

Research article

Bias and heteroscedastic memory error in self-reported health behavior: an investigation using covariance structure analysis

Emil Kupek

Address: Universidade Federal de Santa Catarina, Departamento de Saúde Pública – CCS, Campus Universitario, Trindade, 88.040-900 Florianópolis-SC, Brazil

E-mail: kupek@repensul.ufsc.br

Published: 18 November 2002

Received: 13 June 2002

BMC Medical Research Methodology 2002, **2**:14

Accepted: 18 November 2002

This article is available from: <http://www.biomedcentral.com/1471-2288/2/14>

© 2002 Kupek; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Keywords: bias, heteroscedasticity, covariance structure analysis, measurement errors, validity, sexual behavior, memory error

Abstract

Background: Frequent use of self-reports for investigating recent and past behavior in medical research requires statistical techniques capable of analyzing complex sources of bias associated with this methodology. In particular, although decreasing accuracy of recalling more distant past events is commonplace, the bias due to differential in memory errors resulting from it has rarely been modeled statistically.

Methods: Covariance structure analysis was used to estimate the recall error of self-reported number of sexual partners for past periods of varying duration and its implication for the bias.

Results: Results indicated increasing levels of inaccuracy for reports about more distant past. Considerable positive bias was found for a small fraction of respondents who reported ten or more partners in the last year, last two years and last five years. This is consistent with the effect of heteroscedastic random error where the majority of partners had been acquired in the more distant past and therefore were recalled less accurately than the partners acquired more recently to the time of interviewing.

Conclusions: Memory errors of this type depend on the salience of the events recalled and are likely to be present in many areas of health research based on self-reported behavior.

Background

Scientific research using self-reports for past events is increasingly frequent in areas which need to quantify certain aspects of health behavior whose nature is essentially private and therefore not directly observable, as well as amenable to memory error. Examples range from food consumption to health beliefs and sexual behavior. The latter is the cornerstone for various models of prevalence, incidence and risky behavior related to sexually transmit-

ted diseases among which AIDS has received particular attention. This paper presents a new statistical approach to investigate sources of errors in self-reported behavior, using an example with number of sexual partners.

The accuracy of self-reported number of sexual partners plays a prominent role in the discussions on validity of this method. [1,2] The accuracy is usually expected to increase as the defined time period asked about is closer to

the time of inquiry. This is based on numerous examples of recall difficulties in sex research, notwithstanding many other methodological factors affecting self-reported behavior in this context. [3]

Several studies of the cognitive processes relevant in the survey context pointed to the difficulty of converting episodic memory to an incidence. [4–6] While the former is a typical way of storing information on sexual partners in individual memory, the latter is what scientific research usually aims at. For example, to locate a sexual partner in time, it is often necessary for the respondent to reconstruct the order of many other clues, e.g. the partner's name and most vivid physical and personal characteristics, where did they meet, what was he/she doing at that time, what happened before and after, which year was that. This personal perception of time – defined in terms of ordering personally salient characteristics of events – is very different from the physical time which is independent of the salience clues, e.g. the number of partners in the last year. Under time pressure to respond, guessing may be used as a strategy [7], thus adding to the inaccuracy of answers.

In a large national sex survey in Britain, the issue of recall accuracy might have been related to a large decrease in variances for the number of partners reported ever, in the last five years, in the last two years and in the last year. [8] It was demonstrated with the same data how a few large outliers can affect the mean of the number of partners distribution. [2] This paper builds on these findings and relates the accuracy to the issue of bias in self-reported number of partners. Methodological interest here is in estimating the variance components of number of partners' reports due to true variation on one hand and potential sources of bias on the other. Among the latter, gender, recency of the period asked about, large number of partners and age of the respondent have been mentioned. [8] Gender-specific factors such as underreporting for women versus overreporting for men and likely underrepresentation of women with extremely high number of partners such as commercial sex workers in a large national survey [2] are not examined here as the analysis concentrated on young men aged 26–35 years.

A recent comprehensive review of sources of rater bias [9] provides an excellent systematization of the field but does not deal with self-reports. Hoyt's concept of dyadic variance due to rater's unique perception of targets cannot be applied to the task of reporting the number of sexual partners for varying past periods because there is only one observation per rater-target dyad. When the ratings are based on counts of observed behaviors, explicit nature of this task leaves little scope for different interpretations of the target as opposed to inferential attribute type rating system. [10] It is important to underline that the behavior of

interest here was explicitly defined at the beginning of the self-report questionnaire. [8]

This article is meant to reach both statistically minded and less statistically oriented audience. Because of the latter, mathematical expressions are kept to minimum. The first part of the article briefly describes the survey methodology and statistical methods used to deal with the bias in self-reported number of sexual partners, including the model specification and variable selection criteria. The second part presents results of model selection and bias estimation analyses. Finally, the results are discussed in the light of possible memory distortions such as telescoping effect.

Methods

Subjects

The National Survey of Sexual Attitudes and Lifestyles was undertaken in 1990–91 with primary aims to provide reliable data for modeling the spread of HIV. A random sample of 18876 men and women aged 16–59 living in Great Britain was interviewed and invited to complete a questionnaire containing more sensitive questions. One quarter of them were randomly assigned to fill in the long version of the questionnaire in the sampling phase of the survey. Overall response rate was 63.3%, while acceptance rate among eligible respondents who could be interviewed reached 71.5%.

Multistage cluster sampling was applied to stratified lists of areas in order to arrive at a systematic probability sample of addresses within electoral wards as primary sampling units. The second stage was sampling addresses within wards and the third stage was selecting an individual randomly from each address. As the probability of selecting a survey participant was inversely related to the number of eligible individuals at that address, the household weight was assigned to each respondent to adjust for the differences in selection probability. Regional differences in response rate were also accounted for by introducing another set of weights inversely proportional to the regional response rates. The final weight was calculated by multiplying the two weights and scaling their sum to equal the total sample size. All the analyses in this work used final weights.

