

Interviewer Perceptions of Interview Quality

Kirsten Barrett, Matt Sloan, Debra Wright
Mathematica Policy Research, Inc.

Abstract

The purpose of this analysis was to determine whether interviewer perceptions of respondents' performance are valid indicators of data quality in a survey of persons with mental and physical disabilities. In this paper, we describe interviewer perceptions of respondent participation and examine how interviewer perceptions correspond to data quality indicators. We also examine the relationship between interviewer perceptions and respondent and interview characteristics. We use data from the National Beneficiary Survey, a survey conducted by Mathematica Policy Research and sponsored by the Social Security Administration (SSA), for this analysis. The survey gathers information about health, employment, income, and demographic characteristics. Also included are questions asked of the interviewer about his/her perception of the respondent's intellectual capacity, response accuracy, and comprehension. Our analyses included 5,567 SSA disability beneficiaries.

Chi-square analyses revealed that there were significant relationships between interviewer ratings and several data quality indicators such as cognitive screener performance, item non-response, and codeable verbatim responses. The relationship between other data quality measures, non-differentiation and acquiescence, however, were not significant. Additionally, interviewers perceptions were related to respondent characteristics such as age, race, marital status, health and the mode of interview (CATI vs. CAPI).

The results of this study suggest that interviewer perceptions are consistent with some indicators of data quality and support the inclusion of interviewer ratings within the context of the interview as a valid means to gather information on the quality of data collected

Keywords: Interviewer perceptions, disability, data quality

1. Introduction

Interviewing individuals with disabilities requires increased attention to questionnaire construction and design to ensure ease of participation and to facilitate accurate and truthful responding. Survey researchers often make inferences about the extent to which the

design is successful by considering data quality indicators such as item non-response, non-differentiation, and acquiescence. Interview characteristics such as survey mode, survey duration, and respondent sociodemographic characteristics are also considered when assessing the quality of the survey design. In this paper, we consider a less frequently used source of information about data quality, the interviewer him or herself.

There is a significant amount of research available, albeit with conflicting results, about the impact of interviewer characteristics such as age, race, gender, experience, stereotypes and attitudes on respondent behavior and data quality (Groves et al., 2004; Singer et al., 1983; Clarke et al., 2003; Hill, 2002, Berk & Bernstein, 1988; and Dailey & Clause, 2001). However, minimal attention has been paid to gathering and examining case-level data about respondent performance across an entire survey, as perceived by the interviewer. There is no literature to support or negate the inclusion of questions asked of interviewers about their opinion of the individual respondent's ability to fully engage in the survey process.

Data from interviewers are often collected using qualitative debriefing techniques. Interviewer debriefing can occur following pretesting, iteratively during the data collection period, and/or at the end of the data collection period. Debriefing provides valuable information about various aspects of survey design including question wording, questionnaire pathing, interpretation of key concepts, respondent reactions to particular items, and ease of administration given the mode utilized (Presser et al., 2004, Bieman & Lyberg, 2003).

Less common is the collection of such data using both qualitative and quantitative methods. Campanelli et al. (1991) described how the U.S. Census Bureau used a mixed mode approach to gather data from experienced Current Population Survey (CPS) interviewers. The quantitative component consisted of a questionnaire distributed to experienced interviewers at the time they were convened for focus group debriefing. Questions focused on the performance of respondents on CPS items that the research team considered potentially problematic. Although not stated explicitly, it is assumed that data were provided across all cases completed by the interviewer, rather than on a case-by-case basis. Immediately after collecting the interviewer's data, the

focus group debriefing sessions were conducted. The quantitative component was deemed to be an ‘asset’ in that it allowed interviewers to provide their input without being influenced by other interviewers, as is possible during focus groups (Campanelli et al., 1991).

In this paper, we sought to determine whether or not interviewer perceptions, collected as part of the interview process, could be valid indicators of data quality. If interviewer perceptions are found to be associated with common data quality indicators such as item non-response, non-differentiation, and acquiescence, it would point to the utility of examining cases where the interviewers’ perceptions indicate a potential problem more closely. Alternatively, it might be more efficient to identify problematic cases in batch and then cross-reference them to interviewers’ perceptions, with additional scrutiny given to cases where the perceptions suggest a potential problem with the interview.

In this exploratory study, using data from the National Beneficiary Survey, we (1) describe interviewer perceptions of respondent participation and (2) examine the association between interviewer perceptions and cognitive screener performance, interview and respondent characteristics, and data quality indicators. We conclude with an assessment of the utility of gathering data from interviewers on their perceptions of the respondents’ level of understanding, their intellectual capacity to participate, and their response accuracy.

