

TECHNICAL PAPERS ON HEALTH AND BEHAVIOR MEASUREMENT

TECHNICAL PAPER 53

Falsification in Epidemiologic Surveys: Detection and Remediation

Charles F. Turner, James N. Gribble, Alia A. Al-Tayyib,
James R. Chromy

Paper presented at Conference on Research Integrity sponsored by DHHS Office of Research Integrity, National Institutes of Health, National Science Foundation, American Association for the Advancement of Science, and the Association of American Medical Colleges. Bethesda, MD, November 20, 2000.

Reference Citation

Turner, C.F., J.N. Gribble, A.A. Al-Tayyib, J.R. Chromy (2002) Falsification in epidemiologic surveys: detection and remediation [Prepublication Draft]. *Technical Papers on Health and Behavior Measurement*, No. 53. Washington DC: Research Triangle Institute.

PREPUBLICATION DRAFT

Comments on this draft would be welcome. Send to cft@rti.org

Paper presented at *Research Conference on Research Integrity*. (Conference sponsored by PHS Office of Research Integrity, National Science Foundation, American Association for the Advancement of Science, American Association of Medical Colleges, and the National Institutes of Health.) Bethesda, MD, November 20, 2000.

FALSIFICATION IN EPIDEMIOLOGIC SURVEYS:

Detection and Remediation^a

Charles F. Turner, Ph.D. (a,b), James N. Gribble, Ph.D. (b,c), Alia A. Al-Tayyib, B.A. (b), James R. Chromy, Ph.D. (d)

(a) *CUNY/Queens College and Graduate Center*, (b) *Program in Health and Behavior Measurement, Research Triangle Institute*, (c) *Georgetown University School of Medicine*, (d) *Statistical Research Division, Research Triangle Institute*

Anecdotal reports of data falsification by survey interviewers appear sporadically in the mass media (1), and the topic is alluded to in textbooks on survey research (2). However, a literature search using Medline (3) and informal canvassing of colleagues with decades of experience in epidemiologic and other surveys uncovered no substantial research literature on this important topic. Indeed, written acknowledgement of *specific* instances of falsification are rare. The only published examples of which we are presently aware are: reports from the U.S. Bureau of the Census (4, discussed below), a 1951 study (5) at the National Opinion Research Center cited in the Census publications, and an episode of falsification that occurred in the University of Michigan's National Election Study of 1990 and that is noted in the technical notes accompanying the dataset (6).

Today, we present a case study of the belated detection and remediation of an epidemic of suspected data falsification in a population survey and biological specimen collection conducted in 1997-98 with a large probability sample of Baltimore adults. We present our own painful experience in this study, and we encourage other investigators and research organizations to do likewise. By doing so, we hope to create what does not presently exist --- an empirical literature on data falsification in epidemiologic surveys.

^a Throughout this paper we use the term "suspected falsifiers". We believe the evidence adduced in the field verifications (discussed later) was sufficient to conclude that the interview data submitted by these "suspected falsifiers" might be tainted and that any unverified interviews should be discarded. It is important to recognize, however, that the standards one uses in making judgments designed to protect the integrity of a database are not --- nor should they be --- the same as the standards one would use in a regulatory or other setting where reputations might be damaged or penalties imposed.

There are, for example, sampling protocol violations made by interviewers that will cause an interview to fail field verification and be treated as "falsified". Interviewing the household in apartment 8B, for example, and reporting their data as belonging to the difficult-to-locate residents of 8C would clearly involve a deceit. Such transgressions --- while destructive to the scientific purpose of the research --- are not equivalent to fabricating an entire interview.

We did not attempt to make such distinctions because our time and resources were limited, and our aim was to restore the integrity of the database not to prosecute the interviewers. We suspect, indeed, that it would be impossible in most cases to distinguish "beyond a reasonable doubt" between such protocol violations and falsification --- unless the instances were numerous or other evidence were available. (Even then, adducing evidence to make this distinction beyond a reasonable doubt would be extremely difficult, time-consuming, and expensive.)