The final sample achieved was broadly representative of the target population in terms of gender, age, ethnicity, marital status and socioeconomic status. [8] Internal consistency of answers to different questions and external validation of some self-reports indicated good quality of the data. Item non-response did not exceed 5% for most of the questions (for further details see [11]) and the impact of non-response to the mean of the number of partners'

distribution was likely to be small. [12] The methodology of the survey is discussed in more details elsewhere. [13]

All men aged 26–35 years among the respondents who filled in the long version of the questionnaire and had at least one heterosexual partner were included in this analysis, thus excluding 2.8% of the former who had not had sex by the time of the survey. The long questionnaire contained details on family conditions, sexual attitudes and lifestyle thought to be related to the number of sexual partners.

The age span for this analysis was chosen to minimize the factors related to sexual behavior which might be specific to one birth cohort and absent in the others such as those related to HIV/AIDS, while providing an opportunity to explore the recall accuracy for a variety of sexual lifestyles. Thus age 26–35, the typical age for marriage which is meant to be an exclusive type of sexual partnership, was chosen. As some of the influences on the number of partners were known to be non-linear (e.g. age) and gender-specific, stratified analysis on this basis seemed a more realistic choice than to try to accommodate all these complex effects within one model.

Statistical methods

This analysis is principally interested in estimating accuracy and bias for the number of partners reported in the last year, last two years, last five years, and the period more than five years ago. The latter was calculated as the difference between the lifetime number of partners and the number reported in the last five years as there was no explicit question about the number of partners beyond the last five years period but there was a question about the number of partners ever. This approach allows a clear time ordering of the periods referred to by the questions on number of partners and at the same time preserves the format used in the questionnaire.

Covariance structure analysis [14–16] was used to explore the relationship between reported and estimated true number of partners for the varying length of the past covered by the questions on the number of partners. The analysis is also known as simultaneous equations modeling, path analysis and 'causal modeling' in psychometric, econometric and social sciences but here a generic name is preferred to help recognizing its similarity with the covariance structures analyzed in mixed and/or generalized linear models. Covariance structure analysis (CSA) can model the error structure of both observed and latent variables, as well as the dependency structure due to fractional overlapping of number of partners reported in the last five years, the last two years and the last year. A detailed presentation of CSA is beyond the scope of this work but

can be found in many specialized texts [14–16] and some more recent multivariate statistics textbooks. [17,18]

It is important to underline that CSA model is fitted to the observed covariance matrix. A bit of matrix algebra is needed to state this formally. If S , M and W are observed, CSA model-based and weight matrix respectively, then the CSA fit criterion F , for generalized least squares estimation, is given by

$$F = 0.5 \text{Trace}[(W(S-M))^2] \quad (1)$$

which in the case of maximum likelihood estimation can be written as

$$F = \text{Trace}(SM^{-1}) - n + \log(\det(M)) - \log(\det(S)) \quad (2)$$

where n is the number of observed variables and "det" stands for matrix determinant. [19]

The chi-square statistic for CSA models is defined as the optimum value of the above discrepancy function F multiplied by sample size minus one. [19] It compares the model-based (estimated) covariances with those computed from the observed data i.e. unconstrained covariance matrix. Thus the comparison is formalized as the likelihood ratio test between the two. The degrees of freedom for this test equal the number of constraints imposed on the parameter estimates due to the model proposed. If the latter holds, the chi-square is non-significant and rejects the alternative hypothesis that the similarity between the data-based and model-based covariances is spurious.

The analysis presented used instrumental variables, widely applied in econometrics [20,21], to obtain unbiased regression estimates for the dependent variables of interest, which in turn can be used for further analysis. Instrumental variables are often known confounders such as age, gender and education, whose measurement errors are assumed to be uncorrelated between themselves and with the other variables in the model. Here instrumental variables were used to predict 'true' number of partners and compare it with corresponding self-report. Bias estimates were defined as differences between the two.

There were three basic steps in bias calculation. First, the CSA model parameters were obtained and used to calculate factor scores which represent 'true' number of partners for each respondent. Secondly, as these were given on an arbitrary scale, they were rescaled to the mean and variance of the cube root of reported number of partners (see later in the text about the reasons for this transformation). Thirdly, the difference between reported and rescaled values of 'true' number of partners was taken. Then the mean difference and its standard error was calculated for a com-

Table 1: Number of heterosexual partners reported for different past periods among men aged 26–35 (N = 487).

Period	Mean	Std.dev.	1st qrt.	Median	3rd qrt.	Max.
> 5 years	7.34	11.42	1	3	9	125
< 5 years	3.32	5.32	1	1	3	70
< 2 years	1.78	2.25	1	1	2	25
< 1 year	1.26	1.02	1	1	2	12

Note: Men with no heterosexual partners excluded. Number of partners was unknown for 2 respondents.

parison of various groups of interest. It is important to underline that factor scores can be justified as posterior expectations of the factors. [22] Consequently, bias can be viewed in terms of a posterior means distribution of residuals, i.e. observed minus predicted or expected on the basis of a CSA model.

Model specification

The system of simultaneously estimated regression equations can be represented as following. Let subscripts 0, 5, 2, and 1 denote temporal sequence of the number of partners for more than five years ago, last five years, last two years and last year, respectively. Let uppercase *R* and *T* denote reported and true number of partners respectively. Let lowercase *e* and *d* denote error terms for observed and estimated true values respectively. Let uppercase *I* denote the instrumental variables in the model (indexed from 1 to *n*). Then eight equations

$$R_0 = a_0T_0 + e_0 \quad (3.1)$$

$$R_5 = a_5T_5 + e_5 \quad (3.2)$$

$$R_2 = a_2T_2 + e_2 \quad (3.3)$$

$$R_1 = a_1T_1 + e_1 \quad (3.4)$$

$$T_0 = g_{01}I_1 + g_{02}I_2 + \dots g_{0n}I_n + d_0 \quad (4.1)$$

$$T_5 = g_{51}I_1 + g_{52}I_2 + \dots g_{5n}I_n + d_5 \quad (4.2)$$

$$T_2 = g_{21}I_1 + g_{22}I_2 + \dots g_{2n}I_n + d_2 \quad (4.3)$$

$$T_1 = g_{11}I_1 + g_{12}I_2 + \dots g_{1n}I_n + d_1 \quad (4.4)$$

define the model of interest. The measurement model is given by the first four equations, while the last four describe the relationship between the instrumental variables and the true number of partners for the periods specified. In order to identify the parameters of the model, following assumptions were made:

- a) the latent variables' measurement errors (*d* terms in equations 3.1 to 3.4) are uncorrelated between themselves and with the error terms for the reported values of the number of partners (*e* terms in equations 3.1 to 4.4),
- b) all other possible influences between the variables in the model which are not specified in equations 3.1 to 4.4 are set to zero, and
- c) the variance-covariance matrix of observed variables – including the instrumental ones – represents the true population values.