2. Methods

2.1 Brief Description of Survey.

The National Beneficiary Survey (NBS), conducted by Mathematica Policy Research, Inc. (MPR) and sponsored by the Social Security Administration (SSA), is a nationally representative survey of 18 to 64 year old SSA disability beneficiaries. The 45-minute, dual-mode (CATI/CAPI) survey gathers information about health, employment, income, and demographic characteristics.

The NBS was conducted in 2004 using a mixed mode methodology; CATI with a CAPI follow-up for eligible respondents. A total of 7,603 cases were completed for an overall weighted response rate of 77.6 percent.

For purposes of this analyses, partial completions (n=23), cases with missing data on one or more key interviewer perception items (n=28), and cases completed by proxies (n=1,985) were removed from the analysis file. The total number of cases included in these analyses was 5,567.

2.2 Measures

2.2.1 Interviewer Perception Variable

A key variable discussed in this paper is ‘interviewer perception’. This variable was created based on data from the following three questions asked of the interviewer, immediately after the interview was completed: 1) In general, do you think the respondent was intellectually capable of responding? 2) In general, do you feel the respondent’s answers were reasonably accurate? 3) In general, do you feel the respondent understood the questions? Cases where the interviewer responded ‘no’ to one or more of these questions were coded as ‘unfavorable’ (n=176). The remaining cases were coded as ‘favorable’ (n=5391).

2.2.2 Cognitive Screener.

The cognitive screener is a series of three questions that respondents must answer correctly, within two attempts, to be considered capable of participating in the survey. The questions cover the basic components of informed consent. This is included in the NBS as a quantitative measure to assess the respondent’s ability to participate fully in the survey process. We sought to determine if those passing the screener on their first try were more favorably perceived by the interviewers than those who required two attempts to pass it.

2.2.3 Interview and Respondent Characteristics.

We sought to determine if interviewer perceptions vary based on survey mode, CATI vs. CAPI. We also examined the relationship between interviewer perceptions and respondent demographics including, but not limited to, race, age, education, and marital status.

2.2.4. Health and Disability Status.

There were a number of health and disability items used in the analysis. Included were the following: general health over the course of the past four weeks, age of disability onset (child vs. adult), number of sensory limitations, and presence or absence of mental health issues and mental retardation. We selected these items because they provide an overall measure of general health as well as an assessment of issues that could potentially impact survey participation.

2.2.5. Data Quality Indicators.

We also explored the relationship between interviewer perception and various indicators of data quality including item non-response, uncodeable verbatims, non-differentiation, and acquiescence. These data quality indicators were chosen because they all contribute to non-sampling error. In the case of item non-response and uncodeable verbatims, the effective sample size is reduced. When this occurs on key items used in statistical analyses, it can result in increased non-response bias (Groves et al., 2004). In the case of non-differentiation and acquiescence, there is a decrease in the heterogeneity of responses, making significant differences difficult to detect.

The selection of survey items used to assess data quality were made based on the importance of the topic to researchers, on having an adequate sample size for analysis and, for some quality indicators, based on skip patterns within the instrument itself.

We looked at item non-response on a single income variable and within a series of 33 health questions asked of all respondents. The extent to which open-ended responses were uncodeable was assessed using data from the respondent about the main physical or mental condition that limited his or her ability to work or do other daily activities, a major area of interest for SSA.

The degree of non-differentiation and acquiescence were also assessed. Non-differentiation occurs when respondents do not fully engage in the cognitive processing required to identify differences between questions and thus provides the same response for the majority of questions in a series (Krosnick 1991). Acquiescence is a source of non-sampling error that is introduced when respondents answer in ways that they believe the interviewer wants them to answer, rather than responding with their true opinion or attitude (Biemer & Lyberg 2003).

Non-differentiation and acquiescence were examined using a series of 12 statements that asked employed respondents to indicate the extent to which they agree that certain attributes of their job are favorable. Ratings were provided on a 4-point Likert scale (strongly agree to strongly disagree).

2.3 Statistical Analysis

Basic descriptive statistics were used to describe the characteristics of the sample. Chi-squared analyses were conducted to determine if there were significant associations between interviewer perceptions and

cognitive screener performance, interview and respondent characteristics and data quality.

All analyses were done using SAS (Version 8). In cases where expected counts in one or more cells in the contingency table were less than five, Fisher's Exact Test was used. All data presented in this paper are unweighted. This approach was taken since the interest is methodological and we are not attempting to generalize our findings to the general population of SSA beneficiaries.

