

AN EXPERIMENT TO IMPROVE DRUG USE REPORTS DURING SURVEY INTERVIEWS

Timothy P. Johnson, Michael Fendrich, Seymour Sudman, Joseph Wislar and Elizabeth Severns, Survey Research Laboratory and Department of Psychiatry, University of Illinois at Chicago
Timothy Johnson, Survey Research Laboratory, University of Illinois at Chicago, 412 S. Peoria St., Chicago, IL 60607

Key Words: Drug use, Self-report validation¹

INTRODUCTION

A growing body of research raises questions about the validity of self-reports of substance use behaviors (Harrison and Hughes, 1997; Turner, Lessler and Gfroerer, 1992). Much of what is known about the validity of responses to survey questions is drawn from research on specialized samples in criminal justice populations (Fendrich and Xu, 1994; Feucht, Stephens and Walker, 1994) or from research on clinical samples (Magura, Freeman, Siddiqi and Lipton, 1992; Maisto, McKay and Connors, 1990). Nevertheless, much of what is known about the epidemiology of drug abuse in society depends on general population studies such as the National Household Survey of Drug Abuse (NHSDA; Office of Applied Studies, 1997) and the Monitoring the Future Study (Johnston, O'Malley, and Bachman, 1994). There is a need for a greater understanding of the limits and strengths of survey research for the collection of substance use information in general population studies. Cognitive models of the survey interview process suggest three potential sources of reporting error in self-reports of substance use behaviors. These include errors attributable to poor comprehension, inaccurate recall of information, and deliberate misreporting.

Respondent Comprehension

Belson (1986) and many others have shown that respondents frequently misunderstand survey questions. Survey researchers asking about drug use face particular obstacles with respect to comprehension. Ethnographic work suggests that the names of drugs communicated in questions may not be consistent with names associated with those drugs in the community (Ouellet, Cagle and Fisher, 1997). Street drug terminology changes as new drugs become available and as use patterns change. Questionnaire wording may not convey the same

meaning to respondents that survey researchers assume it does and personal definitions of various drugs may often override those provided in survey questions (Hubbard, Pantula and Lessler, 1992).

The major source of epidemiologic data on drug use in the United States, the NHSDA, relies on a self-administered questionnaire to obtain most of its information about drug use history. This procedure was designed to maximize respondents' privacy in response to "sensitive" questions about drug abuse (Gfroerer, Gustin, and Turner, 1992). Nevertheless, problems with comprehension can persist irrespective of interview mode. Difficulties in question interpretation cannot be easily clarified when self-administered techniques are employed. Similar complaints have been directed at standardized survey interviewing procedures, which are recognized as an integral component of the survey research process in the United States. These standardized procedures, which have been developed and refined over the past 50 years, are designed to insure that all respondents receive near-identical stimuli, in the form of question wording and interviewer behavior (Fowler and Mangione, 1990). Despite their widespread acceptance and practice, these standardized techniques have been criticized for their emphasis on shared stimuli, rather than shared meaning (Suchman and Jordon, 1990). It has been argued that the restrictions placed on the role of the interviewer by these procedures result in an artificial, or unnatural, interaction which inhibits communication and damages the quality of survey data.

Memory Retrieval

The influence of respondent memory is being given increased emphasis in cognitive research of the survey process (Friedman, 1993; Jobe, Tourangeau and Smith, 1993). Important findings, for instance, now suggest that the types of cognitive strategies used to recall information vary depending on the relative frequency and regularity of an event (Blair and Burton, 1987; Menon, 1994). Research in this area has also been successful in using cognitive interventions to assist respondents in accessing relevant health-related memories and improving responses (Jobe, White, Kelley, Mingay and Sanchez, 1990; Loftus, Smith, Klinger and Fiedler, 1992; Means and Loftus, 1991). Some researchers have attempted to use knowledge of these processes to im-