While all of the above assumptions are standard for many covariance structure models [14,15], they are likely to be violated to a degree in observational studies due to confounding variables omitted from the model and model misspecification. [23] The critical issue is the practical impact of the violations regarding the key model parameters of interest which can sometimes be estimated in CSA (an example is given later in the text) or through sensitivity analysis.

It should be noted that the assumptions about the distribution of the error terms in the model leave the covariance matrix of the measurement errors to be estimated, as these are the parameters of interest in this case. The assumption of other influences in the model can be tested by releasing particular restrictions and comparing the restricted and non-restricted model in terms of likelihood ratio under multivariate normality assumption or some other goodness of fit measure [24–27], or simply by univariate t-test for the parameter value when its variance is also estimated by maximum likelihood. [15] Strictly inferential approach is usually tenuous given the number of simplifying assumptions in a complex covariance structure model, particularly for the univariate t-test. [28] With many combinations of the influences between the variables tested and particularly in *post hoc* model modifications based on prior CSA analysis with the same data, the problem of finding a spuriously significant parameter becomes acute and makes hypothesis testing even more dif-

Table 2: Descriptive statistics (%) for men aged 26–35 (N = 489)

Age first sex		Marital status		Educational level		Months living-in	
< 16	21.7	Married	63.9	Degree	19.0	Zero	27.6
16–20	61.2	Cohabit.	8.5	A-level/higher	36.0	One	11.4
21–25	12.5	S/D/W [†]	5.3	O-level/CSE	29.3	2–6	33.0
26–35	3.5	Single	22.3	Other	1.5	7+	26.7
Unknown	1.1			None	14.2	Unknown	1.2

[†] Separated/Divorced/Widowed

difficult. [29] However, some logical ordering of the effects in the model such as clear temporal sequence in this case and/or theoretical grounds can reduce this problem by focusing only on a *a priori* justified subset of the relationships between the variables analyzed.

Transformations and estimation method

Because of some respondents claiming large number of partners, the positive extreme of the distribution is heavy tailed. This adds more of the same problem already present with the lower end as majority of men had few partners. Power transformations indicated the cubic root transformation for number of partners, length of last living-in relationship and age of first sexual intercourse as it reduced relative multivariate kurtosis [30] to the value of 1.47. Two main estimation options were considered: accounting for violation of multivariate normality assumption needed for maximum likelihood (ML) or using asymptotic distribution free (ADF) estimates [30] which do not require this assumption. While the latter are prone to considerable bias unless the sample size is pretty large – of the order of few thousands at least – the former were shown to be reasonably robust with sample sizes of few hundreds. The robustness of CSA parameters and goodness of fit indices is not elaborated here as it is not the central theme of this work but it is discussed in other works. [26,30–38] The strategy adopted was to use both ML and ADF estimates and compare the results.

As there were few cases with missing values on any of the variables included in the CSA, these were excluded from the analysis. SAS software was used. [19]

Instrumental variables' selection

It was known from a previous analysis that a large number of reported sexual partners was associated with respondents' age, gender, marital status, age of first sexual intercourse and social class in a multivariate model. [8] Another analysis showed that in addition to the above effects, smoking and alcohol consumption in young men were associated with increased number of partners in last

five years. [39] Attending STD clinic, bodymass index, educational level, living with both natural parents until the age of 16, the length of last living-in relationship and sexual attitudes from the survey questions 40 and 41 were added to this selection of instrumental variables to be considered. Although many other factors have been shown to influence the number of sexual partners in other publications not cited here, an empirical rather than a purely theoretical basis guided the instrumental variables' selection in this work. It is important to stress that the aim of this work focused on the error part rather than on the structural relationship. Notwithstanding the importance of the latter both for estimating the error components and in its own right, the selection of instrumental variables was not aimed at explaining specific contributions of a wide range of social and psychological factors influencing the number of sexual partners.

Several combinations of instrumental variables were tried to check their impact on the CSA parameters of interest. Likelihood ratio test was used for the final selection of these variables in CSA.

Results

Basic descriptive statistics for the number of partners showed that more partners were reported for the more distant periods of time (table 1).

The latter were also characterized by sharp increase in variability. Essential information about other variables used in the model is presented in table 2.

Covariance structure model

Latent variables (circled on Figure 1) represent the true values estimated in the model. They are measured by observed variables i.e. reported number of partners for different time periods (in rectangles on Figure 1). The lack of accuracy is expressed as measurement errors affecting the number of partners reported. The two-digit subscripts for regression coefficients marked "b" for the temporal se-

