

Bias analysis to improve monitoring an HIV epidemic and its response: approach and application to a survey of female sex workers in Iran

Ali Mirzazadeh,^{1,2,3} Mohammad-Ali Mansournia,¹ Saharnaz Nedjat,⁴ Soodabeh Navadeh,^{1,3} Willi McFarland,² Ali Akbar Haghdoost,^{3,5} Kazem Mohammad¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2013-202521>).

¹Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

²Global Health Sciences, University of California, San Francisco, California, USA

³Regional Knowledge Hub for HIV/AIDS Surveillance, WHO Collaborating Center, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran

⁴Department of Epidemiology and Biostatistics, School of Public Health, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran

⁵Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

Correspondence to

Professor Kazem Mohammad, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran 14155-6446, Iran; mohamadk@tums.ac.ir

Received 17 February 2013

Revised 17 April 2013

Accepted 26 May 2013

To cite: Mirzazadeh A, Mansournia M-A, Nedjat S, et al. *J Epidemiol Community Health* Published Online First: [please include Day Month Year] doi:10.1136/jech-2013-202521

ABSTRACT

Background We present probabilistic and Bayesian techniques to correct for bias in categorical and numerical measures and empirically apply them to a recent survey of female sex workers (FSW) conducted in Iran.

Methods We used bias parameters from a previous validation study to correct estimates of behaviours reported by FSW. Monte-Carlo Sensitivity Analysis and Bayesian bias analysis produced point and simulation intervals (SI).

Results The apparent and corrected prevalence differed by a minimum of 1% for the number of 'non-condom use sexual acts' (36.8% vs 35.8%) to a maximum of 33% for 'ever associated with a venue to sell sex' (35.5% vs 68.0%). The negative predictive value of the questionnaire for 'history of STI' and 'ever associated with a venue to sell sex' was 36.3% (95% SI 4.2% to 69.1%) and 46.9% (95% SI 6.3% to 79.1%), respectively. Bias-adjusted numerical measures of behaviours increased by 0.1 year for 'age at first sex act for money' to 1.5 for 'number of sexual contacts in last 7 days'.

Conclusions The 'true' estimates of most behaviours are considerably higher than those reported and the related SIs are wider than conventional CIs. Our analysis indicates the need for and applicability of bias analysis in surveys, particularly in stigmatised settings.

INTRODUCTION

To monitor the national response and achievement of prevention targets, researchers and public health officials spend considerable resources estimating and tracking HIV risk using biological and behavioural surveys worldwide.¹ However, risk behaviour is reliant upon self-reported information which is highly vulnerable to social desirability response bias.²

In Iran sex work may be particularly stigmatised. Sex between persons who are not married to each other is illegal. Sex work therefore occurs underground and it may be extremely difficult to get accurate reports of behaviour by female sex workers (FSW). We believe the situation is similar in other parts of the world, but perhaps to a greater degree in Iran.

Thus, social stigma surrounding risky behaviours, especially those related to illegal sexual contact, complicate behavioural surveys, affecting the validity of responses and leading to bias in estimating

the prevalence of key behaviours.^{3,4} To obtain valid estimates, researchers should therefore calibrate and triangulate the apparent or self-reported estimates, considering both random and systematic errors. Typically, researchers only report the possible and likely effects of biases on their study briefly in a section on limitations rather than attempting to calculate and correct for biases in their principal findings. Moreover, the literature presents much discussion on how to measure and quantify random error with less discussion on systematic error or bias.^{5,6}

Sensitivity analysis is one naïve method that has been used to measure the amount of uncertainty around effects according to systematic measurement errors. The approach has been advanced to address potential confounding and selection biases, and also to consider the uncertainty about the bias parameters when correcting apparent effects.⁷ Other methods range from probabilistic sensitivity analysis⁸ to semi- and full- Bayesian techniques⁹ which produce more plausible CIs (intervals that include the true effects in the population) for the corrected effects. Although recent efforts are beginning to expand upon these new techniques, there is a gap in applying them, particularly for numerical outcomes. In the field of HIV/AIDS, which is always intimately linked to monitoring and measuring socially stigmatised behaviours, the application of such techniques is notably lacking.

The present study aims to describe probabilistic and Bayesian bias correction techniques for categorical and numerical measures and to empirically apply and compare them on several HIV-related stigmatised risky behaviours in a highly marginalised population. To better understand the pitfalls of the naïve methods and the process of bias adjustment, we also present and compare the above advanced techniques with the naïve methods.

