THE B FILES

Case studies of bias in real life epidemiologic studies

Bias File 8. Don't call my number, anymore! Bias in surveys of sexual behavior

Compiled by

Madhukar Pai, MD, PhD

Jay S Kaufman, PhD

Department of Epidemiology, Biostatistics & Occupational Health

McGill University, Montreal, Canada

madhukar.pai@mcgill.ca & jay.kaufman@mcgill.ca

Bias File 8. Don't call my number, anymore! Bias in surveys of sexual behavior

The story

Surveys of sexual behavior are commonly done, but prone to all sorts of biases and challenges. These surveys are hard to design and implement. For example, non-response bias is a huge concern, especially with surveys that aim to collect data on sensitive and high-risk behavior and practices. As Fenton and colleagues point out in their review (2001), "a key challenge for all sex survey research is to generate unbiased and precise measures of individual and population behaviour patterns. Methods are needed to minimise measurement error which may be introduced by participation bias, recall and comprehension problems, and respondents' willingness to report sensitive and sometimes socially censured attitudes or behaviours."

The study

As an example of a survey done to ascertain sexual behavior at the population level, consider the AIDS and Sexual Behavior in France study, published in 1992 (ACSF, Nature, 2002). This was a massive telephone survey on sexual lifestyles in France, and involved more than 20,000 participants. After pilot research, the telephonic method was selected, and involved more than 100 interviewers. The main results of this survey are shown in the table below.

TABLE 1 Results of 20,055 questionnaires (including 4,820 long questionnaires)

	MEN				WOMEN			
	Age (years)				Age (years)			
	18–24	25–34	35–44	45–69	18–24	25–34	35–44	45–69
Population	1,716	2,232	2,284	3,696	1,670	2,238	2,261	3,958
Age at first intercourse (years)	16.5	16.9	17.6	18.2	17.1	17.9	18.8	20.8
Multipartner — 1 year	27.6	14.1	11.5	8.3	12.1	6.8	5.9	2.9
Intercourse (no.) — 4 weeks (m)	7.6	9.6	9.8	6.8	8.3	8.9	8.5	5.8
Homosexuality — life (%)	2.4	4.2	4.3	4.5	1.2	3.8	2.8	2.4
Intercourse with prostitutes — 5 years	4.7	3.8	3.4	1.8	—	—	—	—
IV drugs — life (%)	0.4	1.2	0.6	0.0	0.2	0.5	0.2	0.0
Condoms — life (%)	73.1	58.1	57.8	48.1	54.7	48.9	49.6	33.9
Condoms — 1 year	58.8	32.3	30.5	17.2	40.8	26.7	23.0	11.1
Monopartner	50.5	26.0	26.5	14.6	37.8	24.1	22.3	10.2
Multipartner (heterosex.)	79.0	69.0	59.7	42.7	62.4	62.0	33.3	33.1

Not surprisingly, the proportion of participants who admitted to using IV drugs was very low. As the authors themselves pointed out, "people who regularly use drugs are the most difficult to contact, and/or most often refuse to participate in any kind of survey or to acknowledge an illegal practice." Social desirability bias is always a concern in these situations. Social desirability bias is the tendency of respondents to reply in a manner that will be viewed favorably by others. This will lead to overreporting good behavior and/or underreporting bad behavior.

There is considerable evidence that interviewer-administered surveys elicit lower self-reports of sensitive behaviors. Self-administration reduces social desirability bias and also provides anonymity. Computerization and audio-assistance may reduce measurement error. Audio presentation

improves question understanding and reduces response error, item refusal, and “don't know” responses. Computerization may rule out inconsistent and "out-of-range" values and is expected to reduce motivational bias (Schroder, 2003).

As an illustration, consider the study by Turner et al (2009) that assessed the impact of Telephone Audio Computer-Assisted Self-Interviewing (T-ACASI) on the reporting of sensitive (mainly heterosexual) behaviours. This technology asks questions by playing digitized voice files, and respondents provide answers using the keypad of their touchtone telephones. In a randomized experiment that was embedded in a telephone survey that drew probability samples of the populations of the USA and Baltimore city, respondents were randomly assigned to have questions asked either by a T-ACASI computer or by a human telephone interviewer. Compared with interviewer-administered telephone surveys, T-ACASI obtained more frequent reporting of a range of mainly heterosexual behaviours that were presumed to be sensitive.

In a previous study, Turner et al. (Science 1998) compared the audio-CASI method with the more traditional self-administered paper questionnaire among 1690 respondents in the 1995 National Survey of Adolescent Males. Estimates of the prevalence of male-male sex, injection drug use, and sexual contact with intravenous drug users were higher by factors of 3 or more when audio-CASI was used. Increased reporting was also found for several other risk behaviors. Some of the key results are shown in the Table below.

Measurement	Estimated prevalence (per 100)		Crude OR	Adj. OR
	Paper SAQ	Audio-CASI		
	Drug use			
Ever taken street drugs using a needle	1.4	5.2	3.85***	3.90*
Injected drugs within last year‡	0.0	0.8	–†	–†
Ever shared needle§	0.1	1.1	9.71**	9.56**
Smoked marijuana daily during last year‖	4.1	6.7	1.69*	2.03*
Used crack/cocaine within last year	3.3	6.0	1.89	1.96
Drank alcohol last year¶	65.9	69.2	1.16	1.29
Drank alcohol weekly last year≠	15.0	19.4	1.34	1.56*
Ever smoked marijuana	41.2	43.3	1.09	1.30*

The bias

Surveys of sexual behavior are prone to a whole range of biases, ranging from sampling bias, participation bias, to misclassification bias (reviewed by Catania et al. 1990; Fenton et al. 2001, and Schroder et al. 2003). As emphasized by Catania and colleagues, "privacy, embarrassment, and fear of reprisals are but a few of the reasons that may motivate people to conceal their true sexual behavior. In contrast, some people may find it rewarding to embellish their actual sexual performance and experience (bragging). These barriers to accurate reporting aside, even highly motivated and uninhibited respondents may have trouble recalling how often and with how many people they have performed specific sexual behaviors." (Catania et al, 1990).

Among other issues, the type of interview is critical; anonymity, privacy and confidentiality are all key components. Interviews can be face-to-face, by telephone, or self-administered (either as questionnaires, or via computer-assisted methods). A growing body of evidence suggests that telephone

surveys are subject to non-trivial reporting biases because they require respondents to disclose sensitive, stigmatized or illicit behaviours to human interviewers. This almost certainly must have influenced some of the data collected in the AIDS and Sexual Behavior in France study. The use of methods such as T-ACASI might have helped overcome some of the issues, and generated a more valid estimate of prevalence of high-risk sexual behaviors. But even methods such as T-ACASI cannot overcome the problems inherent in use of the telephone survey methodology. As pointed out by Kempf et al. (2007) in their excellent review on telephone surveys, the widespread use of answering machines, voicemail, caller ID, and cell phones have contributed to decreasing response rates in telephone surveys. Increasing use of cell phones has made it increasingly difficult to implement traditional telephone surveys using random-digit-dialing to landline telephones.

The lesson

All sample surveys are challenging to design and implement. This is especially true with surveys of sexual behaviors. Every survey method has its own advantages and limitations. This is exemplified by the telephone interview methodology. In addition to the general challenges in survey design, measurement of sexual behavior has its own set of unique problems. Computer assisted techniques, by improving internal consistency and increasing privacy and interviewee control, can potentially improve survey validity. But they do not address all of the problems in sexual behavior surveys.

Sources and suggested readings*

1. Fenton KA, Johnson AM, McManus S, Erens B. Measuring sexual behaviour: methodological challenges in survey research. Sex Transm Infect. 2001 Apr;77(2):84-92.
2. AIDS and sexual behaviour in France. ACSF investigators. Nature. 1992 Dec 3;360(6403):407-9.
3. Turner CF, Al-Tayyib A, Rogers SM, Eggleston E, Villarroel MA, Roman AM, Chromy JR, Cooley PC. Improving epidemiological surveys of sexual behaviour conducted by telephone. Int J Epidemiol. 2009 Aug;38(4):1118-27.
4. Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. Science. 1998 May 8;280(5365):867-73.
5. Catania J et al. Methodological Problems in AIDS Behavioral Research: Influences on Measurement Error and Participation Bias in Studies of Sexual Behavior. Psychological Bulletin 1990, Vol. 108, No. 3,339-362.
6. Schroder KE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: I. Item content, scaling, and data analytical options. Ann Behav Med. 2003 Oct;26(2):76-103.
7. Schroder KE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. Ann Behav Med. 2003 Oct;26(2):104-23.
8. Kempf AM, Remington PL. New challenges for telephone survey research in the twenty-first century. Annu Rev Public Health. 2007;28:113-26.

*From this readings list, the most relevant papers are enclosed.

AIDS and sexual behaviour in France

*ACSF investigators**

The results of a massive telephone survey of sexual lifestyles in France should provide a basis for prevention strategies for AIDS and sexually transmitted diseases.

WITH sexual transmission of the human immunodeficiency virus (HIV) now known to be widespread in northern Europe and North America[1], many countries in these areas have decided to undertake surveys of sexual behaviour in the general population. The aim of these studies is to achieve better-defined strategies for preventing sexually transmitted diseases (STDs) and AIDS, as well as to provide the basis for more accurate models to determine the development of the epidemic[2]. Some preliminary studies have already been conducted in countries such as Norway[3], Denmark[4] and Scotland[5]. Although the study of sexual behaviour has been amply legitimized by the context of AIDS, this legitimacy cannot unfortunately be taken for granted everywhere[6,7].

In France, a survey on sexual behaviour in a representative sample of the general population was carried out in 1970 arising from the increasing use of medical contraception[8]. In response to the spread of the AIDS epidemic, the French government asked the *Agence Nationale de Recherches sur le Sida* (ANRS) to perform a survey of sexual behaviour and AIDS in the general population (called ACSF for *Analyse des Comportements Sexuels en France*). To our knowledge, this study is the most comprehensive of its kind. It has been constructed in a way that can provide a basis for comparisons with other countries, particularly with the British study, which deals with a similar sample in a comparable country[7].

Our work began in July 1989. Three pilot surveys were carried out in the first 2 years to test the questionnaire, decide on the method of investigation (telephone or face-to-face) and to see if sending out a notifying letter affected whether people would participate in the study. The results prompted us to choose the telephone method for the national survey, with a notifying letter being sent beforehand[9,10]. The population base was assembled from a list of all the telephone numbers in France (94% of French households have a telephone). A total of 110 interviewers took part in the survey from September 1991 to February 1992, following a 2-day training period. Each telephone number was tried for up to 12 times.

The first part of the questionnaire, which lasted 15 minutes, was designed to measure the prevalence of a number of risk indicators: homosexual/bisexual intercourse; intercourse with prostitutes over the past 5 years; multiple partners; and drug consumption in the past year. This part of the survey was applied to everyone questioned: a total of 20,055 people aged 18–69 years. An additional questionnaire lasting around 30 minutes was administered to all people replying to the first questionnaire who reported one of the above risk indicators (n = 2,271), and to a control group of people selected at random by their date of birth (n = 2,549)[11]. The longer questionnaire included a sexual biography and details of several psychological and social characteristics.

Acceptability and reliability

Considerable efforts were made to maximize the response rate of the survey and to obtain the most reliable information possible. To convince the interviewees that their answers would remain anonymous, a letter explaining this policy was sent to all the selected households beforehand. All identifying information (name, address, telephone number) was automatically destroyed as soon as the reply to the first question was entered into the computer system.

Once selected, a household was informed that the aim of the survey was to make prevention of disease more effective, and that the survey was being conducted by researchers employed by public-health institutions. The theme of AIDS and sexual behaviour was deliberately not mentioned in the letter to avoid worrying people, to prevent refusals before selection of the interviewee and to prevent people from preparing answers in advance. (The results of our pilot survey indicated that this strategy would be sensible.) Once the eligible person had been randomly selected from within the household, he or she was informed that the questionnaire concerned AIDS prevention and included questions on sexual behaviour. Given that AIDS can be sexually transmitted, the questions referring directly to the disease were placed after questions on sexual behaviour to minimize the possibility of people inaccurately claiming 'socially desirable' behaviour, such as condom use.

A random sample was used, first to ensure compatibility with surveys in other countries, and second because quota sampling leaves a broad freedom of choice in which subjects are interviewed, and generally leads to higher reported proportions of multipartner or homosexual individuals[12]. To minimize recall problems, it was decided to focus on a detailed description of sexual practices and means of protection used in the last two sexual encounters that may have taken place in the year preceding the study[13]. This policy is identical to that adopted in the Norwegian survey[3]. A linguistic analysis of 100 telephone interviews during the pilot study allowed us to adjust the wording of the questionnaire so that the language used would be as clear and neutral as possible, without using words that were considered either too technical or over-vernacular[14].

A total of 40,000 telephone numbers were drawn at random from all the telephone numbers in France, of which 30,157 were potentially eligible. The refusal rate for the households contacted was 11.8%. Subsequently, 11.6% of selected individuals refused to participate. Only 7% of these refusals were prompted by a reluctance to answer questions on sexual behaviour.

If households where no one could be contacted are not taken into account (that is, if we assume that these people would have refused at the same rate as those who were contacted), the refusal rate was 23.5%, very close to the rate

* ACSF (*Analyse des Comportements Sexuels en France*) investigators: Alfred Spira (Director), Nathalie Bajos (Coordinator), André Béjin, Nathalie Beltzer, Michel Bozon, Béatrice Ducot, André Durandeau, Alexis Ferrand, Alain Giami, Augustin Gilloire, Michel Giraud, Henri Leridon, Antoine Messiah, Dominique Ludwig, Jean-Paul Moatti, Lise Mounnier, Hélène Olomucki, Jeanine Poplavsky, Benoit Riandey, Brenda Spencer, Jean-Marie Sztalryd, Hubert Touzard. ACSF is a survey from the *Agence Nationale de Recherches sur le Sida* (ANRS). It is conducted and coordinated under the scientific responsibility of INSERM Unit 292, and the survey associates researchers from INSERM, INED, CNRS and universities Paris V, Paris VII, Paris XI and Paris XIII. This study is supported by ANRS, the *Direction Générale de la Sante* (DGS), the *Comité Français d'Education pour la Santé* (CFES) and the *Agence Française de Lutte contre le Sida* (AFLS — Agency for AIDS Prevention). Correspondence should be addressed to A. Spira at Hôpital de Bicêtre, 78 rue du Général Leclerc, 94275 Le Kremlin–Bicêtre Cedex, France.

TABLE 1 Results of 20,055 questionnaires (including 4,820 long questionnaires)

	MEN Age (years)				WOMEN Age (years)			
	18–24	25–34	35–44	45–69	18–24	25–34	35–44	45–69
Population	1,716	2,232	2,284	3,696	1,670	2,238	2,261	3,958
Age at first intercourse (years)	16.5	16.9	17.6	18.2	17.1	17.9	18.8	20.8
Multipartner — 1 year	27.6	14.1	11.5	8.3	12.1	6.8	5.9	2.9
Intercourse (no.) — 4 weeks (m)	7.6	9.6	9.8	6.8	8.3	8.9	8.5	5.8
Homosexuality — life (%)	2.4	4.2	4.3	4.5	1.2	3.8	2.8	2.4
Intercourse with prostitutes — 5 years	4.7	3.8	3.4	1.8	—	—	—	—
IV drugs — life (%)	0.4	1.2	0.6	0.0	0.2	0.5	0.2	0.0
Condoms — life (%)	73.1	58.1	57.8	48.1	54.7	48.9	49.6	33.9
Condoms — 1 year	58.8	32.3	30.5	17.2	40.8	26.7	23.0	11.1
Monopartner	50.5	26.0	26.5	14.6	37.8	24.1	22.3	10.2
Multipartner (heterosex.)	79.0	69.0	59.7	42.7	62.4	62.0	33.3	33.1

reported in the British survey[7]. It is also similar to those found in other surveys conducted on closely related subjects in France[15]. Of the 20,687 questionnaires started, 632 (3.1%) were not completed. The refusal and drop-out rates did not differ significantly between interviewers. A total of 20,055 questionnaires were available for analysis, corresponding to 9,928 men (including 2,642 long questionnaires of which 1,146 were 'controls') and 10,127 women (including 2,178 long questionnaires of which 1,403 were 'controls').

For questions on sexual practices, the non-response rate varied between 0 and 4.2%, the higher figure being for questions relating to anal sex. For all the other questions, non-response rates were always below 2%, except for questions involving sexual relationships with people of the same sex, where the non-response rate was 1.1% for men but 5.9% for women ($P<0.001$). The main results are summarized in Table 1.

Initial results

The median age at the time of first sexual intercourse was 20 years for men and 22 years for women born in 1922–26, and 17 and 18 years respectively for those born in 1972–74. No increase of age at first sexual intercourse was apparent in younger people who began their sexually active life after the onset of the AIDS epidemic.

No definition of 'sexual intercourse' was provided in the questionnaire. Between 4 and 5% of people who said they lived with a partner nonetheless declared no sexual partners for the reference period. The actual distribution of the number of sexual partners over the preceding year is indicated in Table 2. The difference between men and women is much more pronounced over a whole lifetime than it is for shorter periods of time, as the total number of sexual partners during a lifetime has a mean value of 11.0 for men and 3.3 for women ($P<0.001$), these figures being respectively 1.2 and 0.9 for the past year.

A small difference emerges for the number of occasions of intercourse over the past 4 weeks (8.0 for men and 7.0 for women). The proportion of people who said that they had had intercourse with at least two people during the past year differed between men and women (13.3 and 5.6% respectively, $P<0.001$). This difference between the sexes is reported in all surveys of this type[4,16], and is presumed to stem from both female understatement and male overstatement as well as from a greater diversity of male responses. However, the difference between the sexes seems narrower than it was 20 years ago when the ratio between men and women for the total number of partners was reported to be around 1 to 6 (ref. 8).

The proportion of multipartner subjects (at least two sexual partners over the previous year) decreases sharply with age, especially after 25 years ($P<0.001$). Of the youngest multipartner subjects (18–24 years), the proportion of those declaring at least two sexual partners at the same time (6.9% for men and 5.0% for women) is still lower than in the 35–49-year age group, where it reaches its maximum (33.7% for men and 51.7% for women). Young people change partners more frequently than older people but are less likely to have more than one sexual partner at any one time.

The proportion of multipartner subjects in rural communities (10.4% for men and 4.1% for women) was half that recorded in more built-up areas and the Paris region (18.2% for men and 10.4% for women). The rate of acquisition of new sexual partners is an important factor in the spread of AIDS. Among multipartner heterosexual people, 18.1% of men and 9.5% of women had three or more new partners during this time period, these proportions being respectively 32.2 and 17.7% for homosexual/bisexual men and women.

The proportion of subjects reporting at least one occurrence of intercourse with a person of the same sex during their lifetime was 4.1% for men and 2.6% for women. These proportions were respectively 1.4 and 0.4% for the past 5 years and 1.1 and 0.3% over the past year. These results concur with those found in similar studies[4,17], but the estimates made are strongly affected by the wording of the questions. These practices are reported to be more frequent in large urban areas (4.7 times higher in Paris than in rural communities). Of the people who had had homosexual intercourse at least once, most had had intercourse with people of both sexes (82% of men and 78% of women).

Intercourse with prostitutes was found only in significant proportions among men: 3.3% over the past 5 years. This practice was much more frequent in Paris (6.2%) than in rural communities (1.7%) ($P<0.001$). The proportion of men having their first intercourse with a prostitute has dramatically decreased during the past 20 years: 10% for the men now aged 45–69, compared to only 2% for those aged 20–24.

