



## Parallel accounts? Discrepancies between self-report (diary) and recall (questionnaire) measures of the same sexual behaviour

A. P. M. COXON

*Principal Investigator, Project SIGMA, University of Essex, Professor, Health and Social Services Institute, University of Essex, Colchester, UK*

---

**Abstract** *Questionnaires and diaries have complementary biases and advantages for obtaining information on sexual behaviour but self-completed sexual diaries have the advantage of reducing retrospective bias. In a validation study of homosexual behaviour, sexual diary counts and subsequent questionnaire estimates (together with ratings of the certainty of the estimates) referring to the same month are compared and the discrepancies analyzed. Main findings include: questionnaire data yield consistently higher average estimates than diary counts, but have the same ordinal profile; individual difference (diary-questionnaire) scores show that 55% of questionnaire estimates of acts are higher than diary counts, 20% are identical and 25% are under-estimates; discrepancies are differentially located in different sexual acts. Masturbation and fellatio are systematically over-estimated in questionnaires and anal intercourse without a condom is the major source of inaccuracies, but in different directions: active partners under-estimate and passive partners over-estimate the amount of highest-risk sex. A strategy of joint use is discussed.*

### Introduction

In recent overviews of AIDS and sexuality research, the measurement of sexual behaviour is still seen as the foremost methodological problem confronting not only research on AIDS, but on sexually transmitted infections generally (Boulton, 1994; Catania *et al.*, 1990; 1993; Fortenberry, *et al.*, 1997; Ostrow & Kessler, 1993; Turner *et al.*, 1997). For reasons of practicality and cost, it is inevitable that survey methods form the main mechanism for obtaining reports on sexual behaviour, though the quality, accuracy and validity of the resulting survey data for this purpose have been subject to considerable questioning (see Miller *et al.*, 1990; Johnson *et al.*, 1994, for the USA and Britain, respectively). The main grounds for disputing the accuracy of survey-based estimates of sexual activity are the reactivity of the interview situation (due both to the impact of interviewer bias and to the socially stigmatized nature of the subject matter) and also to the unreliability and distortion of retrospective recall on the part of the subjects. It is argued that these biases are liable to lead to systematic under-estimation of rates of relevant risk activity in the general population

---

Address for correspondence: Professor A. P. M. Coxon, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK. Tel: 01206 872622; Fax: 01206 873410; E-mail: apmc@esscx.ac.uk

(McQueen *et al.*, 1989; Turner *et al.*, 1995) and to inaccurate and cognitively distorted estimates of sexual activity in general (Brewer, 1988; Linton, 1986).

The issue of reactivity can to a considerable extent be overcome by removing the data collection from the face-to-face interview and relying either on an interaction which is not visual (e.g. computer-assisted telephone interviewing (McQueen *et al.*, 1989; Tieleman, 1990, or audi- or telephone-audio computer assisted self-interviewing (O'Reilly *et al.*, 1994; Turner *et al.*, 1997)) or adopting the simpler expedient of using the self-administered questionnaire (SAQ), which in both sexual behaviour studies and in cognate areas such as drug-taking, has been shown to be a less biased method (Sundet *et al.*, 1990; Turner *et al.*, 1992).

Cognitive biases also come into play in most survey methods, especially when subjects are required to count instances of their behaviour, primarily due to factors of memory and recall. How accurately can subjects recall events when called upon to do so some time after their occurrence and what biasing processes come into play when subjects make estimates of rates or frequencies of occurrence? Retrospection has been paid especial attention, both from the point of view of systematic processes of bias in autobiographical recall—forgetting, telescoping, chunking (Croyle & Loftus, 1993; Forsyth *et al.*, 1992; Neisser & Winograd, 1988; Rubin, 1986), together with strategies of aggregation such as rounding and 'grossing up' (Coxon, 1988a). Others have shown how these biases show themselves in behavioural research (Coxon, 1988b; Janson, 1990; Weinhardt *et al.*, 1998; Wu *et al.*, 1988). Retrospection bias is probably the greatest threat to accuracy in sexual behaviour data. But it is not entirely clear how such bias may best be combated, and what other methods might be appropriate for reducing it.

The task of improving the accuracy of measures may suggest that there is some objective 'gold standard' which acts as criterion for assessment of their validity. Ideally, but not realistically, the appropriate *tertium quis* would appear to be the direct (and not laboratory-based) observation of the sexual behaviour, but even this is problematic<sup>1</sup> and yet without this there can be no *unquestioned* criterion of validity. Failing this, the notions of relative precision of a measure and the concordance or convergence of several measures may serve as alternatives. The diary method has been cited as a suitable alternative measure for sexual behaviour in several reviews: Catania *et al.* (1990, p. 174) cite the greater precision of diaries and Fortenberry *et al.* (1997, p. 238) suggest that the concordance of estimates from SAQs and coital diaries would be sufficient to validate each method, and allow the strengths of each to be capitalized upon.

Diaries are used increasingly in social science generally (Butcher & Eldridge, 1990; Corti, 1993; Richardson, 1994; Verbrugge, 1980) and especially in sexual research (Coxon, 1996a; Coxon & Coxon, 1993b; Graham & Bancroft, 1997). Although diaries are subject to a number of shortcomings, especially potential selection bias and difficulties in use for large populations, the advantages of diaries are also very considerable (Catania *et al.*, 1990; Coxon, 1988a). Most importantly, the short time period between the sexual act and its recording makes for greater precision and minimizes retrospection error (Bancroft, 1997; Berk *et al.*, 1995; Coxon, 1992; 1994; Dex, 1991; Freeman *et al.*, 1996; Hilton, 1989; James *et al.*, 1991; Juster, 1985; Kalfs & Saris, 1990; McLaws *et al.*, 1990; Poikolainen & Karkkainen, 1993; Reading, 1983; Robinson, 1985), and diaries also allow more detailed, contextually-specific and naturally expressed accounts of behaviour. It would appear, then, that if it were possible to show convergence between two self-administered methods, the SAQ and the self-administered Sexual Diary (SASD), assessment of sexual behaviour rates would be more firmly, accurately and validly demonstrated.