¹This research was supported by National Institute on Drug Abuse grants #R01DA09285 and #R01DA09286. We also wish to thank staff at the Research Triangle Institute who provided us with technical support for the inclusion of audio-CASI protocols in this study.

prove self-reported drug use. Hubbard et al. (1992) conducted two experiments in which variations of an anchoring manipulation that included the use of a calendar were used to assist respondents in framing their responses. Although findings generally indicated these procedures were not very powerful in improving recall, one experiment was successful in increasing reports of lifetime behaviors, suggesting the need for further investigation. Sudman and Phillips (1995) have also experimented with the use of memory probes designed to improve the reporting of autobiographical behavior by first identifying the cognitive processes used to retrieve information and then asking respondents to modify their methods and re-estimate behavior. Their procedures have not, though, been applied to self-reports of drug use behavior.

Respondent Editing

Underreporting due to question sensitivity and perceived threat by respondents is referred to as response "editing" by cognitive researchers (Strack and Martin, 1987). This is a term describing the process by which subjects may "edit" their answers in order to present interviewers with socially desirable information. As such, it suggests a conscious distortion of the "truth" about use. When editing occurs, respondents intentionally fail to disclose the truth about their behavior. Survey researchers have used indirect evidence to articulate conditions under which truthful responses will be suppressed in surveys. From supplemental survey questions, it is known that race/ethnicity subgroups may vary in their willingness to disclose use of certain substances (Johnston, Bachman, and O'Malley, 1986). From a comparison of drug use rates provided in a longitudinal study, it is known that multiple contacts with the same interviewer may suppress disclosure of cocaine use (Mensch and Kandel, 1988). From comparisons of rates of drug use provided in alternative survey conditions, it is known that respondents are less likely to divulge drug use when interviewed over the telephone than when interviewed face-to-face (Aquilino and LoSciuto, 1990). From studies of the NHSDA (Turner, Lessler, and Devore, 1992), when recent drug use is inquired about, it appears that self-administered questionnaires obtain more valid responses than face-to-face interviews, particularly when asking about more recent experiences and more stigmatizing drugs (Harrison, 1995). And recent studies hold out the possibility that newly-developed computer-assisted technologies, audio-CASI (audio computer-assisted self-interviewing) in particular, may in turn produce even more accurate self-reports of drug use and other sensi-

tive behaviors (Tourangeau and Smith, 1996; Turner et al., 1998).

The existence of response editing has two implications. First, ascertainment techniques which minimize threats to respondent privacy and perceived public disclosure of behavior may increase rates of reporting. The standard NHSDA method that uses a self-administered questionnaire format, along with the use of a sealed envelope and immediate mailing of the survey would appear to maximize privacy (Gfroerer, Gustin, and Turner, 1992). Other methods which would ensure privacy are randomized response, questions about the behavior of close friends, and the use of audio-CASI. In evaluating the strengths and weaknesses of each of these strategies, most studies in the past have relied exclusively on only one to improve the quality of self-reports. Nevertheless, the possibility exists that respondents will vary in their comprehension of different strategies, as well as in their level of comfort with each. For example, private, self-administered questionnaires may prove threatening to respondents who have difficulty reading and computer-assisted technologies may be intimidating to respondents with little direct experience operating computer equipment. The interview threat might be minimized when subjects are offered choices with respect to the mode used to answer questions about sensitive items.

Assessing the Validity of Drug Use Reports

Finally, much of the discussion with respect to the validity of survey results on drug use assumes that there is a bias toward underreporting of substance use in the general population (U.S. General Accounting Office, 1993). We can not assume that drug use questions are perceived as threatening and sensitive by all respondents. Evidence suggests that the perceived sensitivity of drug use questions in survey interviews may vary according to the level of involvement of individuals with substance use and with drug use subcultures. For example, an ethnographic study of drug abusers and drug distributors suggested that respondents initially had a tendency to overestimate their drug use (Fendrich, Goldstein, Tarshish and Bellucci, 1992); retrospective information provided by active drug users may in some cases actually overestimate the extent of current drug involvement. Concerns about response editing and the direction of reporting bias on the quality of self-reports of substance use have lead some reviewers to call for the use of a "gold standard" or physical test for assessing drug use in community surveys (U.S. General Accounting Office, 1993). In this paper, we report results from an experiment designed to address each of the concerns expressed above.