In this type of survey it is very difficult to estimate the proportion of people taking drugs by intravenous injection, this practice being very uncommon. However, one might expect that people who regularly use drugs are the most difficult to contact, and/or most often refuse to participate in any kind of survey or to acknowledge an illegal practice. Even though 14.5% of men and 9.4% of women stated that they had taken a soft or hard drug at least once, the proportion saying they had taken an intravenous drug was very low (0.5% of men and 0.2% of women).

The proportion of subjects saying they had used a condom during intercourse at least once in their lifetime was 56.5% for men and 43.7% for women ($P<0.001$). These prevalences decreased very sharply with age, indicating the most widespread use among the youngest subjects ($P<0.001$). The usage rates over the past year were very high for the youngest subjects (aged 18–19 years): 79.8 and 48.0% for sexually active men and women, respectively. These proportions then fall gradually, indicating highly contrasting behaviour between different age groups and generations. The proportion of new condom users over the past year was significantly higher in women than in men, especially for the younger age groups (4.8% in men and 9.8% in women for 18–24 years, $P<0.05$). The condom was used for contraceptive purposes by 23.6% of men and 8.8% of women in this age group. The data also indicate that condom use is more prevalent in subjects most exposed to the risk of contamination by sexually transmitted diseases (STDs) and AIDS, that is, those

TABLE 2 Distribution of the number of sexual partners during the previous year

	0	1	2–3	4–5	6–14	15 +	Unknown	Total
Age (years)								
Men								
18–19	32.8	42.8	17.2	4.7	2.2	0.3	—	100
20–24	17.9	56.9	17.8	4.3	2.1	0.7	0.3	100
25–29	8.0	74.7	13.5	2.2	1.3	0.3	0.0	100
30–34	5.5	84.1	8.2	1.2	0.7	0.2	0.1	100
35–39	5.0	84.1	9.0	1.1	0.5	0.1	0.2	100
40–44	5.5	85.1	7.4	1.3	0.5	0.2	—	100
45–49	6.5	85.4	6.8	0.8	0.4	0.1	—	100
50–54	6.6	87.8	4.3	1.0	0.3	—	—	100
55–59	8.4	88.2	2.6	0.6	0.2	—	—	100
60–64	17.6	78.3	3.3	0.4	0.3	0.1	—	100
65–69	24.2	73.8	1.8	0.1	0.1	—	—	100
Total	11.1	77.5	8.7	1.6	0.8	0.2	0.1	100
Women								
18–19	41.2	49.6	7.7	1.1	0.4	—	—	100
20–24	20.3	69.7	8.7	1.0	0.3	—	—	100
25–29	9.0	83.6	6.3	0.5	0.5	0.1	—	100
30–34	4.9	89.5	5.0	0.4	0.2	—	—	100
35–39	6.0	89.0	4.2	0.4	0.4	—	—	100
40–44	8.2	88.2	3.4	0.2	—	—	—	100
45–49	9.9	85.8	3.6	0.5	—	—	0.2	100
50–54	15.3	82.1	2.6	—	—	—	—	100
55–59	24.9	73.9	1.2	—	—	—	—	100
60–64	32.5	66.8	0.6	—	—	—	0.1	100
65–69	44.3	55.6	—	—	—	—	0.1	100
Total	17.3	78.0	4.1	0.4	0.2	—	0.0	100

All values in per cent.

with several homosexual or heterosexual partners. Around 75% of multipartner homosexual men aged 18–44 years, and more than 65% of men and 50% of women with multiple heterosexual partners, were found to have used a condom at least once over the past year, but this still leaves a high proportion who do not: one-third of the heterosexual men, half of the heterosexual women and one-quarter of the homosexual/bisexual men did not use condoms at any time during this period.

One feature of our study that is different from any other is the collection of detailed data on condom use during the last occasion of intercourse. Overall, 18% of men and almost 11% of women used a condom during last intercourse. A multivariate analysis indicates that the most important determinants of condom use among those with multipartners are age of the subject (with a much higher use among young people), type of partner (2.5 times higher usage rate with casual partners than with regular ones), length of the relationship (for women, 3.1 times higher condom use for the first intercourse with a new partner) and perceiving that one was personally at risk of contracting AIDS. However, the situation regarding this last variable is complex because among men the reverse is true: condom use is higher in those who are not concerned about getting AIDS. This observation supports the hypothesis that people at high risk have changed their behaviour and are using condoms more often, and therefore no longer feel themselves to be personally at risk[5,18].

Subsequent analyses

The results presented here provide a vast amount of information on the sexual behaviour of the population in France. The most striking finding is the relatively high level of condom use in young people and in those most exposed to the risk of contamination by STDs or by AIDS, although a considerable proportion of these people still do not use condoms, especially with new partners. A more specific survey is now being conducted among young people aged 15–18 years.

The available data should reveal which people protect themselves against the risk of contamination and which do not, taking account of the different types of sexual partners and the practices performed with each of them. Moreover, our study was set up not only to describe the characteristics of these people, but also to try to understand what psychological and sociological mechanisms are at play. These aspects were explored in depth in our questionnaire[11] and we are now analysing them. They include, for example, social norms relating to sexual behaviour; influence of the family environment; strategies for finding sexual partners; verbal communication with a new sexual partner; attitudes about body fluids and so on.

Apart from a few exceptions, sexual behaviour until very recently remained one of the few areas in which there was no systematic information available from surveys conducted on representative samples of the general population in developed countries[12]. The knowledge provided by cross-sectional surveys similar to the one we report here is essential both to optimize strategies for preventing STDs and AIDS and to predict the spread of the AIDS epidemic more accurately.

The rate of transmission of STDs and AIDS in a given population depends on a large number of behavioural variables: the most important of these are the rate of acquisition of new partners, the type of sexual practice and patterns of sexual 'mixing' such as age-dependent partner choice and mixing between populations with differential disease prevalence[19]. The level of protection currently used is insufficient to halt the spread of the HIV epidemic. Small-scale studies, repeated at two-yearly intervals, are planned to monitor future changes in sexual behaviour. Quantitative surveys of this kind have their limitations, however. Qualitative studies are also called for to reach marginal groups of the population and to explore certain kinds of behaviour in depth. Only by combining the two approaches will it be possible to obtain sufficient information to define effective strategies to prevent the spread of STDs and AIDS.

Since submission of this manuscript, the survey which follows on page 410 has been completed[20]. □

1. *AIDS* **6**, 243–247 (1992).
2. Frank, O. *Recherches sur les comportements sexuels et la santé génésique: perspectives actuelles et priorités futures* (Human Reproduction Program, World Health Organization, 1990).
3. Sundet, J. M., Magnus, P., Kvalem, I. L., Gronnesby, J. K. & Bakketeig, L. S. *Health Policy* **13**, 160–167 (1989).
4. Melbye, M. & Biggar, R. J. *Am. J. Epidem.* **135**, 593–602 (1992).
5. McQueen, D. V. & Uitenbroek, D. G. *Health Educ. Res.* **7**, 47–53 (1992).
6. Aldhous, P. *Science* **257**, 25 (1992).
7. Wellings, K. *et al. Nature* **348**, 276–278 (1990).
8. Simon, P., Gondonneau, J., Mironer, L. & Dourlen-Rollier, A. M. *Rapport sur le comportement sexuel des français* (Julliard et Charron, Paris, 1972).
9. ACSF Investigators *Bull. Méthod. Sociol.* **35**, 46–54 (1992).
10. ACSF Investigators *AIDS* **6**, 315–323 (1992).
11. Bajos, N., Spira, A. & ACSF *Sciences Sociales et Santé* **IX**, 58–68 (1991).
12. Catania, J., Gibson, D., Chitwood, D. & Coates, T. *Psychol. Bull.* **108**, 339–362 (1990).
13. Upchurch, D. M. *et al. Am. J. Epidem.* **134**, 1159–1166 (1991).
14. Czaja, R. *Int. Quart. Commun. Health Educ.* **8**, 23–32 (1988).
15. Leridon, H. & Toulemon, L. *Population* **2**, 777–812 (1991).
16. Lagrange, H. *Population* **2**, 249–278 (1991).
17. Rogers, S. M. & Turner, C. F. *J. Sex Res.* **28**, 491–519 (1991).
18. Pollak, M. in *Assessing AIDS Prevention* (eds Paccaud, F., Vader, J. & Gutzwiller, F.) 305 (IUMSP, Lausanne, 1992).
19. Anderson, R. M. & May, R. M. *Nature* **333**, 514–519 (1988).
20. Johnson, A. M. *et al. Nature* **360**, 410–412 (1992).

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2009;38:1118–1127
doi:10.1093/ije/dyp210

Improving epidemiological surveys of sexual behaviour conducted by telephone

Charles F Turner,[1,2]* Alia Al-Tayyib,[2,3] Susan M Rogers,[2] Elizabeth Eggleston,[2] Maria A Villarroel,[2,4] Anthony M Roman,[5] James R Chromy[6] and Phillip C Cooley[7]

Accepted 9 April 2009

Background This study assesses the impact of Telephone Audio Computer-Assisted Self-Interviewing (T-ACASI) on the reporting of sensitive (mainly heterosexual) behaviours.

Methods A randomized experiment was embedded in a telephone survey that drew probability samples of the populations of the USA ($N = 1543$) and Baltimore city ($N = 744$). Respondents were randomly assigned to have questions asked either by a T-ACASI computer or by a human telephone interviewer.

Results Compared with interviewer-administered telephone surveys, T-ACASI obtained more frequent reporting of a range of mainly heterosexual behaviours that were presumed to be sensitive, including recency of anal sex [adjusted odds ratio (A-OR) = 2.00, $P < 0.001$), sex during menstrual period (A-OR = 1.49, $P < 0.001$), giving oral sex (A-OR = 1.40, $P = 0.001$) and receiving oral sex (A-OR = 1.36, $P = 0.002$), and sexual difficulties for the respondent (A-OR = 1.45, $P = 0.034$) and their main sex partner (A-OR = 1.48, $P = 0.0$). T-ACASI also obtained less frequent reporting that respondent had a 'main sex partner' (A-OR = 0.56, $P = 0.011$) and discussed contraception prior to first sex with that sex partner (A-OR = 0.82, $P = 0.094$). For both males and females, T-ACASI obtained more frequent reports of first vaginal sex occurring at early ages (before ages 12 through 15). 'For males only', T-ACASI also elicited more frequent reports that first vaginal sex had 'not' occurred at later ages (i.e. by ages 20 through 24).

Conclusion T-ACASI increases the likelihood that survey respondents will report sensitive heterosexual behaviours.

Keywords Population surveys, methodology, sexual behaviours, STD risk behaviours, T-ACASI, computerized surveys, sexually transmitted infections

[1] Queens College and the Graduate Center, City University of New York, Flushing, NY, USA.
[2] Program in Health and Behavior Measurement, Research Triangle Institute, Washington, DC, USA.
[3] Denver Public Health, Denver, CO, USA.
[4] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.
[5] Center for Survey Research, University of Massachusetts at Boston, Boston, MA, USA.
[6] Statistics and Epidemiology Division, Research Triangle Institute, Research Triangle Park, NC, USA.
[7] Research Computing Division, Research Triangle Institute, Research Triangle Park, NC, USA.
* Corresponding author. Queens College, City University of New York, Flushing, NY 11367, USA.
E-mail: cfturner@popepi.org

Introduction

Because of the substantial costs associated with sending field interviewers to tens of thousands of households, the earliest comprehensive AIDS surveys in developed nations used 'telephone' survey techniques.[1–3] A growing body of evidence suggests, however, that telephone surveys are subject to non-trivial reporting biases because they require respondents to disclose sensitive, stigmatized or illicit behaviours to human interviewers.[4–6]

In 1996, we reported the first test of Telephone Audio Computer-Assisted Self-Interviewing (T-ACASI).[7] This technology asks questions by playing digitized voice files, and respondents provide answers using the keypad of their touchtone telephones. We found that T-ACASI interviews were feasible, well tolerated by respondents and they could reduce underreporting of sensitive or stigmatized sexual behaviours. This technology was subsequently tested in a randomized experiment embedded in the 1996–98 Urban Men's Health Survey (UMHS); it found that representative samples of men who have sex with men were more likely to report use of a range of illegal drugs, concern about their current drug use and exchange of money or drugs for sex when interviewed by T-ACASI rather than human telephone interviewers.[8] Subsequent work by our group and collaborators has found that T-ACASI increased reporting of sensitive and stigmatized behaviours including illicit (but not licit) drug use, same-gender sexual behaviours, STD histories and 'unpopular attitudes' in probability samples of the US population and teen smoking in a regional sample.[9–13]

Work by other researchers has yielded generally supportive evidence.[14–15] Of particular note is the independent development of T-ACASI technology by Lau and collaborators,[16] who conducted a randomized experiment in 2001 comparing T-ACASI with traditional interviewer-administered telephone interviewing in a survey of Hong Kong adults. Lau *et al*. reported that for the 13 questions asked of male respondents, T-ACASI elicited 'more frequent' reporting of sex in the last 6 months with female sex workers and non-regular female sex partners, lifetime experience of male–male sex, and 'less frequent' reporting of having a regular sex partner. Female respondents were asked 11 questions, and those interviewed in the T-ACASI mode were 'more likely' to report some risk or a high risk of contracting HIV from their husband and having a one-night stand or HIV test in the past 6 months; they were 'less likely' to report having a husband, having intercourse with their husband in the past 6 months and being able to insist that their husband use condoms if they suspected that he had sex with female sex workers.

The U.S.A. National STD and Behavior Measurement Experiment (NSBME) was designed to assess the impact of T-ACASI on reporting of a wide range of sensitive characteristics and behaviours in a probability sample of US adults aged 18–45. This article reports results for the 29 NSBME questions that asked respondents about their (mainly) heterosexual experiences, practices and problems.

Methods

The protocol for this research was approved and supervised by Institutional Review Boards for the Protection of Human Subjects at the Research Triangle Institute (RTI) and the University of Massachusetts at Boston.

Sample design and execution

The NSBME was embedded in a telephone survey of a probability sample of women and men aged 18–45 years residing in US households with working landline telephones. The survey was conducted between September 1999 and April 2000. Two sample strata were recruited for this survey measurement experiment: (i) a sample of the telephone-accessible US household population aged 18–45 (national stratum) and (ii) a parallel sample of the telephone-accessible population of the city of Baltimore, MD (Baltimore stratum). [In 2000, telephone service was available in 96.7% of US household and 93.0% of Baltimore households with a householder aged 18–45 (Tabulated from: 2000 Census Summary File 3, available at http://factfinder.census.gov).]

For the national stratum, 14 250 telephone numbers were generated, and 12 322 telephone numbers (86.5%) were successfully screened for eligibility. Of these screened telephone numbers, 2183 were found to be residential numbers with one or more eligible English-speaking respondents aged 18–45. One eligible household member of these households was randomly selected for participation in the survey (without substitution). Of the 2183 target respondents, 1452 in the national strata completed interviews (66.5%), and 91 respondents (4.2%) completed partial interviews that included at least one substantive questionnaire section. (Interviews were considered 'complete' if the respondent completed the 101st of 123 questions in the male version of the questionnaire and the 103rd of 125 questions in the female version, excluding closeout questions on reactions to the survey.) A maximum of 91 calls per household were made to screen households and complete an interview in the national stratum.

For the Baltimore stratum, 7498 telephone numbers were generated and 6326 (84.4%) were successfully screened for eligibility. Screening identified 1072 households with an eligible respondent, and 697 of these eligible respondents completed interviews (65.0%). An additional 47 respondents (4.4%) completed partial interviews. A maximum of 82 calls per household were made to screen households and complete an interview in the Baltimore stratum.

Using a professionally endorsed methodology,[17] we calculated the survey response rates for the NSBME to be 62% for the telephone interviewer-administered questioning (T-IAQ) condition and 53% for the T-ACASI condition in the national stratum. In the Baltimore stratum, these response rates were 56% for the T-IAQ condition and 50% for the T-ACASI condition (see ref.[18], pp. 23–29). These calculations take account of the joint effects of failures to: (i) reach households on the telephone, (ii) screen the households to identify all eligible adults and (iii) interview the randomly selected eligible adult. Additional details of NSBME survey sample design and execution have been published elsewhere.[10,18,19]

Interview modes

Telephone numbers were randomly assigned to the T-IAQ or T-ACASI conditions prior to their release to the telephone survey unit. Following screening and recruitment into the study, telephone interviewers at the Center for Survey Research (University of Massachusetts, Boston) conducted the survey either by asking the respondent questions and recording their answers (T-IAQ condition) or by transferring the respondent to a T-ACASI system developed by Cooley *et al*.[20,21]

Survey measurements

The NSBME included 125 questions of which 29 focused on sexual experience, sexual practices, condom use and problems with sexual satisfaction and sexual arousal. Most NSBME questions were adapted from past large-scale surveys of sexual behaviour in the United States and the UK. This was done to ensure the generalizability of NSBME results to contemporary research efforts. The complete wordings and sources of the survey questions used in this article can be found Appendix A1 (available as Supplementary Data at *IJE* online).

Heterosexual focus

To the extent possible given the survey questions, this article focuses on heterosexual behaviour. (We have previously reported NSBME results on same-gender sexual behaviors.[10]) Many questions reported in this article refer 'explicitly' to heterosexual experiences. Other questions—such as number of sex partners—do not explicitly restrict reporting to heterosexual experiences. To (imperfectly) focus our analyses, we exclude respondents who reported never having heterosexual vaginal sex and who also report that they were mostly or exclusively attracted to same-gender partners.

Statistical analysis

Our analyses of the NSBME are intended to determine whether T-ACASI increases the willingness of respondents to report sensitive, stigmatized or embarrassing behaviours. To address this research question, we combine the national and Baltimore sample strata. The combined sample strata are treated as a population that has been randomly allocated to one of two experimental conditions: T-ACASI or T-IAQ interview mode. Data in these analyses are unweighted, and our statistical analyses assess the likelihood that observed fluctuations in survey responses across the two interview modes arose by chance from the random allocation of respondents to one of the two experimental groups. (All estimates reported in this article are sample not population prevalences.)

The survey measurements reported in this article are either binary, ordinal or metric. Tests of the equivalence of the response distributions obtained in the T-ACASI and T-IAQ conditions were performed by fitting logit models to predict the binary measurements, ordered logit models to predict the ordinal measurements and multiple linear regression to predict the metric measurements. Our analyses present both raw coefficients representing the impact of T-ACASI (i.e. the experimental manipulation: T-ACASI vs T-IAQ) and adjusted coefficients that represent the impact of T-ACASI after incorporating statistical controls for a wide range of sociodemographic variables. Our final analyses tested for variation in the impact of T-ACASI on reporting by male and female respondents. For each measurement, we estimated (i) the impact of T-ACASI and female gender on response distributions and (ii) the impact of T-ACASI, female gender and the interaction of female-by-T-ACASI on response distributions. We report results for these analyses when the *P*-value of the coefficient for the interaction term was $\leqslant 0.10$.