METHODS

Data and study subjects

From April to July 2010, 872 FSW were recruited in the first national biobehavioural survey to track the risk of HIV in key populations. They were recruited from 21 health facilities in 14 cities by convenience sampling. At each facility one trained female interviewer identified FSW routinely getting services and, after verbal informed consent, those meeting the eligibility criteria and agreeing to participate were enrolled in the survey. Risk behaviour

questions were asked by the interviewer in a face-to-face interview (FTFI) and answers were recorded. All procedures were approved by the Research Review Board of the Kerman University of Medical Sciences.

Completed questionnaires were entered into STATA V10 for data cleaning, quality checks and data analysis. The point estimates and CI for each risky behaviour were calculated by the standard survey analysis package. These estimates were called the apparent prevalence (AP) measures because they were as reported by FSW.

Adjustment methods for systematic error in reported categorical behaviour measures

To adjust estimates for misclassification bias influencing the validity of reported behaviours by the FSW, we acquired bias parameters (sensitivity (Se) and specificity (Sp)) from a previous validation study presented in detail elsewhere.² The validation study was an independent survey conducted among 63 FSW in Tehran and Kerman (ie, they were not FSW who participated in the larger national survey) in which the participants were interviewed by FTFI as well as by in-depth interviews (IDI). These FSW were consecutively approached at the same clinics in the two cities and enrolled in the validation study using the same eligibility criteria. The interviewers conducted an IDI with each of the 63 FSW as a cognitive cross-check of their answers collected in the FTFI. The IDI was an open-ended interview and began with mutual trust building questions about their general living conditions, health status and social welfare needs as a consultative interview. This time the interviewer did not apply the questionnaire as step-by-step reading of the questions but rather followed a natural discussion leading to more private topics according to the participants' comfort and lead. IDI responses were recorded by short notes and later transferred and aligned to the FTFI questionnaire in a separate column after finishing the IDI. The IDI was considered as the 'gold standard' for the purposes of this study and the bias parameters were calculated accordingly. Our assumption is that the responses given to

professional interviewers in an open discussion—with a longer period to establish trust, check internal consistency and with the final answer based on the interviewers' judgement—would provide a more truthful response than a structured series of close-ended questions.

The bias parameters are presented as Se and Sp for categorical behaviour variables (table 1). We applied two methods of bias analysis: a naïve sensitivity analysis and a Bayesian approach.

In the naïve method we used the following equation to correct the biased estimates:

$$(TP)True\ Prevalence = (AP + Sp - 1)/(Se + Sp - 1) \quad (1)$$

Since we do not know the true amount of the parameters in the above model, instead we used the estimates:

$$\widehat{TP} = (\widehat{AP} + \widehat{Sp} - 1)/(\widehat{Se} + \widehat{Sp} - 1)$$

Because of the uncertainty around the parameters (ie, Se and Sp), we re-ran the above model with different combinations of Se and Sp within the range of their lower and upper 95% CI (see online supplementary appendix 1). This provided a bound CI around the true prevalence (TP) of each risk behaviour. We called this method a naïve sensitivity analysis approach to bias analysis.

To apply the Bayesian methods for bias correction we assumed that the bias parameters (Se and Sp) probabilities have β distributions. These are presented in online supplementary appendix 2 (with the shape parameters of a.Se, b.Se for sensitivity and a.Sp, b.Sp for specificity, respectively). β Distributions are a type of probability distribution that are frequently used to describe prior uncertainty about prevalence, Se or Sp in the Gibbs sampler.¹⁰ We defined the β distribution by two parameters, α and β , with $\alpha=y+1$ and $\beta=n-y+1$, where y is the number of events out of n trials. This was done for Se and Sp separately and their β distributions were defined for every risky behaviour. We have assumed that there is no informative prior distribution for the TP of risky behaviours and so we assign a uniform prior distribution to TP. To obtain posterior

Table 1 Bias parameters in behaviours between IDI and FTFI with female sex workers in the validation study (N=63), Iran, 2010–2011