3. Results

Interviewers perceived the vast majority of respondents as being intellectually capable of participating in the survey (98.4 percent), as giving reasonably accurate responses (98.2 percent) and as understanding the questions being asked (99 percent). Approximately 3 percent of respondents (n=176) were identified as having difficulty in one or more of these three areas. In most cases (n=118), a problem was identified in one area only, primarily accuracy of answers provided (n=56). The remaining respondents had difficulty in two (n=33) or three areas (n=25).

3.1 Interview and Respondent Characteristics

3.1.1. Cognitive Screener Performance

Respondents that required two chances to pass the cognitive screener were nearly two times more likely to have unfavorable ratings than were their counterparts who passed the cognitive screener on the first attempt ($\chi^2=16.623$, $p<.0001$).

3.1.2. Interview Mode

Interviewers were more likely to rate the interview as favorable when the survey was conducted in CATI rather than CAPI ($\chi^2=7.822$, $p=.005$).

3.1.3 Respondents' Sociodemographic Characteristics

Significant associations were found between respondent age, race, marital status, and education and interviewer perceptions. Younger respondents were more likely to receive unfavorable ratings compared to older respondents ($\chi^2=7.70$, $p=.05$). Similarly, respondents who were non-white ($\chi^2=17.537$, $p=.0002$), respondents who were not currently married ($\chi^2=10.308$, $p=.0058$), and respondents with lower levels of education ($\chi^2=23.44$, $p<.0001$) were all more likely to be perceived as having problems with the interview.

Table 2—Interviewer Perceptions and Respondent Health and Disability Status
(Because of rounding, not all rows sum to 100%.)

	<i>Favorable Perception</i>	<i>Unfavorable Perception</i>	<i>p-value</i>
<i>General Health</i>			
Good to excellent	96.0% (1742)	4.0% (72)	p=.0105
Poor to fair	97.3% (3642)	2.7% (101)	
<i>Disability Onset</i>			
Younger than 18	95.2% (1370)	4.8% (69)	p<.0001
18 or older	97.4% (4009)	2.6% (107)	
<i>Severe Sensory Limitations</i>			
None	97.0% (4794)	3.0% (146)	p=.007
One	96.02% (531)	4.0% (22)	
More than one	89.8% (44)	10.2% (5)	
<i>Mental Impairments (all)</i>			
No	97.1% (3930)	2.9% (117)	p=.282
Yes	96.6% (1426)	3.5% (51)	
<i>Mental Retardation</i>			
No	97.1% (5133)	2.9% (154)	p=.006
Yes	90.5% (67)	9.5% (7)	

3.1.4 Respondents' Disability-Related Characteristics

Associations between respondent health and disability status and interviewer perceptions were also examined. As can be seen in Table 2, respondents with disability onset prior to age 18 ($\chi^2=16.751$, $p<.0001$) and those with more than one sensory limitation were significantly more likely to be perceived as having problems with the interview ($\chi^2=9.921$, $p=.007$).

Respondents with mental impairments were only slightly more likely to be rated as having an unfavorable interview ($\chi^2=1.159$, $p=.282$). However, when parsing out respondents with mental retardation, these respondents were significantly more likely to have unfavorable ratings, 10 percent vs. 3 percent ($p=.006$). Interestingly, those who rated their general health as fair to poor ($\chi^2=6.541$, $p=.011$) were significantly more likely to be perceived as having a favorable interview.

3.2 Data Quality Indicators

3.2.1. Non-Response (Item-Level).

Income is a critical variable of interest for this study, and is often left unanswered by respondents; therefore, we used this variable to examine the relationship between interviewer perception and item non-response. The item asks, "What was your total income in 2003, before taxes or other deductions? Please include money you received from all sources." Item non-response was defined as a "don't know" or "refused" response.

Respondents perceived as having a problem with the interview were significantly more likely to respond with a 'don't know' or 'refused' response on the income item as compared to those with a 'problem-free' interview, 49 percent vs. 34 percent. ($\chi^2=16.035$, $p<.0001$).

3.2.2. Item Non-Response (Section Level).

In the preceding section, item non-response at the question-level was examined. However, item non-response on a single item may be less problematic than item non-response across an entire section or across an entire survey. Therefore, we also examined item non-response across a series of questions about health and functional ability.