All statistical analyses were carried out using Stata SE, versions 8 and 10.[22]

Results

Sample equivalence

Previously published analyses of the NSBME tested the equivalence of T-IAQ and T-ACASI samples by gender, age, marital status, education, race/ethnicity, region, urbanization and sample strata. No comparison produced evidence of non-equivalence with a *P*-value $\leqslant 0.286$.[19]

Reports of sexual experience

For the nine binary measurements shown in the top panel of Table 1, the odds ratios (ORs) indicate that T-ACASI obtained increased reporting of the presumed 'sensitive' answer, i.e. never having sex of any type or heterosexual vaginal sex, not having a main sex partner, and having one-night stands, coerced sex, paid sex, another sex partner while married or in a 'committed relationship' (extra-relationship sex). These results are statistically reliable for reports of never having sex of

Table 1 Sexual experience reported by subjects interviewed by T-ACASI or by T-IAQ

Measurement	T-ACASI (%)	T-IAQ (%)	ORs			
			Crude	*P*	Adjusted[a]	*P*
Binary variables						
Never 'had sex with anyone'	5.8	2.9	2.05	0.001	1.76	0.048
Never had heterosexual vaginal sex	3.4	2.9	1.17	>0.500	1.38	0.416
Currently have a main sex partner	89.0	93.8	0.53	<0.001	0.56	0.011
Ever had a one-night stand	45.4	37.3	1.40	<0.001	1.35	0.011
Had extramarital (or extra-relationship) sex partners during last marriage or committed relationship[b]	13.4	8.8	1.61	0.001	1.61	0.012
Ever forced to have sex	21.9	17.1	1.36	0.006	1.54	0.005
Ever forced someone to have sex	1.7	1.3	1.25	>0.500	1.83	0.253
Ever had sex with a prostitute[c]	7.2	5.9	1.23	0.251	1.24	0.412
Did someone ever pay you for sex	4.2	3.7	1.14	>0.500	1.34[d]	0.412
(Base *N*)	854–954	1036–1178				
Measurement	**T-ACASI (%)**	**T-IAQ (%)**	**Regression coefficient**			
			Crude	*P*	Adjusted[a]	*P*
Metric variables						
Number of lifetime sex partners (if not zero)[e]						
0	na	na	0.28	>0.50	1.06	0.27
1	14.1	14.4				
2–3	23.1	24.8				
4–5	20.1	19.2				
6–10	19.1	18.9				
11–20	12.4	11.4				
21–50	7.8	7.8				
51+	3.6	3.6				
(Base *N*)	897	1121				
Number of sex partners in the past year[e]						
0	9.8	8.1	−0.01	>0.50	0.10	0.244
1	68.4	76.5				
2	11.7	8.6				
3–4	7.1	3.6				
5+	3.0	3.2				
(Base *N*)	898	1136				
Number of sex partners in the past month[e,f]						
0	16.2	15.5	−0.01	>0.50	−0.01	>0.500
1	77.9	80.9				
2	4.8	2.3				
3+	1.1	1.3				
(Base *N*)	810	1041				
Number of new sex partners in the past year[e]						
0	59.9	76.7	0.34	<0.001	0.43	<0.001
1	24.2	14.2				
2	8.9	4.6				

(Continued)

Table 1 Continued

Measurement	T-ACASI (%)	T-IAQ (%)	Regression coefficient			
			Crude	*P*	Adjusted[a]	*P*
3–4	2.9	2.6				
5+	4.1	1.9				
(Base *N*)	805	1029				
Number of new sex partners in the past 5 years (if zero in the past year)[e]						
0	60.2	61.3	0.21	0.04	0.31	0.011
1	15.4	16.9				
2	9.1	11.1				
3–4	8.6	6.4				
5+	6.8	4.3				
(Base *N*)	573	876				
Frequency of heterosexual vaginal sex in the past 7 days[e]						
0	48.7	45.2	−0.02	>0.500	0.10	0.464
1	18.5	21.9				
2	14.9	13.1				
3–5	13.9	16.4				
6+	4.0	3.4				
(Base *N*)	827	1051				

Unweighted data from 2000 NSBME: national and Baltimore sample strata combined. Sample excludes respondents who report never having had heterosexual vaginal sex and who also report sexual attraction only or mostly to persons of the same gender.

[a]Adjusted odds ratios and regression coefficients were calculated by adding independent variables to control for gender, race (Black vs non-Black), Hispanic ethnicity, age in years, education in years, marital status (married or cohabiting vs not), region of the country (six regions), urbanicity (four categories: 21 largest metropolitan areas; jurisdictions with 85 000 or more households; 20 000–84 999 households, and less than 20 000 households), and sample strata (National vs Baltimore). The adjustment for sample strata was dropped during model estimation due to multicollinearity with the other adjustment variables.

[b]Excludes persons who were never married or involved in committed relationship.

[c]Men were much more likely than women to report paying for sex (15.4 vs 0.3%), but the T-ACASI effect for men was not statistically reliable (16.8% in T-ACASI vs 14.3% in T-IAQ; $P = 0.31$).

[d]Adjusted odds ratios and regression coefficients were estimated for pooled sample of male and female respondents. This estimated effect evidenced a statistically borderline interaction ($P = 0.06$) with gender, i.e. estimates for males and females were not equivalent. See the text for discussion.

[e]To conserve space, this table collapses categories for reporting of many partners (e.g. 50–99 and 100+ lifetime partners) since they were infrequently reported. Regression analysis used all response categories. For variables that were coded as ranges, e.g. '6–10' partners, respondents were assigned the midpoint of the interval. For final categories such as 100+ partners, respondents were assigned value of 120% of the highest value of base category, e.g. 100+ was coded to 120 partners for regression analysis.

[f]Subjects who reported no partners in the past year were not asked this question.

any type [adjusted OR (A-OR) = 1.76, $P = 0.048$], *not* having a 'main sex partner' (A-OR = 0.56, $P = 0.011$), having one-night stands (A-OR = 1.35, $P = 0.011$), extra-relationship sex (A-OR = 1.61, $P = 0.012$) and being forced to have sex (A-OR = 1.54, $P = 0.005$). For the six metric variables shown in the bottom panel of Table 1, T-ACASI had a statistically noteworthy impact only on reporting of 'new' sex partners in the past month (adjusted coefficient = 0.43 partners, $P < 0.001$) and past 5 years (adjusted coefficient = 0.31 partners, $P = 0.011$). Tests for variation in the T-ACASI mode effect between men and women yielded only one statistically borderline interaction from 15 tests, which is roughly consistent with expectations for the results of 15 independent tests of the null hypothesis (with $\alpha = 0.05$) when no true effects exist.

Age at sexual debut

Figure 1 plots the odds ratio (T-ACASI vs T-IAQ) that male and female respondents would report heterosexual intercourse before specific ages between 12 and 24 years, and it displays one of our major findings. For both men and women, T-ACASI respondents were significantly more likely to report that their heterosexual debut occurred before the ages of 12, 13, 14 and 15 years. For women, the odds ratios range from 2.74 ($P = 0.014$) for reporting sexual debut

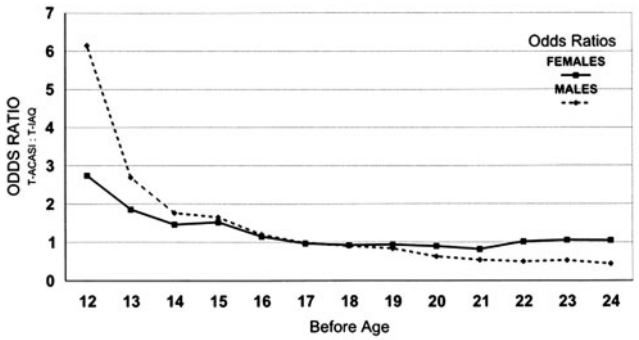

Figure 1 Odds ratio (T-ACASI: T-IAQ) for reporting of first heterosexual vaginal sex before specific ages by gender

before age 12 to 1.52 ($P = 0.013$) for reporting sexual debut before age 15. For men these odds ratios range from 6.15 ($P < 0.001$) for debut before age 12 and 1.65 ($P = 0.004$) for sexual debut before age 15. For women, the range of odds ratios for reporting sexual debut before ages between 16 and 24 are statistically indistinguishable from 1.0 ($P = 0.24$–0.91). For men, however, T-ACASI also elicits significantly increased reporting of 'not' having had heterosexual vaginal sex by ages 20 through 24 (ORs = 0.44–0.63; $P = 0.016$–0.054). (To conserve journal space, additional details of these results are presented in Appendix A2, available as Supplementary Data at *IJE* online.)

Sexual practices

Table 2 presents results for 10 questions asking about respondents' sexual practices, condom use and sexual communication. T-ACASI elicited more frequent reporting of recent active and passive heterosexual oral sex, heterosexual anal sex and heterosexual vaginal sex during a woman's menstrual period (A-ORs = 1.36–2.00; $P = 0.002$ to <0.001). T-ACASI also decreased reporting of consistent condom use in the past month (OR = 0.80, $P = 0.012$), but the effect did not persist when the ordered logit model was expanded to include our sociodemographic controls (A-OR = 0.87, $P = 0.259$). We note, however, that respondents were more likely to tell a human interviewer that they used condoms 'every time' they had sex in the past month (8.4% in T-ACASI condition vs 16.4% in T-IAQ, A-OR = 0.44, $P < 0.001$). T-ACASI also elicited fewer reports that respondents had talked with their most recent sex partner about contraception and more reports that they had used withdrawal as a contraceptive method, although these results were of borderline statistical significance (A-ORs = 0.82 and 1.23, $P = 0.094$ and 0.076, respectively). Finally, we note that T-ACASI elicited reports of more frequent sharing of information about numbers of past sex partners with new sex partners acquired during the preceding year (A-OR = 1.82 for asking for information from new partner, $P = 0.007$; and A-OR = 1.57 for providing information to a new partner, $P = 0.031$).

There were also noteworthy variations in the impact of T-ACASI on male and female reporting of recency of giving oral sex to heterosexual partners and exchanging information on sexual partners (data not shown). T-ACASI elicited more reports from women of recently giving oral sex to a male partner (A-OR = 1.55, $P < 0.001$), while the parallel T-ACASI effect for males is smaller and statistically unreliable (A-OR = 1.16, $P = 0.24$). This result is seen clearly in the percentage of men and women reporting that they had 'never' given oral sex to a heterosexual partner. For women, T-ACASI decreased the reporting of 'never' giving oral sex from 27.2 to 14.9%, whereas for men the reduction was more modest, from 16.8 to 12.0%. T-ACASI also did not have a noteworthy impact on women's reports of sharing information on numbers of past partners with a new sex partner (OR = 0.89, $P > 0.50$ for asking; OR = 1.02, $P > 0.50$ for telling). T-ACASI, however, increased the odds that males would report more frequent sharing of this information with new sex partners (OR = 1.98, $P = 0.002$ for asking; OR = 1.76, $P = 0.008$ for telling). T-ACASI increased the percentage of men reporting 'always asking' new partners for this information from 14.1 to 30.0% and the percentage reporting 'always providing' this information from 17.2 to 28.3%.

Sexual problems

Table 3 describes the impact of T-ACASI on reporting of respondent's and partner's sexual problems and the respondent's ease or difficulty becoming sexually aroused when 'hav[ing] any kind of sex'. T-ACASI significantly increased the odds that respondents would report sexual problems for their partners (A-OR = 1.48, $P = 0.047$) and for themselves (A-OR = 1.45, $P = 0.034$). T-ACASI also shifted responses on ease of sexual arousal towards 'very easy' and away from 'very difficult' (A-OR = 0.59, $P < 0.001$). Thus, the percentage of respondents reporting that it is very easy for them to become sexually aroused increases from 39.6% when questioning is done by a human interviewer to 54.3% when questioning is done by a T-ACASI computer. Tests for gender-by-T-ACASI interaction indicate, however, that the impact of T-ACASI is stronger for males (A-OR = 0.38, $P < 0.001$) than for females (A-OR = 0.79, $P = 0.046$; with $P < 0.001$ for test of the interaction). This is seen most clearly in reporting that sexual arousal is 'very easy'. The percentage of males giving this answer increases from 53.6% in the T-IAQ condition to 76.2% in T-ACASI, whereas for females the parallel increase is from 29.8 to 38.5%.

Discussion

A growing number of studies report that computer-assisted self-interviewing (CASI) technologies obtain

Table 2 Heterosexual practices reported by subjects interviewed by T-ACASI and the T-IAQ

Measurement	T-ACASI (%)	T-IAQ (%)	ORs			
			Crude	*P*	Adjusted[a]	*P*
Recency: gave oral sex						
Never	13.7	22.9	1.37	<0.001	1.40[b]	0.001
>5 years	5.1	5.4				
1–5 years	11.9	9.3				
6–12 months	8.4	7.0				
1–6 months	17.6	18.3				
8–30 days	18.8	16.9				
7 days	24.5	20.2				
(Base *N*)	(898)	(1092)				
Recency: received oral sex						
Never	8.9	15.2	1.31	0.001	1.36	0.002
>5 years	5.1	5.9				
1–5 years	11.0	11.5				
6–12 months	8.3	7.6				
1–6 months	21.1	19.2				
8–30 days	20.6	18.3				
7 days	25.0	22.3				
(Base *N*)	(900)	(1092)				
Recency: anal sex						
Never	63.3	74.6	1.7	<0.001	2.00	<0.001
>5 years	10.6	7.9				
1–5 years	9.3	7.6				
6–12 months	5.4	3.2				
1–6 months	5.3	3.2				
8–30 days	3.6	2.3				
7 days	2.4	1.3				
(Base *N*)	(900)	(1120)				
Recency: sex during menstrual period						
Past 30 days	38.9	47.1	1.43	<0.001	1.49	<0.001
1–3 months ago	23.7	23.7				
4–6 months ago	7.1	4.8				
7–12 months ago	6.5	5.7				
>12 months ago	10.1	10.4				
Never	13.8	8.3				
(Base *N*)	(893)	(1106)				
Condom use in the past month						
Never	58.0	56.1	0.80	0.012	0.87	0.259
Rarely	8.9	6.3				
Some of the times	9.5	6.0				
Half of the time	2.1	3.7				
Most of the times	8.8	5.1				
Almost every time	4.3	6.5				
Every time	8.4	16.4				
(Base *N*)	(810)	(1042)				

(continued)

Table 2 Continued

Measurement	T-ACASI (%)	T-IAQ (%)	ORs			
			Crude	*P*	Adjusted[a]	*P*
Ever a time you thought you should use a condom but did not, % Yes	35.2	38.7	0.86	0.100	0.88	0.273
(Base *N*)	(896)	(1134)				
Before you had intercourse with your current partner did you and partner talk about contraception? % Yes	56.1	60.0	0.85	0.073	0.82	0.094
(Base *N*)	(897)	(1126)				
Ever practice withdrawal as method of contraception? % Yes	59.9	58.9	1.04	>0.500	1.23	0.076
(Base *N*)	(890)	(1111)				
How often in the past year have you asked a new sex partner about the number of past sex partners he or she has had?						
Never	32.5	34.7	1.37	0.041	1.82[b]	0.007
Sometimes	16.4	28.9				
Almost always	15.5	8.4				
Always	35.6	28.0				
(Base *N*)	(323)	(239)				
How often in the past year have you told a new sex partner about the number of past sex partners you have had?						
Never	34.3	36.8	1.36	0.042	1.57[b]	0.031
Sometimes	16.9	27.2				
Almost always	13.9	9.6				
Always	34.9	26.4				
(Base *N*)	(367)	(239)				

Unweighted data from 2000 NSBME: National and Baltimore strata combined. Sample excludes respondents who report never having had heterosexual vaginal sex and who also report sexual attraction only or mostly to persons of the same gender.

[a]Adjusted odds ratios were calculated by adding independent variables to control for gender, race (Black vs non-Black), Hispanic ethnicity, age in years, education in years, marital status (married or cohabiting vs not), region of the country (six regions), urbanicity (four categories: 21 largest metropolitan areas; jurisdictions with 85 000 or more households; 20 000–84 999 households, and less than 20 000 households), and sample strata (National vs Baltimore). The adjustment for sample strata was dropped during model estimation due to multicollinearity with the other adjustment variables.

[b]Adjusted odds ratios were estimated for pooled sample of male and female respondents. These estimated effects evidenced a statistically significant or borderline interaction with gender, i.e. estimates for males and females were not equivalent. See text for discussion.

increased reporting of sensitive and stigmatized behaviours and, more recently, attitudes and opinions. The present study provides a substantial demonstration of the superiority of T-ACASI to traditional telephone interviewing in obtaining reports of sensitive behaviours. For 18 of 29 measurements of (mainly) heterosexual behaviours, T-ACASI obtained statistically reliable or borderline differences in the response distributions from the measurements made by human telephone interviewers. In the vast majority of cases, T-ACASI elicited more frequent reporting of the behaviours presumed to be more sensitive. So, for example, the percentage of respondents reporting ever having heterosexual anal sex increased from 25.5% when questioning was done by human telephone interviewers to 36.7% when questioning was done by a T-ACASI computer (OR = 1.70, $P < 0.001$). T-ACASI also elicited: (i) more frequent reporting of recent active and passive heterosexual oral sex, and of heterosexual vaginal sex during a woman's menstrual period; (ii) reporting of a larger number of 'new' sexual partners in the past month and past year; (iii) reporting an earlier age of sexual debut; (iv) more frequent reporting of one-night stands, forced sex, sexual problems of respondents and their partners, and additional sex partners while married or in a 'committed relationship'; (v) more frequent reporting of a lack of sexual experience of any kind

Table 3 Heterosexual problems reported by subjects interviewed by T-ACASI and the T-IAQ

Measurement	T-ACASI (%)	T-IAQ (%)	ORs			
			Crude	*P*	Adjusted[a]	*P*
Physically or emotionally difficult for partner to have satisfying sex?[b]	10.7	8.5	1.29	0.097	1.48	0.047
(Base *N*)	(869)	(1062)				
Physically or emotionally difficult for you to have satisfying sex?	13.2	9.7	1.42	0.011	1.45	0.034
(Base *N*)	(952)	(1136)				
Easy or difficult for you to get sexually aroused?						
Very easy	54.3	39.6	0.62	<0.001	0.59[c]	<0.001
Somewhat easy	35.8	52.0				
Somewhat difficult	8.3	6.3				
Very difficult	1.6	2.1				
(Base *N*)	(866)	(988)				

Unweighted data from 2000 NSBME: national and Baltimore strata combined.

[a]Adjusted odds ratios were calculated by adding independent variables to control for gender, race (Black vs non-Black), Hispanic ethnicity, age in years, education in years, marital status (married or cohabiting vs not), region of the country (six regions), urbanicity (four categories: 21 largest metropolitan areas; jurisdictions with 85 000 or more households; 20 000–84 999 households, and less than 20 000 households), and sample strata (National vs Baltimore). The adjustment for sample strata was dropped during model estimation due to multicollinearity with the other adjustment variables.

[b]This question refers to either the respondent's current 'main sex partner' or—if there was no main sex partner—'the last partner with whom you had an ongoing sexual relationship that lasted at least a month'.

[c]Adjusted odds ratios were estimated for pooled sample of male and female respondents. This estimated effects evidenced a statistically significant ($P < 0.001$) with gender, i.e. estimates for males and females were not equivalent. See the text for discussion.

by men and women; (vi) inexperience with heterosexual vaginal sex among men aged $\geqslant 20$; and (vii) less frequent reporting that condoms were used 'every time' respondents had sex in the past month.

The foregoing results are generally consistent with the investigators' expectation that the privacy afforded by T-ACASI would increase reporting of potentially sensitive or embarrassing behaviours (e.g. extramarital sex) and decrease reporting of socially approved behaviours (e.g. consistent condom use).