METHODS

An experimental interview format was developed and contrasted with current standard procedures, here defined as procedures used as part of the NHSDA. The experimental format consisted of additional procedures designed to address potential comprehension, recall and editing concerns. There were four general differences between the experimental and standard interview formats. The first of these was designed to minimize comprehension difficulties. To do so, interviewers in the experimental format were trained to employ conversational interviewing during the first module which inquired about lifetime drug use. Specifically, interviewers were taught and given permission to deviate from specific question wording whenever respondents provided verbal or nonverbal feedback indicative of comprehension difficulties. Interviewers assigned to the control format were trained to conduct structured survey interviews only.

A second difference was the use of skip patterns when inquiring about drug use behaviors as part of the experimental format in an effort to reduce respondent burden. Consistent with NHSDA procedures, the control format asked all respondents to answer all questions.

Third, during the second module of the interview, structured memory probes were employed to follow-up survey questions concerned with recent (i.e., last 30 day) use of specific drugs. These probes were designed to identify the specific recall strategies used by respondents and, based on this information, assist respondents in improving the quality of their responses. For example, respondents who estimated drug use frequency using a rate (a schematic recall strategy) would be probed to determine if adjusting rates were necessary because of atypical reference periods. These specific procedures have been previously employed by Sudman and Phillips (1995). The control condition did not employ memory probes.

Fourth, respondents assigned to be interviewed via the experimental procedures were offered the opportunity to answer survey questions (during the second and subsequent questionnaire modules) via their choice of audio-CASI or CAPI (i.e., computer-assisted personal interviews). Respondents were given this choice with the expectation that the opportunity to select the method of reporting with which they were most comfortable would encourage respondents to provide accurate information. The control format, again consistent with NHSDA procedures at the time the study was conducted, required that respondents complete self-administered questionnaires via paper-and-pencil methods.

A sample of respondents was selected within the city of Chicago using area probability methods. Eligible adults aged 18-35 were sampled at random from within households in neighborhoods with above-average rates of admissions to state substance abuse treatment programs. Respondents within selected blocks were randomly assigned to either the experimental or control interview format. Data collection was conducted between February and September 1997. Each interviewer was randomly assigned to and trained to conduct interviews using either the experimental or control format only. At the end of each interview, all respondents were offered \$10 to provide a small hair sample (see Fendrich, Johnson, Sudman, Wislar and Spiehler, In-press, for details of the procedures used to collect hair samples). Hair samples were subsequently assayed by the United States Drug Testing Laboratories in Des Plaines, Illinois, where they were tested and confirmed for the presence of cocaine and several other illicit drugs. A total of 570 interviews were completed: 271 using the experimental format, and 300 using the control format. The CASRO response rates were 47.9% and 47.7% for the experimental and control formats, respectively. The proportion of all respondents willing and able to supply an adequate hair sample was 56%.

Two sets of outcome measures were examined: (1) willingness to report cocaine/crack use, and (2) concordance between self-reports of cocaine use and hair-test results. The proportions of respondents willing to report lifetime, past year, and past 30-day cocaine use were compared across interview formats. Concordance between self-reports and hair-test results by interview format were assessed using proportions and Kappa statistics of agreement. Group differences in willingness to report cocaine use and concordant reports were compared using logistic regression models that controlled for respondent gender, age, race/ethnicity, education and employment status.