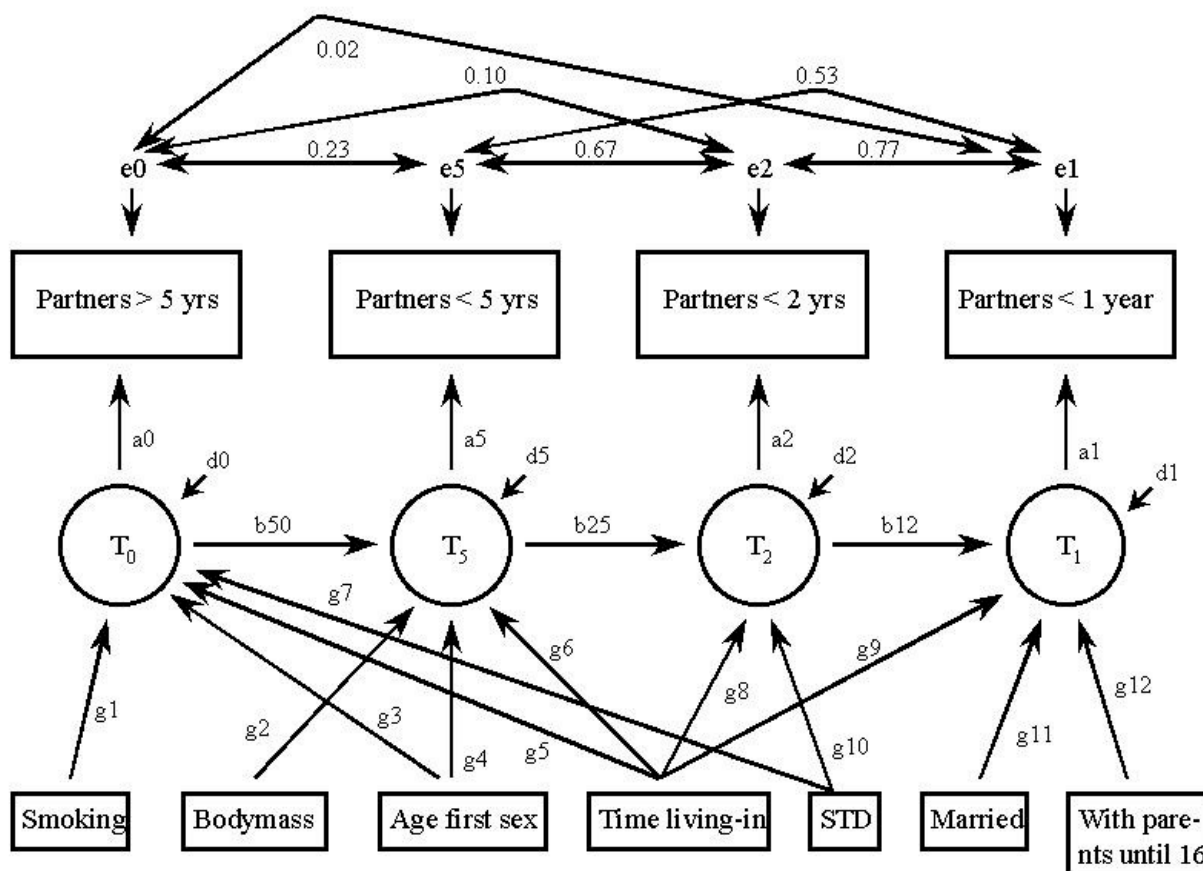


Figure 1
Correlations among measurement errors

quence of 'true' number of partners indicate the time frame of dependent (first digit) and independent variable (second digit).

The differences between ML and ADF estimates were in the second decimal place for few parameters, so it did not have any practical impact on the model interpretation. Therefore only ML estimates are reported here (table 3). The chi-square values for overall goodness of fit test was 13.3 with nine degrees of freedom and associated probability of 0.15, thus indicating the acceptability of the model. Root mean square error approximation [25] was 0.03 with 90% confidence interval upper bound of 0.06, which is considered a good fit. Goodness of fit index and its adjusted version [15] were 1.0 and 0.97 respectively. Probability of close fit [25,27] was 0.82. This alternative measure of fit assumes approximate rather than exact model fit. The latter is unrealistic to expect even if no structural misspecifications are present as measurement errors due to sampling variation would still indicate some discrepancy between the population and data based esti-

mates. Normalized residuals did not exceed the range from -2 to 2, thus showing good fit on the level of individual variables in the model. All these indices confirmed the acceptability of the model.

The magnitudes of error variances for reported number of partners were ordered as expected – the largest one for most distant past was more than six times the one for the last year (table 3). The correlations between measurement errors were considerably larger between adjacent time periods as compared to those further apart (Figure 1). High correlation of .77 was found between measurement errors for the number of partners in the last two years and in the last year. This was expected as the carry-over effect should be large for the large overlap of one year period between the questions on number of partners. This effect was somewhat smaller for the number of partners in the last two and in the last five years. When the time periods for the number of partners did not overlap (the last five years versus before that), the correlation between the measurement errors was .23.

Table 3: Maximum likelihood parameter estimates (standard error) for the model presented in Figure 1

Endogenous		Instrumental		Variances		Covariances	
code	estimate	code	estimate	code	estimate	code	estimate
a0	0.817 (0.053)	g01	-4.355 (0.028)	e0	0.537 (0.033)	c05	0.066 (0.013)
a5	0.864 (0.187)	g02	0.086 (0.033)	e5	0.147 (0.009)	c02	0.022 (0.010)
a2	0.772 (0.073)	g03	-0.137 (0.059)	e2	0.091 (0.006)	c25	0.078 (0.006)
a1	0.783 (0.050)	g04	0.689 (0.133)	e1	0.083 (0.005)	c01	0.005 (0.009)
b50	0.408 (0.029)	g51	-0.243 (0.073)	T0	0.504 *	c15	0.058 (0.006)
b25	0.709 (0.112)	g52	-0.014 (0.007)	T5	0.100 *	c12	0.067 (0.005)
b12	0.808 (0.049)	g53	0.914 (0.271)	T2	0.030 *		
		g21	0.058 (0.031)	T1	0.008 *		
		g22	-0.101 (0.049)				
		g11	0.090 (0.024)				
		g12	0.069 (0.038)				
		g13	-0.077 (0.031)				

* standard errors not estimated

Among exogenous variables, the length of the last living-in relationship was found to influence the estimated true number of partner for all the questions analyzed. The age of first sexual intercourse was a highly significant determinant of the number of partners more than five years ago and in the last five years but not afterwards (table 3 and Figure 1). For men aged 26–35, being married at the time of the interview and not living with both natural parents until the age of 16 was associated with increased number of partners reported in the last year.

The length of last living-in relationship was a significant predictor of number of partners for any past period. Longer live-in relationships were associated with fewer partners in the last five years and before that and larger number of partners in the last two years and last year (cf. the change of sign for the parameters g7 and g8 with respect to g5 and g6 in table 3). This change in the direction of influence paralleled the change in percentage of those reporting more than one partner while in living-in relationship which rose from 6% in the last five years to 18% in the last two years and 23% in the last year. Thus the living-in relationships in more distant past could have had the opposite effect of the more recent ones after adjusting for the other influences in the model.