Risk behaviours Categorical	Bias parameters	
	Sensitivity (95% CI)*	Specificity (95% CI)†
Ever married	92.0 (80.8 to 97.8)	91.7 (61.5 to 99.8)
Ever used drugs	90.2 (79.8 to 96.3)	100.0 (15.8 to 100)
Not tested for HIV, last 12 months	93.3 (77.9 to 99.2)	84.8 (68.1 to 94.9)
Not receiving the result of the HIV test	66.7 (22.3 to 95.7)	100.0 (86.3 to 100)
History of genital ulcer or discharge, last 12 months	63.9 (46.2 to 79.2)	100.0 (87.2 to 100)
Non-condom use last sex act	87.5 (67.6 to 97.3)	91.7 (77.5 to 98.2)
Ever associated with a venue to sell sex	52.4 (29.8 to 74.3)	97.6 (87.1 to 99.9)
Numerical	Mean difference between IDI and FTFI (95% CI)‡	Pearson correlation coefficient (95% CI)§
Number of sexual contacts, last 7 days	1.48 (0.7 to 2.3)	−0.2 (−0.4 to 0.1)
Number of non-condom use acts, last 7 days	0.40 (0.0 to 0.7)	−0.2 (−0.5 to 0.1)
Age at first sex act for money, drugs or shelter	0.05 (−1.1 to 1.2)	−0.5 (−0.7 to −0.3)
Number of clients, last 7 days	0.80 (0.2 to 1.5)	−0.3 (−0.6 to −0.1)
Number of days in last week with sex with a client	0.90 (0.2 to 1.5)	−0.4 (−0.6 to −0.1)
Income through last commercial sex act (US\$)	1.40 (−0.6 to 3.4)	−0.2 (−0.4 to 0.1)

*Sensitivity is the conditional probability, p (classified as FSW having the risky behaviour for HIV|truly having the risky behaviour).

†Specificity is the conditional probability, p (classified as FSW without the risky behaviour for HIV|truly don't have the risky behaviour).

‡Positive mean difference represents under-reporting in FTFI.

§Pearson correlation coefficient between FTFI and the mean difference between IDI and FTFI. FSW, female sex workers; FTFI, face-to-face interview; IDI, in-depth interview.

Table 2 Apparent and true prevalence of risky behaviours by naïve sensitivity and Bayesian methods in female sex workers participating in a national behavioural survey (N=872), Iran, 2010

Risky behaviours prone to social desirability bias (categorical variables)	Apparent prevalence		True prevalence (naïve sensitivity)		True prevalence (Bayesian)	
	%	95% CI	%	95% CI	%	95% CRI
Ever married	81.5	71.5 to 88.6	87.5	73.7 to 101.7*	87.7	77.7 to 97.9
Ever used drugs	71.6	59.7 to 81.1	79.4	*-787.5 to 3150.0*	69.6	22.9 to 87.5
Not tested for HIV, last 12 months	64.1	51.9 to 74.7	62.6	48.7 to 80.8	64.0	51.6 to 78.4
Not receiving the result of the HIV test	21.0	9.7 to 39.8	31.5	9.0 to 93.6	36.8	14.0 to 82.0
History of genital ulcer or discharge, last 12 months	52.9	39.3 to 66.1	82.8	61.3 to 120.1*	82.1	65.3 to 98.0
Non-condom use last sex act	36.7	28.0 to 46.8	36.1	19.4 to 52.5	35.8	22.9 to 48.5
Ever associated with a venue to sell sex	35.5	25.8 to 46.6	66.3	37.1 to 133.7*	68.0	44.9 to 95.7

*Impossible values.
CRI, credible interval.

distributions for TP, positive predictive value (PPV) and negative predictive value (NPV), we ran a model (see online supplementary appendix 2) in WinBUGS V1.4¹¹ with 50 000 iterations and report the posterior mean and 95% credible intervals (CRI) (by percentile 2.5% and 97.5%) for the TP, PPV and NPV from iterations 40 001–50 000. The results are summarised in table 2 (column 3) and figure 1.

Adjustment for systematic error in reported numerical behaviour measures

In order to correct the reported numerical behaviours (eg, number of sex acts in last 7 days) for measurement error, we also acquired the bias parameters from the validation study, this time as mean differences between the findings in the IDI and the routine FTFI and also the Pearson correlation coefficient between the differences (IDI – FTFI) and the reported amount in the FTFI.² The bias parameters for the numerical measures are presented in table 1. Monte-Carlo Sensitivity Analysis (MCSA) was used to correct the bias in numerical risk behaviours.

MCSA model

The theory behind MCSA is that values from the bias parameters (in this case, the difference between observed and true

numerical behaviour means and the correlation coefficient between them) are randomly selected from their assigned probability distributions and then used to solve the bias correction equations to produce bias-adjusted estimates.