For the purposes of protecting the integrity of the database, however, such distinctions are unimportant since the remedy is the same in either case, i.e., expunge the interview data and recollect the interviews from the sampled households.

STUDY.

The 1997-98 Baltimore STD and Behavior Survey (BSBS) [20] was designed to survey STD-related and other sensitive behaviors and to collect urine specimens for STD testing from a probability sample of adults in Baltimore, Maryland. This NIH-funded research was designed to:

(1) assess -- using a randomized experiment -- the impact of a new method of conducting survey interviews, audio computer-assisted self interviewing (audio-CASI), upon the reporting of STD-related and other sensitive behaviors (7,8);

(2) estimate the prevalence of untreated gonococcal and chlamydial infections using nucleic acid amplification tests (NAAT) on urine samples collected from a probability sample of Baltimore adults; and

(3) assess the patterns of STD-related behaviors in this population and their association with the likelihood of contracting and carrying an untreated gonococcal or chlamydial infection.

Survey Procedures.

Data collection for this project was conducted by the Research Triangle Institute, one of the nation's leading survey organizations. In most respects, operational procedures for this survey paralleled those used in other epidemiologic surveys. A Survey Director managed the project, and two experienced survey operations staff served as Data Collection Managers. Thirty-six interviewers were recruited as part-time workers. To reduce incentives for cutting corners in their work, interviewers were paid by the hour not by the completed interview.

All interviewers received two days of training on the procedures to be used in the survey. Interviewer paper work (records of household screenings, paper survey questionnaires, consent forms, etc.) was submitted for review by data collection managers who were permanent salaried employees of the survey organization. Data collection managers scheduled telephone conferences to discuss the progress and quality of work with each field interviewer on an approximately weekly basis. Data collection managers were supervised by the survey director who, in turn, reported to the principal investigator (PI) and his research team. The survey organization's standard quality control procedures required independent verification of a subsample of all survey interviews. The PI and research team scheduled conference calls with the data collection staff on a regular basis to review the progress and problems of data collection.

SURVEY PROGRESS AND PROBLEMS.

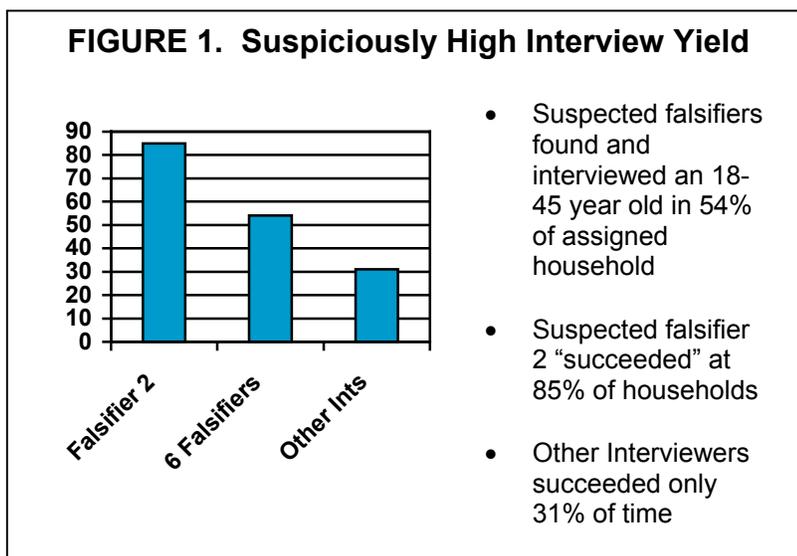
The survey began in January, 1997, and it was expected to take approximately six months to complete 1,200 interviews. At the outset of the survey, quality control procedures detected one interviewer who had submitted interviews that appeared to be falsified. This interviewer was terminated, and the "interviews" he collected were removed from the database. The other field interviewers, however, were not informed of this incident.

After six months of data collection, only 585 interviews had been completed, and study costs greatly exceeded budget estimates. During the summer of 1997, interviewers were reconvened for additional training, which emphasized, among other things, ways to encourage respondent cooperation. Survey operations proceeded through the end of the year, when evidence emerged that led to the discovery of an epidemic of suspected interview falsification.