RESULTS

Respondents assigned to the experimental and control interview formats were very similar. The sample as a whole was 39.1% male, 52.6% African-American, 20.2% Hispanic, 27.2% White-non-Hispanic and other groups, 22.9% with less than a high school degree, 32.6% high school graduates, 44.5% with college experience, 60.3% employed, 41.5% aged 18-25, 26.9% aged 26-30 and 31.6% aged 31 and older. The demographic composition of the samples assigned to the experimental and control interview formats differed significantly only in terms of education. Specifically, persons assigned to the control format were more likely to have post-high school education and persons in the

experimental condition were more likely not to have completed high school ($p = .02$).

Comparisons of willingness to report cocaine use by interview format are presented in Table 1. No significant differences in lifetime, past year and past 30-day reports of cocaine/crack use were found between interview formats. A trend towards increased proportions reporting use across each reference period among those assigned to the control interview format, however, was observed.

Table 1. Self-Report Prevalence of Cocaine/Crack Use by Interview Format ($N=570$).

	<u>Control</u>	<u>Experimental</u>
Lifetime	20.3%	16.7%
Past Year	9.7%	5.9%
Past 30 days	6.7%	3.7%

Interview format comparisons were also made after controlling for gender, race/ethnicity, age, education, and employment status using logistic regression. Contrary to expectations, respondents interviewed using the experimental format were found to be less willing to report past year and past 30-day cocaine/crack use when adjustments were made for these variables. Adjusted odds ratios and confidence intervals for these comparisons are presented in Table 2.

Table 2: Adjusted¹ Odds and 95% Confidence Intervals of Experimental Format Reports of Cocaine/Crack Use ($N=564$).

	<u>Odds Ratio</u>	<u>(CI)</u>
Lifetime	0.73	(0.46-1.17)
Past Year	0.51	(0.26-1.00)*
Past 30 days	0.44	(0.19-1.00)*

¹Adjusted for gender, age, race/ethnicity, education and employment status.

* $p \leq .05$.

Cocaine/crack concordance by interview format was next examined (Table 3). Because hair assays only provide information regarding drug use during the previous three to six months, only last year and last 30 self report concordances were examined. The differences in concordance (i.e., percentage of reports consistent with hair test results) by interview format were nonsignificant for these two reference periods. The self-reports of those assigned to the experimental interview format, though, were slightly more concordant with the hair-test assays than were those of persons assigned to the con-

rol condition. Uncorrected kappa statistics for both reference periods and interview formats are also included in Table 3. Last year and past 30-day self-reports and hair assays were concordant at only low levels. For both of these reference periods, the reports of persons interviewed via the control format were slightly more concordant with hair-assay results, compared to persons interviewed using the experimental format. Logistic regression models that adjusted for demographic variables confirmed the absence of differences in concordance across interview formats (Table 4).

Table 3. Cocaine/Crack Concordance by Interview Format ($N=322$)¹.

	<u>Control</u>		<u>Experimental</u>	
	<u>(N=151)</u>		<u>(N=171)</u>	
	<u>%</u>	<u>(N)</u>	<u>%</u>	<u>(N)</u>
Past year	66.2	(100)	71.9	(123)
Kappa	0.202**		0.134*	
Past 30 days	67.5	(102)	73.7	(126)
Kappa	0.223***		0.170***	

¹Includes all cases that contributed a useable hair sample.

* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

Table 4: Adjusted¹ Odds and 95% Confidence Intervals of Experimental Format Cocaine/Crack Use Concordance ($N=320$).

	<u>Odds Ratio</u>	<u>(CI)</u>
Past Year	1.32	(0.78-2.24)
Past 30 days	1.44	(0.82-2.51)

¹Adjusted for gender, age, race/ethnicity, education and employment status.