As an example, we observed that, on average, FSW report their ‘number of sex acts in the last week’ as 1.48 (0.7 to 2.3) fewer than in reality.² We called this Delta (D). The true mean (TM) would be a linear function of apparent mean (AM), which is the mean of the reported behaviour and the amount of bias (D):

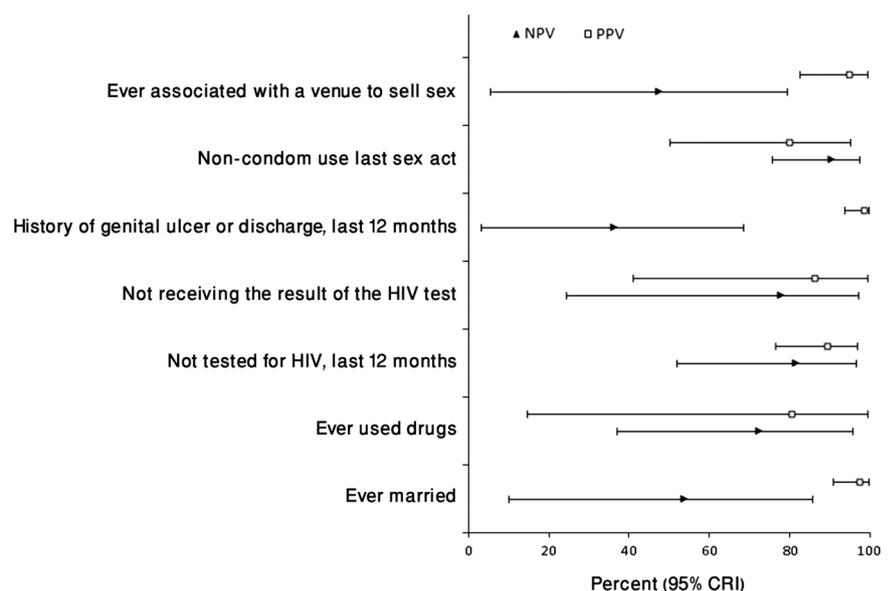
$$TM = AM + D$$

Since we do not know the true amount of the parameters (AM and D) in the above equations, instead we used the estimates:

$$\widehat{TM} = \widehat{AM} + \widehat{D} \quad (2)$$

To estimate the TM from the known normal probability distribution of AM and D (in the validation study), considering both random and systematic errors, we randomly picked two numbers and solved equation 2. Finally, we made an average of the entire simulated TM as the adjusted true estimate. We did this simulation by running a model in WinBUGS (see online supplementary appendix 3).

Figure 1 Posterior negative predictive value (NPV) and positive predictive value (PPV), point estimate and 95% credible interval for behaviours reported by female sex workers (FSW) in face-to-face interviews, Iran, 2010 (N=872). NPV (lower grey bars) is the conditional probability, p (truly do not have the risky behaviour|classified as FSW without the risky behaviour). PPV (upper white bars) is the conditional probability, p (truly having the risky behaviour|classified as FSW having the risky behaviour).



The model randomly selects a number from the distribution of bias parameter (D), with the mean and precision (1/variance) of 1.4 and 6.2 for ‘number of sex acts in the last week’, and adds it to a randomly selected measure from the distribution of ‘number of sex acts in the last week’ (AM), with the mean and precision of 2.2 and 6.7 reported by FSW recruited in the national survey. We repeated this for 50 000 iterations and, in every iteration, we summed these two numbers.

AM and D are correlated, and this should be considered when simulating the data. Suppose that the Pearson correlation coefficient between AM and D is R. In order to produce the two correlated distributions (for AM and D), we first created two independent random variables (X and Y) which have standard normal distributions. Given the Pearson correlation coefficient of R, we computed Z as:

$$Z = (R \times X) + \sqrt{(1 - R^2)} \times Y$$

Z has a standard normal distribution and is correlated with X with the correlation coefficient of R. The last step is to transform X and Z standard distributions into AM (mean=AM_{mean} and standard division=AM_{SE}) and Delta (mean=D_{mean} and standard division=D_{SE}) by:

$$\begin{aligned} AM &= AM_{\text{mean}} + AM_{\text{SE}} \times X \\ D &= D_{\text{mean}} + D_{\text{SE}} \times Z \end{aligned}$$

We have AM and D (correlated with the Pearson coefficient of R) and now we can calculate the true mean based on equation 2. The best estimate for R is calculated from the validation study (table 1). We applied the Fisher z transformation to reproduce its normal distribution in WinBUGS (see online supplementary appendix 3). At each iteration we drew a new value for R (consistent with the evidence from the validation study).