In November of 1997, one interviewer submitted work that raised the suspicions of data collection managers because of the large number of interviews completed in a relatively short period of time. Re-contacts of the households where the interviews were supposedly conducted revealed that no interviews took place. Furthermore, discussions with data collection staff revealed that — without the knowledge of the PI or research team --- quality control procedures had been discontinued for the majority of interviewers in May, 1997.

A preliminary review by the research team and senior management of the survey organization identified a number of survey interviewers whose work appeared suspicious based upon factors such as the ratio of completed to assigned interviews.

Figure 1 presents the proportion of assigned addresses that yielded a completed "interview" for the six interviewers whose cases were dropped from the final database due to highly suspicious findings ("suspected falsifiers"). For comparison, Figure 1 also shows the completion rate for the other interviewers, and for "suspected falsifier 2" who appeared to have extraordinary success in obtaining interviews.



It will be seen from Figure 1 that our other interviewers succeeded in obtaining interviews at only 31 percent of their assigned residences.

The 31 percent interview rate achieved by our other interviewers reflects the fact that:

1. Some assigned addresses were vacant or demolished residences ;
2. Some households refused to permit household screening;
3. Some households were ineligible because they did not include an adult aged 18 to 45; and
4. Some adults selected for the survey refused to participate.

In contrast to the 31 percent interview rate achieved by other interviewers, interviewers suspected of falsifying reported obtaining interviews at 54 percent of their addresses. Indeed, one particularly "successful" interviewer reported obtaining completed interviews at 85 percent of the assigned addresses.

Our preliminary review provided dramatic evidence suggesting that data collection may have been subverted by falsification on the part of some interviewers. It also provides our first

example of the sort of data that should have been routinely analyzed during the course of data collection to permit early identification of suspicious interviewer behavior. While such data might not definitively prove that an interviewer was falsifying, it would serve to identify interviewers whose work should be subject to heightened verification efforts.

FIELD VERIFICATIONS.

We subsequently undertook a large-scale field verification effort to detect and eliminate from the database any interviews we suspected could be falsified. One hundred percent of the interviews submitted by interviewers judged to be "suspicious" were subject to independent verification. In addition, 40 percent of the interviews conducted by all other interviewers were selected for independent verification (9). Verification was performed by telephone, if a telephone number was available, or by a trusted and experienced interviewer who had not worked on the study, if telephone contact could not be made. Interviewers whose work appeared suspicious based upon 40 percent verification were subject to 100 percent verification.

After this verification process was complete, the PI and research team reviewed the verification data and all written survey materials for each interview (e.g., screening forms, paper self-administered questionnaires, and forms used in literacy testing) and each household screening. Based on this evidence, the research team concluded that six field interviewers had submitted cases that appeared to be falsified for 49 percent of the 451 interviews they submitted. We also concluded that 26 percent of the interviews submitted by these interviewers were true interviews and 25 percent could not be classified (10). For the six interviewers suspected of falsifying, the proportion of interviews verified as "true interviews" ranged from 1 percent to 72 percent.

REMEDIATION.

Given these findings, we adopted a strategy designed to:

- purge the survey database of all taint of falsification, and
- expeditiously complete the survey by collecting real data to substitute for the data we suspected were falsified.

Since we were surveying a probability sample of the population of residences in Baltimore, remediation was more straightforward than it might have been in other cases. In particular, we had a full record of the addresses where household screenings and interviews (if eligible) needed to be conducted to replace the data we suspected to be falsified. Second, since our sample was constructed using probability sampling, it was possible to draw a random subsample of these addresses for re-fielding so that the study could be completed in an efficient and timely manner. (This subsampling did, of course, inflate the variances of our survey estimates somewhat.)

All interviews obtained by any interviewer under suspicion were discarded if positive evidence could not be adduced that the interview had, in fact, been conducted (N = 348). Their household screenings that did not result in an interview were also discarded (N = 386). To recover from this data loss and to preserve the sample as a probability sample of the Baltimore population, all addresses assigned to interviewers under suspicion became candidates for refielding (unless the interview data had been verified). A randomly selected subsample of 50

percent (11) of these addresses was assigned for refielding by interviewers whose work had proven trustworthy.