But the direct comparison of SAQs and SASDs is no simple matter, and the

few studies which have attempted to show convergence of diary and survey estimates (Fortenberry *et al.*, 1997; Gold & Rosenthal, 1995; McLaws *et al.*, 1990; Phellas, 1994; Poikolainen & Karkkainen, 1993) have tended to suffer from methodologically serious shortcomings of one sort or another, and in particular (i) that the data referred to two different reference time periods, making it possible that the behaviour *was* different in those intervals, or (ii) comparison was only made of aggregate rates rather than on differences at the individual level.

Whilst the relative agreement between questionnaires and diaries has been shown (Huysman, 1996; McLaws, 1990; Reading, 1983), these authors also report that estimates derived from diaries were generally *lower* than estimates reported in the interview or questionnaire (Huysman *et al.*, 1996; McLaws *et al.*, 1990; Phellas, 1994). In a study most similar to the present one, McLaws *et al.* (1990) report that discrepancies also differed systematically by the activity involved, not only in terms of the size of discrepancies, but in some cases also the direction (sign) of the discrepancies. Moreover, with specific reference to sexual behaviour, they report (*ibid.*, pp. 275–276) that diary and interview estimates are in good agreement for more infrequent (homo)sexual activities, but in poor agreement for frequent ones, and an earlier SIGMA study (Phellas, 1994)<sup>2</sup> found that the only significant estimate which was lower in the interview than in the subsequent month's diary data was receptive anal intercourse—an ominous finding, consistent with McLaws *et al.*, but suggesting that interviews (in this case) were least accurate when reporting rates of the most significant risk behaviour, and that such under-reporting was systematic.

In this paper three questions are therefore addressed:

- Is there agreement at the profile (aggregate) level between the two methods (SAQ and SASD) when estimates are made referring to the *same* time period?
- How sure do the respondents feel about making their retrospective questionnaire estimates of the frequency of the sexual behaviour which has been previously recorded in their diaries?
- At the individual level: (a) are there systematic tendencies to under-over-report? and (b) are divergences greater for some sexual acts than for others?

## Method

The research reported here is part of a set of validation studies of the method of sexual diaries by Project SIGMA, Essex.<sup>3</sup> The sexual diary method has been developed and used since 1984 in conjunction with interview and serological testing of gay and bisexual men in England and Wales (Coxon, 1988a; 1992; 1994; 1996a; 1996b).

The standard Project Diary Kit was used to obtain the basic SASD diary data (Coxon, 1992; SIGMA, 1993; 1996a). The *diary* consisted of four one-week pages, where the diarist wrote his entry on a daily basis for each sexual 'session', being reminded on each page about the characteristics of the sexual behaviour which were to be recorded.<sup>4</sup> Sexual partners were entered as a numbered list, together with their characteristics, such as age and HIV status.

## Subjects

Recruitment to the Project SIGMA Sexual Diary studies has been via the five-wave Longitudinal Panel (Coxon, 1996a; Davies *et al.*, 1993), supplemented by biennial appeals for volunteers, made in the British gay press. The data for this study were obtained in the 1992/3 study. Following an appeal in three gay publications for volunteer diarists,<sup>5</sup> 178 men wrote

in for a diary kit. Of these, just under one-half (86) completed and returned their month-long sexual diary. Immediately the completed diary was returned, they were sent (without prior warning) a new letter and form asking them to estimate the number of times they had done each of a set of 14 sexual acts during the period of their diary, together with a request to give a five-point rating of how certain they were of each estimated number. An incentive of £2 was offered, to be donated to the AIDS charity Terrence Higgins Trust on successful completion and return. Seventy-four of the 86 diarists (86% participation rate) sent back a set of estimates of frequency of behaviour and associated certainty ratings. The purpose of this supplementary exercise was to obtain estimates of their sexual activity in self-administered questionnaire format from the same men who had completed (and returned) their sexual diary, and referring to the same time period.

### Measures

The three main measures are:

- The *diary count*: obtained by counting *post factum* (by means of specially-written software (Coxon & Coxon, 1993c)) the number of times that each of the 14 sexual acts occurred in the subjects' encoded diaries.
- The corresponding (retrospective) *questionnaire estimate*: made by each subject of the number of times in the diary-month they had done each of the acts, together with
- The associated five-point<sup>6</sup> graded *certainty rating*: of how sure the diarist was about the numerical estimate of each behaviour, which he had just made in the questionnaire.

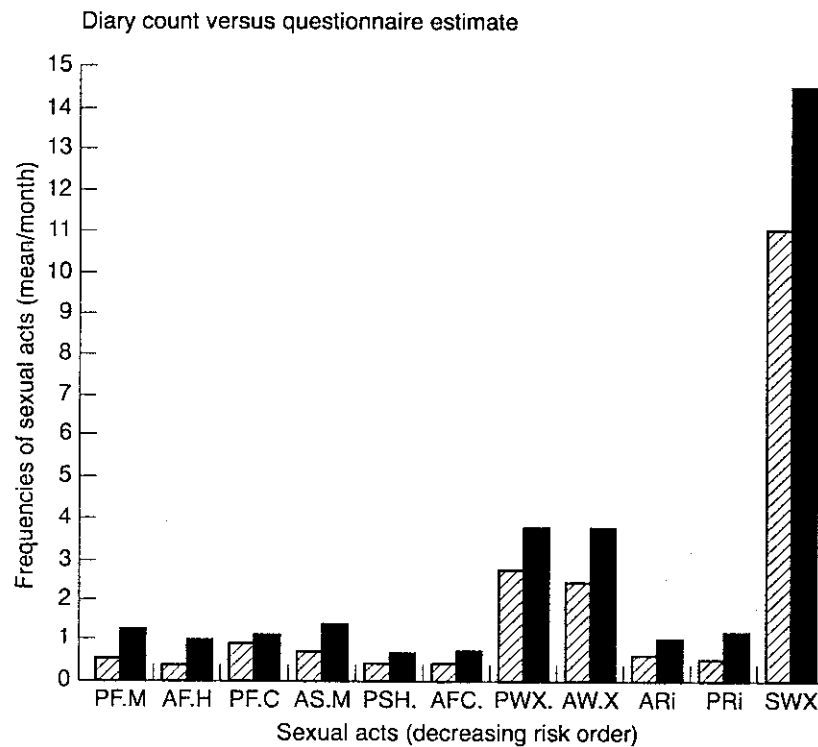
The three main measures refer to the same set of 14 sexual acts. The sexual acts consist of:

- (1) The active and passive<sup>7</sup> variants of four homosexual behaviours: three common acts carried to ejaculation:<sup>8</sup> masturbation, fellatio and (unprotected) anal intercourse together with one relatively uncommon act ('rimming'<sup>9</sup>—anilingus or oral-anal contact).
- (2) Active and passive variants of fellatio where the inserter 'pulled out' before ejaculating.
- (3) Active and passive anal intercourse with the use of a condom.
- (4) Solitary masturbation to ejaculation, together with
- (5) Vaginal intercourse (which does not occur in these data).

These particular acts (specified in Figure 2, column vii) were chosen to vary acts of different frequency (common versus uncommon) (i) to include acts of questionable (and currently questioned) safety (ii), to include high-risk behaviours with and without protection (iii), to include the single most common activity (iv) and the most significant heterosexual activity (v). For purposes of simplification, this set is reduced to 11 in subsequent analysis<sup>9</sup> and is specified in Figures 1 and 2.

### *Discrepancy measure*

In order to examine the discrepancies between a subject's diary and questionnaire estimates, a simple measure of difference was used (D-Q), defined by subtracting a subject's retrospective questionnaire estimate (Q) of a given sexual act from the corresponding count derived from his diary record (D). If this value is zero, the diary count and the questionnaire estimate are the same; if negative, the questionnaire estimate is the higher and if positive, the diary count is the higher.



No.	Code	Sex act (risk order)
1	SWX	Wanking yourself till you come
2	PRI	Being rimmed by a guy
3	ARI	Rimming a guy
4	PWX.	Being wanked by a guy till you come
5	AW.X	Wanking a guy till he comes
6	AFC.	Fucking a guy; you wear a condom
7.	PSH.	Being sucked by a guy till you come in his mouth
8	AS.M	Sucking a guy till he comes in your mouth
9	PF.C	You are fucked by a guy; he wears a condom
10.	AFH.	Fucking a guy; you come in him
11.	PF.M	You are fucked by a guy; he comes in you

FIG. 1. Aggregate estimates of sexual behaviour (one month). Average number of sexual acts per month: sexual diary count and retrospective estimated number in questionnaire (▨ diary count; ■ questionnaire estimate).

## Results

In the following analysis, the 11 sexual acts will be presented in 'risk order'—risk to the subject, that is. The risk order is induced by imposing three constraints:

- (1) *behaviour*: that anal intercourse > (is more risky than) fellatio > oral-anal > masturbation;
- (2) *modality*: that ejaculation of semen 'in' the diarist (receptive anal and oral intercourse) > ejaculation 'on' the diarist;
- (3) *prophylaxis*: no condom use > condom use.

The resulting full rank order<sup>10</sup> of risk is given in column (i) of Figure 2. Such a risk order is used to recognize how the main variables vary with risk.

(i) Risk rank	(ii) Diary count (D)	(iii) Questionnaire estimate (Q)	(iv) Relative difference (D-Q)/D	(v) Certainty rating	(vi) SIGMA code	(vii) Sexual act	
						Behaviour/modality	Outcome
11 high	0.65 (2.08)	1.40 (4.15)	-1.15	4.42 (1.28)	PF.M	Were you fucked by a guy?	He came in you?
10	0.54 (2.15)	1.08 (3.23)	-1.00	4.42 (1.27)	AFH.	Did you fuck a guy?	You came in him?
9	1.05 (2.15)	1.34 (3.06)	-0.28	4.11 (1.45)	PF.C	Were you fucked by a guy?	He wore a condom?
8	0.76 (1.75)	1.44 (3.65)	-0.89	4.13 (1.32)	AS.M	Did you suck a guy?	Till he came in your mouth
7	0.49 (1.35)	0.67 (1.59)	-0.37	4.04 (1.40)	PSH.	Were you sucked by a guy?	Till you came in his mouth?
6	0.49 (1.37)	0.73 (1.70)	-0.52	4.34 (1.31)	AFC.	Did you fuck a guy?	You wore a condom?
5	2.75 (4.05)	3.75 (6.47)	-0.36	3.22 (1.40)	PWX.	Were you wanked by a guy?	Till you came?
4	2.41 (3.40)	3.75 (5.18)	-0.56	2.97 (1.38)	AW.X	Did you wank a guy?	Till he came?
3	0.63 (1.34)	1.01 (1.77)	-0.60	4.14 (1.35)	ARi	Did you rim a guy?	
2	0.49 (1.23)	1.09 (1.90)	-1.22	4.05 (1.35)	PRi	Were you rimmed by a guy?	
1 low	10.93 (9.64)	14.37 (13.07)	-0.31	3.07 (0.85)	SWX	Did you wank yourself?	Till you came?

(upper figure in summary data cells is the mean, the lower figure (in brackets) is the standard deviation)

FIG. 2. Summary diary count, retrospective estimate and certainty ratings data by risk order.

#### *Do the diary and questionnaire estimates differ?*

The relevant summary data relating the diary counts and the questionnaire estimates are illustrated in Figure 1 and the summary data are given in Figure 2, arrayed in order of risk. The most striking characteristic of the profiles of SAQ and SASD *average* frequencies is that, as in previous studies, diary counts are lower than questionnaire estimates, and are so for all sexual acts. The largest average differences are not surprisingly for the most common activities (masturbation), and the biggest relative differences (relative to the frequency of occurrence, see column iv) include the most risky activity: unprotected anal intercourse. But the overall profiles are very similar indeed ( $r = 0.99$ ), though this masks some rank differences ( $\tau = 0.76$ ).