The above model was also used for other numerical behaviour measures. Finally, for all the numerical measures we calculated the mean and 95% simulation intervals (by percentile 2.5% and 97.5%) from iterations 40 001–50 000 and the results are summarised in table 3 (column 2 with MCSA label).

RESULTS

The 872 FSW recruited in the national behavioural survey were of average age 31.7 (95% CI 29.7 to 33.7) years (table 4). According to self-report, 81.5% (95% CI 71.5% to 88.6%) had ever been married while only 35.9% (95% CI 25.7% to 47.5%) were still in a marital union. The education level tended to be low, with most not having completed secondary education.

Table 4 Demographic characteristics of female sex workers recruited for a national behavioural survey (N=872), Iran, 2010

Variables	Estimates	95% CI
Mean age (years)	31.7	29.7 to 33.7
Ever married (%)	81.5	71.5 to 88.6
Current marital status (%)		
Single	19.8	12.5 to 30.0
Married	35.9	25.7 to 47.5
Divorced	22.2	16.8 to 28.8
Other	22.1	16.1 to 29.6
Education (%)		
Illiterate	14.7	8.3 to 24.6
Literate/primary school	30.1	21.8 to 39.9
Middle school	26.6	22.5 to 31.2
High school	24.4	16.6 to 34.4
University degree	4.2	2.0 to 8.9

In table 2 the AP (ie, reported in FTFI) and TP (ie, corrected for bias by naïve sensitivity and Bayesian methods) for categorical risky behaviours are presented. For ‘never tested for HIV in last 12 months’ and ‘non-condom use last sex act’, the corrected estimates were about 1–2% lower than those reported by the FSW. For all other behaviours the TP was higher than that reported. The differences between the AP and TP varied from <1% for ‘non-condom use sex act’ to 33% for ‘ever associated with a venue to sell sex’. CIs were wider with both the naïve and Bayesian approaches than with the analysis which only considered random error. For four behaviours the upper CI calculated by the naïve sensitivity method exceeded the plausible limit (ie, >100%), while all the Bayesian interval bounds fell within plausible ranges. The adjusted point estimates for both the naïve and Bayesian methods were similar for all behaviours except for ‘ever used drugs’ (79.4% vs 69.6%) and ‘not receiving the result of the HIV test’ (31.5% vs 36.8%).

The PPV and NPV (calculated from the posterior distribution) of the responses to the FTFI for seven categorical behaviours are illustrated in figure 1. The PPV varied from the lowest level of 80.5% (95% CRI 54.7% to 95.0%) for ‘non-condom use last sexual act’ to 97.4% (95% CRI 90.4% to 99.8%) for ever being married.

The NPV of the behavioural questionnaire for a history of symptoms of sexually transmitted infections and ‘ever associated with a venue to sell sex’ had the lowest levels of 36.3% (95% CRI 4.2% to 69.1%) and 46.9% (95% CRI 6.3% to 79.1%),

Table 3 Apparent and true mean of some stigmatised numerical risky behaviours by MCSA, female sex workers participating in a national behavioural survey (N=872), Iran, 2010

Risky behaviours prone to social desirability bias (numerical variables)	Apparent mean		True mean by MCSA	
	Mean	95% CI	Mean	95% SI*
Number of sexual contacts, last 7 days	2.3	1.5 to 3.1	3.8	2.8 to 4.8
Number of times condoms used, last 7 days	2.6	1.8 to 3.3	2.9	2.2 to 3.6
Age at first sex act for money, drugs or shelter	24.8	23.0 to 26.6	24.9	23.3 to 26.4
Number of clients, last 7 days	1.5	0.9 to 2.1	2.3	1.6 to 3.0
Number of days last week doing commercial sex	2.9	2.3 to 3.4	3.7	3.1 to 4.3
Income through last commercial sex act (US\$)	20.4	12.0 to 28.8	21.8	14.0 to 29.4

*Bounds are 2.5% and 97.5% percentiles. MCSA, Monte-Carlo sensitivity analysis; SI, simulation interval.

respectively. The best NPV was recorded for 'non-condom use last sexual contact' at 90.3% (95% CRI 76.4% to 97.8%).