In total, there were 33 items within the section that were asked of all respondents. Among those with favorable interviews, 92.1 percent had zero "don't know" or "refused" responses across the items as compared to 84.7 percent of those with unfavorable ratings ($\chi^2=112.258$, $p<.0001$).

Table 3– Interviewer Perceptions and Non-Differentiation
(Because of rounding, not all rows sum to 100%.)

	<i>Agree to Strongly Agree</i>			<i>p-value</i>
	<i>3 or less</i>	<i>4 to 8</i>	<i>9 to 12</i>	
Favorable Perception	5.0% (27)	33.3% (179)	61.7% (332)	p=.174
Unfavorable Perception	11.7% (2)	47.1% (8)	41.2% (7)	
	<i>Disagree to Strongly Disagree</i>			
	<i>3 or less</i>	<i>4 to 8</i>	<i>9 to 12</i>	<i>p-value</i>
Favorable Perception	63.2% (340)	32.3% (174)	4.5% (24)	p=.120
Unfavorable Perception	41.2% (7)	47.1% (8)	11.8% (2)	
	<i>Don't Know or Refused</i>			
	<i>3 or less</i>	<i>4 to 8</i>	<i>9 to 12</i>	<i>p-value</i>
Favorable Perception	99.8% (537)	.19% (1)	0% (0)	p=1.00
Unfavorable Perception	100% (17)	0% (0)	0% (0)	

3.2.3. *Uncodable Verbatims.*

Since disability is a defining characteristic of the target population, it is a major topic of interest for SSA and independent researchers alike. Many data elements within NBS address disability and related topics. One such item asks respondents to identify their disabling condition. Respondents are asked, “What physical or mental condition is the main reason you are limited?” Interviewers are instructed to probe so that the respondent provides an actual condition, rather than the symptoms or resulting deficits thereof. Verbatim responses are subsequently assigned numeric code. There are some DK and ‘refused’ responses and some that are not codeable, based on quality.

Respondents who received an unfavorable rating were four times more likely to have uncodeable disability data as compared to their counterparts perceived as having a favorable interview, 1 percent vs. 5 percent ($\chi^2=35.664, p<.0001$).

3.2.4. *Non-Differentiation.*

As described previously, to assess the extent of non-differentiation in NBS, we selected a series of 12 work-related items asked of those currently employed, excluding those who were self-employed and those for whom none of the 12-items were coded as ‘not applicable (n=555)’. Items address topics including, but not limited to, pay, benefits, autonomy, job security, potential for growth, and level of enjoyment and interest in work. Responses are given on a 4-point Likert ranging from strongly agree to strongly disagree. DK and refused options are also available.

There were no statistically significant associations between interviewer perceptions and the extent to which a respondent was likely to agree or disagree or respond with ‘DK’ or ‘R’ across a series of 12 items. However, based on a review of Table 3, it appears that those with favorable ratings were more likely to agree or strongly agree with the vast majority of items than their counterparts with unfavorable perceptions and that those with unfavorable ratings by the interviewer were nearly three times more likely to disagree or disagree with the vast majority of items as compared to respondents with favorable perceptions.

3.2.5. *Acquiescence.*

The findings illustrated in Table 3 can also be used to assess the degree of acquiescence. Respondents would display evidence of acquiescing if they ‘agreed’ or ‘strongly agreed’ with most or all the statements pertaining to elements of job satisfaction. While those with favorable ratings had higher degrees of acquiescence than those with unfavorable ratings, 62 percent vs. 41 percent, this difference was not statistically significant.

4. Discussion

In this analysis, we found that the vast majority of respondents were perceived by interviewers as being intellectually capable of participating, as being able to comprehend the survey questions, and as being accurate in their responses. There were a small number of respondents who were perceived as having difficulty with the interview (n=176). This is not unexpected since all respondents in this analysis passed the cognitive screener. If the cognitive screener

identified the sample member or, less commonly, a potential proxy as incapable of completing the survey, then a proxy (or another proxy) was selected. There were 1,985 cases with proxy respondents in round one of the NBS.

We examined the association between interviewer perceptions and cognitive screener performance, interview and respondent characteristics, and data quality indicators. In this discussion, we touch on the key findings in each area. Then, we conclude with an assessment of the utility of the interviewer perception items in NBS.