DISCUSSION

Our analysis suggests that the experimental interview format assessed in this study failed to improve either the willingness of respondents to report cocaine use or the consistency between reports and hair-test results. This experimental approach may have in fact actually produced lower prevalence estimates of past year and past 30-day cocaine use, although we speculate that this may be a consequence of the use of skip patterns in the experimental format. Because our study was not constructed as a factorial experiment, but rather as an evaluation of a set of innovations that were ex-

pected to collectively improve respondent reporting, it is difficult to isolate the specific effects of individual components of our intervention. Additional reports by our study team, to be available shortly, will attempt to examine the effects of individual components of our intervention. Future research should continue to investigate the usefulness of what we believe to be several promising innovations, although perhaps individually rather than as a group.

There are several threats to the internal and external validity of these findings that may limit their generalizability. Only persons aged 18-35 who lived in neighborhoods with above-average substance use were eligible to be interviewed. These eligibility factors, designed to produce a random sample with above-average drug use behaviors, also unfortunately contributed to the study's relatively low response rate. In addition, we have only reported results for cocaine/crack use here, as this was the only substance that could be assayed with a high degree of confidence and for which sufficient numbers in our sample tested positive affording us reasonable power to detect group differences. In addition, hair testing as a form of validation remains controversial for a variety of reasons, including possible race differences in test sensitivity (Fendrich, Johnson, Wislar and Sudman, In-press). That only 56% of our respondents were both willing and able to provide an adequate hair sample, of course, also limits the generalizability of our findings.

This study suggests that it is difficult to improve on the survey procedures currently being used to conduct the National Household Survey on Drug Abuse. Future large-scale epidemiologic studies of drug abuse that seek to deviate from these standards would be well-advised to conduct careful pilot and validation work before introducing radical departures from these procedures.

Regardless of the data collection methods used, however, our findings also suggest that self-reported drug use in community surveys serves as a poor indicator of actual behavior. Continued research regarding the methods for enhancing the accuracy of drug use reports, given their critical importance to both researchers and policy makers, is clearly warranted.

REFERENCES

Aquilino WS and LoSciuto LA (1990) Interview mode effects in drug use surveys. *Public Opinion Quarterly* 54: 362-395.

Belson WA (1986) *Validity in Survey Research*. Aldershot, England: Gower.

Blair E and Burton S (1987) Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research* 14: 280-288.

Briggs CL (1986) *Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. New York: Cambridge University Press.

Fendrich M, Goldstein P, Tarshish C and Bellucci PA (1992) Longitudinal measurement of substance use in ethnographic samples. *Journal of Community Psychology* 20: 326-342.

Fendrich M, Johnson T, Sudman S, Wislar J and Spiehler V (In-press) The validity of drug use reporting in a high-risk community sample: A comparison of cocaine and heroin survey reports with hair tests. *American Journal of Epidemiology*.

Fendrich M, Johnson T, Wislar JS and Sudman S (In-press) The feasibility of hair testing in a household survey on drug abuse. In T Mieczkowski (Ed.) *Drug Testing Methods: Assessment and Evaluation*. Boca Raton, FL: CRC Press.

Fendrich M and Xu Y (1994) The validity of drug use reports from juvenile arrestees. *International Journal of the Addictions* 29: 971-985.

Feucht TE, Stephens RC and Walker ML (1994) Drug use among juvenile arrestees: A comparison of self-report, urinalysis and hair assay. *Journal of Drug Issues* 24: 99-116.

Fowler FJ and Mangione TW (1990) *Standardized Survey Interviewing*. Newbury Park, CA: Sage.

Friedman WJ (1993) Memory for the time of past events. *Psychological Bulletin* 113: 44-66.

Gfroerer J, Gustin J and Turner CF (1992) Introduction. Pp. 3-12 in CF Turner, JT Lessler and JC Gfroerer (Eds.) *Survey Measurement of Drug Use*. Rockville, MD: National Institute of Drug Abuse.

Harrison LD (1995) The validity of self-reported data on drug use. *Journal of Drug Issues* 25:91-111.

Harrison L and Hughes A (1997) *The Validity of Self-Reported Drug Use*. Rockville, MD: National Institute on Drug Abuse.