#### *How sure are subjects about their estimates?*

Several suggestions have been made about the likely accuracy of recall of sexual data which may be illuminated by the certainty ratings data. The most usually cited hypothesis is:

- That less common acts are recalled more accurately than more common ones (due presumably to the greater 'surprisal value' of rare acts).

The relative frequency with which homosexual acts occur in sexually active gay populations is basically the same, whether the reference is to the number of *acts* or the number of *individuals* who do it (Coxon, 1996a, pp. 64–86; Davies *et al.*, 1993, p. 106), and this makes it simple to define the 'more common acts' as masturbation and oral sex. This hypothesis holds quite well: the less common the act, the more sure subjects are of its accuracy ( $r = -0.68$ ,  $\tau = -0.43$ , between averages).

More relevant from the point of view of making epidemiological predictions is the question of whether certainty of estimates is related to the riskiness of the sexual act: is it the case that the more risky the act, the higher the certainty of the estimate? Again the answer is yes ( $r = 0.15$ ) and although the correlation of average certainty and average riskiness is only modest, the Multisample Median test and Jonkheere's Trend test give highly significant ( $p < 0.001$ ) differences in certainty ratings between the sex act medians. Thus diarists are most sure about their estimates of those sex acts which are of the highest risk. In particular (following the median judgements),

- They are virtually all certain how often they had engaged in anal intercourse (both insertive and receptive anal intercourse, whether or not protected).
- They are relatively unsure about how often they had engaged in fellatio.
- They are only moderately sure of the number of times they engaged in masturbation with others, and not at all sure of their estimate of solitary masturbation.

#### *Analysis of individual differences*

Having confirmed that the profiles of the average frequencies of sexual acts for the diary and questionnaire methods are relatively similar, it is important now to examine discrepancies between questionnaire estimates and diary counts *at the individual level*.

First, what is the gross overall pattern of discrepancy between diary count and questionnaire estimate (summed over all individuals and acts)? It is negative and averages just over one act, i.e. there is an overall tendency for individuals to give questionnaire estimates which are higher than diary counts, as there is in the aggregate data.

To what extent do *individuals* differ in their discrepancies? Calculating the average discrepancy score for each individual (over all the sex acts he engages in), the distribution of individual differences (presented in Figure 3) is clearly skew, dominated by rather small differences around zero (median:  $-0.75$ , IQR: 2.2) and there is a long tail toward negative values. This takes no account of the fact that the individual sexual repertoires differ considerably in the range of acts they include, nor of the differences between the acts: the overall set is highly dominated by masturbation.

In order to gain further insight about how discrepancies are distributed, it is useful to define the typology of discrepancy scores (D–Q), as shown in Figure 4. Irrelevant discrepancies are simply included for completeness of the typology. 'Mistaken' and 'forgotten' discrepancies, where one or other frequency is zero, are presently ignored in calculations. Applying the typology first to the sum total of discrepancies, 55% are over-reports, 20% are identical and 25% are under-reports, i.e. twice as many judgements are over-estimates than are under-estimates, and only one-fifth of questionnaire estimates are identical to diary counts.

The usual individual *tendency* in discrepancies is mostly in one direction—predominantly

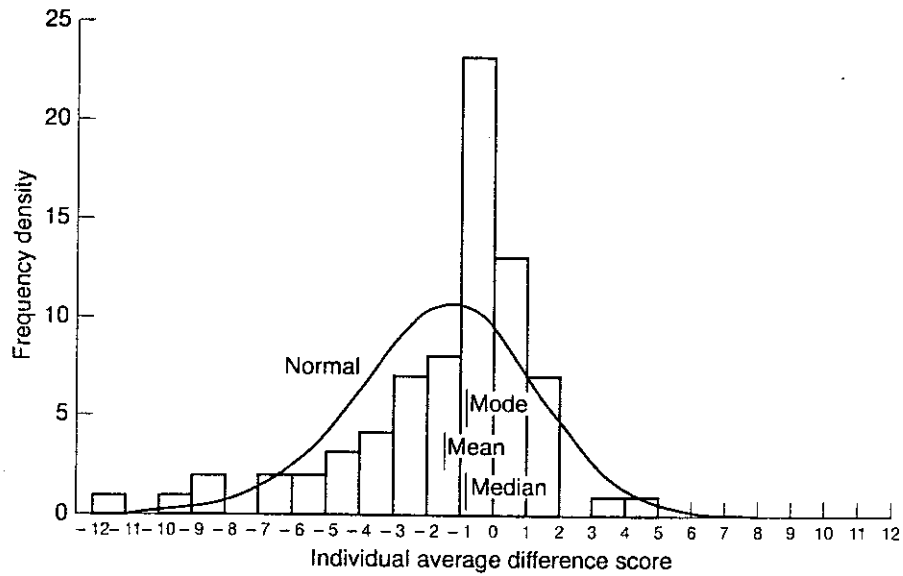


FIG. 3. Overall individual differences between diary and questionnaire estimates (over all sexual acts).

to under-report or predominantly to over-report. Very few indeed have a mixture of under- and over-estimation—almost all err in one direction only, and by far the majority of discrepancies are negative. This shows quite clearly that:

- There is a strong tendency for an individual to be *either* an under-estimator (43%) or an over-estimator (26%).
- That few (17%) have a mixture of positive and negative discrepancies.
- That very few individuals indeed give consistently identical estimates—only one person has zero discrepancies for every sexual act he engages in—and few (14%) are even close to zero, and they have a very restricted sexual repertoire.

Pattern of discrepancy [D,Q]	Description of discrepancy	Type
[0,0]	<i>(The sexual act concerned ...)</i> Does not feature in diary and zero estimate in questionnaire	Irrelevant
[0,*]	Does not feature in diary, but questionnaire estimate is over zero	Mistaken
[* ,0]	Diary count over zero, but questionnaire estimate is zero	Forgotten
[D > ]	Diary count is greater than questionnaire estimate	Under-report
[D = Q]	Diary count is same as questionnaire estimate	Identical
[D < Q]	Diary count is less than questionnaire estimate	Over-report

(\*signifies a value over zero).

