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. 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 censored attitudes or behaviours. 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, 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 and Sweden 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) 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 indicators or modules) may be added to probability sample general social surveys.

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, 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 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. Cohort studies enable estimation of disease incidence and monitoring of behavioural risk over time. 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
alternative mechanism for prospective behaviour monitoring. In London, annual surveys of homosexual men in social venues, STD clinics, and Gay Pride events 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. 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 and to determine the role of sexual networks in maintaining epidemic transmission. These studies have highlighted the importance of “core groups” 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. Studies exploring the social context of sexual behaviour—for example, the importance of San Francisco “bath houses” where homosexual men had large numbers of anonymous sexual contacts, were key to understanding the early evolution of the AIDS epidemic. Qualitative research has enabled the exploration of concepts within communities 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. 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.

**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, 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.

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. Similarly, telephone samples often underrepresent young people and poorer populations.

**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. 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. 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. 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. 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 50 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 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 50 drew similar conclusions.

**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. 46, 49, 51–53, 55 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, 58 and the time frame for recall. 59 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. 51–63 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. 54–66, 68

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 58 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. 52–69 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 Reinsch found that 60% of a sample of college students did not consider oral sex alone to be “having sex.” Development work for NATSAL 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. 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. 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 to report more sexual information to female interviewer. Additionally, there is little control over how, in what order, or by whom the questionnaire is completed.

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. 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.

Other methods of validation include interviews with the respondents and their sexual partners separately. 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. Padian et al 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.

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 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 of parameter estimates could have been a problem in one part of an individual’s questionnaire are checked for logical agreement with related topics. 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. 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.

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. 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. 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.

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 of parameter estimates could have been a problem in one part of an individual’s questionnaire are checked for logical agreement with related topics. 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. 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.

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 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-retest reliability). In order 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.

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. 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. 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. 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.

**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. Using appropriate and comprehensible language and terminology is important. Binson and Catania 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. This technique has been shown to elicit higher reporting of sensitive behaviours; 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. 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” and ACSF used “technical anatomical terms.”

Finally, care in the ordering of questions is also important. Spencer et al. 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 and the American NHSLS, 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), other national sex surveys and others. 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 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. 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. reported increased disclosure of censured behaviours (for example, homosexual experience) in self completion compared with face to face questioning. Davoli et al. reported good correlation between self completion and face to face interviews among Italian adolescents for reported coital experiences 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 (CASE) 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, Toumande and Smith 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. 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. Turner et al. reported significant audio-CASI effects for the reporting of several sensitive behaviours. 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. 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.\textsuperscript{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.\textsuperscript{43} In a study among commercial sex workers, Ramjee \textit{et al.}\textsuperscript{24} 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 \textit{et al.}\textsuperscript{25} found most respondents preferred using the diary to the recall questionnaire, their sample of homosexual men, like Coxon’s,\textsuperscript{19} 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.\textsuperscript{26}

\section*{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.

\begin{thebibliography}{99}
\end{thebibliography}
Measuring sexual behaviour

52 Delamater J. Methodological issues in the study of
51 James NJ, Bignell CJ, Gillies PA. The reliability of self
45 Catania JA, Gibson DR, Marin B, et al.
44 Catania JA, Coates TJ, Stall R, et al.
39 Strassberg D, Lowe K. Volunteer bias in sexuality research.
35 Mitchell K, Wellings K, Elam G. NSSAL II survey design
31 Spencer L, Faulkner A, Keegan J. et al.
30 Parker RG, Carballo M. Qualitative research on homosexual
28 Power R. The role of qualitative research in HIV/AIDS [edi-
27 Rothenberg RR, Potterat JJ, Woodhouse DE. Personal risk
8 Martin J, Ashworth K, Heath A. “Question wording e
7 Strassberg DS, Kelly JA, Bloom FR, et al.
6 Coates RA, Shiboski SC, Glass SO, et al.
5 Paul C, Dickson N, Davis PB, et al.
4 Zenilman JM, Weisman CS, Rompalo AM, et al.
3 Abraham HD, Degli ES, Marino L. Seroprevalence of hepa-
1 Becker S, Begum S. Reliability study of reporting of days

24 Catania JA, Binson D, Canchola J, et al.
22 Birkmeyer JD, Smith RA, et al. Study of infection with HIV
21 Ostrow DG, Vanable PA, McKirnan DJ, et al. Hepatitis and
20 Seal DW, Kelly JA, Bloom FR, et al.
19 Kaulth MR, St Lawrence JS, Kelly JA. Reliability of retrospective
17 Catania JA, Binson D, Canchola J, et al.
12 Biggar RJ, Melbye M. Responses to anonymous question-
4 Capaldi D. The reliability of retrospective reports for tim-
1 Bellak AS, Herson M. Self-report inventories in behavioural

www.texttransinf.com

Sex Transm Infect: first published as 10.1136/sti.77.2.84 on 1 April 2001. Downloaded from http://sti.bmj.com/ on November 10, 2023 by guest. Protected by copyright.