OUTCOME.

Survey data collection was finally completed in September, 1998. Our 1997 Baltimore STD and Behavior Survey thus became the 1997-98 BSBS. The survey organization absorbed the fieldwork and associated costs of the remediation efforts, and the research team sacrificed uncountable hours of labor and anguish. Indeed, the Principal Investigator regards 1998 as the most harrowing year of his professional career.

Nonetheless, having removed the suspicious data from the database and refielded a random subsample of the addresses for interviews that could not be verified, the survey did provide good coverage of the sampled population. A total of 3,182 addresses were selected for interview and 2,727 (85.7 percent) were successfully screened. Screening identified a total of 1,224 English-speaking adults between the ages of 18 and 45 eligible for interviewing (12). Interviews were completed with 1,014 (82.8 percent) of these eligible adults (13).

LESSONS.

Overview.

This expensive survey effort came perilously close to disaster. Data collection was almost complete when the suspiciously high productivity of one interviewer triggered concerns about interview falsification. This led to discovery of the unexplained discontinuation of most interview verification efforts. Our subsequent field verification efforts documented an epidemic of apparent interview falsification that would have subverted most scientific uses of these survey data. Fortunately, the survey was still in progress, so it was possible to remedy the problem by removing what appeared to be falsified or unverifiable data and re-collecting new data from the sampled address.

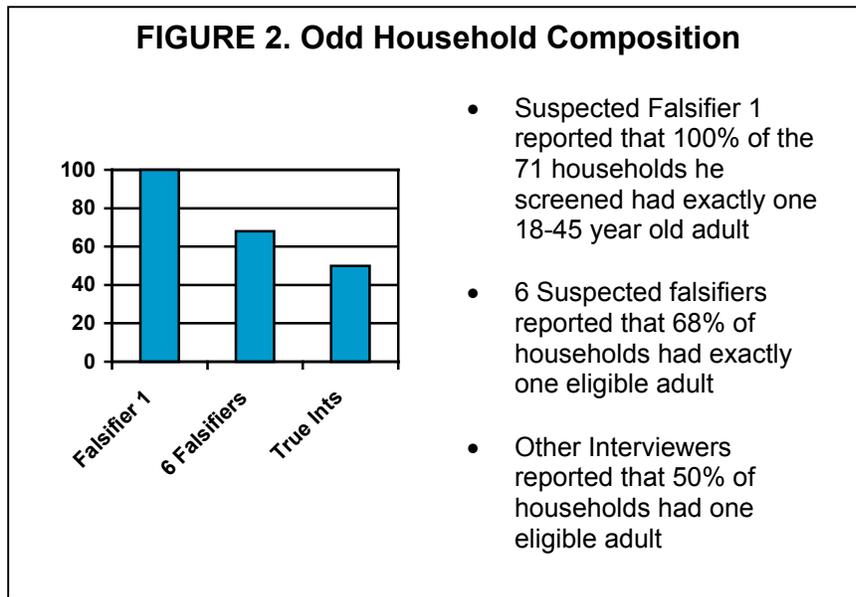
Our research team and the survey organization hope the research community can learn from our painful experience. This presentation is one attempt to foster that learning. We hope that other investigators and survey organizations will share their experiences in the detection and remediation of interview falsification in epidemiologic and other large-scale surveys.

As McNemar (14) noted more than five decades ago: "All measurement is befuddled with error." Those measurements that rely upon human data collectors are vulnerable to all manner of human foibles. We have spent decades attempting to understand and minimize the bias in respondents' self-reports of illicit and stigmatized behaviors (15). Given the important role surveys now play in epidemiologic research, similar attention should be paid to the study of conditions that discourage falsification by survey interviewers and techniques that permit rapid detection of such falsification when it occurs.

Clues.