FIG. 4. The typology of discrepancy scores.



**Table 1.** *Individuals' sexual behaviours: types of discrepancies between diaries and questionnaires*

	Did not do in month (%)	Mistaken/ forgotten	Under-report	Equal estimates	Over-report	Total (%)
Masturbation (with others) ( $p < 0.01$ , paired <i>t</i> -test)	24	31	40	14	15	100
Fellatio (NS)	65	30	34	22	14	100
Anal intercourse:						
with condom (NS)	68	30	12	36	22	100
active, no condom ( $p < 0.001$ )	80	36	50	14	0	100
passive, no condom ( $p < 0.01$ )	75	31	11	26	32	100

Finally, it remains to see the ways in which the pattern of an individual's discrepancies are related to particular sexual acts. The profile of discrepancies for a number of relevant sexual acts is given in Table 1. The patterns for each act are quite distinct:

- *Masturbation*: is typified chiefly by under-reporting; little difference between active and passive modalities; diary counts are significantly higher than questionnaire estimates ( $p < 0.001$ ).
- *Fellatio*: Somewhat more under- than over-reporting; slightly more difference between modalities. Diary counts usually greater than questionnaire estimates (NS).
- *Anal intercourse*:
  - with condom: Highest identical reporting; little difference between modalities. (NS differences between diary counts and questionnaire estimates.)
  - no condom: Very different pattern compared to condom-protected-anal intercourse, and major differences between active and passive modalities:
    - active: Very large degree of under-reporting; no over-reporting.
    - passive: Considerable identical and over-reporting.

Masturbation and fellatio thus present a distinct pattern of discrepancies: questionnaire estimates are usually lower than diary counts. But the major, significant and important discrepancies in reporting occur for anal intercourse, especially for its riskiest unprotected variants, and for both its active and passive modalities. Compared to these, the discrepancies for other behaviours are unimportant. The pattern which is most distinctive and different is that for anal intercourse without a condom, which is highly dissimilar both to anal intercourse with a condom *and* to receptive anal intercourse without a condom.<sup>11</sup> Here is the Achilles' heel of the SAQ compared to the SASD.

### Conclusions and discussion

The diary method has been used as an appropriate method for collecting sequential and time-linked data on sexual behaviour since 1982, and has proved to be a reliable tool for the study of sexual behaviour in a number of pioneering studies, resistant to interview and retrospection bias and yielding detailed and often unique information. But it is undoubtedly a demanding method (of both respondents and analysts), subject to selection bias and (as it stands) it is unlikely to be a substitute for survey methods.

The study reported here compares two methods for measuring the sexual behaviour of gay men—self-administered diary and subsequent questionnaire estimates—both of which refer to the same index month. The analysis confirms that the sexual diary method is likely to be more accurate than retrospective questionnaire data, but that they are strongly related. In particular:

- (1) Diarists attribute highest certainty in their questionnaire estimates of sexual behaviour to the less common sexual acts, and in particular for higher-risk acts which involve anal intercourse. This is confirmed in the smaller relative discrepancies associated with these same acts.
- (2) At the individual level, *post factum* questionnaire estimates also yield much the same aggregate profile as diary counts, but they consistently yield higher average frequencies.
- (3) Most individuals have a consistent pattern of over-identical/under-reporting estimates across different sexual acts.
- (4) The main important exception occurs for the highest-risk sexual activity: anal intercourse without a condom which differs markedly according to whether it refers to insertive (active) or receptive (passive) variants. The discrepancies may in part reflect a sort of 'social desirability', or gendered, effect: active partners tend to under-estimate, and passive partners tend to over-estimate the amount of highest-risk sex<sup>12</sup> in their questionnaire accounts. If the diary is taken as the more precise method, SAQ estimates of insertive anal intercourse would exhibit a considerably lower incidence and estimates of receptive anal intercourse would assume a higher incidence than is the case.

How representative or generalizable are these results and how do these results apply to the issue of obtaining better data on sexual behaviour? First, issues of representativeness. The answer depends in part on known facts about samples of gay men:

- That prevalence is hard to establish in random sample studies due to interviewer effect, covert practice and adjudged deviance of homosexuality (Johnson *et al.*, 1994, pp. 189–193; Kalton, 1993).
- That samples of gay men based upon volunteer subjects are known to be subject to over-representation of younger men, of those living in the metropolis, of those overt about their sexuality and of those with higher educational qualifications (Coxon, 1995; Davies, 1986; Fitzpatrick *et al.*, 1990; Hunt *et al.*, 1991).
- That within the SIGMA samples, those completing SASDs are likely to be slightly more sexually active (an empty diary or one consisting simply of solitary masturbation is unlikely to be volunteered or returned), and not be in a regular relationship (Coxon, 1990; 1996a).

The current sample is volunteer based and is small in size, but it is at least as representative as the main SIGMA sample and earlier diary studies, and differs primarily in being smaller in size, almost certainly due to 'volunteer fatigue', since the numbers of those responding to Project SIGMA appeals has declined sharply from 1987 to the present. However, when the aggregate sexual behaviour profiles from the various sexual diary data sets are compared, whether emanating from sub-samples within the SIGMA panel (Coxon, 1996a; Phellas, 1994) or from the biennial appeals—they yield markedly similar results (compare, for instance, Coxon, 1994, pp. 136–137). More relevantly, the finding that diary data estimates of rates of the main sexual behaviours are *in aggregate* systematically lower than questionnaire data holds for all SIGMA data sets where it is possible to compare interview

data with diary data for the same panel members, even though the reference month may be different. The earlier study of divergences in SIGMA panel members' diary and interview data (Phellas, 1990) had also detected that estimates of anal intercourse exhibited systematic individual differences in reporting, relative to other behaviours, but the fact that estimates for receptive and insertive variants of unprotected anal intercourse were biased in opposite directions was not discovered until the present analysis.