As illustrated in table 3, after adjusting for systematic errors by the MCSA technique, estimates of all numerical risky behaviour variables increased from a minimum of 0.1 year for 'age at first sex act for money' to a maximum of 1.5 for 'number of sexual contacts in last 7 days'. In all cases the correction for measurement errors yielded higher estimates.

DISCUSSION

Bias analysis quantifies the influence of systematic error on the estimation of association in an epidemiology study.¹² We have elaborated further the applicability of bias analysis in cross-sectional behavioural surveys to provide the best estimates of risky behaviour prevalence rather than associations. Both our naïve and Bayesian techniques for adjustment of measurement error produce point estimates for sensitive risk behaviours which are considerably higher than those reported by FSW in FTFIs. To our knowledge, adjustment techniques for bias, particularly for numerical outcomes, are very seldom elaborated upon in the literature.^{7 9 13 14} We propose a model for adjusting measurement errors which is applicable to surveys of high-risk populations measuring highly stigmatised behaviours.

Bias analysis has been proposed mainly to obtain an estimate of the direction and the magnitude of the bias and to reduce the human tendency of overconfidence in findings from a particular study. In line with Lash *et al.*,¹⁵ we wish to emphasise the third objective of the bias analysis which is to guide new data collection. We have demonstrated in a behavioural survey (in this case, an HIV-related survey) that answers to sensitive questions are affected differently and to a considerably high level because of the stigma surrounding them. How large the effects are and how to adjust for them highlight the need for more careful attention in the data collection process and consideration of whether additional validation processes should be implemented. As an empirical example, our study audience is people working in the field of public health and HIV to encourage calibration steps.

Our general additive model (equation 2), considering both random and systematic types of errors, appears to be a simple but effective model to provide adjusted estimates for numerical variables. The steps are similar to the analytical approach described by Lash *et al.*,¹² applying the distribution of bias parameters to take into account the bias and random errors and summarising the final corrected estimate as a point with its simulation interval.¹⁵

We recognise that the two approaches have limitations. First, in the naïve method it was possible to have implausible values (ie, upper limits >100%) for the corrected prevalence when using Se and Sp in equation 1 since the uncertainty around AP was not considered properly. In contrast, the Bayesian technique considers the random errors around the AP as well as the uncertainty around the bias parameters (Se and Sp) to produce the adjusted estimates conditional to the prior TP which is bound by 0% and 100%.⁹ This condition on the outcome removes the need for $Se+Sp-1$ to be positive.¹⁴

Second, we note that the TP of 'ever used drugs' corrected by the Bayesian technique was considerably lower than the naïve method. This was probably due to the large degree of uncertainty around one of the bias parameters for this variable (ie, Sp 15.8–100%). The effect on the prevalence point estimate was masked in the naïve method, indicating that, for bias analysis, we need a perfect or at least a most probable guess of the bias parameters.¹³ For the plausible range of the TP, the uncertainty

around the bias parameters (Se, Sp) translated into impossible values with an unintermittable range for the naïve method.

Third, we have assumed the bias parameters to be equal for all subgroups of the survey, but this may not always be true. Moreover, considering a uniform prior distribution for the TP could be improved by replacing it with more plausible or realistic distributions. Unfortunately, access to such extra information is typically very limited. We therefore recommend that a sensitivity analysis should be performed to see how the final bias-adjusted estimates vary according to changes in the prior distributions of the target estimates, that the exact distributions of bias parameters (instead of approximate parametric distributions) should be employed and that different link functions between random and systematic error be used in the sensitivity analysis formula.

Finally, some may find the corrections for numerical variables too small and therefore not necessary to be considered. However, the magnitude of bias in most variables we assessed can result in substantial changes affecting the potential of the epidemic. For example, the number of sexual contacts in the last week was corrected from 2.3 to 3.8 per week. As a corrected rate, expanding the time frame illuminates the difference between the biased and corrected estimates, which is about 78 sexual contacts per year.

As part of any well-planned epidemiological study, researchers need to design and implement the bias analysis carefully, ensuring that the information on bias parameters is valid and sufficiently precise when used to adjust the observed estimates and also that the bias adjustment equations are properly chosen.^{12 15} Investigators planning behavioural surveys aimed at collecting stigmatised behaviours must also consider threats to the validity of responses and plan for a systematic validation study, even as a small subsample of the whole study, to collect the required data on bias parameters. Having a quality assurance strategy for HIV serological testing to make sure the test results are accurate is well-accepted practice. The same concept should apply to behavioural data upon which prevention planning and impact assessment depend.