4.1 Respondent and Interview Characteristics

4.1.1. *Interviewer Perceptions and Cognitive Screener Performance.*

With regard to cognitive screener performance, those needing more than one attempt to pass the cognitive screener were significantly more likely to be perceived by the interviewer as having some difficulty participating in the survey. This is an interesting result because it provides some degree of face validity for the cognitive screener questions included in NBS. The inclusion of the cognitive screener questions is consistent with recommendations from a growing number of researchers that an a priori determination should be made regarding the respondent's ability to provide accurate responses to survey questions (Seiber, 2001). Including such questions goes beyond the more common practice of using passive consent or the interviewer's unilateral judgment as a means to determine if the respondent is capable of participating in telephone interviews, and sometimes in-person interviews.

4.1.2. *Interviewer Perceptions and Interview Mode.*

With regard to interview mode, those who completed the survey in CAPI were significantly more likely than CATI completers to be perceived as having problems comprehending questions and/or responding accurately to questions. Because mode of interview is related to impairment type as well as demographic characteristics such as education, it is likely that interviewers were accurately recording greater difficulty with the interview among these respondents.

However, it may be that in-field interviewers are more likely to perceive respondents as unfavorable because they have added observational information about the respondent. When "seeing" a respondent having apparent difficulty with the survey, interviewers may be more likely to report that the person had difficulty

with comprehending questions and/or responding accurately, even if the observed behavior was unrelated to cognitive processes.

4.1.3. *Sociodemographic and Disability-Related Characteristics.*

Respondents who were younger, non-White, not currently married, and had lower levels of education were significantly more likely to be perceived as having difficulty participating in the survey. Gender and employment status were not associated with interviewer perceptions of survey difficulty.

Because age, race, and marital status are correlated with education, it is not possible to parse out any independent relationships between respondent characteristics and interviewer perceptions in these analyses. Additionally, these variables are correlated with mode of interview and impairment. While it is possible that CAPI interviewers could have made judgments of respondents' abilities based on some these variables, CATI interviewers could not have. Additional analyses controlling for the impact of demographic variables on impairment type and mode are needed to disentangle these effects.

4.1.4. *Disability-Related Characteristics.*

With regard to disability-related characteristics, those who acquired their disability prior to the age of 18, those with severe sensory limitations, and those with mental retardation were significantly more likely to be perceived by the interviewer as having difficulty with the survey. Approximately 30 percent of respondents with disability onset prior to the age of 18 reported that their disability began at 3 years of age or less. It may be that many of these respondents had development disabilities with resulting cognitive impairments, thus making survey participation difficult. More in-depth study into the relationship between and among respondent diagnosis, age of disability onset, and interview difficulty would clarify this.

Interestingly, those rating their general health as poor to fair were less likely to be perceived as having difficulty with the survey as compared to their counterparts with good to excellent health. This seems counterintuitive. It would seem that those who were not in good health would have more difficulty with the survey than those who were feeling well. However, it should be noted that the general health question contained within NBS asks respondents to provide a rating that reflects their general health over the past four weeks, rather than at the time of the interview itself. It is plausible that respondents reported poor to

fair health were either having a ‘good day’ or their medical condition did not impact their ability to participate.

With regard to mental impairment, we thought that respondents with these conditions would be perceived as having more difficulty with survey participation. Overall, they did not. This may be because those with the most severe mental impairments had been screened out and a proxy identified. However, when limiting the mental impairments to mental retardation only, we did find that those with this condition were significantly more likely to have an unfavorable rating. This has implications for interviewer training. Since respondents report their main limiting condition early in the survey, those who cite mental retardation may benefit from interviewing strategies that could ease their participation. This may include providing additional response time, offering continued praise and reassurance, liberal re-reading of questions, and frequent offers of rest breaks.

4.2 Data Quality

The final area of interest was to determine if interviewer perception was associated with data quality indicators such as extent of non-response at the item and section level, extent of uncodeable verbatims, and degree of non-differentiation and acquiescence. With regard to item non-response, although we did find significant differences between groups in terms of the percent responding DK or “refused” on the income item, both those perceived as having survey difficulty and those perceived favorably had relatively high levels of non-response, 49 percent vs. 34 percent respectively.

With non-response across a range of items, those perceived as having difficulty with the survey were more likely to respond with DK or “refused” across the 33 items asked of everybody. However, among the group perceived as having difficulty, 85 percent had no DK or refused responses across the 33 items and an additional 13 percent had between 1 and 7 DK or refused responses across the 33 items. Although statistically significant, the difference between groups may be of little practical significance.

Respondents perceived as having difficulty with the survey were more likely to provide uncodeable disability verbatims than their counterparts not having difficulty (5 percent vs. 1 percent). This is an important finding since disability is a major area of interest for SSA. Such items may require proxy or assisted responses to obtain sufficient information or

may need to be supplemented with data from administrative sources.