Two results, however, were inconsistent with our initial expectations. With the deluge of advertising for medications to treat erectile dysfunction, we had expected T-ACASI to elicit more frequent reporting of problems with sexual arousal. We found, however, that T-ACASI substantially increased the odds that men would report that it was easy for them to become aroused. (A parallel effect was not found for females.) On reflection, we suspect that our initial expectation caused us to overlook the relative youth of our sample (ages 18–45) and the potential negative aspects of either having or admitting to having a low threshold for sexual arousal (e.g. embarrassing erections at inappropriate times or worries about being perceived as sexually impulsive).

We were also surprised and remain perplexed that T-ACASI increased the odds that men would report sharing their sexual histories with their last new sex partner. Since having a large number of sex partners is a risk factor for STIs, it is often recommended that this information be shared in new sexual partnerships. Our result would suggest that T-ACASI induced respondents to provide a 'more socially desirable' response. This result is, however, consistent with another NSBME finding reported elsewhere that respondents in T-ACASI reported 'more frequent discussions' of their sex life with their main partner.[11] It is possible that our surprise at these results is due to our own misunderstanding of the social dynamics of reporting on the frequency of sexual communications. Reporting such sexual communication—rather than the absence of such communication—may be the more sensitive or embarrassing response for the majority of the population. This is obviously speculation on our part, but the topic should merit further research in the future.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

National Institutes of Health grant (R01-MH56318 and R01-HD31067) (to C.T.).

Acknowledgements

The authors wish to thank Joseph Catania for collaboration in early phases of the design of this study and for access to the public use dataset from his 1996 National Survey of Sexual Health. The authors also wish to thank the many other people who made major contributions to this work during the proposal and design phase, most importantly, James Gribble.

Conflict of interest: None declared.

References

[1] Catania JA, Coates TJ, Stall R *et al*. Prevalence of AIDS-related risk factors and condom use in the United States. *Science* 1992;**258:**1101–6.

[2] ACSF Investigators. AIDS and sexual behaviour in France. *Nature* 1992;**360:**407–9.

[3] Davis PB, Yee RL, Chetwynd J, McMillan N. The New Zealand partner relations survey: methodological results of a national telephone survey. *AIDS* 1993;**7:**1509–16.

[4] Turner CF, Forsyth BH, O'Reilly J *et al*. Automated self-interviewing and the survey measurement of sensitive behaviors. In: Couper M, Baker R, Bethlehem J *et al*. (eds). *Computer-Assisted Survey Information Collection*. New York: Wiley, 1998.

[5] Tourangeau R, Smith TW. Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opin Q* 1996;**60:**275–304.

[6] Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science* 1998;**280:**867–73.

[7] Turner CF, Miller HG, Smith TK, Cooley PC, Rogers SM. Telephone audio computer-assisted self-interviewing (T-ACASI) and survey measurements of sensitive behaviors: preliminary results. In: Banks R, Fairgrieve J, Gerrard L (eds). *Survey and Statistical Computing*. Chesham, Bucks, UK: Association for Survey Computing, 1996.

[8] Gribble JN, Miller HG, Cooley PC, Catania JA, Pollack L, Turner CF. The impact of T-ACASI interviewing on reported drug use among men who have sex with men. *Subst Use Misuse* 2000;**35:**869–90.

[9] Turner CF, Villarroel MA, Rogers SM *et al*. Reducing bias in telephone survey estimates of the prevalence of drug use: a randomized trial of telephone audio-CASI. *Addiction* 2005;**100:**1432–44.

[10] Villarroel MA, Turner CF, Eggleston EE *et al*. Same-gender sex in the USA: impact of T-ACASI on prevalence estimates. *Public Opin Q* 2006;**70:**166–96.

[11] Villarroel MA, Turner CF, Rogers SM *et al*. T-ACASI reduces bias in STD measurements: the National STD and Behavior Measurement Experiment. *Sex Transm Dis* 2008;**35:**499–506.

[12] Harmon T, Turner CF, Rogers SM *et al*. Impact of T-ACASI on survey measurements of subjective phenomena. *Public Opin Q* (in press).

[13] Currivan DB, Nyman AL, Turner CF, Biener L. Does telephone audio computer-assisted self-interviewing improve the accuracy of prevalence estimates of youth smoking? Evidence from the UMass Tobacco Study. *Public Opin Q* 2004;**68:**542–64.

[14] Moskowitz JM. Assessment of cigarette smoking and smoking susceptibility among youth: telephone computer-assisted self-interviews versus computer-assisted telephone interviews. *Public Opin Q* 2004;**68:**565–87.

[15] Corkey R, Parkinson L. A comparison of four computer-based telephone interviewing methods: getting answers to sensitive questions. *Behav Res Methods Instrum Comput* 2002;**34:**354–63.

[16] Lau JT, Tsui HY, Wang QS. Effects of two telephone survey methods on the level of reported risk behaviours. *Sex Transm Inf* 2003;**79:**325–31.

[17] American Association for Public Opinion Research [AAPOR] (2006). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates, 4th edn*. Lenexa, KSs: AAPOR.

[18] Roman A. Survey of AIDS and other social issues: field report, August 2000. *Technical Papers in Health and Behavior Measurement, No. 82*. Washington DC: Program in Health and Behavior Measurement, Research Triangle Institute, 2008. Available at: http://dragon.soc.qc.cuny.edu/Staff/turner/TechPDFs/82_NSBME_FieldReport.pdf (1 May 2009, date last accessed).

[19] Villarroel MA, Turner CF, Eggleston EE *et al*. Sample design and demographic characteristics of respondents in the 1999–2000 National STD and Behavior Measurement Experiment (NSBME). Published as *Technical Papers in Health and Behavior Measurement, No. 69*, Washington DC: Program in Health and Behavior Measurement, Research Triangle Institute, 2006. Available at: http://dragon.soc.qc.cuny.edu/Staff/turner/TechPDFs/69_NSBMEsample.pdf (1 May 2009, date last accessed).

[20] Cooley PC, Miller HG, Gribble JN, Turner CF. Automating telephone surveys: using T-ACASI to obtain data on sensitive topics. *Comput Hum Behav* 2000;**16:**1–11.

[21] Cooley PC, Turner CF. Implementing audio-CASI on Windows platforms. *Comput Hum Behav* 1998;**14:**195–207.

[22] StataCorp. *STATA Statistical Software for Windows, Versions 6.0 & 8.0*. College Station, TX: StataCorp, 2000 & 2003.

Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology

C. F. Turner,* L. Ku, S. M. Rogers, L. D. Lindberg, J. H. Pleck, F. L. Sonenstein

Surveys of risk behaviors have been hobbled by their reliance on respondents to report accurately about engaging in behaviors that are highly sensitive and may be illegal. An audio computer-assisted self-interviewing (audio-CASI) technology for measuring those behaviors was tested with 1690 respondents in the 1995 National Survey of Adolescent Males. The respondents were randomly assigned to answer questions using either audio-CASI or a more traditional self-administered questionnaire. Estimates of the prevalence of male-male sex, injection drug use, and sexual contact with intravenous drug users were higher by factors of 3 or more when audio-CASI was used. Increased reporting was also found for several other risk behaviors.

A small number of national surveys using large, representative samples of the United States and other populations have attempted to assess the sexual and drug-using behaviors that risk transmission of the human immunodeficiency virus (HIV) that causes acquired immunodeficiency syndrome (AIDS) (*1*). Because sexual activity is typically initiated in adolescence or early adulthood and because that period for many young people is characterized by greater amounts of experimentation, partner change, and risk taking than in later years, research programs with a focus on the behaviors of adolescents and young adults are of particular importance (*2*). AIDS is not, however, the only threat faced by young people. Interpersonal violence represents an even greater threat of morbidity and mortality (*3*); consequently, interpersonal violence has also received greater scrutiny in recent surveys of the adolescent and young adult population. Yet concerns have surfaced regarding the quality of survey measurements of such sensitive behaviors (*4, 5*). The concern most often cited is that respondents may be reluctant to report accurately about sensitive or stigmatized behaviors in which they have engaged. (Reservations are particularly strong in situations in which teenage respondents are asked to tell an adult interviewer whether, for example, they have had a certain sexual experience or used a certain drug.) In one study (*5*), it was estimated that survey measurements that relied upon women's responses to questions about their abortion history during a face-to-face interview captured only 35% of the abortions performed in the United States (a conclusion based on aggregate data supplied by abortion providers for the 1984–1987 period).

The level of privacy that an interviewing mode affords a respondent can dramatically affect survey measurements of sensitive behaviors (*5–7*). Traditionally, surveys have attempted to encourage more accurate reporting of sensitive behaviors by allowing respondents to complete a paper-and-pencil self-administered questionnaire (paper SAQ), which they can seal in an envelope and return to the interviewer. Paper SAQs, however, require that respondents be sufficiently literate to complete a written questionnaire. In addition, because a respondent's answers together with their identification number are recorded on a paper form, some respondents may remain suspicious about the privacy of their responses. Another criticism of paper SAQs is that extensive use of contingent questioning (that is, branching or skip patterns) is difficult. Even literate respondents may have trouble following instructions for navigating through a complex self-administered form (*8*).

Audio-CASI technology. A computer-driven technology has been developed that can administer complex survey questionnaires in an audio format and record respondents' answers without the direct participation of a survey interviewer (*9*). This approach, audio computer-assisted self-interviewing (audio-CASI), allows respondents to listen over headphones to spoken questions that have been digitally recorded and stored on a laptop computer. To answer, respondents press numbered keys on the computer keyboard. Questions are also displayed on the computer's screen, and respondents may respond to the visual presentation of the question rather than waiting until the audio reading has been completed. Respondents can thus answer questions in complete privacy, even if their reading ability is limited. Because survey data are stored on the drives of the computer, they are less vulnerable to inadvertent disclosure to interviewers or others. Audio-CASI also provides (*10*) (i) a completely standardized measurement system—every respondent (in a given language) hears the same question asked in exactly the same way; (ii) computer-controlled branching through complex questionnaires and automated consistency and range checking; and (iii) efficient multilingual administration of surveys (*10, 11*).

Here we report the use of this technology in a large-scale national survey, the 1995 National Survey of Adolescent Males (NSAM), which included a randomized experiment to assess the effect of this technology on the measurement of sensitive behaviors. This study was begun in January 1995 and it measured HIV-risk behaviors, drug use, and interpersonal violence among young men in the United States.

Since 1988, NSAM has tracked the sexual, contraceptive, HIV-risk, and other behaviors of a national probability sample of young men who were 15 to 19 years of age in 1988 (*12*). Past data have presented some intriguing methodological puzzles. A case in point is the measurement of male-male sexual contacts. Because of the potential sensitivity of reporting such behavior, the relevant NSAM questions were presented in a paper SAQ instead of being administered by the interviewer. We expected that use of this more private mode of survey administration would encourage more accurate reporting of the behavior. We found, however, that the proportion of males reporting such contacts in 1988 (2.2%) was lower than would be expected on the basis of surveys of adult men, which have recorded retrospective reports of adolescent male-male sexual behaviors of 4 to 9% (*13*). Further evidence of the difficulty of making such measurements was provided by the rescissions that occurred in these NSAM reports between 1988 and 1991. Eleven of the 30 young men who reported having had an oral or anal male-male contact in the

C. F. Turner and S. M. Rogers are with the Program in Health and Behavior Measurement at the Research Triangle Institute, 1635 M Street, NW, Washington, DC 20036, USA. L. Ku, L. D. Lindberg, and F. L. Sonenstein are with the Urban Institute, 2100 M Street, NW, Washington, DC 20037, USA. J. H. Pleck is with the Department of Human Development and Community Studies, University of Illinois, Urbana, IL 61801, USA.

*To whom correspondence should be addressed.

1988 NSAM survey reported that they had never had such contacts when they were interviewed again in the 1991 survey. [Paper SAQs were used in prior rounds of the survey (*14*).] These considerations and our desire to increase the actual and perceived privacy of the 1995 NSAM interview context motivated us to use audio-CASI technology for measurements of the most sensitive behaviors among the new cohort of 15- to 19-year-olds recruited in the 1995 round of the survey.

Survey sample and experiment. We drew a multistage area probability sample from the population of males ages 15 to 19 who were living in households in the continental United States between February and November 1995 (*15*). Black and Hispanic males were oversampled. To obtain the final sample, we screened 54,265 housing units out of the 56,199 that were eligible for screening. The process identified 2240 eligible males; we completed interviews with 1729 of them. After adjusting for the 3.4% of housing units that could not be screened, we had an overall response rate of 75% (*16*). (The major reasons for nonresponse were refusal by the respondent, refusal by the parent of the potential respondent, and unavailability of the potential respondent after repeated visits.) Of the 1729 young men who completed the interview, 1690 (97.7%) also completed a self-administered interview containing questions on topics that were potentially quite sensitive (including same-gender sex, drug use, and violent behaviors).

Sample weights were used to adjust for the different initial probabilities of sample selection and subsequent nonresponse. Poststratification adjustments were applied to align the NSAM sample estimates with 1995 population estimates from the Bureau of the Census (*17*). To represent accurately the effect of this sample weighting and the stratified and clustered design used in drawing the 1995 NSAM sample, we used statistical software (*18*) that adjusted variances for our sample estimates to reflect the complex design. Our statistical inferences are intended to generalize to the universe of (noninstitutionalized) U.S. males ages 15 to 19 (Table 1).

By design, the sample included almost as many Hispanic and black respondents as white ones (Table 1). That pattern reflects our oversampling of those minorities and

Table 1. Selected social and demographic characteristics of the 1995 NSAM respondents who completed self-administered interviews. (Wgt. %, weighted percent.)

Characteristic	*N*	Wgt. %*
Race-ethnicity		
Black	474	14.0
White	605	68.4
Hispanic	539	12.5
Other	54	5.1
Age		
15 years	391	21.2
16 years	371	20.0
17 years	354	20.4
18 years	317	19.1
19 years	239	19.3
Last year of school completed†		
Currently enrolled		
8th grade or less	262	14.1
9th to 11th grade	943	54.0
12th grade	116	7.3
Some college	45	3.9
Not currently enrolled		
8th grade or less	13	0.6
9th to 11th grade	145	9.7
12th grade	122	8.9
Some college	11	1.5
Parents' education‡		
8 years or less	112	2.8
9 to 11 years	128	6.4
12 years	587	34.9
13 to 15 years	285	18.7
16 or more years	466	37.2
Currently married	18	1.5
Unweighted *N*	1672	

*Estimates are weighted to correct for unequal probabilities of selection and nonresponse. †Highest grade of school or year of college completed. ‡Highest grade or year of school completed by any parent or stepparent living in the household.

Table 2. Alternate estimates of prevalence of male-female and male-male sexual behaviors among 1995 NSAM respondents obtained by using different methods of questioning. For Tables 2 to 5 the odds ratio (OR) is adjusted (Adj.) for covariates [race (white, black, other as residual category), whether they have health insurance, age, whether they currently attend school, and whether they had sexual intercourse with a female], as reported in the interviewer-administered portion of the survey.

Measurement	Estimated prevalence (per 100)		Crude OR	Adj. OR
	Paper SAQ	Audio-CASI		
Male-female sexual contacts				
Ever had sex with a prostitute†	0.7	2.5	3.65***	4.24***
Ever been paid for sex‡	1.6	3.8	2.36*	2.60
Sexual intercourse with female within last year§‖	49.6	47.8	0.93	1.24
5+ lifetime female partners§‖	15.8	18.8	1.23	1.57*
Condom use at last sex (among males reporting sex)§	64.4	64.0	0.98	1.01
Ever had anal intercourse w/female	10.3	11.4	1.13	1.26
Ever made girl pregnant§‖	7.9	6.5	0.81	0.98
Ever fathered a child§‖	4.6	2.4	0.51	0.59
Ever had vaginal, oral, or anal intercourse with female¶	68.1	63.9	0.83	0.81
Male-male sexual contacts				
Ever masturbated another male	1.4	2.6	1.94	2.25*
Ever been masturbated by another male	0.9	3.5	3.79**	4.23*
Ever had receptive oral sex with another male (your mouth on his penis)	0.5	2.3	5.08**	5.68*
Ever had insertive oral sex with another male (your penis in his mouth)	1.1	3.1	2.83*	3.50*
Ever had insertive anal sex with another male (your penis in his rectum or butt)	1.0	1.9	1.85	2.41
Ever had receptive anal sex with another male (his penis in your rectum or butt)	0.1	0.8	7.91***	7.85**
Any male-male sex	1.5	5.5	3.84***	4.20***

*P = 0.15 or less for two-tailed test of null hypothesis that OR = 1.0. **P = 0.05 or less for two-tailed test of null hypothesis that OR = 1.0. ***P = 0.01 or less for two-tailed test of null hypothesis that OR = 1.0. †Although we have listed contact with a prostitute under male-female behaviors, the question was not gender-specific. It is possible that some contacts were with a male prostitute. ‡A total of 59 respondents reported ever being paid for sex; of those, 88% reported being paid by a female or females, 7% by a male(s), and 5% by both male(s) and female(s). An additional 11 respondents in the paper SAQ (not included in the Table 2 estimate) reported they had never been paid for sex, yet noted the gender of that person(s) in the subsequent question. §This question from the experiment repeats a question on heterosexual contact that was in the interviewer-administered portion of the survey. Because respondents may have felt compelled to answer consistently, responses to this question could be subject to a consistency bias that might have attenuated the effect of the interview mode. ‖In the SAQ, respondents who reported no sexual activity were instructed to skip a series of questions on specific sexual practices. For this analysis, these respondents were recorded as not reporting this particular behavior. ¶Estimated prevalence is for responding yes to one (or more) of four questions asking whether a respondent had engaged in vaginal sex, insertive or receptive oral sex with a female, or anal sex with a female.

permits a relatively precise characterization of any differences in reported risk behaviors that may exist among the groups or in their reaction to the technology being tested in our experiment. (All subsequent tabulations use weighted data.)

All respondents initially completed an interviewer-administered survey that contained a range of questions on their relations with their peers; the sources of their education and information on sexuality and contraception; their perceptions of and attitudes toward condoms, contraceptives, and the consequences of unintended pregnancy; their heterosexual and contraceptive experiences and their knowledge and perceptions of sexually transmitted diseases (STDs); and their social background and demographic characteristics. The interviewer-administered portion of the survey lasted an average of 68 min.

Next, respondents completed a self-administered interview. This interview contained 99 of the most sensitive questions being asked in the survey, including questions on the respondent's male-male sexual experiences, drug use, STD history, experiences with violence and weapons, heterosexual oral and anal sex, and sexual orientation (*19*). To assess the effects of audio-CASI versus more traditional paper SAQs on the reporting of such information, we embedded an experiment in this part of the survey. Participants were randomly assigned (at a ratio of 4:1) to a condition in which questions were administered with audio-CASI ($N = 1361$) or to a condition that used a paper SAQ ($N = 368$). Of the 1729 eligible survey respondents, 1690 completed the experiment. Because of computer problems and interviewer mistakes, a small number of those respondents ($N = 18$) were not interviewed in the mode that had been randomly assigned. These cases were excluded from our analyses. Thus, the base N for our analyses is 1672 (*20*).