The earliest sexual lifestyle in a temporal sequence can be regarded as baseline. No effect between the baseline and the number of partners for non-adjacent time periods was found significant (details not shown).

Another line of model modifications was testing two basic assumptions of the model obtained. Firstly, alternative model with the assumption of correlated "d" terms (unique factor variances for estimated 'true' number of partners) was tested. Structural relationships were fixed to the values obtained in table 3 to allow sufficient degrees of freedom for all the error terms' variances and covariances to be estimated. The chi-square value for this model (table 4) was 13.3 with 14 degrees of freedom. The measurement error estimates and their covariances were very similar to those obtained from the previous model. The unique variances of estimated true number of partners – the "d" terms – were approximately ten times smaller than the measurement errors estimates and their covariances were well within the range of the sampling variation around the zero value. The impact of these covariances on the overall goodness of fit was so small that fixing all of them to zero achieved practically the same value of chi-square as the model including these additional six parameters, thus giving clear preference to the simpler model.

Secondly, alternative model allowing the error in reported age of first sex to correlate with error terms of the reported number of partners was estimated. This modification was clearly rejected (chi-square (12) = 105.45, p = .0001). Estimated measurement error terms for self-reported number of partners ("e" terms) for >5, <5, <2 and <1 year ago, as well as their standard errors, were 0.750 (0.047), 0.151 (0.009), 0.095 (0.006) and 0.086 (0.005) respectively. The estimates of the correlation between these errors were almost identical to those in table 3. The

Table 4: Maximum likelihood parameter estimates (standard error) for a modification of the model presented in Figure 1 where "d" terms are allowed to be correlated

Variances				Covariances			
"e" terms		"d" terms		"e" terms		"d" terms	
code	estimate	code	estimate	code	estimate	code	estimate
e0	0.533 (0.023)	d0	0.006 (0.015)	c05	0.066 (0.013)	d05	-0.002 (0.006)
e5	0.145 (0.004)	d5	0.004 (0.004)	c02	0.022 (0.010)	d02	-0.001 (0.004)
e2	0.090 (0.002)	d2	0.004 (0.001)	c25	0.078 (0.006)	d25	-0.002 (0.003)
e1	0.080 (0.001)	d1	0.006 (0.001)	c01	0.005 (0.009)	d01	-0.001 (0.003)
				c15	0.058 (0.006)	d15	-0.000 (0.002)
				c12	0.067 (0.005)	d12	-0.001 (0.001)

measurement error estimate for the reported age of first sex was almost five times less than the smallest error term for the reported number of partners. Therefore it seems reasonable to conclude that both overall goodness of fit and specific error terms estimates give no support for the hypothesized relationship between error terms of reported number of partners and age of first sex with these data. Even if such remote hypothesis were accepted, the shift in estimates of error terms' variances and covariances would have probably been fairly small and of no practical importance in this case.

Bias analysis

The largest estimated bias was found among the men who reported large number of partners in the last five years, last two years and last year (Figure 2). The confidence intervals for those reporting ten or more partners in the last two years and last year were notably widened. It should be noted that only a small fraction of men aged 26–35 reported ten or more partners in last five years. Hence overall overreporting in this stratum is likely to be small for most practical purposes. A small percentage of men reporting ten or more partners in the last five years (8.2%), last two years (2.2%) or last year (0.2%) were estimated to have overreported between 3 and 4 partners on average, while overreporting for those who reported three to nine partners did not exceed one (Figure 2). On the other hand, two thirds of all men who reported one or two partners in last five years were likely to underreport slightly.

Estimated bias was very small with respect to age, age of first sex, visiting STD clinic, education, social class, marital status, length of living-in relationship, reporting a partner outside living-in relationship, smoking and alcohol consumption after adjusting for the effect of the number of

partners reported (details not shown). An interaction effect between number of partners on one hand and age of first sex and visiting an STD clinic on the other was also investigated. Negative moral connotation of STD and common British misconception that a person below the age of consent is himself/herself committing an offence [8] led to the hypothesis that social desirability of the answers to these questions could have biased the reports on number of partners. Self-reporting of socially disapproved behavior such as drug use has been shown to be biased downwards as a function of the time elapsed in repeated surveys. [40] However, no support for the above hypotheses was found with these data (details not shown) as the number of partners effect was far too dominant.

Discussion

Overreporting the number of partners was particularly large among men who reported five or more partners for more recent periods (Figure 2). This may be a result of so-called telescoping effect where events from the past are shifted towards more recent period during the recall process due to the compression of the time scale or because of heteroscedastic random measurement error due to the increase of error variance for more distant events. [41] In the latter case, more accurately recalled recent events result in reported event density such as locating sexual partners in time being skewed towards the present, i.e. the time of interviewing. Thus the net effect is overreporting the number of partners in last five years due to the redistribution of some partners acquired more than five years ago within the last five years period. This is consistent with Figure 2 and the estimates of measurement error variances for varying periods in the past (table 3).

