "To Mix or Not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics*, vol. 21, 2005, pp. 233–255.
- Holbrook, A. L., M. C. Green, and J. A. Krosnick. "Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Bias." *Public Opinion Quarterly*, vol. 67, 2003, pp. 79–125.
- Jackle, A., C. Roberts, P. Lynn. "Assessing the Effect of Data Collection Mode on Measurement." *International Statistical Review*, vol. 78, 2010, pp. 3–20.
- Jordan, L. A., A. C. Marcus, and L. G. Reeder. "Response Styles in Telephone and Household Interviewing: A Field Experiment." *Public Opinion Quarterly*, vol. 44, 1980, pp. 210–222.
- Kreuter, R., S. Presser, and R. Tourangeau. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effect of Mode on Question Sensitivity." *Public Opinion Quarterly*, vol. 72, 2008, pp. 847–865.
- Krosnick, J. A., A. L. Holbrook, M. K. Berent, R. T. Carson, W. M. Hanemann, R. J. Kopp, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, W. R. Moody, M. C. Green, and M. Conaway. "The Impact of 'No Opinion' Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly*, vol. 66, 2002, pp. 371–403.
- Locander, W. B., and J. P. Burton. "The Effect of Question Form on Gathering Income Data by Telephone." *Journal of Marketing Research*, vol. 13, 1976, pp. 189–192.
- Mitchell, S., A. Ciemnecki, K. CyBulski, and J. Markesich. "Removing Barriers to Survey Participation for Persons with Disabilities." Ithaca, NY: Cornell University ILR School, 2006.
- Pierzchala, M., D. Wright, C. Wilson, and P. Guerino. "Telephone Collection as Part of a Multi-Mode Survey." 2005. Paper presented at the Second International Conference on Telephone Survey Methodology, Miami, 2006.
- Schwarz, N. "Questionnaire Design: The Rocky Road from Concepts to Answers." In *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York, NY: Wiley, 1997.
- Sykes, W., and M. Collins. "Effects of Mode of Interview: Experiments in the U.K." In *Telephone Survey Methodology*, edited by R. M. Groves, P. P. Biemer, L. Lyberg, J. T. Massey, W. L. Nicholls, and J. Waksberg. New York, NY: Wiley, 1988.

Thornton, C., G. Livermore, D. Stapleton, J. Kregel, T. Silva, B. O'Day, T. Fraker, W. G. Revell Jr., H. Schroeder, and M. Edwards. "Evaluation of the Ticket to Work Program: Initial Evaluation Report." Prepared for the Social Security Administration. Washington, DC: Mathematica Policy Research, 2004.

Walsh, E., and G. Khatutsky. "Mode of Administration Effects on Disability Measures in a Sample of Frail Beneficiaries." *The Gerontologist*, vol. 47, 2007, pp. 838–844.

About the Series

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information, contact Debra Wright, senior researcher, at dwright@mathematica-mpr.com.

www.mathematica-mpr.com

**Improving public well-being by conducting high-quality,
objective research and surveys**

PRINCETON, NJ - ANN ARBOR, MI - CAMBRIDGE, MA - CHICAGO, IL - OAKLAND, CA - WASHINGTON, DC

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.