Statistical analyses. In estimating the prevalence of reported behaviors, we focused on variations between modes of interviewing (audio-CASI versus paper SAQ). Because respondents were randomly assigned to the two interviewing conditions, the crude ratio of the odds of reporting a particular behavior in each condition provides an unbiased measure of the effect of audio-CASI on such reports relative to the more traditional paper SAQs. Our target behaviors were dichotomous variables (for example, did versus did not have male-male sex); therefore, we used log-linear modeling and logistic regression procedures to gauge the statistical significance of any effect that was observed and to test for variations in effect across subpopulations (*21*). Because we wished to make inferences about the likelihood that a given effect would hold in the population at large, our analyses used the sampling weights described earlier, and our statistical tests took account of weighting and the NSAM's complex sample design (*22*). The P values shown in Tables 2 through 5 test the null hypothesis that given the odds ratio observed in this experiment, the "true" odds ratio would have been 1.0 (that is, no effect of audio-CASI) if the same experiment was repeated among an infinite number of similarly designed samples of the national population (*23*).

Although the crude odds ratio is an unbiased indicator of the effect of the survey technology, more precise measures were also calculated. Adjusted odds ratios may correct for random perturbations across experimental conditions in the distribution of factors correlated with the targeted risk behavior (*24*).

Estimated prevalence of risk behaviors. As shown in Table 2 (*25, 26*), the mode of survey administration made little difference in estimates of the prevalence of male-female sexual contacts. The only significant effect was for reporting of sexual contact with a prostitute. Respondents who were interviewed with audio-CASI were 3.6 times as likely (crude odds ratio) to report such contact as respondents who were interviewed with the paper SAQ (2.5% versus 0.7%) (*27*).

Although the use of audio-CASI had little effect on reporting of male-female adolescent sexual behavior, it produced highly significant increases in reports of male-male sex. Overall, respondents were almost four times as likely to report some type of male-male sex in the audio-CASI mode compared with the paper SAQ (5.5% versus 1.5%). Estimated odds ratios varied from 1.85 to 7.91 across the six measurements of passive and active masturbation and oral and anal sex with males. The individual results are not all significant, but the pattern is consistent; that is, audio-CASI was more likely than the paper SAQ to elicit a report that the respondent had engaged in male-male sex. In addition, the observed fourfold increase in the reported prevalence of male-male contact among adolescents is statistically significant ($P < 0.001$), and the estimated prevalence of such contact (5.5%) is more consistent with prevalences derived from retrospective reports provided by adult men reporting on their sexual behaviors during adolescence (4 to 9%) (*13*).

Table 3 presents parallel results for the reporting of drug use alone and as an adjunct to sexual behaviors. For the drug use

Table 3. Alternate estimates of prevalence of drug use, per se, and drug use during sex derived by using different methods of questioning. Results are from the 1995 NSAM.

Measurement	Estimated prevalence (per 100)		Crude OR	Adj. OR
	Paper SAQ	Audio-CASI		
Drug use				
Ever taken street drugs using a needle	1.4	5.2	3.85***	3.90*
Injected drugs within last year‡	0.0	0.8	–†	–†
Ever shared needle§	0.1	1.1	9.71**	9.56**
Smoked marijuana daily during last year‖	4.1	6.7	1.69*	2.03*
Used crack/cocaine within last year	3.3	6.0	1.89	1.96
Drank alcohol last year¶	65.9	69.2	1.16	1.29
Drank alcohol weekly last year#	15.0	19.4	1.34	1.56*
Ever smoked marijuana	41.2	43.3	1.09	1.30*
Drug use and sex (among those having sex)††				
Ever had sex with someone who shoots drugs	0.2	2.8	13.84**	17.06**
You/your partner drunk or high at last heterosexual intercourse	15.3	34.8	2.95***	3.04*
Always/often drunk or high during heterosexual intercourse last year	2.2	10.8	5.52***	5.69***
You/your partner had been drinking at time of last heterosexual intercourse	13.9	25.4	2.10***	2.14***
You/your partner used drugs at time of last heterosexual intercourse	9.7	15.8	1.74*	1.89*

*$P = 0.15$ or less for two-tailed test of null hypothesis that OR = 1.0. **$P = 0.05$ or less for two-tailed test of null hypothesis that OR = 1.0. ***$P = 0.01$ or less for two-tailed test of null hypothesis that OR = 1.0. †Odds ratio not calculated because of zero cell value. (Fisher's exact test for unadjusted frequency had a p value of 0.13 with normalized weighted data (without adjustments for impact of complex sample design and weighting). ‡Never injected drugs (question 17) recoded to no use within last year. §Never injected drugs (question 17) recoded to never shared needle. ‖Never smoked marijuana (question 13) recoded to no use within last year. ¶Never drank alcohol (question 9) recoded to no alcohol use last year. #Never drank alcohol (question 9) recoded to no alcohol use last week. ††Due to questionnaire skip patterns, these questions were asked only of respondents who reported ever having sexual intercourse, including vaginal, oral, or anal intercourse with a female.

behaviors that are most commonly associated with transmission of HIV—that is, injection of illicit drugs and use of crack (a refined, smokable form of cocaine) or cocaine—the audio-CASI mode elicited substantially more reporting. Overall, the audio-CASI measurements yielded an estimate of the prevalence of injection drug use (5.2%) that is substantially higher than any prior population estimate. The estimated prevalence derived from the more traditional paper SAQ measurements was 1.4% (*28*).

Behaviors that commingled drug use and sex are another category that evinced a strong effect from the use of the audio-CASI technology (*29*). Respondents assigned to the audio-CASI condition who reported that they were sexually active were more likely than their counterparts in the SAQ condition to report that they or their partners were drunk or high when they last had heterosexual intercourse (odds ratio = 2.95); they were also more likely to report that they were "always" or "often" drunk or high when they had heterosexual sex during the past year (odds ratio = 5.5). In the audio-CASI mode, respondents who reported having sex were much more likely than SAQ respondents to report that they had had sex with someone who injected drugs (odds ratio = 13.8).

The pattern of higher estimates of prevalence in the audio-CASI mode is repeated in the NSAM measurements of violent behaviors (Table 4). Respondents in the audio-CASI mode were more likely to report that they had carried a gun in the past 30 days (12.4 versus 7.9% in the SAQ mode) (*30*) and to report that they had carried a knife or razor in the same period (27.1 versus 19.4%). Similar increases were found in reports of threats of violence, both those made by the respondent and those made against him.

Variations in effects across social and demographic subgroups. To determine whether the increase in reporting we observed with audio-CASI might vary across race-ethnicity or age groups, we fit a hierarchical series of log-linear models (*31*) to the four-way tables for each behavior reported in Tables 2 through 4 for which there was a significant difference ($P < 0.05$) in estimated prevalence across modes. The four-way table broke down the estimated prevalence of each risk behavior by three variables: mode of interview (audio-CASI versus paper SAQ), age (15 to 16 versus 17 to 19), and race-ethnicity (black, Hispanic, white, or other) (*32*). These analyses revealed little evidence of significant variation in the effect of the survey mode across those age and race-ethnicity groups.

Although we did not measure respondent literacy directly, we were able to identify respondents who were "not at grade level" in school or who had dropped out of high school without graduating. Respondents were coded as "at approximate grade level" if they were age 15 and had completed 8th grade or higher; age 16 and had completed 9th grade or higher; age 17 and had completed 10th grade or higher; age 18 and had completed 11th grade; or any age and had completed 12th grade. All other respondents were coded as "below grade level." To detect variations in how audio-CASI influenced reporting in the latter group, we focused on the reporting of drug use, the combination of drug use and sex, male-male sex, and violence.

For most behaviors, no substantial differences were observed in the effect of audio-CASI across these two education strata. However, in four instances (reporting of any male-male sex, injection drug use, daily marijuana use in the past year, and crack or cocaine use in the past year), there were noteworthy differences in the effect of the survey mode. For reporting of the three drug use behaviors, audio-CASI had a strong effect (odds ratios = 2.8 to 14.1) for the subpopulation of respondents who were high school graduates or who were at their approximate grade level (Table 5). No such effect was found for the subpopulation who had dropped out of school or were still in school but seriously behind their normal grade level. For the three drug use behaviors, the paper SAQ yielded estimates that were about equivalent to those obtained with audio-CASI technology (odds ratios = 0.5 to 1.5). For the reporting of male-male sex, there were parallel increases in reporting when audio-CASI was used, but the effect was more pronounced for males who were behind in school. No male who was behind in school reported any of the six male-male sexual behaviors when completing the paper SAQ. In the audio-CASI condition, 6.2% of respondents not at grade level reported one (or more) types of male-male sexual contact. For males who were at grade level a more attenuated effect was observed [1.8% in the paper SAQ versus 5.3% in audio-CASI (*33*)].

For reports of serious drug use, this pattern of differential sensitivity to interview mode might imply that the segment of the population that is "on track" educationally may perceive that they have more to lose by discovery of their past drug use. Alternatively, there may be differences across groups in peer norms about the acceptability of such behavior. Similar differences in peer norms could account for a differential sensitivity to reporting of male-male sexual contacts. Data from representative national samples of adults surveyed in 1990 to 1994 indicate that tolerance of same-gender sex increases with the educational attainment of the respondent (*34*). Although such speculations have some intuitive appeal, they do not account for the failure to find parallel results for questions other than the four noted here.

Other results. Nonresponse to individual questions has been a persistent problem in self-administered interviews that use paper SAQs (*35*). Respondents in the audio-CASI mode in our experiment were less likely than respondents in the SAQ mode to use the "don't know" or "refuse to answer" options. (The audio-CASI technology provided labeled keys for those responses.) Of the 58 questions that were designed to be answered by all respondents, an average of only 0.5% of respondents in the audio-CASI mode did not answer them, compared with an average of 2.3% of respondents in the paper SAQ mode.

Discussion. The results of this large-scale experiment indicate that respondents

Table 4. Alternate estimates of prevalence of violent behaviors derived by using different questioning methods. Data are from the 1995 NSAM.

Measurement	Estimated prevalence (per 100)		Crude OR	Adj. OR
	Paper SAQ	Audio-CASI		
Someone threatened to hurt you in past year	25.7	34.3	1.51***	1.61***
You threatened to hurt someone in past year	17.1	26.1	1.72***	1.89***
Carried a gun in past 30 days	7.9	12.4	1.66**	1.76**
In physical fight within last year	38.9	43.8	1.22	1.29
You pulled knife or gun on someone in past year	6.2	8.9	1.50*	1.75**
Someone pulled knife or gun on you in past year	16.9	21.1	1.32*	1.47*
Carried a knife or razor in past 30 days	19.4	27.1	1.55**	1.68**

*P = 0.15 or less for two-tailed test of null hypothesis that OR = 1.0. **P = 0.05 or less for two-tailed test of null hypothesis that OR = 1.0. ***P = 0.01 or less for two-tailed test of null hypothesis that OR = 1.0.

in the 1995 NSAM's new cohort, a probability sample of U.S. males aged 15 to 19, were much more likely to report risky behaviors when they were interviewed with audio-CASI measurement technology than when interviewed with the more traditional paper SAQ. The magnitude of the observed effect was substantial for many behaviors, particularly those that were illicit or highly stigmatized. Thus, the estimated prevalence of past-year use of crack or cocaine almost doubled, from 3.3 to 6.0%, and the estimated proportion of adolescent males who reported ever taking street drugs with a needle more than tripled, from 1.4 to 5.2%. Although we have some concern about the interpretation of the injection drug use result (*36*), overall the audio-CASI estimates were considerably higher for illicit and stigmatized behaviors. These NSAM audio-CASI estimates were also higher than estimates derived from the federal government's 1995 National Household Survey on Drug Abuse. Estimates derived from that survey are that the prevalence of reported crack or cocaine use during the past year among males aged 15 to 19 was 2.6% and that the prevalence of reported injection drug use in the same group was 0.3% (*37*).

Varying the interview mode also affected the reporting of the co-occurrence of drug use and sexual behavior. The proportion of sexually active adolescent males who reported that they or their partner had been drunk or high at last intercourse more than doubled in the audio-CASI mode (from 15.3% in the paper SAQ condition to 34.8%). In addition, indicators of interpersonal violence increased markedly.

In planning this experiment, we were aware of prior research (*5*–*7*) showing that a private mode of survey response such as paper SAQs encouraged more complete reporting of sensitive, stigmatized, and illicit behaviors. We anticipated that the additional benefits of computerization and an audio format for the self-administered interview would bring only modest improvements in reporting. The large differences we found were surprising. If this measurement technology provides a more accurate picture of patterns of sexual behaviors, drug use, and violence among adolescent males in the United States, then the risks they are encountering are substantially greater than was previously supposed. This conclusion assumes, of course, that the observed increase in reporting of these behaviors is, in fact, more accurate reporting. That assumption follows standard practice in much social measurement research (*38*); nonetheless, it remains an assumption. Yet in the present instance, there is empirical evidence to support it. The large increases in the reporting of male-male sexual contacts under audio-CASI (to 5.5% for any contact) make these measurements more compatible with estimates derived from adults' reports of their adolescent behaviors (*39*).

The fact that audio-CASI both increases reporting of male-male contacts and makes our estimates more consistent with those obtained from retrospective reports of adults encourages the conclusion that this technology is reducing the underreporting bias known to affect such measurements. In addition, the technology appears to have a more pronounced effect on the reporting of behaviors that are particularly sensitive, stigmatized, or subject to serious legal sanctions, compared with less sensitive areas of conduct. Thus, our analyses of the NSAM data found few reliable effects of audio-CASI on the reporting of male-female sexual contacts; the sole exception was for reporting of sexual contact with prostitutes. That behavior is arguably a more sensitive topic for adolescents than are other male-female sexual contacts. Similarly, audio-CASI substantially increased reporting of all types of male-male sexual contacts and drug use. Here again, exceptions occurred for reports of more common and, we suspect, less sensitive behaviors such as the use of alcohol and marijuana.

Those results and complementary evidence emerging from smaller studies using local samples (*40*) lead us to conclude that the estimates presented here are more accurate than previous estimates derived from data collected in less private interview contexts. These estimates present a disturbing picture of the biological and social risks that confront young males in the United States at the end of the 20th century.

Table 5. Selected estimates of prevalence of male-male sex and drug use by mode of interview and whether respondent was at "grade level." Respondent is coded as "at grade level" if by age 15 they have completed grade 8, completed 9th grade by 16, 10th grade by 17, 11th grade by 18, or completed high school. All other respondents and those that have not completed 8 years of school are "below grade level."

Measurement	At or above approximate grade level				Below grade level			
	Paper SAQ	Audio-CASI	Crude OR	Adj. OR	Paper SAQ	Audio-CASI	Crude OR	Adj. OR
Any male-male sex	1.8	5.3	3.03**	3.22***	0.0	6.2	–†	–†
Ever taken street drugs using a needle	0.3	4.5	14.11***	16.67***	6.1	8.7	1.48	1.22
Smoked marijuana daily during last year‡	1.9	5.1	2.78**	3.51**	13.5	14.6	1.10	1.34
Used crack/cocaine within last year	0.9	5.7	6.67***	7.15***	13.8	7.4	0.50	0.50

*P = 0.15 or less for two-tailed test of null hypothesis that OR = 1.0. **P = 0.05 or less for two-tailed test of null hypothesis that OR = 1.0. ***P = 0.01 or less for two-tailed test of null hypothesis that OR = 1.0. †Odds ratio not computed due to zero cell entries; P = 0.011 by Fisher exact test. ‡Never smoked marijuana (question 13) recoded to no use within the last year.

NOTES AND REFERENCES

1. F. L. Sonenstein, J. H. Pleck, L. Ku, *Fam. Plann. Perspect.* **21**, 132 (1989); J. A. Catania *et al.*, *Science* **258**, 1101 (1992); E. Laumann, J. Gagnon, R. Michael, S. Michaels, *Social Organization of Sexuality* (Univ. of Chicago Press, Chicago, IL, 1994); A. Johnson, J. Wadsworth, K. Wellings, J. Field, S. Bradshaw, *Sexual Attitudes and Lifestyles* (Blackwell, London, 1994); P. B. Davis, R. L. Yee, J. Chetwynd, N. McMillan, *AIDS* **7**, 1509 (1994); K. Tanfer, *Fam. Plann. Perspect.* **25**, 83 (1993).
2. See, for example, H. G. Miller, C. F. Turner, L. E. Moses, Eds., *AIDS: The Second Decade* (Report of the National Academy of Sciences–National Research Council Committee on AIDS Research and the Behavioral, Social, and Statistical Sciences; National Academy Press, Washington, DC, 1990), pp. 147–152.
3. This situation appears to be unique to the United States. Among young males 15 to 24 years of age, the United States has the highest homicide rate among 22 industrialized nations [M. L. Rosenberg, *J. Health Care Poor Underserved* **6**, 102 (1995); H. Saner and P. Ellickson, *J. Adolesc. Health* **19**, 94 (1994); J. M. Bergstein, D. Hemenway, B. Kennedy, S. Quaday, R. Ander, *J. Trauma* **41**, 794 (1996); P. F. Adams, C. A. Schoenborn, A. M. Moss, C. W. Warren, L. Kann, *Vital Health Stat.* **10**, 1 (1992)].
4. R. C. Lewontin, *N.Y. Rev. Books*, 20 April 1995, p. 24.
5. E. F. Jones and J. D. Forrest, *Demography* **29**, 113 (1992).
6. C. F. Turner, R. D. Danella, S. M. Rogers, *Sex. Transm. Dis.* **22**, 173 (1995).
7. C. F. Turner, J. T. Lessler, J. Devore, in *Survey Measurement of Drug Use: Methodological Issues*, C. F. Turner, J. T. Lessler, J. C. Gfroerer, Eds. (Government Printing Office, Washington, DC, 1992), pp. 177–220.
8. See, for example, C. Jenkins, paper presented at the meeting of the American Association for Public Opinion Research, Norfolk, VA, 15 to 18 May 1997.
9. G. Johnston, demonstration for a seminar at the National Center for Health Statistics, Hyattsville, MD, 28 January 1992; J. M. O'Reilly and C. F. Turner, presentation to the Washington Statistical Society, 18 March 1992; J. M. O'Reilly, M. Hubbard, J. Lessler, P. Biemer, C. F. Turner, *J. Off. Stat.* **10**, 197 (1994).
10. C. F. Turner *et al.*, in *Survey Research Using Automated Data Collection Systems*, M. Couper *et al.*, Eds. (Wiley, New York, in press); C. F. Turner, H. G. Miller, S. M. Rogers, in *Researching Sexual Behavior*, J. Bancroft, Ed. (Indiana Univ. Press, Bloomington, in press); C. F. Turner, S. M. Rogers, T. P. Hendershot, H. G. Miller, J. P. Thornberry, *Public Health Rep.* **111**, 276 (1996).
11. Multilingual audio-CASI administration was not used in NSAM but has been tested in preliminary studies [T. P. Hendershot, S. M. Rogers, J. P. Thornberry, H. G. Miller, C. F. Turner, in *Health Survey Research*