In the preliminary section, reasons were adduced as to why SASDs might be considered to give more accurate estimates than SAQs, but it was stated that, in the absence of a behavioural *tertium quis* for homosexual behaviour, no determination of greater accuracy is possible. But the results presented here can also be used the other way round—in effect, SASDs can be used to validate or check on survey data.<sup>13</sup> It would be difficult, costly<sup>14</sup> and probably impractical to use the method of SASDs to make national or large-scale studies of sexual behaviour, and the SAQs and interviews are much better adapted to that. However, if a random sub-sample of such a large-scale study were to use SASDs, it would be possible not only to check on patterns of under-over-reporting, but also to examine the contexts in which sexual behaviour (and especially risk behaviour) take place, and which have specific relevance to explaining it. And this is, after all, a major advantage of the method of SASDs.

## Notes

- [1] Recording real-time sexual activity has major problems of reactivity (if observers are present), and of selection bias if recruited from volunteers, and the behaviour is in any event continuous in form, rather than being re-interpreted into categorical form, as self-reports of sexual behaviour invariably are.
- [2] The data were a balanced random sample from the core SIGMA data set, consisting of 1,035 month/diaries obtained from panel members drawn from ten sites throughout England and Wales in the period 1987–1993.
- [3] Project SIGMA (Socio-sexual Investigations of Gay Men and Aids). The research in this paper was funded by the Medical Research Council and by the Department of Health (UK). Work on Sexual Diary software and on the validity and reliability of the sexual diary method reported here is funded under a series of grants from the Department of Health and reported in Coxon; (1996a; 1996b). The views expressed in this paper are those of the author and not necessarily those of the Council or of the Department.
- [4] These read: 'Remember, each session should include: The Time, The Place, The Partners (from partner list). Then, describe the session in your own words. Remember to mention exactly what happened to the "come" (ejaculate) and always mention the use of condoms. List any accompaniments you use (poppers, lubricants, drugs, sex toys ...) and mention how much you drank each day if it is associated with sex.
- [5] Proportionate responses were: from BOYZ (64%), *Capital Gay* (18%) (both weekly free newspapers, the latter now defunct) and *Gay Times* (17%) (monthly). The first has a younger readership and is primarily London-based and 'scene'-oriented, the second is also London-based, but more serious and with a slightly older readership, and the third has a more substantial and long-established national readership.
- [6] The wording of instructions and format of the questions were as follows:

---

HOW OFTEN do you reckon you did each of the following sexual acts?

(write in the box the number of times you estimate)

... then for each number, underline the category below that best indicates how sure you are.

In the month of the diary, how often ...

1. Wanking a guy till he comes □

(I am: CERTAIN/VERY SURE/MODERATELY SURE/NOT VERY SURE/NOT AT ALL SURE of this number)

---

- [7] The term 'modality' is used to differentiate active and passive variants of the same behaviour. 'Active' refers to *ego* doing an act to *alter*; 'passive' to *alter* doing an act to *ego*; it is not identical to the insertor/insertee distinction.
- [8] Masturbation, fellatio and anal intercourse together account for an estimated 81% of homosexual acts (Coxon, 1996a, p. 167), anilingus accounts for 2%.
- [9] The acts of fellation referring to 'pulling out' were found to be ambiguous with respect to ejaculation and were removed from further analysis. (For instance: 'Sucking a guy: he pulls out before he comes'—does *alter* ejaculate

- after withdrawing, or does he withdraw *in case* he ejaculates?) Vaginal intercourse was not reported and was also removed.
- [10] The order cannot be reduced to a strict order, due to the relatively arbitrary placing of oral-anal contact (ARI and PRI), where there is considerable risk of transmission of Hepatitis B, but little of HIV.
- [11] Using the Index of Dissimilarity (Blau & Duncan, 1967), the most similar profiles are for masturbation and fellatio (8%) and for receptive anal intercourse without a condom and anal intercourse with a condom (16%). The maximum dissimilarities are (insertive anal intercourse without a condom and anal intercourse with a condom) and (insertive anal intercourse without a condom and receptive anal intercourse without a condom) at 44%.
- [12] Regression of diary counts on interview estimates (not presented here) also shows that active/insertive and passive/receptive variants work in different ways: insertive anal intercourse discrepancies are basically higher by a constant, whereas receptive anal intercourse (the more risky behaviour) discrepancies differ by a multiplier.
- [13] I am grateful to an anonymous referee for making this point.
- [14] Altruism is a potent incentive to participate by those who are already motivated to do so; monetary incentives are more successfully used to persuade those who would not otherwise participate (see Church, 1993).