- Hubbard M, Pantula J and Lessler J (1992) Effects of decomposition of complex concepts. Pp. 245-266 in CF Turner, JT Lessler and JC Gfroerer (Eds.) *Survey Measurement of Drug Use*. Rockville, MD: National Institute of Drug Abuse.
- Jobe J, Tourangeau R and Smith AF (1993) Contributions of survey research to the understanding of memory. *Applied Cognitive Psychology* 7: 567-584.
- Jobe J, White AA, Kelley CL, Mingay DJ and Sanchez MJ (1990) Recall strategies and memory for health-care visits. *The Milbank Quarterly* 68: 171-189.
- Johnston LD, Bachman JG and O'Malley PM (1986) *Monitoring the Future: Questionnaire Responses from the Nation's High School Seniors, 1985*. Ann Arbor, MI: Institute for Social Research.
- Johnston LD, O'Malley PM and Bachman JG (1994) *National Survey Results on Drug Use from the Monitoring the Future Study, 1975-1993*. Rockville, MD: National Institute on Drug Abuse.
- Loftus EF, Smith KD, Klinger MR and Fiedler J (1992) Memory and mismemory for health events. Pp. 102-137 in JM Tanur (Ed.) *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Magura S, Freeman RC, Siddiqi Q and Lipton DS (1992) The validity of hair analysis for detecting cocaine and heroin use among addicts. *International Journal of the Addictions* 27: 51-69.
- Maisto SA, McKay JR and Connors GJ (1990) Self-report issues in substance abuse: State of the art and future directions. *Behavioral Assessment* 12: 117-134.
- Means B and Loftus EF (1991) When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology* 5: 297-318.
- Menon G (1994) Judgments of behavioral frequencies. Pp. 161-172 in N Schwarz and S Sudman (Eds.) *Autobiographical Memory and the Validity of Retrospective Reports*. New York: Springer-Verlag.
- Mensch B and Kandel D (1988) Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Public Opinion Quarterly* 52: 100-124.
- Mishler EG (1986) *Research Interviewing: Context and Narrative*. Cambridge: Harvard University Press.
- Office of Applied Studies (1997) *National Household Survey on Drug Abuse Main Findings 1995*. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- Ouelett LJ, Cagle HH and Fisher DG (1997) "Crack" versus "rock" cocaine: The importance of local nomenclature in drug research and education. *Contemporary Drug Problems* 24: 219-237.
- Strack F and Martin LL (1987) Thinking, judging, and communicating: A process account of context effects in attitude surveys. Pp. 123-148 in HJ Hippler, N Schwarz and S Sudman (Eds.) *Social Information Processing and Survey Methodology*. New York: Springer-Verlag.
- Suchman L and Jordan B (1990) Interactional troubles in face-to-face survey interviews. *JASA* 85: 232-244.
- Sudman S and Phillips J (1995) The use of interactive cognitive interviewing to improve the quality of reporting of autobiographical behavior. Pp. 216-220 in *Proceedings of the International Conference on Survey Measurement and Process Quality*. Alexandria, VA: American Statistical Association.
- Tourangeau R and Smith TW (1996) Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 60: 275-304.
- Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH and Sonenstein FL (1998) Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science* 280: 867-873.
- Turner CF, Lessler JT and Devore JW (1992) Effects of mode of administration and wording on reporting of drug use. Pp. 177-219 in CF Turner, JT Lessler and JC Gfroerer (Eds.) *Survey Measurement of Drug Use*. Rockville, MD: National Institute of Drug Abuse.
- Turner CF, Lessler JT and Gfroerer JC (1992) *Survey Measurement of Drug Use: Methodological Studies*. Rockville, MD: National Institute of Drug Abuse.
- U.S. General Accounting Office (1993) *Drug Use and Measurement: Strengths, Limitations, and Recommendations for Improvement*. GAO/PEMD-93-18. Washington, DC: US General Accounting Office.