Methods, R. Warnecke, Ed. (National Center for Health Statistics, Hyattsville, MD, 1996), pp. 165–169].
12. L. Ku, F. L. Sonenstein, J. H. Pleck, *Fam. Plann. Perspect.* **26**, 246 (1994); *Am. J. Public Health* **83**, 1609 (1993); J. H. Pleck, F. L. Sonenstein, L. Ku, *Fam. Plann. Perspect.* **25**, 106 (1993); L. Ku, F. L. Sonenstein, J. H. Pleck, *ibid*. **24**, 100 (1993).
13. The 1992 National Health and Social Life Survey (NHSLS), for example, found that 9.1% of U.S. men in 1992 reported having had male-male contacts after puberty but only 4.9% reported such contacts after age 18. Those results imply that 4.2% of men had male-male contacts that were restricted to adolescence and some portion of the remaining 4.9% would have begun such contacts during adolescence. In a 1970 Kinsey Institute survey of a representative sample of U.S. men, 81% of the men reporting some history of male-male contact said that their first male-male experiences occurred before age 19 (*6*). It should also be noted that other analyses of the 1970 Kinsey Institute survey indicated that 8.4% of men reported some male-male contact that did not persist beyond age 14 [R. E. Fay, C. F. Turner, A. D. Klassen, J. H. Gagnon, *Science* **243**, 338 (1989)]. The 1992 NHSLS did not assess prepubertal contacts.
14. The design of the survey was changed in 1991 to make it more gender neutral in discussing sexual contacts and included a SAQ on male-male sexual contacts that was given before asking about male-female contacts. This change was made so that respondents would not infer that we expected only heterosexual contacts to be reported.
15. The sample was drawn as a probability subsample of an area probability sample selected for another study conducted in 1992. Details of the multistage sampling procedures used in selecting the NSAM sample can be found in Research Triangle Institute, *National Survey of Adolescent Males: Field Report* (Research Triangle Institute, Research Triangle Park, NC, 1995), pp. 3.2–3.6. (Our analysis is based on a final data set that includes a small number of NSAM cases that were received after the field report was completed.)
16. An additional 12 units (0.02%) were excluded because of mismatches between the original sample frame information and interviewer determination as to whether the unit was a household or group quarters.
17. Bureau of the Census, *Preliminary Projections: Noninstitutional Population by Age, Sex, Race and Hispanic Origin* (U.S. Department of Commerce, Washington, DC, May 1995).
18. B. Shah, B. G. Barnwell, P. N. Hunt, L. M. LaVange, *Software for Survey Data Analysis* (*SUDAAN*) (Research Triangle Institute, Research Triangle Park, NC, 1996).
19. After completing the self-interview portion of the survey, respondents who were age 18 and older were asked to provide a urine sample to be tested for chlamydia and gonorrhea. The results of those tests are not discussed here.
20. Nonresponse to individual questions may reduce the *N* for individual items. However, item nonresponse rates were quite low, averaging 1.3% for the measurements shown in Tables 2 through 5. Nonrespondents to individual items have been excluded from our analyses. In addition, it should be noted that some analyses purposely restrict the population base. So, for example, our estimate of the percentage of males using "condoms at last sex" (Table 2) is restricted to those males who reported having had vaginal, oral, or anal sex with a female. Population restrictions are noted in the tables.
21. L. A. Goodman, *J. Am. Stat. Assoc.* **63**, 1091 (1968); *Analyzing Qualitative/Categorical Data* (Haberman, Cambridge, 1978); *Analysis of Categorical Data* (Academic Press, New York, 1978); Y. Bishop, S. Feinberg, P. Holland, *Discrete Multivariate Analysis: Theory and Practice* (MIT Press, Cambridge, 1975).
22. An alternative approach would be to treat the sample of 1672 respondents as a closed population that had been randomly assigned to one of two interviewing conditions. In that approach, neither weights nor special statistical algorithms would be required. Statistical tests using that approach would provide an appropriate answer to the question of whether audio-CASI had had a significant effect on that particular sample of respondents. The approach is reasonable, but it does not allow one to generalize the results to the national population.
23. *P* values are based on the χ^2 statistic reported by SUDAAN (*18*).
24. As a check on the adequacy of the randomization, we compared the response distributions of 62 predominantly yes-no nonattitude questions measured in the prior interviewer-administered section of the questionnaire for respondents assigned to the two experimental conditions. We were initially perplexed to find that people who were assigned to the paper SAQ condition were more likely than those in the audio-CASI condition to report in the interviewer-administered section of the survey that they were virgins. They did not, however, show significant differences on demographic variables or other sexual behaviors (for example, employment during the past year, timing of last heterosexual intercourse). Our examination indicated that this result appears to have been a random occurrence. Of the 62 comparisons, 2 were found to be significant at the $P < 0.05$ level. On the basis of probability theory, we would expect random assignment to produce (on average) at least three significant ($P < 0.05$) results in 62 independent tests. We thus conclude that the randomization was successfully executed. The adjusted odds ratios shown in Tables 2 through 5 provide statistical adjustments for observed random differences in the composition of the sample used in the two experimental conditions. These adjusted ratios were derived from logistic regression models that used the following as control variables: race (black, white, other), respondent's age, whether the respondent was attending school, whether the respondent reported during the earlier interviewer-administered section of the survey that he had ever had sex with a female, and whether the respondent was covered by health insurance. The first three characteristics were selected to take account of differences among subpopulations in our targeted behaviors; the last two variables were chosen on the basis of preliminary statistical analyses that indicated random variations between the experimental and control conditions for the two characteristics (of 62 that were examined). Use of the adjusted rather than crude odds ratios does not markedly affect our conclusions (Tables 2 to 5).
25. In a few instances noted in Table 2, respondents in the experiment were asked again about heterosexual contacts that they had previously reported on in the interviewer-administered portion of the survey. Because respondents may have felt compelled to answer consistently rather than "honestly," responses to the questions that were asked again could be subject to a consistency bias that might have attenuated the effect of the interview mode. Moreover, there is a weak suggestion in these results that audio-CASI may have diminished "overreporting," which is sometimes thought to afflict measurements of highly normative behaviors, for example, reports by males of some male-female sexual experience. That effect, however, is not statistically significant and is weak compared with the effects observed on reporting of male-male sex and the other behaviors discussed [see also S. Newcomer and J. R. Udry, *J. Adolesc. Res.* **2**, 419 (1988)].
26. Since the original tabulation of estimates for Tables 2 to 5 in this article, we detected a minor data anomaly that has been corrected in the public use data set. For four variables, the value 999 was an allowable (albeit unusual) response to questions on: (i) the number of times in past 12 months someone pulled a gun, knife, or razor on respondent; (ii) the number of times the respondent threatened to hurt someone during the last 12 months; (iii) the respondent's total number of lifetime female sexual partners; and (iv) the respondent's total number of female sex partners during the past 12 months. Similarly, the value 99 was an allowable response to a fifth question on the number of times the respondent was in a fight during the last year. Seven respondents gave one or more of these answers (999 or 99) to these questions. In our original tabulations, these extreme values were treated as missing data codes and excluded from the analysis. Correcting this error alters the reported percentages for two mode comparisons by more than 0.1 percentage points. For reporting of five or more lifetime (female) sex partners (Table 2), the results are 15.8% (SAQ) and 19.0% for audio-CASI (versus 15.8% and 18.8%, if the value 999 were treated as missing). Similarly, for respondents' reports of having threatened to hurt someone in the past year (Table 4), the results are 17.1% for SAQ and 26.3 % for audio-CASI (versus 17.1% and 26.1%, if the value 999 were treated as missing). In manual editing of paper SAQs for data entry, all values of 400 or more for these variables were treated as invalid and coded as "missing data." Thus comparison of paper SAQ estimates to audio-CASI estimates in which 999 codes are treated as missing values—as was done in Tables 2 through 5—is appropriate. If the additional values 400 to 998 were treated as missing values for the audio-CASI measurements, the audio-CASI estimates would not change by more than 0.1%.
27. As noted in Table 2, it is possible that the prostitute was male. It should also be noted that a suggestive mode effect ($P < 0.15$) was found in Table 2 for respondents who reported that they had been paid to have sex.
28. It should also be noted that the only internally consistent reports of injection drug use in the past year were obtained from nine males in the audio-CASI mode. One male in the paper SAQ mode gave an inconsistent report of injection drug use. Respondents in the paper SAQ condition were asked a series of questions on injecting drugs. First, they were asked if they had ever taken street drugs by using a needle. Next, they were asked how frequently they had injected drugs during the past year, whether they had ever shared a needle, and how often they had sterilized previously used needles. In the audio-CASI condition, respondents who reported they had never injected drugs were skipped out of the remaining questions. However, one respondent completing the paper questionnaire first reported never injecting drugs but later reported doing so "a few times" during the past year.
29. The series of questions on commingling of sex and alcohol or drug use was preceded by a gate question asking "Have you ever had sexual intercourse, including vaginal, oral, or anal intercourse, with a female?" Respondents who answered "no" to that question were skipped past the series of questions on drug and alcohol use during sex. Analyses comparing responses to this question with those to questions asked separately about oral, vaginal, and anal sex suggest that some respondents may have interpreted the gate question as asking about the co-occurrence of all three types of sexual behavior. Thus, only 60.6% of paper SAQ and 54.0% of audio-CASI respondents replied "yes" to this question. In comparison, in response to four questions asking separately about vaginal sex, insertive and receptive oral sex, and anal sex with females, 68.1% of paper SAQ and 63.9% of audio-CASI respondents indicated that they had engaged in one (or more) types of sex.
30. This finding is not likely to reflect the carrying of guns for hunting: 13.4% of males living in urban areas reported that they had carried a gun during the prior 30 days compared with 11.8% of males living in rural areas.
31. Exploratory analyses were conducted with weights that had been normalized so that the weighted *N* equaled the sample *N*. The effect of the complex sample design was not taken into account in these initial exploratory analyses that fit a series of hierarchical log-linear models. Subsequently, odds ratios were recalculated for results that evidenced substantial disparities across schooling levels in the estimated effect of interview mode on survey response. These recalculated odds ratios and associated *P* values (Table 5) used algorithms that took account of the complex sample design.
32. More precisely, the four categories were Hispanics, non-Hispanic whites, non-Hispanic blacks, and other.
33. There was also a less substantial difference in the effect of mode on reporting that someone had pulled a gun on the respondent in the past year. For re-

spondents who were at grade level, there was an insignificant mode effect (crude OR = 1.06; 17.5% in paper SAQ versus 18.4% in audio-CASI), whereas a significant ($P < 0.05$) mode effect in the same direction was found for respondents who were not at grade level (crude OR = 2.22; 15.3% in paper SAQ versus 28.5% in audio-CASI).

34. The National Opinion Research Center (NORC) General Social Survey asked the following question of national samples drawn annually between 1988 and 1994: "What about sexual relations between two adults of the same sex. Do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?" Among males aged 18 and older (N = 2877), 86% of those who had not completed high school responded "always wrong" compared with 83% of those who left school after 12th grade, 72% of those who had completed 13 to 15 years of school, and 57% of those who had completed 16 or more years of education. (Estimated percentages are weighted to account for varying probabilities of selection into the sample.) Data were extracted and tabulated from the NORC cumulative data file [J. A. Davis and T. W. Smith, *General Social Surveys, 1972–1994* (machine-readable data file) (NORC, Chicago, IL, 1972–1994)].
35. S. M. Rogers and C. F. Turner, *J. Sex Res.* **28**, 491 (1991); M. Witt, J. Pantula, R. Folsom, B. Cox, in (7), pp. 85–108.
36. In particular, we note that the question on injection drug use was worded: "Have you ever taken street drugs, like heroin or cocaine, using a needle? This includes 'shooting up' and 'skin popping'" (emphasis in original). It is possible that in formulating a response, audio-CASI respondents may have focused on the initial clause "ever taken street drugs, like heroin or cocaine" in the audio format of this question, whereas respondents in the paper SAQ condition may have been more likely to notice the subsequent "using a needle" clause and the explanatory phrases "shooting up" and "skin popping" when reading the paper version of this question.
37. These estimates are based on weighted tabulations from the public use data set for the 1995 National Household Survey on Drug Abuse (NHSDA). The NHSDA is sponsored by the Substance Abuse and Mental Health Services Administration of the Department of Health and Human Services. In 1995, the NHSDA surveyed a population sample of more than 17,000 household residents in the United States. Questions on illicit drug use were administered with paper SAQs. Our tabulations are based on the sample of 1624 males ages 15 to 19 who were included in the NHSDA survey.
38. See, for example, N. Bradburn, S. Sudman, and Associates, *Improving Interview Method and Questionnaire Design* (Bass, San Francisco, 1979); J. J. Waterton and J. C. Duffy, *Int. Stat. Rev.* **52**, 173 (1984).
39. We would also note that a parallel study of the effect of audio-CASI on abortion reporting found that with this technology the known underreporting of this procedure decreased. In that instance, there was an external standard (reports by abortion providers) to provide assurance that increased reporting was more accurate reporting [H. G. Miller, S. M. Rogers, J. Gribble, C. F. Turner, in *The Science of Self Report* (provisional title), A. Stone, Ed. (Lawrence Erlbaum Associates, Mahwah, NJ, in press)].
40. R. Tourangeau and T. Smith, *Public Opin. Q.* **60**, 275 (1995); C. F. Turner, H. G. Miller, T. K. Smith, P. C. Cooley, S. M. Rogers, in *Survey and Statistical Computing 1996*, R. Banks, J. Fairgrieve, L. Gerrard, Eds. (Association for Survey Computing, Chesham, Bucks, UK, 1996).
41. This research was supported by NIH grant R01-HD30861 with funding from the National Institute of Child Health and Human Development (NIH-NICHD), the National Institute of Mental Health (NIH-NIMH), the Office of Population Affairs, and the Centers for Disease Control and Prevention. We thank H. Zelon, F. Mierzwa, J. Chromy, D. Allen, B. Forsyth, and S. Williams for their contributions and A. Harrell, H. Miller, R. Folsom, and J. Gfroerer for helpful comments.

19 November 1997; accepted 19 March 1998

Research methods

Series editors
J M Stephenson,
A Babiker

Measuring sexual behaviour: methodological challenges in survey research

Kevin A Fenton, Anne M Johnson, Sally McManus, Bob Erens

Introduction

The study of sexual behaviour lies at the heart of understanding the transmission dynamics of sexually transmitted infections (STIs). Academic investigation into sexual behaviour dates back to the 18th century and, over time, has employed a variety of approaches including the medical and psychiatric investigation of sexual disorders, anthropological investigations, and survey research based largely on volunteer samples. More recent studies, driven largely by the public health response to HIV/AIDS, have focused on large scale probability sample survey research.[1–5] Key areas of inquiry have shifted towards describing population patterns of risk behaviours for STI/HIV transmission, understanding how epidemics of STIs are generated, and informing disease control strategies.

Sexual behaviour is a largely private activity, subject to varying degrees of social, cultural, religious, moral and legal norms and constraints. A key challenge for all sex survey research is to generate unbiased and precise measures of individual and population behaviour patterns. Methods are needed to minimise measurement error which may be introduced by participation bias, recall and comprehension problems, and respondents' willingness to report sensitive and sometimes socially censured attitudes or behaviours.[6 7] This paper briefly considers the role of different types of study in understanding STI epidemiology. It then focuses on potential sources of measurement error in survey research and strategies for assessing and limiting them.

Types of study

The type of study chosen will depend on the purpose of the investigation. However, studies generally fall into four main groups: general population surveys, studies on population subgroups, partner and network studies, ethnographic and qualitative studies.

GENERAL POPULATION PROBABILITY SAMPLE SURVEYS

Cross sectional population surveys aim to describe the overall distribution of behaviours in populations. By using probability sampling techniques and maximising response rates, large scale behavioural surveys can provide robust estimates of the prevalence of behaviours and their determinants in the population. However, they are frequently not large enough to determine the prevalence of behaviours among small population subgroups (for example, homosexual men) or among individuals with relatively rare experiences (for example, injecting drug use) which may be particularly important in transmission of infection. Since cross sectional surveys provide a snapshot in time, multiple surveys are required to measure and monitor behaviour change over time. Data from Switzerland[8] and Sweden[9] have shown temporal changes in partner change and condom use over time. In Britain, although two successive national surveys of sexual attitudes and lifestyles (NATSAL)[3 10] have been carried out a decade apart, there are few robust data for the interim period. In order to supplement data from intermittently commissioned large scale sex surveys, sexual behaviour questions (as key indictors or modules) may be added to probability sample general social surveys.[11 12]

SURVEYS ON SMALL SUBGROUPS AT HIGH RISK

Sexual behaviour studies often focus on epidemiologically important core groups that maintain STI transmission in the population such as commercial sex workers, homosexual men, injecting drug users, and STD clinic attenders. With very rare exceptions,[13] difficulties in accessing these groups make probability sampling costly and challenging, and more cost effective sampling strategies are required, including advertising, snowballing, recruiting from STD clinics, social and commercial venues. However, findings from these studies may not be representative of the wider target population. Thus, homosexual men who attend STD clinics have higher risk behaviours than those who do not[14] and STD clinic surveys will therefore tend to overestimate the prevalence of these behaviours.

Prospective monitoring of behaviours in high risk groups may be achieved through cohort investigations or serial surveying. Probability samples from the general population can also be followed up to provide repeated behavioural measurements over time.[15] Cohort studies enable estimation of disease incidence and monitoring of behavioural risk over time.[15–19] In these instances, attributing lifestyle changes to behavioural interventions can be difficult, since significant age confounding (associated with decreasing sexual activity) may occur. Attrition rates can also be problematic in cohort studies, if those with high risk behaviours are more likely to drop out, leaving more compliant individuals. Behavioural surveillance, involving serial cross sectional surveys of a target group using the same sampling strategy over time, provides an

Department of Sexually Transmitted Diseases, Royal Free and University College Medical School, London WC1E 6AU, UK
K A Fenton
A M Johnson

The National Centre for Social Research, London EC1V 0AX, UK
S McManus
B Erens

Correspondence to:
Dr Kevin Fenton, Department of Sexually Transmitted Diseases, Royal Free and University College Medical School, Mortimer Market Centre, off Capper Street, London WC1E 6AU, UK
kfenton@gum.ucl.ac.uk

Accepted for publication
15 December 2000

alternative mechanism for prospective behavioural monitoring.[12] In London, annual surveys of homosexual men in social venues, STD clinics,[14 20] and Gay Pride events[21] use a stable set of behavioural indicators—for example, unprotected anal intercourse in the past 3 months, which are then monitored repeatedly. Both have demonstrated increasing risk behaviour among homosexual men and have provided useful behavioural trend data to inform public health interventions.

PARTNER AND NETWORK STUDIES

Partner studies are concerned with studying transmission probabilities for STIs and their association with specific sexual behaviours. In the 1980s, a series of partner studies examined the transmission probability of heterosexual transmission of HIV.[22 23] These relied on detailed behavioural data to exclude other sources of exposure than the index case, and to identify risk factors for transmission. These studies established the role of unprotected vaginal intercourse in heterosexual transmission; the protective role of condoms; the increased risk of unprotected anal intercourse; and the poor association between the number of acts of intercourse and the probability of transmission. Other studies have utilised partner notification data to estimate transmission probabilities for STIs[24] and to determine the role of sexual networks in maintaining endemic STI transmission.[24–26] These studies have highlighted the importance of "core groups"[27] and of particular individuals within networks, in maintaining chains of transmission. Such studies are however highly intensive, with many practical difficulties. Nevertheless, epidemiological research on STI transmission is increasingly focusing on the importance of understanding mixing matrices, particularly in "core" populations. More detailed considerations of these important developments are beyond the scope of this paper.

ETHNOGRAPHIC AND QUALITATIVE STUDIES

Ethnographic and qualitative studies on sexual behaviour have made significant contributions to our understanding of STI transmission dynamics.[28] Studies exploring the social context of sexual behaviour—for example, the importance of San Francisco "bath houses"[29] where homosexual men had large numbers of anonymous sexual contacts, were key to understanding the early evolution of the AIDS epidemic.[30] Qualitative research has enabled the exploration of concepts within communities[31 32] and revealed behaviours or cultural factors which are relevant for developing prevention strategies. For example, understanding the relevance of and preference for "dry sex" in different African communities has been an important consideration in developing vaginal microbicides.[33 34] Qualitative research has also been used to inform the design and development of quantitative research instruments and methods. Cognitive and in-depth interviewing have been used to inform the use of appropriate language in surveys and to identify factors which influence willingness to report such as privacy, sex of interviewers, and use of computer assisted self completion interviews.[29 32 35]

Sources of measurement error in sexual behaviour survey research

All epidemiological research aims to achieve accuracy in estimation. This requires minimising measurement error, which may occur at any stage of the survey from sample selection, to questionnaire content, design, and administration. Potential sources are discussed in detail below.