Comparison of work submitted by interviewers suspected of falsification (based on field verifications) to that of other interviewers provides guidance on the sorts of clues one might use to detect falsification. As we previously noted, the "work" submitted by interviewers who failed field verification showed an unexpectedly high yield of interviews per assigned sample address (see Figure 1). A thoughtful falsifying interviewer might, of course, reduce their "success" rate to better mimic the impact of vacant households, households with no eligible adults, interview refusals, etc. However, this information is not always available to interviewers. Furthermore, our experience suggests that some interviewers made no conscious effort in this regard and

thus submitted work with interview rates that are simply unbelievable, e.g., a purported 85% interview success rate when only households with an 18-45 year old resident were eligible to participate in the survey.

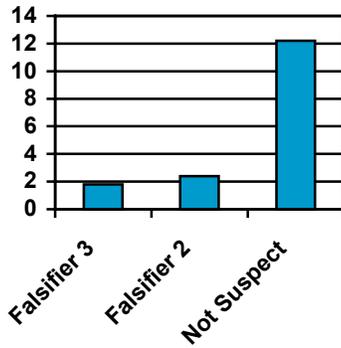


Similarly, examination of the household compositions reported during screening (see Figure 2) reveals substantial variation between suspected falsifiers and other interviewers (17). Over all reported interviews, interviewers who were not suspected of falsification found that only 50 percent of households had one eligible adult. Suspected falsifiers found that 68

percent of their “interviewed households” had a single eligible adult. Indeed, one suspected falsifier reported that in all 71 of his “interviews”, the household contained exactly one eligible adult. Here again, we have results which are simply unbelievable. If household composition reports had been monitored, they would have prompted heightened field verification efforts examining this interviewer’s work.

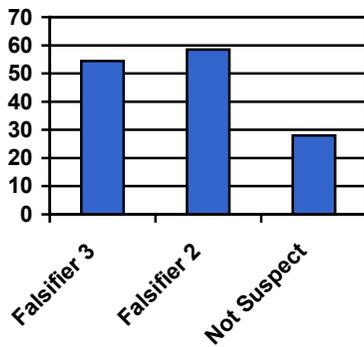
While interview success rates and household structure information will be available in most surveys, there will often be additional information that can be gleaned from unusual “response” patterns reported for substantive questions. In our survey, respondents were asked about a range of sexual and other sensitive behaviors. To provide “realistic” responses to such questions, a falsifying interviewer would need to accurately anticipate the actual distribution of responses by the population at large. Obviously many, if not most, interviewers will lack a clear idea of the patterns of sexual activity reported by actual respondents. Thus, falsified responses may be detectable because they vary substantially from reality --- or from the results obtained by other interviewers. Figures 3 through 5 provide such examples for reporting of the number of lifetime sex partners and for the proportion of respondents who report they have not had sex in the past 7 days. In each instance, there were suspected falsifiers whose “interviews” depicted substantially more sexually activity than was reported to interviewers who were not suspected of falsifying.

FIGURE 3. Falsifier's World View: Only 1 Lifetime Partner



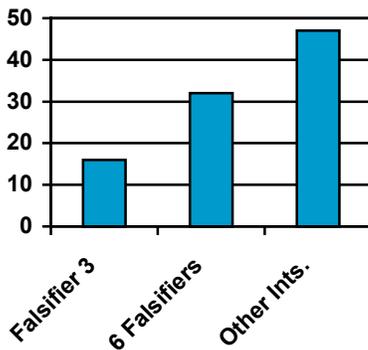
- Suspected Falsifiers 2 and 3 found 2%.
- Interviewers not suspected of falsification found 12% had 1 partner.