Non-differentiation and acquiescence across a series of 12 items related to job characteristics differed between groups, although they were not statistically significant. Those with favorable ratings were more likely to agree or strongly agree with the vast majority of items than their counterparts with unfavorable perceptions, 62 percent vs. 41 percent. Conversely, those with unfavorable ratings by the interviewer were nearly three times more likely to disagree or disagree with the vast majority of items as compared to respondents with favorable perceptions, 12 percent vs. 5 percent. Given the low number of cases in the unfavorable perception group (n=17), these findings should be interpreted with caution and considered as preliminary in nature. Further research is needed.

5. Conclusion

We were interested in determining whether interviewer perceptions or ratings were valid indicators of the quality of data collected for a national survey of SSA beneficiaries.

Interviewer observations were associated with cognitive screener performance, impairment, and the need to conduct an in-person CAPI interview. Additionally, interviewer observations were associated with item non-response and quality of verbatim responses. Given these findings, we conclude that it is useful to collect data from interviewers using a quantitative measure within the context of the interview. Cases flagged by interviewers as problematic would seem to be a good starting point for investigation in the data cleaning process.

It should be recognized that we found minimal variability in responses to the interviewer perception items. This was not unexpected since a cognitive screener was included. Although proxies were allowed for a number of reasons, the majority of the nearly 2,000 proxy respondents were the result of sample members’ inability to pass the cognitive screener or the sample members’ level of cognitive impairment per an informant. If the cognitive screener was not utilized it is likely that we would see more variation in response to the interviewer observation items. Still, we suspect that perception items that include a broader array of response options (rather than yes/no) may be more useful in picking up variation in respondent ability.

This analysis is not without its limitations. With only 176 cases in the group perceived as ‘unfavorable’, the analytic options were limited. Further, the perception

questions are broad and how interviewers interpreted their meaning is unclear. In terms of data quality indicators, there was a limited pool of survey items to pick from. We attempted to pick questions that had a high degree of topic relevance and that had an adequate sample size by perception group. As a result, the data quality indicators are somewhat isolated from each other and do not provide a comprehensive understanding of how data quality may have varied in a particular topic area such as disability, employment, or earnings. Finally, many of the variables we examined are related, making it difficult to parse out relationships between interviewer observations and other variables. Further analysis controlling for demographic variables is needed to clarify this.

References

- Biemer, PP & Lyberg, LE. (2003). Introduction to survey quality. Hoboken,NJ: John Wiley and Sons, Inc.
- Campanelli, PC, Martin, EA., & Rothgeb, JM. (1991). The use of respondent and interviewer debriefing studies as a way to study response error in survey data. *The Statistician*, 40(3): 253-264.
- Berk, ML. And Bernstein, AB. (1988). Interviewer characteristics and performance on a complex health survey. *Social Science Research*,17(3), 239-251.
- Biemer, PP & Lyberg, LE. (2003). Introduction to Survey Quality. Hoboken,NJ: John Wiley and Sons, Inc.
- Clarke et al. (2003). An investigation into expectation-led interviewer effects in health surveys. *Social Science and Medicine*, 56, 2221-2228.
- Dailey, RM. & Clause, RE. (2001). The relationship between interviewer characteristics and physical and sexual abuse disclosures among substance users: a multilevel analysis. *Journal of Drug Issues*, (31): 867-888
- Groves, RM, Fowler, FJ, Couper, MP, Lepkowski, JM, Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. Hoboken,NJ: John Wiley and Sons, Inc.
- Hill, M. (2002). Race of the interviewer and perception of skin color: evidence from the multi-city study of urban inequality. *American Sociological Review*, 67(1):_99-108.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5, 213–236.
- Mitchell, S., Ciemnecki, A., CyBulski, K., Markesich, J. (2006). Removing Barriers to Survey Participation for Persons with Disabilities. Report to the Rehabilitation Research and Training Center on Disability Demographics and Statistics. Princeton, NJ: Mathematica Policy Research, Inc.
- Presser, S, Couper, MP, Lessler, JT, Martin, E., Martin, J., Rothgeb, JM, & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68, 109-133.
- Seiber, JE. (2001). Summary of human subjects protection issues related to large sample surveys. U.S. Department of Justice, Bureau of Justice Statistics.
- Singer, E, Frankel, M, Glassman, MB. (1983). The effect of interviewer characteristics and expectations on response. *The Public Opinion Quarterly*, 47, 68-83.