SAMPLING PROCEDURES

Many early sexual behaviour studies, including those of Kinsey,[36 37] relied on volunteer samples with little attempt to achieve representativeness of the demographic and behavioural characteristics of the target population. A number of studies have since shown that volunteers tend to be more sexually experienced, sensation seeking, and unconventional, and to have more relaxed sexual attitudes and behaviours than those randomly recruited from the general population.[38–40]

Random probability sampling methods can reduce volunteer bias by yielding unbiased samples of the target population. Commonly used sampling frames for general population surveys include electoral registers, postcode files, and telephone numbers; however, all may systematically underrepresent certain groups whose behaviours may differ from the general population. In many countries, no sampling frames of households, addresses, or individuals exist. A common strategy in these circumstances is to use a multistage clustered sampling technique in which census enumeration areas are first selected, all contained households listed, and then sampled. Homeless and prison populations are missed in most population samples, yet they have high prevalence of epidemiologically important behaviours such as injecting drug use or commercial sex.[41] Similarly, telephone samples often underrepresent young people and poorer populations.[42]

RESPONDENT VARIABLES

Survey non-response and representativeness

Achieving good response rates in sex survey research is essential to improve the representativeness of the survey and reduce participation bias (see below). Obtaining a representative sample increases our ability to make robust inferences about the source population—that is, to generalise survey findings. Generally, between 25–35% of people refuse to engage in telephone or face to face interviews designed to investigate sexual attitudes and lifestyles, and non-return rates of 40% in postal surveys of this nature are common.[38] However, others have argued that non-response rates are no greater for sex research than for other studies of sensitive issues, which would suggest that the *sexual* nature of the questionnaire does not necessarily bias the responses.[4 43] Survey non-response may become more problematic if

public interest in survey participation declines, particularly in studies perceived to be intrusive, sensitive, or of no immediate relevance. Reasons for non-participation vary but include non-contact with selected addressees, refusals in person or by proxy, respondent being ill or unable to speak the appropriate language. Methods that rely on high levels of literacy may also exclude groups particularly vulnerable to poor sexual health outcomes. Refusal to participate may occur at any stage of the interview but is most likely at the point of initial contact or invitation.[6] In the National AIDS Behavioral Survey, over 80% of refusals occurred before respondents heard that the survey concerned AIDS related issues.[44]

Participation bias

Participation bias describes error arising from systematic differences in the characteristics (for example, sexual behaviour) of those who agree to participate in a study compared with those who do not. Even in well designed studies, achieving response rates in excess of 80% may be difficult, although higher response rates are often achieved in developing countries.[2] Therefore participation bias has the potential to introduce significant error in measuring estimates of behavioural risk. Participation bias has been documented in a variety of sexual behaviour studies, and is associated with the respondents' characteristics (for example, sex, age, social class), beliefs, and sexual behaviour.[45] Clement[46] argues that the more intrusive a survey, the higher the barrier to intimacy, and the more likely we are to encounter participation bias that overestimates variability and frequency of sexual behaviour (since those with conservative or normative lifestyles are less likely to participate). However, Biggar and Melbye[47] found little difference in the sexual behaviour of those who responded early and late to a sexual behaviour survey, and Laumann *et al*[5] drew similar conclusions.

Item response bias is another type of participation bias in which respondents refusing to answer a particular question(s) are systematically more or less likely to have experience of the relevant behaviour. Copas *et al*[48] found older age, problems of comprehension, and ethnicity to be associated with refusal to complete more detailed and sensitive questions contained in a self completion booklet in the British NATSAL survey, but concluded that those who declined to answer the more intimate questions were, if anything, likely to be at lower HIV risk. Dunne *et al*[49] reached similar conclusions with a cohort study of twins, but concluded that the effect on most measures was small. In both cases, participation bias may have led to an overestimation of HIV risk behaviours which counteracts the observed tendency for survey respondents to minimise or underreport the frequency and diversity of their sexual behaviour.[39 42 50]

Reporting and recall bias

Sexual behaviour is most commonly studied using self reported recall of behaviours across some retrospective time frame. Even among respondents who attempt to "accurately" report their past behaviours, problems with recall can distort the reported incidence and frequency of specific behaviours.[6 42 51–53] Studies have found that the reliability of self reported sexual behaviour varies with a variety of factors including age,[54–56] ethnicity,[57] the number of sexual partners,[42] and the time frame for recall.[45] *Incidence reports* (for example, first sexual intercourse) are generally more reliably reported than *frequency reports* (for example, number of partners, frequency of sex). The reliability of frequency reports decreases with longer recall periods and more frequent behaviours (for example, vaginal sex).[54 58–60] Other reliability studies have found that recall of the number of partners tended to be less variable than the number of acts.[61–63] In general, longer recall intervals result in either underreporting or inaccurate recall of sexual practices and partners, because a more elaborate reconstruction of events rather than a simple scanning of more recent events is required.[64–66]

Sex related bias in self reported behaviours may also occur. In a closed population with a balanced sex ratio, men and women should report the same population mean number of partners over a defined period. However, men consistently report a higher mean number of partners in nearly all surveys.[67] Wadsworth *et al*[68] explored this relation in data from NATSAL and concluded that the discrepancy could be reduced but not eliminated by accounting for age mixing in partnership formation, underrepresentation of prostitutes, and modest assumptions about response bias introduced by lower response rates among men than women. Similarly, evidence from other surveys indicates that men and women may differ in what they count as "sex," with men more likely to include non-penetrative sex than women.[62 69] However, it is likely that there remains some social desirability bias in the direction of overreporting by men and/or underreporting by women.

Other examples of social desirability bias include the general tendency for women to underreport their premarital sexual experiences.[70 71] In the 1980s, Potterat[72] and Stoneburner *et al*[73] showed that HIV positive military personnel were initially more likely to report sexual encounters with prostitutes to be the source of infection than in later interviews with civilian counsellors when they were more willing to admit to homosexual exposure. Social desirability bias may also be influenced by data collection modes, with self completion modules typically eliciting higher rates of sensitive behaviours than face to face interviews (see below).

QUESTIONNAIRE DESIGN, CONTENT, AND DELIVERY

The design, content, and mode of administration of the survey questionnaire, whether by interviewer or self completed, may contribute

to measurement error. Pen and paper methods may exclude those with poor literacy, and long questionnaires may lead to poor data quality with missing data and inconsistent answers. Detailed behavioural surveys may require elaborate skip and filtering instructions, which are difficult to follow. Words that might be considered offensive and "big words" may lead to significant item non-response and, as the meanings and use of terms used in surveys vary across sexes and cultures, they should never be assumed. For example, Sanders and Reinisch[69] found that 60% of a sample of college students did not consider oral sex alone to be "having sex." Development work for NATSAL[31] encountered different assumptions about the nature of a "sexual partner." Some married respondents felt the term was too casual to refer to their married partner, while single respondents thought it implied a steady relationship rather than a casual encounter. A sexual partner was carefully defined to all respondents in NATSAL, as were all behaviours reported in the survey.

Although postal self completion surveys are less expensive, and may reach respondents in rural areas or who are hard to find at home, most studies have found response rates to be poorer on postal surveys than interviewer administered surveys, despite reminders.[42] While respondents have time to reflect on their answers, there is no motivational effect of the interviewer. Additionally, there is little control over how, in what order, or by whom the questionnaire is completed.

Face to face (and to some extent, telephone) contact with respondents is often used in sex survey research. Interviewers can explain the rationale and format of a survey directly, and they may have a motivating effect on the respondent, by providing full, clear definitions, probing ambiguous responses, or querying inconsistent answers.[74] However, interviewers can also introduce reporting bias, leading to reduced disclosure of socially proscribed attitudes or behaviours (even when done in coded fashion). Research has shown that people tend to report more sexual information to female interviewers, and that in this regard, women may be more influenced than men by interviewer sex differences.[52 67] Delamater[51] found that females were more likely to underreport proscribed behaviours to male interviewers than to female interviewers whereas Johnson and Delamater[75] found male interviewees with good rapport with the interviewer also reported more frequent sexual activity.

Assessing measurement error

RESPONSE RATES AND REPRESENTATIVES

Strategies for assessing the extent and magnitude of participation bias remain relatively undeveloped. Checking the overall study response rates provides some indication of the representativeness and the likely magnitude of participation bias in the survey. However, formal assessment of sample representativeness usually involves comparing demographic characteristics such as age, sex, socioeconomic group, and geographic location with census data or other large scale studies on less sensitive topics.[3 76] Data from probability sample surveys consistently suggest that non-responders are more likely to be male, older, urban residents, with lower educational attainment than responders, with no consistent relation being noted with marital status, occupational status, and ethnicity.[38 42 48] NATSAL obtained a 65% response rate and the achieved sample was broadly representative of the population of Great Britain aged 16–59 years. In common with other surveys, response rates were lower among men than women, and those least likely to respond were in the oldest age group. Parameter estimates could have been affected if recruited males were younger (therefore reporting more sexual activity) and if non-participation was related to sexual behaviours.

VALIDITY CHECKS

Validity describes the extent to which an instrument measures what it purports to measure. It is extremely difficult to determine the absolute validity of self reported sexual behaviours and therefore a number of indirect measures (internal and external) are used instead. External validation of reports may be achieved by using independent data sources as external references. For example, in NATSAL, self reported abortion showed a good approximation to national statutory reports, although there was some evidence of underreporting of STD clinic attendance.[3] Similarly, data from studies among high risk population subgroups may be triangulated for consistency with similar information on the overall spectrum of behaviour from general population surveys. Validation of survey results with those obtained from in-depth interviewing has also been used.[77]

Other methods of validation include interviewing the respondents and their sexual partners separately.[36] These reports may vary with the stability of the relationship, degree of substance abuse, type of sexual behaviour within the relationship, and time interval asked about.[6] Padian *et al* [78] found high levels of agreement in couples with one HIV infected partner on levels of frequency of sex, sex practices, and condom use. Others have found only fair agreement in couples attending STD clinics, which tends to decrease as recall periods increase.[79]

Biological methods using incident STIs or urinary testing for HIV, *Chlamydia trachomatis*, and pregnancy are being increasingly used to assess the validity of self reports. However further evaluation of this strategy is needed. Zenilman *et al* [80] in an STD clinic population, found similar levels of incident STI in "always" condom users to "never users" suggesting evidence of reporting bias (assuming high condom effectiveness in preventing STIs).

INTERNAL CONSISTENCY

The internal consistency of questionnaire responses, where responses to questions asked in one part of an individual's questionnaire are checked for logical agreement with related questions, may be used to assess the reliability

and validity of self reports. NATSAL[3] included 158 consistency checks, and around 80% of respondents had no inconsistencies. Where differences occurred in different parts of the interview, the most common inconsistencies were greater reporting of multiple heterosexual partners and of homosexual experiences in questions completed in a self completion booklet compared with those in face to face interviews.

TEST-RETEST RELIABILITY

Readministration of the same items after a brief time interval has been used to assess optimal recall time frames or the stability of responses (*test re-test reliability*)[42 58 64] and to compare different techniques for enhancing memory. This provides an index of the stability of people's estimates of their sexual behaviours over time. A variety of studies have examined the reliability of reports of a range of behaviours across different populations. Factors increasing reliability include age (adolescents have higher test-retest coefficients than adults), rarity of events, incidence reports compared with frequency reports, and shorter period of recall.[42 45 58 64] In 1990, Catania argued that existing test-retest data represented a "mixed bag" and called for studies which examine reliability for different reporting periods across specific sexual behaviours, in different population subgroups.

Reducing measurement error

IMPROVE SAMPLE DESIGN

In a probability sample survey, increasing the size of the study can reduce sampling error and increase study precision (thereby providing more robust parameter estimates). However, this must be balanced against increasing research costs. Stratifying the sample, or sorting the sampling frame before selection, ensures that the sample proportion from any particular stratum equals the population proportion. Variable sampling fractions can also be applied to increase the sample size of small groups of particular interest—for example, to achieve acceptable confidence intervals for estimates based on different ethnic or regional groups, and to increase the precision of estimates by oversampling more variable strata. Weighting can be applied to correct for different selection probabilities resulting from the use of variable sampling fractions or to control for random variations in the sample numbers across strata.

REDUCE PARTICIPATION BIAS

Any intervention that improves response rates will reduce participation bias. Respondent callbacks, re-invitations to participate, and postal reminders have been used to obtain interviews with the selected participant. Laumann *et al*[5] used incremental payments to encourage participation in those initially declining to participate. Interviewer characteristics and training, and the perceived public health importance of the survey topic may also influence response rates.[81] Methods that make the interview process less invasive or more private (for example, use of computer assisted self interviewing techniques) may reduce participation bias since embarrassment and worries about confidentiality, often of primary concern to participants, are reduced.

However, even if very high response rates were achieved, estimates of rarer behaviours remain sensitive to participation bias and there are no simple techniques to reduce their effect in analysis. If the demographic differences between the sample and the population are known then statistical weighting techniques can be used to adjust for differential non-response. Typically, results are weighted to the known demographic structure (age, marital status, region, etc) of the target population to provide population estimates. However, this method assumes that the prevalence of behaviours is the same as in responders (at least within demographic classes). It cannot overcome participation bias that arises independently of demographic factors. Alternatively, special studies with non-participants may be undertaken to characterise the magnitude of, and subsequently adjust for, participation bias.[38 48] A sensitivity analysis approach may then be employed to calculate and present parameter estimates, which take into account different assumptions of this (participation bias) effect.[48]

IMPROVE QUESTIONNAIRE DESIGN AND CONTENT

The terms used to describe or investigate sexual behaviour may influence respondents' willingness to participate in the study or to provide accurate and reliable answers. Items should be specific, clear, and use defined time periods to inquire about sexual behaviour. They should also avoid acquiescence bias (implying a "mid point" or "norm") and undue embarrassment.[82]

Using appropriate and comprehensible language and terminology is important. Binson and Catania[83] state that one approach to establishing appropriate language is to ask each respondent to select the sexual terminology they would prefer the interviewer to use.[36 37 74] This technique has been shown to elicit higher reporting of sensitive behaviours[83]; however, tailoring language to each respondent is less feasible on a large scale, heterogeneous, general population sample. It also places demands on the interviewer, and may create problems in quantifying precise and standardised behaviours. Spencer *et al*[31] also found general population respondents felt awkward about providing their own definitions for sexual practices. While colloquial or street language has been found suitable for specific populations, such as bar attending homosexual men, drug users, and prostitutes, general population surveys have tended towards the formal. NATSAL development work found a strong preference for "formal rather than street language"[31] and ACSF used "technical anatomical terms."[50]

Finally, care in the ordering of questions is also important. Spencer *et al*[31] found that both interviewers and respondents preferred the questionnaire to begin with neutral questions,

leading in to more intimate and sensitive ones once rapport had been developed. General questions also provided a "contextual framework" into which life events could be situated to aid recall. However, beginning with first sexual experiences may be particularly sensitive if the age was perceived by the respondent to be very early or late, or involved abuse. In NATSAL[3] and the American NHSLS,[5] attitude questions are asked towards the end of the interview and after the sexual behaviour questions to avoid possible reinforcement of social norms in reporting on partners and practices.

TELEPHONE INTERVIEWING

Telephone surveys have gained increasing popularity over the past two decades and are a mainstay of market oriented research. Telephone interviews were used for the French (ACSF),[4] other national sex surveys and others.[62 84–86] Telephone interviewing allows for an unclustered sample at a lower cost than could be achieved face to face. It allows faster data collection, greater control over and monitoring of the interview process. However, telephone interviews need to be shorter, require simple questions, and do not allow the use of show cards or long lists. It may also be more difficult to guarantee privacy as other household members may be listening in. Nevertheless, in the French survey, Bajos and Spira[87] compared telephone interviewing and face to face interviewing with pen and paper self completion and found that questions were "more easily answered" and answers were more coherent in the telephone study. New systems are available for both private call-in and call-out telephone interviews. With a call-in system, respondents telephone a live interviewer; with call-out, live interviewers screen households and recruit participants. Some of the questionnaire is administered directly, with respondents transferred to an automated system for the sensitive sections.

SELF COMPLETION QUESTIONNAIRES

Self completion questionnaires reduce the need for respondents to disclose sensitive behaviours to the interviewer and may result in more valid reports than interviews.[6] Paper self completions should be simple and short with limited filtering and few open ended questions. Combinations of pen and paper self completion and interviewer techniques have been used in many of the large surveys and combine the benefits of face to face interviewing with the privacy of self completion for more sensitive questions. Johnson *et al*[3] reported increased disclosure of censured behaviours (for example, homosexual experience) in self completion compared with face to face questioning. Davoli *et al*[88] reported good correlation between self completion and face to face interviews among Italian adolescents for reported coital experience and age at first intercourse; however, interviews underreported coitus and overreported condom use when administered before the questionnaire. Despite good reproducibility, social desirability bias had occurred.

COMPUTER ASSISTED INTERVIEWS

In the past decade there have been major developments in the use of technologies for undertaking computer assisted personal interviews (CAPI) and self completion interviews. Face to face and telephone interviews are undertaken with responses keyed directly into computers by interviewers. Computer assisted self interviews (CASI) are increasingly being used where the respondents key their response to questions on the screen directly into a laptop computer. These methods are well suited to complex questionnaires since skips and routing can be automatically programmed without respondents having to follow complex instructions on paper.

In audio-CASI, respondents listen to prerecorded questions on headphones and key in appropriate responses. All respondents can hear the same standardised delivery of questions (with voice quality, not computer generated words). Audio-CASI helps overcome literacy problems and can provide prerecorded questionnaires in different languages and can also be used for telephone interviews. In comparing CAPI, CASI, and audio-CASI, Tourangeau and Smith[89] found audio-CASI elicited highest mean number of reported partners and highest reporting of anal sex. They found that respondents felt a greater sense of privacy, that CASI gave the study an air of "legitimate and scientific value," and that audio input (whether on face to face or audio-CASI) facilitated comprehension. Des Jarlais *et al*[90] assessed audio-CASI as a method of reducing underreporting of HIV risk behaviour among injecting drug users and noted significantly increased reporting of HIV risk and sensitive behaviours, such as borrowing or renting used injecting equipment, in audio-CASI than in face to face interviews.

Studies comparing CASI with identical questions using pen and paper self completion have demonstrated the potential of CASI to improve the quality of data, and to increase respondents' willingness to report sensitive behaviours.[91 92] Turner *et al*[92] reported significant audio-CASI effects for the reporting of several sensitive behaviours. However, their sample was restricted to adolescent males, many from disadvantaged backgrounds, and the study used audio-CASI to get over potential literacy problems in this group. Johnson *et al*,[10] in a methodological experiment in a British general population sample, found no consistent evidence of increased reporting of risk behaviour when comparing CASI with pen and paper self completion, although item response and data consistency were improved using CASI. Method effects may be related to the degree of perceived social censure of particular behaviours and these vary between cultures and demographic groups.