In conclusion, we believe that adjusting for measurement bias in sensitive behavioural studies will yield a higher prevalence of risky behaviours most of the time which, given the social desirability response bias, is more likely to approach the truth. We further believe that both Bayesian and MCSA methods provide more plausible corrected estimates and therefore recommend their use in the analysis of bias and its effects on key measures of risk in surveys of hidden, hard-to-reach and marginalised populations. In addition to other types of study designs such as case-control studies, cross-sectional surveys can also benefit from applying bias analysis.

What is already known on this subject

- ▶ Policymakers and health authorities need accurate estimates of risky behaviours to better monitor the response to an HIV epidemic.
- ▶ HIV self-reported risk behaviours are mostly stigmatised issues and are prone to under-reporting (bias).
- ▶ The literature presents much discussion on how to measure and quantify random error with less discussion on systematic error or bias.

What this study adds

- ▶ There is a clear need for and applicability of bias analysis in surveys, particularly in stigmatised settings.
- ▶ The two proposed bias analysis methods can be applied to produce bias-adjusted estimates for both categorical and numerical HIV self-reported risk behaviours.

Acknowledgements The authors are indebted to their colleagues from the medical universities for helping in recruitment, interviewing and supervising the national biobehavioural survey in female sex workers.

Contributors AM contributed to the conception and design of the study; acquisition, analysis and interpretation of the data; drafting the article and revising based on reviewer comments. M-AM made substantial contributions to the conception, analysis and interpretation. SaN contributed to the conception and design of the study, interpretation of the data and revised the article for important intellectual content. SN contributed to the analysis and interpretation of the data and revised the article for important intellectual content. WM contributed to the conception, design, analysis and interpretation of data and revised the article for important intellectual content. AAH contributed to the conception of the study, interpretation of the data and revised the article for important intellectual content. KM contributed to the conception and design of the study, interpretation of data and revised the article for important intellectual content.

Funding This work (as a PhD thesis for AM) was supported jointly by HIV Research Trust (grant no. HIVRT11-052), Tehran University of Medical Sciences (grant no. 240/1626) and Regional Knowledge Hub for HIV/AIDS Surveillance—WHO collaborating center based at Kerman University of Medical Sciences (grant no 90/122).

Competing interests None.

Ethics approval All procedures were approved by the Research Review Board of the Kerman University of Medical Sciences.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Ghys PD, Jenkins C, Pisani E. HIV surveillance among female sex workers. *AIDS* 2001;15(Suppl 3):S33–40.
- 2 Mirzazadeh A, Haghdoost AA, Nedjat S, *et al.* Accuracy of HIV-related risk behaviors reported by female sex workers, Iran: a method to quantify measurement bias in marginalized populations. *AIDS Behav* 2013;17:623–31.
- 3 Cullinane J. The domestication of AIDS: stigma, gender, and the body politic in Japan. *Med Anthropol* 2007;26:255–92.
- 4 Tucker JD, Ren X. Sex worker incarceration in the People's Republic of China. *Sex Transm Infect* 2008;84:34–5; discussion 6.
- 5 Low PC. Low P-values or narrow confidence intervals: which are more durable? *Epidemiology* 2001;12:291–4.
- 6 Rothman KJ, Greenland S, Lash TL. Precision and statistics in epidemiologic studies. In: *Modern epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins 2008:149–68.
- 7 Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003;14:459–66.
- 8 Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol* 2005;34:1370–6.
- 9 MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian? *Epidemiology* 2012;23:151–8.
- 10 Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol* 1995;141:263–72.
- 11 Lunn DJ, Thomas A, Best N, *et al.* WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000;10:325–37.
- 12 Lash TL, Fox MP, Fink AK. *Applying quantitative bias analysis to epidemiologic data*. New York: Springer, 2009.
- 13 Gustafson P, Le N, Saskin R. Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics* 2001;57:598–609.
- 14 Rothman KJ, Greenland S, Lash TL. Bias analysis. In: *Modern epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins 2008:354–6.
- 15 Lash TL, Ahern TP. Bias analysis to guide new data collection. *Int J Biostatistics* 2012;6:2.