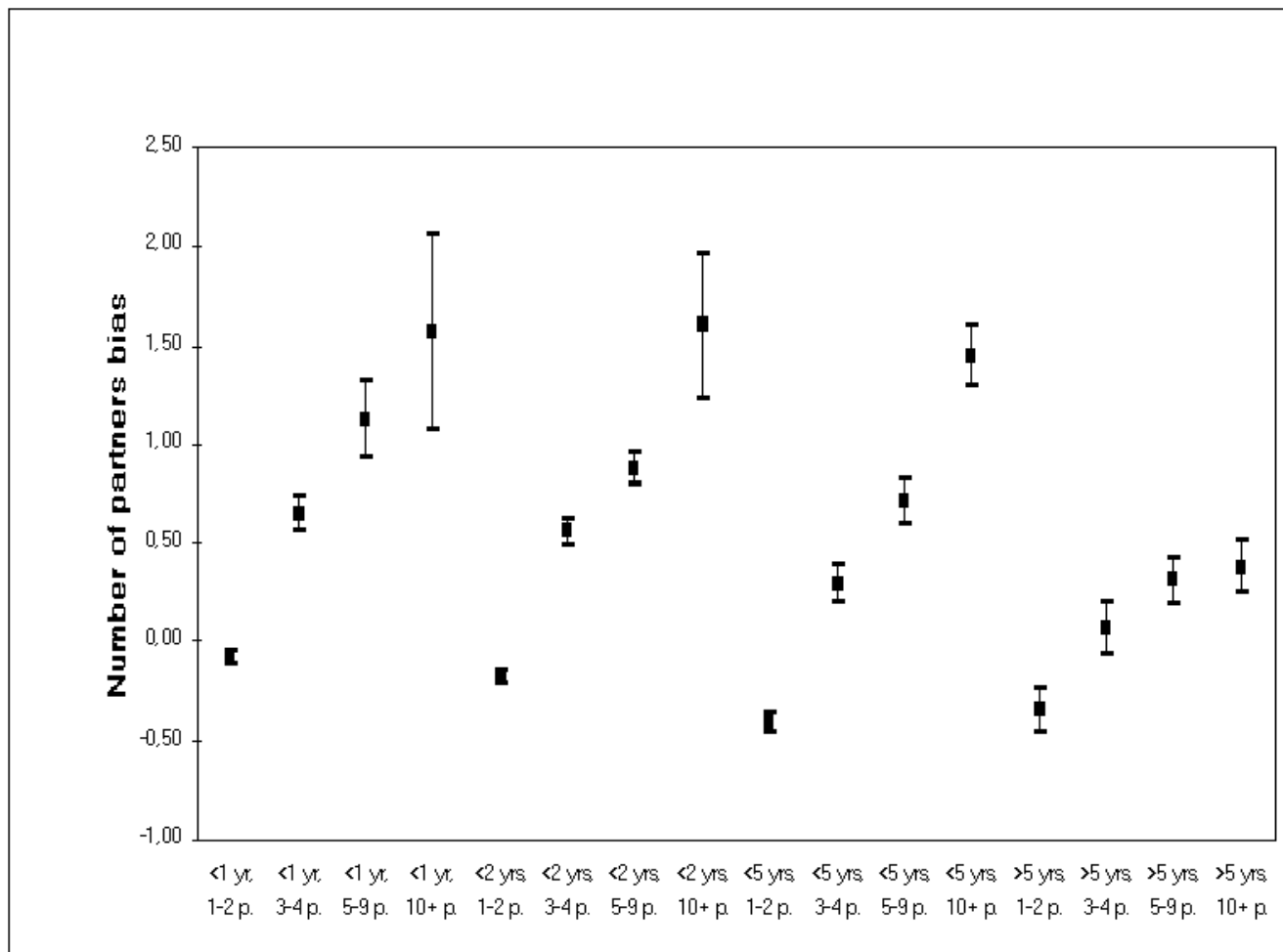


Figure 2
Bias estimates by period and reported number of partners Note: Words "year", "years" and "partners" abbreviated with "yr", "yrs" and "p.", respectively. Vertical axis (number of partners bias) is given on the cubic root scale.

The overall telescoping effect depends on both the rate of increase in error variance and true distribution of events. [41] As most of the partners were acquired in more distant past period when the recall was less accurate, this could explain the bias for men who reported three or more partners in the last five years, last two years and the last year (Figure 2). The bias clearly increases with the number of partners. Men with one or two partners for any period in the past were unlikely to have had serious difficulties in recalling the numbers accurately, so no heteroscedasticity effect could have contributed to the bias estimates. This again supports the view that the shift of recalled density for the number of partners towards the date of the interview can be explained in terms of heteroscedastic random error.

The bias estimation involved several simplifying assumptions. They seem reasonable in a rough and ready analysis with these data but may be too simplistic in other situations, particularly with small samples. The key problem is that the interval scale property of the population factor scores cannot be guaranteed by their empirical normal distribution in one sample – only ordinal scale can be assumed. [22] However, with reasonably large sample the asymptotic equivalence of the scales should be satisfactorily approximated. Another problem is quantification of complex CSA model uncertainty which factor scores are conditional upon. No simple confidence limits can be provided as there are many parameters, possibly estimated with variable precision, so that power considerations become fairly complicated. [42] Bayesian approach seems a natural framework for including this type of knowledge into prior distribution specification. However, many CSA

model parameters would considerably increase the computational burden in this case, which is a likely reason why so far only limited use of Bayesian statistics has been found in this area. [43] Bootstrapping CSA parameters [14] and calculating corresponding factor scores may be an alternative way of obtaining more realistic confidence intervals for bias scores. Gibbs sampling can also be used to construct the confidence limits of posterior means for residual scores of interest. [44] Other important developments in this field include an adaptation of CSA to deal with count data [45] and to treat counts as ordinal scale variables in CSA. [46]

The limitations of the use of instrumental variables were often understated in the literature, principally that for many realistic data in the econometrics, social sciences and psychology it is extremely difficult to ascertain the independence of the instrumental and endogenous variables in the model [21,48] because of confounding. Therefore at least a theoretical justification or preferably some empirical evidence for this assumption should be provided. [47] Testing robustness of the model obtained could be done by assessment of goodness of fit and parameter modifications when some model assumptions are altered, which may require fixing some parameters on the basis of the previous model in order to identify and estimate the others. This strategy is as reasonable as the certainty about the former but provides a simple measure of shift in the parameters of interest under various violations of model assumptions which may arise in practice. Alternatively, a full sensitivity analysis [48] may be used.

For the model presented, long term health problems and social class are examples of potential confounders both on theoretical and empirical grounds [8] which may influence selected instrumental variables and the number of partners reported. However, their addition to the model hardly changed the estimates of error terms variances and covariances. While it is impossible to ascertain that the same holds for the unknown confounders, this is an empirical issue which can be further explored by examining the robustness of the model with the data from other studies of a similar kind.