SEXUAL DIARIES

Sexual diaries have been proposed as a means of improving reliability of reported behaviours. If kept regularly they can allow prospective collection of data and minimise problems

associated with long term recall.[42] Verbal diaries, regularly collected by an interviewer, have also been used with poorly literate respondents. This may be particularly useful given that recall of sexual partners is more likely to be cited as a difficulty by the most sexually active respondents, and that infrequent practices are easier to remember than frequent ones.[93] In a study among commercial sex workers, Ramjee *et al*[94] found a significantly greater mean number of clients, condoms used, vaginal acts and anal acts reported in diary format compared with recall questionnaire. While McLaws *et al*[93] found most respondents preferred using the diary to the recall questionnaire, their sample of homosexual men, like Coxon's,[95] may have been particularly well motivated. The burden of a regular diary may be too time consuming a task to expect of most respondents, and measuring behaviours may in turn produce changes in the behaviour being measured (monitoring effects). Consequently McLaws concluded that data collected by recall were, in fact, more consistently reliable than data collected by diary.[93]

Conclusions

Reliable data on sexual behaviour remain difficult to collect. Nevertheless, many of the methodological challenges of sexual behaviour research are common to other areas of self reported behaviour including diet, smoking, and alcohol consumption. Improvements in social research methods provide a number of strategies for reducing measurement error. Computer assisted techniques, by improving internal consistency and increasing privacy and interviewee control, offer exciting possibilities for improving survey validity. So too does our increasing ability to triangulate survey results with focused qualitative investigations and a variety of social research and surveillance data. Increasingly available non-invasive diagnostic techniques provide biological outcome measures, which in turn offer new opportunities for studying the relation between behaviours and STI epidemiology.

Continued methodological research is needed to better identify the sources and magnitude of measurement error. Achieving high response rates in population based studies remains a challenge, despite technological developments, increasing public discourse about sex, and greater awareness of sexual health matters. In many developed countries, this is further compounded by a reduction in the perceived threat posed by the HIV/AIDS epidemic, undoubtedly a stimulant for much progress over the past two decades. As a result, waning public interest and changing political prioritisation can only serve to increase these difficulties. Spiralling research costs mean that large scale studies of sexual behaviour are becoming less attractive to policy makers. Cost effective and robust strategies for monitoring sexual behaviour are required, and behavioural surveillance programmes (ongoing population based prospective monitoring of sexual behaviour) are urgently needed. A potential way to develop this surveillance in the United Kingdom and elsewhere may involve adding a small module of key sexual behaviour questions to other routine surveys (for example, general health surveys). Such surveillance programmes would not obviate the need for targeted or in-depth studies of sexual behaviours but would, in concert, continue to increase our understanding of disease epidemiology and strategies to promote sexual health.

1 Catania JA, Moskowitz JT, Ruiz M, *et al.* A review of national AIDS-related behavioral surveys. *AIDS* 1996;**10**:S183–90.
2 Cleland J, Ferry B. *Sexual behaviour and AIDS in the developing world.* London: Taylor Francis, 1995.
3 Johnson AM, Wadsworth J, Wellings K, *et al. Sexual attitudes and lifestyles.* Oxford: Blackwell Scientific Press, 1994.
4 ASCF principal investigators and their associates. Analysis of sexual behaviour in France (ACSF). A comparison between two modes of investigation: telephone survey and face-to-face survey. *AIDS* 1992;**6**:315–23.
5 Laumann EO, Gagnon JH, Michael RT, *et al. The social organization of sexuality: sexual practices in the United States.* Chicago and London: University of Chicago Press, 1994.
6 Catania JA, Binson D, Van der Straten A, *et al.* Methodological research on sexual behavior in the AIDS era. *Annual Review of Sex Research* 1995;**6**:77–125.
7 McLaws M, Oldenburg B, Ross MW, *et al.* Sexual behaviour in AIDS-related research: reliability and validity of recall and diary measures. *Journal of Sex Research* 1990;**27**:265–81.
8 Hausser D, Zimmerman E, Dubois-Arber F, *et al.* Evaluation of the AIDS Prevention Strategy in Switzerland: third assessment report, 1989–1990. Lausanne: Institut Universitaire de Medicine Sociale et Preventive, 1991.
9 Herlitz CA, Steel JL. A decade of HIV/AIDS prevention in Sweden: changes in attitudes associated with HIV and sexual risk behaviour from 1987 to 1997. *AIDS* 2000;**14**: 881–90.
10 Johnson AM, Copas A, Field J, *et al.* Do computerised self-completion interviews influence the reporting of sexual behaviours? A methodological experiment. Proceedings of the Thirteenth Meeting of the International Society for Sexually Transmitted Diseases Research, Denver Colorado, 11–14 July, 1999.
11 Smith TW. A methodological analysis of the sexual behaviour questions on the general social surveys. *Journal of Official Statistics* 1992;**8**:309–25.
12 World Health Organization and Joint United Nations Programme on HIV/AIDS. *Second generation surveillance for HIV: the next decade.* Geneva: WHO, 2000.
13 Schneider J, King J, Macnab GM, *et al.* Hepatitis-B surface antigen and antibody in black and white patients with venereal diseases. *Br J Vener Dis* 1977;**53**:372–4.
14 Nardone A, Dodds JP, Mercey DE, *et al.* Active surveillance of sexual behaviour among homosexual men in London. *Communicable Disease and Public Health* 1998;**1**: 197–201.
15 Choi K-H, Catania JA. Changes in multiple sexual partnerships, HIV testing and condom use among US heterosexuals 18 to 49 years of age, 1990 and 1992. *Am J Public Health* 1996;**86**:554–6.
16 Adib SM, Joseph JG, Ostrow DG, *et al.* Relapse in sexual behavior among homosexual men: a 2-year follow-up from the Chicago MACS/CCS. *AIDS* 1991;**5**:757–60.
17 Hutchinson GA, Simeon DT. HIV infection rates and associated factors in high risk patients admitted to a psychiatric hospital in Trinidad and Tobago. *West Indian Med J* 1999; **48**:129–31.
18 Loveday C, Kaye S, Tenant-Flowers M, *et al.* Viral load rebound and viral resistance are independent of each other in a cohort of patients with HIV disease on open zidovudine therapy. *Lancet* 1995;**345**:820–4.
19 Ng'weshemi JZL, Ties Boerma J, Pool R, *et al.* Changes in male sexual behaviour in response to AIDS epidemic: evidence from a cohort study in urban Tanzania. *AIDS* 1996; **10**:1415–20.
20 Dodds JP, Nardone A, Mercey DE, *et al.* Increase in high risk sexual behaviour among homosexual men, London 1996–8: cross sectional, questionnaire study. *BMJ* 2000; **320**:1510–1.
21 Weatherburn P, Stephens M, Reid D, *et al.* Vital statistics. *Findings from the national gay men's sex survey 1999.* London: Sigma Research , 2000, chaps 1–70.
22 Carael M, Van De Perre PH, Lepage PH, *et al.* Human immunodeficiency virus transmission among heterosexual couples in central Africa. *AIDS* 1988;**2**:201–5.
23 De Vincenzi I, for the European Study Group on Heterosexual Transmission of HIV. A longitudinal study of human immunodeficiency virus transmission by heterosexual partners. *N Engl J Med* 1994;**331**:341–6.
24 Ghani A, Ison CA, Ward H, *et al.* Sexual partner networks in the transmission of sexually transmitted diseases. An analysis of gonorrhoea cases in Sheffield, UK. *Sex Transm Dis* 1996;**23**:498–503.

25 Day S, Ward H, Ghani A, *et al.* Sexual histories, partnerships and networks associated with the transmission of gonorrhoea. *Int J STD AIDS* 1998;**9**:666–71.
26 Ghani A, Swinton J, Garnett GP. The role of sexual partnership networks in the epidemiology of gonorrhoea. *Sex Transm Dis* 1997;**24**:1–12.
27 Rothenberg RR, Potterat JJ, Woodhouse DE. Personal risk taking and the spread of disease: beyond core groups. *J Infect Dis* 1996;**174**:S144–9.
28 Power R. The role of qualitative research in HIV/AIDS [editorial]. *AIDS* 1998;**12**:687–95.
29 McKusick L, Horstman W, Coates TJ. AIDS and sexual behavior reported by gay men in San Francisco. *Am J Public Health* 1985;**75**:493–6.
30 Parker RG, Carballo M. Qualitative research on homosexual and bisexual behavior relevant to HIV/AIDS. *J Sex Res* 1990;**27**:497–525.
31 Spencer L, Faulkner A, Keegan J. *Talking about sex*. London: SCPR, 1988.
32 Elam G, Fenton K, Johnson A, *et al. Exploring ethnicity and sexual health.* London: Social and Community Planning Research, 1999:1–116.
33 Civic D, Wilson D. Dry sex in Zimbabwe and implications for condom use. *Soc Sci Med* 1996;**42**:91–8.
34 Pool R, Whitworth JA, Green G, *et al.* An acceptability study of female-controlled methods of protection against HIV and STDs in south-western Uganda. *Int J STD AIDS* 2000;**11**:162–7.
35 Mitchell K, Wellings K, Elam G. NSSAL II survey design and implementation report of the qualitative stage. Report to Medical Research Council. London: MRC, June, 1998: 1–40.
36 Ostrow DG, Vanable PA, McKirnan DJ, *et al.* Hepatitis and HIV risk among drug-using men who have sex with men: demonstration of Hart's law of inverse access and application to HIV. *J Gay Lesbian Med Assoc* 2000;**3/4**:136.
37 Seal DW, Kelly JA, Bloom FR, *et al.* HIV prevention with young men who have sex with men: what young men themselves say is needed. *Aids Care Psychol Socio Med Asp AIDS HIV* 2000;**12/1**:26.
38 Dunne MP, Martin NG, Bailey JM, *et al.* Participation bias in a sexuality survey: psychological and behavioural characteristics of responders and non-responders. *Int J Epidem* 1997;**26**:844–54.
39 Strassberg D, Lowe K. Volunteer bias in sexuality research. *Arch Sex Behav* 1995;**24**:369–82.
40 Catania J, McDermott L, Pollack L. Questionnaire response bias and face to face interiew sample bias in sexuality research. *J Sex Res* 1986;**22**:52–72.
41 Bird AG, Gore SM, Burns SM, *et al.* Study of infection with HIV and related risk factors in young offenders' institution. *BMJ* 1993;**307**:228–31.
42 Catania JA, Gibson DR, Chitwood DD, *et al.* Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull* 1990;**108**:339–62.
43 Biggar RJ, Brinton LA, Rosenthal MD. Trends in the number of sexual partners among American women. *J Acquir Immune Defic Syndr* 1989;**2**:497–502.
44 Catania JA, Coates TJ, Stall R, *et al.* Prevalence of AIDS-related risk factors and condom use in the United States. *Science* 1992;**258**:1101–6.
45 Catania JA, Gibson DR, Marin B, *et al.* Response bias in assessing sexual behaviors relevant to HIV transmission. *Evaluation and Program Planning* 1990;**13**:19–29.
46 Clement U. Surveys of heterosexual behaviour. *Annual Review of Sex Research* 1990;**1**:45–74.
47 Biggar RJ, Melbye M. Responses to anonymous questionnaires concerning sexual behavior: A method to examine potential biases. *Am J Public Health* 1992;**82**:1–7.
48 Copas A, Johnson AM, Wadsworth J. Assessing participation bias in a sexual behaviour survey: implications for measuring HIV risk. *AIDS* 1997;**11**:783–90.
49 Dunne MP, Martin NG, Bailey JM, *et al.* Participation bias in a sexuality survey: Psychological and behavioural characteristics of responders and non-responders. *Int J Epidem* 1997;**26**:844–53.
50 Horner RD, Oddone EZ, Matchar DB. Theories explaining racial differences in the utilization of diagnostic and therapeutic procedures for cerebrovascular disease. *Millbank Q* 1995;**73**:443–62.
51 James NJ, Bignell CJ, Gillies PA. The reliability of self reported sexual behaviour. *AIDS* 1991;**5**:333–6.
52 Delamater J. Methodological issues in the study of pre-marital sexuality. *Sociol Methods Res* 1974;**3**:30–61.
53 Anderson B, Broffitt B. Is there a reliable and valid self-report measure of sexual behaviour? *Arch Sex Behav* 1988;**17**:509–25.
54 McFarlane M, St Lawrence JS. Adolescents' recall of sexual behaviour: consistency of self-report and effect of variations in recall duration. *J Adolesc Health* 1999;**25**: 199–206.
55 Clark LR, Brasseux C, Richmond D, *et al.* Are adolescents accurate in self-report of frequencies of sexually transmitted diseases and pregnancies? *J Adolesc Health* 1997;**21**:91–6.
56 Capaldi D. The reliability of retrospective report for timining first sexual intercourse for adolescent males. *J Adolesc Health* 1999;**11**:375–87.
57 Rogers J. The rescission of behaviours: inconsistent responses in adolescent sexuality data. *Soc Sci Res* 1982;**11**: 280–96.
58 Saltzman SP, Stoddard AM, McCusker J, *et al.* Reliability of self-reported sexual behavior risk factors for HIV infection in homosexual men. *Public Health Rep* 1987;**102**: 692–7.
59 Bellak AS, Herson M. Self-report inventories in behavioural assessment. In: Cone JD, Hakins RP, eds. *Behavioral assessment: new directions in clinical psychology.* New York: Brunner Mazel, 1977.
60 Reading A. A comparison of the accuracy and reactivity of methods for monitoring male sexual behaviour. *J Sex Res* 1983;**36**:11–23.
61 Becker S, Begum S. Reliability study of reporting of days since last sexual intercourse in Matlab, Bangladesh. *J Biosoc Sci* 1994;**26**:291–9.
62 Jeannin A, Konings E, Dubois-Arber F, *et al.* Validity and reliability in reporting sexual partners and condom use in a swiss population survey. *Eur J Epidemiol* 1998;**14**: 139–46.
63 Van Duynhoven YTHP, Nagelkerke NJD, Van De Laar MJW. Reliability of self-reported sexual histories: test-retest and interpartner comparison in a sexually transmitted diseases clinic. *Sex Transm Dis* 1999;**26**:33–42.
64 Blake SM, Sharp SS, Temoshok L, *et al.* Methodological considerations in developing measures of HIV risk-relevant behaviours and attitudes; an empirical illustration. *Psychology and Health* 1992;**6**:265–80.
65 Coates RA, Soskolne CL, Calzavara L, *et al.* The reliability of sexual histories in AIDS-related research: evaluation of an interview-administered questionnaire. *Can J Public Health* 1986;**77**:343–8.
66 Kauth MR, St Lawrence JS, Kelly JA. Reliability of retrospective assessments of sexual HIV risk behavior: a comparison of biweekly, three-month, and twelve-month self-reports. *AIDS Educ Prev* 1991;**3**:207–14.
67 Catania JA, Binson D, Canchola J, *et al.* Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behaviour. *Public Opin Q* 1996;**60**:345–75.
68 Wadsworth J, Johnson AM, Wellings K, *et al.* What's in a mean?—an examination of the inconsistency between men and women reporting sexual partnerships. *J Roy Stat Soc* 1996;**159**:111–23.
69 Sanders SA, Reinisch JM. Would you say you "had sex" if...? *JAMA* 1999;**281**:275–7.
70 Abraham HD, Degli ES, Marino L. Seroprevalence of hepatitis C in a sample of middle class substance abusers. *J Addict Dis* 2000;**18**:77–87.
71 Appleby J. Poor control of AIDS spending. *BMJ* 1990;**301**: 509.
72 Potterat JJ, Phillips L, Muth JB. Lying to military physicians about risk factors for HIV infections. *JAMA* 1987;**257**: 1727.
73 Stoneburner RL, Chiasson MA, Solomon K, *et al.* Risk factors in military recruits positive for HIV antibody. *N Engl J Med* 1986;**315**:1355.
74 Gribble JN, Miller HG, Rogers SM, *et al.* Interview mode and measurement of sexual behaviors: methodological issues. *J Sex Res* 1999;**36**:16–24.
75 Johnson W, Delamater J. Response effects in sex surveys. *Public Opin Q* 1976;**40**:165–81.
76 Michael R, Ganon J, Laumann E, *et al. Sex in America: a definitive survey.* Boston: Little, Brown, 1994.
77 Konings E, Bantebya G, Carael M, *et al.* Validating population surveys for the measurement of HIV/STD prevention indicators. *AIDS* 1995;**9**:375–82.
78 Padian NS, Shiboski SC, Glass SO, *et al.* Heterosexual transmission of human immunodeficiency virus (HIV) in Northern California: results from a ten year study. *Am J Epidemiol* 1997;**146**:350–7.
79 Ellish NJ, Weisman CS, Celentano D, *et al.* Reliability of partner reports of sexual history in a heterosexual population at a sexually transmitted diseases clinic. *Sex Transm Dis* 1996;**23**:446–52.
80 Zenilman JM, Weisman CS, Rompalo AM, *et al.* Condom use to prevent incident STDS: the validity of self-reported condom use. *Sex Transm Dis* 1995;**22**:15–21.
81 Herlitz CA, Steel JL. A decade of HIV/AIDS prevention in Sweden: changes in attitudes associated with HIV and sexual risk behaviour from 1987 to 1997. *AIDS* 2000;**14**: 881–90.
82 Martin J, Ashworth K, Heath A. "Question wording effects on the British General Election Studies" Joint Centre for Survey Methods and Joint Unit for the Study of Social Trends, Occasional Paper, Number 1. 1991.
83 Binson D, Catania JA. Respondents understanding of the words used in sexual behavior questions. *Public Opin Q* 1998;**62**:190–208.
84 Bray F, Chapman S. Community knowledge, attitudes and media recall about AIDS, Sydney 1988 and 1989. *Aust J Public Health* 1991;**15**:107–13.
85 Paul C, Dickson N, Davis PB, *et al.* Heterosexual behaviour and HIV risk in New Zealand: data from a national survey. *Aust J Public Health* 1995;**19**:13–8.
86 Davis PB, Yee RL, Chetwynd J, *et al.* The New Zealand partner relations survey: methodological results of a national telephone survey. *AIDS* 1993;**7**:1509–16.
87 Bajos N, Spira A. Analysis of sexual behaviour in France (ACSF). A comparison between two modes of investigation: telephone survey and face-to-face survey. *AIDS* 1992;**6**:315–23.

88 Davoli M, Perucci CA, Sangalli M, *et al.* Reliability of sexual behavior data among high school students in Rome. *Epidemiology* 1992;**3**:531–5.
89 Tourangeau R, Smith TW. Asking sensitive questions. The impact of data collection mode, question format, and question context. *Public Opin Q* 1996;**60**:275–304.
90 Des Jarlais DC, Paone D, Milliken J, *et al.* Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: a quasi-randomised trial. *Lancet* 1999;**353**:1657–61.
91 Weeks MF. Computer assisted survey information collection: a review of CASIC methods and their implications for survey operations. *Journal of Official Statistics* 1992;**8**:445–65.
92 Turner CF, Ku L, Rogers SM, *et al.* Adolescent sexual behaviour, drug use and violence: increased reporting with computer survey technology. *Science* 1998;**280**:867–73.
93 McLaws M, Oldenburg B, Ross MW, *et al.* Sexual behaviour in AIDS-related research: reliability and validity of recall and diary measures. *J Sex Res* 1990;**27**:265–81.
94 Ramjee G, Weber AE, Morar NS. Recording sexual behaviour: comparison of recall questionnaires with a coital diary. *Sex Transm Dis* 1999;**26**:374–80.
95 Coxon APM. Parallel acocunts? Discrepancies between self-report (diary) and recall (questionnaire) measures of the same sexual behaviour. *AIDS Care* 1999;**11**:221–34.