FIGURE 4. Falsifier's World View: 11+ Lifetime Sex Partners



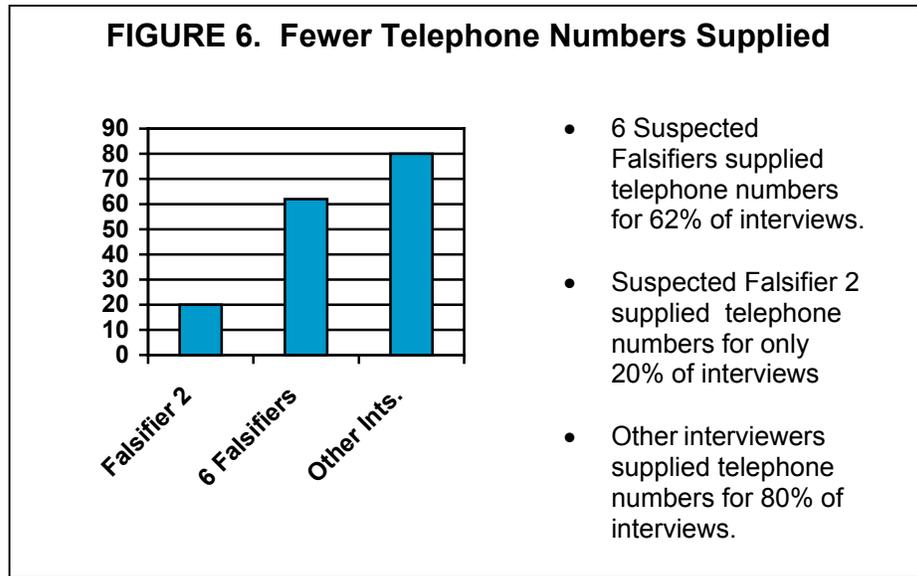
- Suspected Falsifiers 2 and 3 found 59% and 54% had 11+ lifetime sex partners
- Interviewers not suspected of falsification found 28% had 11+ lifetime sex partners.

FIGURE 5. Falsifier's World View: No Sex in Last 7 days?



- 6 Falsifiers found 32%
- Falsifier 3 found only 16%
- Other Interviewers found 47% had no sex in past 7 days
-

Our final example (Figure 6) involves an indicator directly relevant to verification. Field surveys often request that survey respondents provide telephone numbers. This permits telephone recontacts for many reasons, including verification that an interview has been conducted. As Figure 6 shows, there were systematic differences between suspected falsifiers and other interviewers in the proportion of interviews in which a telephone number was reported for the respondent (80 versus 62 percent). Indeed, the worst of the suspected falsifiers (16) provided telephone numbers for only 20 percent of his interviews.



IMPLICATIONS.

We believe that our experience has four major implications for other investigators who employ survey organizations to conduct large-scale epidemiologic surveys.

1. We strongly recommend that Principal Investigators take responsibility and budget time for themselves and for their research team to validate the survey interviews themselves. As scientists, we believe it is a mistake to delegate verification entirely to a data collection unit. Scientists often have better tools and substantive knowledge appropriate to the detection of suspicious patterns of interview data. (The examples above speak well to this issue.) In addition, there may sometimes be different incentives for the research team and the data collection unit to uncover problems. The research team must live with the data and its problems for many years after data collection ends. They will be particularly well served by the detection of survey problems while they can still be remedied.
2. Data analysis should begin as soon as practical after data collection commences. Widespread adoption of computerized data collection tools (CAPI, Audio-CASI, etc.) and daily uploading of field interview data by modem or other means, makes it possible for the research team to analyze data as it is collected. The earlier data analysis commences, the sooner clues to potential falsification (and other data collection problems) can be detected.
3. Where possible, the PI should opt for on-site supervision of the interviewing staff. In the 1997-98 BSBS, data collection managers were located at the survey organizations' headquarters in North Carolina while interviewers conducted the survey in Baltimore City. Although the research team cannot be sure, we suspect that better communication and more rapid identification of problems might have occurred if there had been regular in-person contact between supervisory and interviewing staff. (In this

study, which was restricted to a single city, on-site supervision would have been feasible. Obviously that would not be possible in field surveys of national or regional populations.)

4. Investigators should insist on the rapid re-assembly of research materials --- both electronic data and paper forms --- during data collection. Review of complete records can improve the odds of detecting suspicious results, e.g., an interviewer screening form recording a dozen different household visits on different dates --- all recorded in the same unusual ink color used on the paper forms completed by the survey respondent.