## References

- BANCROFT, J. (1997). *Researching sexual behaviour: methodological issues*. Bloomington, Ill: Indiana University Press.
- BERK, R., ABRAMSON & OKAMI, P. (1995). Sexual activities as told in surveys. In: P. R. ABRAMSON & S. O. PINKERTON (Eds), *Sexual nature, sexual culture*. Chicago: Chicago University Press.
- BLAU, P.M. & DUNCAN, O.D. (1967). *The American occupational structure*. Chichester: John Wiley.
- BOULTON, M. (Ed.) (1994). *Challenge and innovation: methodological advances in social research on HIV/AIDS*. London: Taylor and Francis.
- BREWER, W. F. (1988). Memory for randomly sampled autobiographical events. In: U. NEISSER & E. WINOGRAD (Eds), *Remembering reconsidered: ecological and traditional approaches to the study of memory* (pp. 21-90). Cambridge: Cambridge University Press.
- BUTCHER, R. & ELDRIDGE, J. (1990). The use of diaries in data collection. *The Statistician*, 24-41.
- CATANIA, J., TURNER, H., PIERCE, R.C., GOLDEN, E., STOCKING, C., BINSON, D. & MAST, K. (1990). Response bias in surveys of AIDS-related sexual behaviour. In: D. G. OSTROW & R. C. KESSLER (Eds), *Methodological issues in AIDS behavioural research* (pp. 132-158). New York: Plenum Press.
- CHURCH, A.H. (1993). Estimating the effect of incentives on mail survey response rates: a meta-analysis. *Public Opinion Quarterly*, 57, 62-79.
- CONRATH, D.W., HIGGINS, C.A. & MCCLEAN, R.J. (1983). A comparison of the reliability of questionnaire versus diary data. *Social Networks*, 5(3), 315-321.
- CORTI, L. (1993). Using diaries in social research. *Social Research Update*, (2), 1-4.
- COXON, A.P.M. (1988a). 'Something sensational... The sexual diary as a tool for mapping detailed sexual behaviour. *Sociological Review*, 36(2), 353-367.
- COXON, A.P.M. (1988b). The numbers game: lifestyles, epidemiology of AIDS and social science. In: P. AGGLETON & H. HOMANS (Eds), *Social aspects of AIDS*. London: Falmer Press.
- COXON, A.P.M. (1990). The effect of age and relationship on gay men's sexual behaviour: a preliminary analysis of sexual diary data. Cardiff: Project SIGMA Working Paper No. 13 (ISBN 1 872956 03 3).
- COXON, A.P.M. (1994). Diaries and sexual behaviour: the use of sexual diaries as method and substance in researching gay men's response to HIV/AIDS. In: M. BOULTON (Ed.) *Challenge and innovation: methodological advances in social research on HIV/AIDS*. (pp. 125-148). London: Taylor & Francis.
- COXON, A.P.M. (1995). Networks and sex: some social contexts of gay men's response to HIV/AIDS. In: R. PARKER & J. GAGNON (Eds), *Conceiving sexuality: approaches to sex research in a postmodern world*. London: Routledge.
- COXON, A.P.M. (1996a). *Between the sheets: sexual diaries and gay men's sex in the era of AIDS*. London: Cassell.
- COXON, A.P.M. (1996b). Comparative validity of interview and diary methods for assessing rates of (homo)sexual behaviour in Aids research. Colchester: Project SIGMA, Essex, working paper 96-93.
- COXON, A.P.M. & COXON, N.H. (1993a). Risk in context: the use of sexual diary data to analyse sequences of homosexual risk behaviour. In: H. TEM BRUMMELHUIS & G. HERDT (Eds), *Culture and sexual risk: anthropological perspectives on AIDS* (pp. 205-225). Amsterdam: Gordon and Breach.
- COXON, A.P.M. & COXON, N.H. (1993c). *SDA<sup>TM</sup>: sexual diary analysis: documentation of programs*. Colchester: Project SIGMA, Essex.
- COXON, A.P.M. & COXON, N.H. *et al.* (1993b). Sex role separation in sexual diaries of homosexual men. *AIDS*, 7(6), 877-882.

- COXON, A.P.M., WEATHERBURN, P., HUNT, A.J. & DAVIES, P.M. (1992). The structure of sexual behaviour. *Journal of Sexual Research*, 29(1), 61-83.
- CROYLE, R.T. & LOFTUS, E.F. (1990). Recollection in the Kingdom of AIDS. In: D. OSTROW & R. C. KESSLER (Eds) *Methodological issues in AIDS behavioral research* (pp. 163-180). New York: Plenum Press.
- DAVIES, P.M. (1986). *Some problems in defining and sampling non-heterosexual males*. Cardiff: Project SIGMA working paper 3.
- DAVIES, P.M., HICKSON, F.C.I., WEATHERBURN, P., HUNT, A.J., BRODERICK, P.J., COXON, A.P.M., MCMANUS, T.J. & STEPHENS, M.J. (1993). *Sex, gay men and AIDS*. London: Falmer Press.
- DEX, S. (1991). *The reliability of recall data: a literature review*. Colchester: ESRC Research Centre on Micro-Social Change, working paper 11.
- FITZPATRICK, R., BOULTON, M. & HART, G. (1990). Variation in sexual behaviour in gay men. In: P. AGGLETON, P. M. DAVIES & G. HART (Eds), *AIDS: individual, cultural and policy dimensions* (pp. 121-132). London: Falmer Press.
- FORSYTH, B.H., LESSLER, J.T. & HUBBARD, M.L. (1992). Cognitive evaluation of the questionnaire. In: C. F. TURNER, J. T. LESSLER & J. C. GFROERER (Eds), *Survey measurement of drug use: methodological issues of drug use: methodological Studies* (pp. 13-53). Washington, DC: US Department of Health and Human Services.
- FORTENBERRY, J.D., CECIL, H., ZIMET, G.D. & ORR, D.P. (1997). Concordance between self-report questionnaires and coital diaries for sexual behaviours of adolescent women with sexually transmitted infections. In: J. BANCROFT, *Researching sexual behaviour: methodological issues* (pp. 237-250). Bloomington, Ill: Indiana University Press.
- FREEMAN, E.W., DERUBEIS, R.J. & RICKELS, K. (1996). Reliability and validity of a daily diary for pre-menstrual syndrome. *Psychiatry Research*, 65, 97-106.
- GOLD, R.S. & ROSENTHAL, D.A. (1995). Preventing unprotected anal intercourse in gay men: a comparison of two intervention techniques. *International Journal of STD & AIDS*, 6, 89-94.
- GRAHAM, C. & BANCROFT, J. (1997). A comparison of retrospective interview assessment versus daily ratings of sexual interest and activity in women. In: J. BANCROFT, *Researching sexual behaviour: methodological issues* (pp. 227-236). Bloomington, Ill: Indiana University Press.
- HILTON, M.E. (1989). A comparison of a prospective diary and two summary recall techniques for recoding alcohol consumption. *British Journal of Addiction*, (84), 1085-1092.
- HUNT, A.J., DAVIES, P.M., WEATHERBURN, P., COXON, A. P. M. & MCMANUS, T. J. (1991). Sexual partners, penetrative sexual partners and HIV risk. *AIDS*, 5(6), 723-728.
- HUYSMAN, F., LAMMERS, J., RENCKSTORF, K. & WESTER, F. (1996). *Diary measures versus survey measures of mass media use: a validation study*. Nijmegen: Department of Communications.
- JAMES, N.J., BIGNELL, C.J. & GILLIES, P. A. (1991). The reliability of self-reported sexual behaviour [short communication]. *AIDS*, 5(3), 333-335.
- JANSON, C.-G. (1990). Retrospective data, undesirable behaviour and the longitudinal perspective. In: D. MAGNUSSON & L. R. BERGMAN (Eds), *Data quality in longitudinal research*. Cambridge: Cambridge University Press.
- JOHNSON, A., WADSWORTH, J., WELLINGS, K., FIELD, J. & BRADSHAW, S. (1994). *Sexual attitudes and lifestyle*. London: Blackwell.
- JUSTER, F.T. (1985). The validity and quality of time-use estimates obtained from recall diaries. In: F. T. JUSTER & F. P. STAFFORD (Eds), *Time, goods and well-being* (pp. 63-88). Ann Arbor, MI: Survey Research Center.
- KALFS, N. & SARIS, W.E. (1990). Mode effects in time diary research. *Kwantitative Methoden*, (7), 65-86.
- KALTON, G. (1993). Sampling considerations in research on HIV risk and illness. In: D. G. OSTROW & R. C. KESSLER (Eds), *Methodological issues in AIDS behaviour research* (pp. 53-75). New York: Plenum Press.
- LINTON, M. (1986). Ways of searching and the contents of memory. In: D. C. RUBIN (Ed.), *Autobiographical memory* (pp. 50-67). Cambridge: Cambridge University Press.
- MCLAWS, M.L., OLDENBURG, B., ROSS, M. W. & COOPER, D.A. (1990). Sexual behaviour in AIDS related research: reliability and validity of recall and diary measures. *Journal of Sex Research*, 27(2), 265-281.
- MCQUEEN, D., GORST, T., NISBET, L., ROBERTSON, B., SMITH, R. & UITENBROEK, D. (1989). *A study of lifestyle and health*. Edinburgh: Research Unit in Health and Behavioural Change.
- MILLER, H.G., TURNER, C.F. & MOSES, L.E. (Eds) (1990). *AIDS: the second decade*. Washington, DC: National Academy Press.
- NEISSER, U. & WINOGRAD, E. (Eds) (1988). *Remembering reconsidered: ecological and traditional approaches to the study of memory*. Cambridge: Cambridge University Press.
- O'REILLY, J., HUBBARD, M., LESSLER, J., BIEMER, P. & TURNER, C.F. (1994). Audio computer-assisted self-interviewing: new technology for data collection on sensitive issues and special populations. *Journal of Official Statistics*, 10, 197-214.
- OSTROW, D.G. & KESSLER, R.C. (Eds) (1990). *Methodological issues in AIDS behavioral research*. New York: Plenum Press.
- PHELLAS, C.N. (1994). The reliability of recall and diary measures of sexual behaviour in the HIV and social context. MSc thesis, London: City University (SSRU).