Colinearity among predictors may also influence the choice of instrumental variables. Generalized inverse of a Hessian matrix may be a simple computational solution for this problem and is available in many statistical software for CSA but there are other strategies if this one fails. [49] However, a sensible selection of instrumental variables on sound theoretical basis should precede more complicated ways of dealing with this issue.

Memory error and personal salience of the partners were reported as important sources of bias for the number of

partners in recent period of time in various studies. [1,3,8] Personal salience is likely to be more influential for more distant past events. For example, first sexual intercourse is a relatively vivid memory for most respondents as an event of personal importance despite the time elapsed, while some more recent partnerships which bear little relevance to the current emotional life may have faded away. Personal salience is usually culturally and/or ethnically contextualized, e.g. machismo or male adolescents overreporting of number of sexual partners could be a way of asserting their perceived social role. [3] However, this survey was conducted at the time of great concern for and media attention to AIDS epidemic in Great Britain. In addition, the age range 26–35 years does not include adolescents.

Large number of partners is difficult to report accurately, particularly if the time frame is long. The difficulty arises from the effort to convert episodic memory which operates with discrete time intervals into continuous time scale from which incidence is derived. [4–6,41] Motivation to search through complex patterns of sexual behavior for people with large number of partners is likely to be lower in the general population in comparison with some special groups at risk of AIDS such as gay men who may perceive the study as personally relevant. [3] The finding that overreporting is particularly large among men who reported highest number of partners in last year highlights the need for caution in interpreting mean values for a heterogeneous population. It also calls for some more focused research into sexual behavior of people with large number of partners.

The potential of CSA in the context of reliability analysis has been recognized [9,41,50,51] but is still largely of illustrative nature. When the contrast between exploratory and confirmatory models [52] is seen in relative rather than in absolute terms, i.e. dependent on the amount of knowledge accumulated in a particular area of research, a more realistic expectation of the interpretational gains based on these models may help assessing their performance in a wider range of applications. In this context, it is important to emphasize that CSA estimates of the true values and measurement errors of self-reports are not based on any new evidence with respect to the data at hand. The estimated bias of self-report calculated in this way should be seen as a possible scenario which is supported by the data analyzed, rather than as any sort of direct validation by some additional independent information. The latter provides much stronger evidence for generalizability of the findings than the former. However, the feasibility of gathering new data may well be limited. CSA can highlight the likely sources of bias which can then become a matter of a separate study. From this point of view, empirical cross-validation in terms of new data and statistical modeling of the existing data can be seen as complemen-

tary rather than mutually exclusive ways of addressing the issue of bias in self-reports.

Conclusions

CSA provided evidence of the impact of the telescoping effect on bias in self-reported number of sexual partners. Considerable positive bias was found for a small fraction of respondents who reported ten or more partners in the last year, last two years and last five years. This is consistent with the effect of heteroscedastic random error where majority of partners had been acquired in more distant past and therefore recalled less accurately than the partners acquired more recently to the time of interviewing. Memory errors of this type depend on the salience of the events recalled and are likely to be present in many areas of health research based on self-reported behavior.

Competing interests

None declared.

Authors' contributions

The single author of the manuscript carried out all the phases necessary to complete the work: he conceived of the study, drafted the manuscript, performed the statistical analysis and wrote the interpretation and discussion of the results.