To these recommendations we add one further insight derived from our own experience and from the more extensive studies conducted at the U.S. Bureau of the Census, albeit with very different sorts of surveys. The 1997-98 BSBS required recruitment of a large number of interviewers in a small city. It was impossible to fully staff this survey using only the cadre of professional interviewers who typically staff the surveys conducted by major survey organizations. There simply were not enough such interviewers available. In this regard, paradoxically, large local surveys are a more difficult undertaking than large national surveys. (National surveys have only a fraction of their sample in any given city, and thus they are unlikely to exhaust the supply of professional interviewers.)

All seven (18) interviewers suspected of falsification in the 1997-98 BSBS were newly or recently recruited to survey work. Experience at the Census Bureau (19) is consistent with our experience. Studies of falsification in the monthly Current Population Survey and annual National Crime Survey find that 0.4 percent of interviewers in these surveys falsified interviews. These national surveys employ a stable workforce of professional interviewers. In contrast, the Census Bureau reports falsification rates of 6.5 percent for the Household Vacancy Survey which is conducted once every three to four years in New York City by a temporary staff of interviewers. These findings should prompt investigators to insist that heightened quality control be used with new interviewers.

REFERENCES AND NOTES.

1. Mann, C.C. (2000) How Many is Too Many. *Atlantic Monthly*, February 1993.
2. See, for example, C. Moser and G. Kalton (1971) *Survey Methods in Social Investigation*. New York: Basic Books, p. 45; Babbie, E. (1973) *Survey Research Methods*. Belmont, Calif.: Wadsworth, p. 183-184; C. Blackstrom and G. Hursh-Cesar (1981) *Survey Research*. (second edition). New York: Wiley, p. 297-298.
3. Medline searches were conducted using MESH terms: (1) survey interview and fraud; (2) survey interview and falsification; and (3) survey data and falsification. These searches yielded one article that reported on fictitious interview data: Inciardi, J.A. (1981) Fictitious data in drug abuse research. *International Journal of the Addictions*, 16:377-380. (A fourth search using the terms: survey data and fraud yielded a large number of citations to surveys of health care fraud.)
4. See Schreiner, I. Pennie, K. and Newbrough, J. (1988) Interviewer falsification in Census Bureau surveys. *Proceedings of the American Statistical Association (Survey Research Methods Section)*, 491-496. Forsman, G. and Schreiner, I. (1991) The design and analysis of reinterview: An overview. In Biemer, P. et al., eds. (1991) *Measurement Errors in Surveys*. New York: Wiley. Biemer, P. and Stokes, S.L. Optimal design of quality control samples to detect interviewer cheating. *Journal of Official Statistics*, 5: 23-39.
5. Sheatsley P.B. (1951) An analysis of interviewer characteristics and their relationship to performance, Part II. *International Journal of Opinion and Attitude Research*, 5: 79-94. (Cited in Biemer and Stokes, see note 5).
6. Subsequent to the original release of those data, the researchers deleted 20 (of 2,000) cases because they were "fabricated interviews or [interviews with] ineligible respondents". This is noted in the "Help.Txt" file on the data CD-ROM. This file contains the advisory: "Note that the 1990 data file (ICPSR #9548) now includes 1,980 cases rather than the 2,000 cases in the ICPSR releases preceding the issuance of this CD. [20 cases from 1990 were discovered, prior to the release of 1992 data, to be fabricated interviews or ineligible respondents." See, Inter-university Consortium for Political and Social Research (1995) *American National Election Studies, 1948-1994*. (Data CD-ROM) Ann Arbor, Mich.: Institute for Social Research, University of Michigan.
7. Turner, C.F., L. Ku, S.M. Rogers, L.D. Lindberg, J.H. Pleck, and F.L. Sonenstein (1998) Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, 280:867-873.
8. Turner, C.F., B.H. Forsyth, J. O'Reilly, P.C. Cooley, T.K. Smith, S.M. Rogers, and H.G. Miller. (1998) Automated self-interviewing and the survey measurement of sensitive behaviors. In Couper, M., et al., eds., *Computer-Assisted Survey Information Collection*. New York: Wiley and Sons.
9. All interviews (N = 1,003) in the study database through December 8, 1997 were subject to verification at rates of either 100 percent (FIs under suspicion) or 40 percent (FIs not under suspicion). Subsequent interviews by interviewers whose work passed this verification were verified at a reduced rate (15 to 20 percent). Over the course of the study, 56 percent of all completed survey interviews were subject to independent verification.