- POIKOLAINEN, K. & KARKKAINEN, P. (1993). Diary gives more accurate information about alcohol consumption than questionnaire. *Drug and Alcohol Dependence*, 11, 209-216.
- READING, A.E. (1983). A comparison of the accuracy and reactivity of methods of monitoring male sexual behavior. *Journal of Behavioral Assessment*, 5(1), 11-23.
- RICHARDSON, A. (1994). The health diary: an examination of its use as a data collection method. *Journal of Advanced Nursing*, (19), 782-791.
- ROBINSON, J.P. (1985). The validity and reliability of diaries versus alternative time-use measures. In: F. T. JUSTER & F. P. STAFFORD (Eds) *Time, goods and well-being* (pp. 33-62). Ann Arbor, MI: Survey Research Center.
- RUBIN, D.C. (Ed.) (1986). *Autobiographical memory*. Cambridge: Cambridge University Press.
- SIGMA (1993). *Notes on keeping a sexual diary*. Colchester: Project SIGMA, Essex.
- SIGMA (1996). *Diary Documentation Kit*. Colchester: Project SIGMA, Essex.
- SUNDET, J., MAGNUS, P., KVALEM, I. & BAKKETEIG, L. (1990). Self-administered anonymous questionnaires in sexual behaviour research: the Norwegian experience. In: M. HUBERT (Ed.) *Sexual behaviour and risks of HIV infection* (pp. 79-96). Brussels: Publications des Faculté Universitaires Saint-Louis.
- TIELMAN, R. (1990). Telephone surveys in comparison with other methods of psycho-social AIDS research. In: HUBERT (1990), pp. 97-103.
- TURNER, C.F., DANELLA, R. & ROGERS, S. (1995). Sexual behaviour in the United States, 1930-1990: trends and methodological problems. *Sexually Transmitted Diseases*, 22(3), 173-190.
- TURNER, C.F., LESSLER, J.T. & DEVORE, J. (1992). Effects of mode of administration and wording on reporting of drug use. In: C. F. TURNER, J. T. LESSLER & J. C. GFROERER (Eds), *Survey measurement of drug use: methodological studies*. Washing, DC: US Department of Health and Human Services.
- TURNER, C.F., LESSLER, J.T. & GFROERER, J.C. (Eds) (1992). *Survey measurement of drug use: methodological studies*. Washington, DC: US Department of Health and Human Services.
- TURNER, C.F., MILLER, H.G. & ROGERS, S.M. (1997). Survey measurement of sexual behavior: problems and progress. In: J. BANCROFT (Ed.) *Researching sexual behaviour: methodological issues* (pp. 37-60). Bloomington, Ill: Indiana University Press.
- VERBRUGGE, L.M. (1980). Health diaries. *Medical Care*, 18, 73-95.
- WEINHARDT, L.S., FORSYTH, A.D., CAREY, M.P., JAWORSKI, B.C. & DURANT, L.E. (1998). Reliability and validity of self-reported measures of HIV-related sexual behaviour: progress since 1990 and recommendations for research and practice. *Archives of Sexual behaviour*, 27(2), 155-180.
- WU, M.L., WHITTEMORE, A.S. & JUNG, D.L. (1988). Errors in reported dietary intakes. *American Journal of Epidemiology*, 128, 1137-1145.