References

- Boulton M: **The methodological imagination.** In: *Challenge and innovation: Methodological advances in social research on HIV/AIDS* (Edited by: Boulton M) London, Taylor & Francis 1994, 1-21
- Wadsworth J, Johnson AM, Wellings K, Field JJ: **What is in a mean – an examination of the inconsistency between men and women reporting sexual partnerships.** *J R Stat Soc A* 1996, **159**:111-123
- Catania JA, Gibson DR, Chitwood DD, Coates DJ: **Methodological problems in AIDS behavioural research: Influences of measurement error and participation bias in studies of sexual behaviour.** *Psychol Bull* 1990, **108**:339-362
- Friedman WJ: **Memory for the time of past events.** *Psychol Bull* 1993, **113**:44-66
- Gaskell G, Wright D, O'Muircheartaigh C: **Reliability of surveys.** *Psychologist* 1993, **6**(11):500-503
- Krosnick JA: **Response strategies for coping with the cognitive demands of attitude measures in surveys.** *Appl Cognit Psychol* 1991, **5**:213-236
- Nadeau R, Niemi RG: **Educated guesses: The process of answering factual knowledge questions in surveys.** *Public Opinion Quart* 1995, **59**:323-346
- Johnson AM, Wadsworth J, Wellings K, Field J: **Sexual Attitudes and Lifestyles.** Oxford, Blackwell Scientific Publications 1994
- Hoyt TW: **Rater bias in psychological research: When is it a problem and what can we do about it?** *Psy Method* 2000, **5**:64-86
- Hoyt WT, Kerns MD: **Magnitude and moderators of bias in observer ratings: a meta-analysis.** *Psy Method* 1999, **4**:403-424
- Kupek E: **Determinants of item non-response in a large national sex survey.** *Arch Sex Behav* 1998, **27**:581-594
- Kupek E: **Estimation of the number of sexual partners for the non-respondents to a large national survey.** *Arch Sex Behav* 1999, **28**:233-242
- Wadsworth J, Field J, Johnson AM, Bradshaw S, Wellings K: **Methodology of the National Survey of Sexual Attitudes and Lifestyles.** *J R Stat Soc Ser A Stat Soc* 1993, **156**:407-421
- Dunn G, Everitt B, Pickles A: **Modelling covariances and latent variables using EQS.** London, Chapman & Hall 1993
- Joreskog KG, Sorbom D: **LISREL VI.** Uppsala, Uppsala University Press 1981
- Scott Long J: **Covariance Structure Models: An Introduction to LISREL.** London, Sage Publications 1990
- Johnson RA, Wichern DW: **Applied multivariate statistical analysis.** Englewood Cliffs, Prentice Hall Int 1988
- Sharma S: **Applied multivariate techniques.** New York, John Wiley & Sons 1996
- Institute SAS: **Proc CALIS. SAS/STAT (R) User's Guide. Version 6** (Edited by: SAS Institute) Cary, NC, SAS Institute 1989, **2**:245-366
- Maddala GS: **Econometrics.** New York, McGraw-Hill 1977
- Heckman J, Robb R: **Alternative methods for solving the problem of selection bias in evaluating the impact of treatment on outcomes.** In: *Drawing inferences from self-selected samples* (Edited by: Wainer H) Berlin, Springer-Verlag 1986
- Bartholomew DJ: **Foundation of factor analysis: Some practical implications.** *British J Math Statist Psych* 1985, **38**:1-10
- Wood PW: **The effect of unmeasured variables and their interactions on structural models.** In: *Latent variables analysis* (Edited by: von Eye A, Clogg CC) London, Sage Publications 1994, 109-130
- Bollen KA, Stine RA: **Bootstrapping goodness-of-fit measures in structural equation models.** In: *Testing structural equation models* (Edited by: Bollen KA, Scott Long J) Newbury Park, Sage Publications 1993, 111-135
- Browne MW, Cudeck R: **Alternative ways of assessing model fit.** In: *Testing structural equation models* (Edited by: Bollen KA, Scott Long J) Newbury Park, Sage Publications 1993, 136-162
- Chou CP, Bentler PM, Satorra A: **Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study.** *British J Math Statist Psych* 1993, **44**:347-357
- MacCallum RC, Roznowski M, Mar CM, Reith JV: **Alternative strategies for cross-validation of covariance structure models.** *Multivariate Behav Res* 1994, **29**:1-32
- Gonzalez R, Griffin D: **Testing parameters in structural equation modeling: every "one" matters.** *Psy Method* 2001, **6**:258-269
- MacCallum RC, Roznowski M, Necowitz LB: **Model modification in covariance structure analysis: the problem of capitalization on chance.** *Psychol Bull* 1992, **111**:490-504
- Browne MW: **Asymptotically distribution-free methods for the analysis of covariance structures.** *British J Math Statist Psych* 1984, **37**:62-83
- Boomsma A: **The robustness of maximum likelihood estimation in structural equation models.** In: *Structural Modelling by Example* (Edited by: Cuttance P, Ecob R) Cambridge, Cambridge University Press 1987, 160-188
- Browne MW, Shapiro A: **Robustness of normal theory methods in the analysis of linear latent variable models.** *British J Math Statist Psych* 1988, **41**:193-208
- Satorra A, Bentler PM: **Model conditions for asymptotic robustness in the analysis of linear relations.** *Comput Stat Data Analysis* 1990, **10**:235-249
- Satorra A, Bentler PM: **Corrections to tests statistics and standard errors in covariance structure analysis.** In: *Latent variables analysis* (Edited by: von Eye A, Clogg CC) London, Sage Publications 1994, 399-419
- Bentler PM, Newcomb MD: **Linear structural modeling with nonnormal continuous variables. Application: Relations among social support, drug use, and health in young adults.** In: *Statistical models for longitudinal studies of health* (Edited by: Dwyer JH, Feinlieb M, Lippert P, Hoffmeister H) Oxford, Oxford University Press 1992, 132-160
- Gerbing DW, Anderson JC: **Monte Carlo evaluations of goodness-of-fit indices for structural equation models.** In: *Testing structural equation models* (Edited by: Bollen KA, Scott Long J) Newbury Park, Sage Publications 1993, 40-65
- Muthen BO: **Goodness of fit with categorical and other non-normal variables.** In: *Testing structural equation models* (Edited by: Bollen KA, Scott Long J) Newbury Park, Sage Publications 1993, 205-234
- Tanaka JS: **Multifaceted conceptions of fit in structural equation models.** In: *Testing structural equation models* (Edited by: Bollen KA, Scott Long J) Newbury Park, Sage Publications 1993, 10-39
- Kupek E: **Sexual attitudes and number of partners in young British men.** *Arch Sex Behav* 2001, **30**:13-27

40. Fendrich M, Vaughn CM: **Diminished lifetime substance use over time: an inquiry into differential underreporting.** *Public Opinion Quart* 1994, **58**:96-123
41. Pickles A, Pickering K, Taylor C: **Reconciling recalled dates of developmental milestones, events and transitions: a mixed generalized linear model with random mean and variance functions.** *J R Statist Soc A* 1996, **159**:225-234
42. Saris WE, Satorra A: **Power evaluations in structural equation models.** In: *Testing structural equation models* (Edited by: Bollen KA, Scott Long J) Newbury Park, Sage Publications 1993, 181-204
43. Raftery AE: **Bayesian model selection in structural equation models.** In: *Testing structural equation models* (Edited by: Bollen KA, Long JS) Newbury Park, Sage Publications 1993, 163-180
44. Goldstein H, Spiegelhalter D: **League tables and their limitations: Statistical issues in comparisons of institutional performance.** *J R Statist Soc A* 1996, **159**(3):385-443
45. Cameron AC, Trivedi PK: **Regression analysis of count data.** Cambridge, Cambridge University Press 1998
46. Muthen LK, Muthen B: **Mplus user's guide.** Los Angeles, Muthen & Muthen 1998
47. Heckman J: **Comment.** *J Amer Statist Assoc* 1996, **91**:459-462
48. Angrist JD, Imbens GW, Rubin DB: **Identification of causal effects using instrumental variables.** *J Amer Statist Assoc* 1996, **91**:444-455
49. Wothke W: **Nonpositive definite matrices in structural modeling.** In: *Testing structural equation models* (Edited by: Bollen KA, Long JS) Newbury Park, Sage Publications 1993, 256-293
50. Dunn G: **Design and analysis of reliability studies.** *Stat Methods Med Res* 1992, **1**:123-157
51. Plummer MT, Clayton DG: **Measurement error in dietary assessment: an investigation using covariance structure models.** *Stat Med* 1993, **12**:925-948
52. Joereskog KG: **Testing structural equation models.** In: *Testing structural equation models* (Edited by: Bollen KA, Scott Long J) Newbury Park, Sage Publications 1993, 294-316

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/2/14/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

http://www.biomedcentral.com/info/publishing_adv.asp