10. The 6 interviewers suspected of falsification submitted a total of 451 interviews. When verification efforts were complete, we concluded that: 116 (26%) of these interviews were true interviews; 223 (49%) of these interviews appeared to be either fabricated interviews or interviews conducted at households other than the one assigned; 111 (25%) of these interviews could not be classified (e.g., resident had moved); and 1 interview was coded as "field verification pending". (For the purpose of remediation, this case was treated the same as interviews that could not be classified.) Failure to classify an interview typically resulted from inability of the verification interviewer to contact the persons residing in the household on the date of the survey interview, e.g., because the former residents had moved. (Note also that this tabulation includes the 6 interviewers detected using the expanded verification procedures. The interviewer described on page 2 whose data were removed from the database at outset of the study is not included in this tabulation.)

11. A total of 734 interviews and screenings were discarded, and they would become candidates for re-fielding. 13 of these cases were dropped from the re-fielding for logistical reasons. A 50 percent random sample of the remaining 721 sample units were re-fielded.

12. Only one randomly-selected adult per household was eligible to participate in the survey.

13. The 1,014 cases in the final database include 10 cases which were submitted by "true" interviewers but failed verification. The research team concluded that these cases were likely to be "ordinary" survey errors (made by the original interviewer or the verification interviewer) or cases in which the household informant(s) did not provide consistent information at the initial contact and recontact. "Ordinary" survey errors would include instances in which the original interviewer (or verification interviewer) mistakenly went to the wrong street address. Inconsistent responses by household informants might include cases in which the informant disguised the identity of household residents when the interviewer attempts to roster the household members. In preparing publications, the research team intends to repeat key analyses with and without these 10 cases to insure (prior to publication) that findings are not substantially affected by the data obtained from these cases.

14. McNemar, Q. (1946) Opinion-attitude methodology. *Psychological Bulletin*, 43: 294.

15. See, for example, notes 2-3 above, and Gribble, J.N., Miller, H.G., Catania, J.A., Pollack, L., Turner, C.F. (2000) The Impact of T-ACASI Interviewing on Reported Drug Use among Men Who Have Sex With Men. *Substance Use and Misuse* 35: 869-890; Miller, Turner, Moses, eds. (1990) Methodological issues in AIDS Surveys. Chapter 5 in *AIDS: The Second Decade*. Washington D.C.: National Academy Press; Turner, C.F., Lessler, J., and Gfroerer, J., eds. (1992) Survey Measurement of Drug Use: Methodological Studies. Washington D.C.: Government Printing Office. Rogers, S.M., Gribble, J.N., Turner, C.F., and Miller, H.G. (1999) Entretiens autoadministrés sur ordinateur et mesure des comportements sensibles (Computerized self-interviewing and the measurement of sensitive behaviors). *Population* 54: 231-250; Des Jarlais, D.C., D. Paone, J. Milliken, C.F. Turner, H.G. Miller, J.N. Gribble, Q. Shi, H. Hagan, S.R. Friedman. (1999) Audio-computer interviewing to measure risk behavior for HIV among injecting drug users: A quasi-randomized trial. *Lancet* 353(9165): 1657-1661.

16. These comparisons exclude data from one sampling strata in which we screened for males residing in a sample of Census tracts that had 95 to 100 percent black residents in 1990. (**Note.** This is being checked to make sure it is accurate.)

17. That is, the suspected falsifier with the lowest rate of interviews verified as true interviews.

18. This includes the 6 interviewers who failed field verification and the 1 interviewer who was suspected of falsification and dismissed at the beginning of the survey.

19. See Forsman and Shreiner, reference note 4, pages 293-295.

20 Turner CF, Rogers, SM, Miller HG, et al. (2002) Untreated gonococcal and chlamydial infection in a probability sample of adults. *JAMA*, 287: 726-733.