

# Accuracy of HIV-Related Risk Behaviors Reported by Female Sex Workers, Iran: A Method to Quantify Measurement Bias in Marginalized Populations

## Precisión de las conductas de riesgo relacionadas con el VIH reportadas por trabajadoras del sexo en Irán: un método para cuantificar el sesgo de medición en las poblaciones marginadas

Ali Mirzazadeh · Ali Akbar Haghdoost ·  
Saharnaz Nedjat · Soodabeh Navadeh ·  
Willi McFarland · Kazem Mohammad

© Springer Science+Business Media, LLC 2012

**Abstract** We quantified discrepancies in reported behaviors of female sex workers (FSW) by comparing 63 face-to-face interviews (FTFI) to in-depth interviews (IDI), with corroboration of the directions and magnitudes of reporting by a panel of psychologists who work with FSW. Sensitivities, specificities, positive and negative predictive values (PPV and NPV) were assessed for FTFI responses using IDI as a “gold standard”. Sensitivities were lowest in reporting symptoms of sexually transmitted infections

(63.9 %), finding sex partners in venues (52.4 %) and not receiving HIV test results (66.7 %). Specificities (all >83 %) and PPVs (all >74.0 %) were higher than NPV. FSW significantly under-reported number of clients, sexual contacts and non-condom use sex acts with clients and number of days engaging in sex work in the preceding week. This study provides a quantified gauge of reporting biases in FSW behaviors. Such estimates and methods help better understand true HIV risk in marginalized populations and calibrate survey estimates accordingly.

A. Mirzazadeh · S. Navadeh · K. Mohammad (✉)  
Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences,  
PO Box 14155-6446, Tehran, Iran  
e-mail: mohamadk@tums.ac.ir

A. Mirzazadeh  
e-mail: ali.mirzazadeh@hivhub.ir

S. Navadeh  
e-mail: s.navadeh@hivhub.ir

A. A. Haghdoost  
Regional Knowledge Hub for HIV/AIDS Surveillance,  
Kerman University of Medical Sciences, Kerman, Iran  
e-mail: ahaghdoost@hivhub.ir

S. Nedjat  
Department of Epidemiology and Biostatistics, School of Public Health, Knowledge Utilization Research Center,  
Tehran University of Medical Sciences, Tehran, Iran  
e-mail: nejatsan@tums.ac.ir

W. McFarland  
San Francisco Department of Public Health,  
San Francisco, CA, USA  
e-mail: willi.mcfarland@sfdph.org

**Resumen** Se cuantificaron discrepancias en la notificación de conductas de trabajadoras sexuales (TS) mediante comparación de 63 entrevistas cara a cara (ECC) con entrevistas en profundidad (EEP), con la corroboración de las direcciones y magnitudes de informes de especialistas. Se evaluaron la sensibilidad, especificidad, valores predictivos positivo y negativo (VPP y VPN) de las ECC usando la EEP como “estándar de oro”. La sensibilidad fue más baja en la notificación de síntomas de infecciones de transmisión sexual (63.9 %), de búsqueda de parejas sexuales (52.4 %) y no recibir los resultados de la prueba del VIH (66.7 %). La especificidad (>83 %) y VPP (>74.0 %) fueron más altos que el VPN. Las TS infrareportaron significativamente el número de clientes, contactos sexuales, actos sexuales sin preservativos y el número de días trabajados sexualmente la semana anterior. Este estudio proporciona un indicador cuantitativo de sesgos en la información sobre los comportamientos de las TS. Estas estimaciones y métodos ayudan a entender mejor el riesgo al VIH de las poblaciones marginadas y calibrar las estimaciones correspondientes.

**Keywords** Female sex workers · Validity · HIV · Risk behaviors · Bias · Iran

**Palabras clave** Trabajadoras sexuales · Validación · VIH · Conductas de riesgo · Sesgo · Irán

## Introduction

Global and national responses to the HIV epidemic are based on evidence from HIV prevalence studies and surveys of risk behaviors [1]. Most countries routinely and systematically gather behavioral data which rely on self-reported measures from general populations and populations at higher risk of HIV, including female sex workers (FSW) [2]. In Iran, as in many places around the world, behaviors leading to HIV infection may be stigmatized, illegal or both; therefore, such behavioral data are vulnerable to under-reporting for fear of legal incrimination, discrimination and societal condemnation.

The Iranian Ministry of Health has registered more than 23,000 cases of HIV/AIDS as of November, 2011 [3]; however, the true number of persons living with HIV/AIDS in the country is estimated to be five times as many [4]. The most common mode of transmission, for 56 % of cases, is through high risk injection, followed by 34 % through heterosexual contact, and 10 % through male–male sex [5]. The proportion of reported female-to-male transmitted HIV cases has been substantially increasing [3]. The existence of sex work had previously been denied in Iran by officials and by society in general [6]. Authorities of the Ministry of Health now estimate there are 30,000–60,000 FSW in Iran [7]. Published data explicitly about the context of sex work in Iran have been emerging in recent years [8–11]. Two broad topologies of sex work in Iran can be described. At the low socio-economic level are FSW selling sex for their own and their families' basic survival needs. They are networked to other FSW and also access social support services from public centers and health institutions. High socio-economic level FSW are less networked, harder to reach, work independently and access private health services for medical care.

Recently, the Iranian government has acknowledged FSW as one of the groups most vulnerable to HIV who urgently need prevention and care services. This change in strategy and approach, we believe, likely reflects the evidence of increasing sexual transmission especially to women and to increased use of these data for advocacy and rapport-building with high-level leaders and health policy decision-makers. The recent evidence of the rise of HIV epidemic among women and the increased potential for expansion to other groups, such as FSW, may be fostering more open talk about affected sub-populations. As a result,

more than twenty specialized centers have developed to provide a minimum package of services to FSW. The service package includes basic primary and reproductive health care, HIV testing and counseling, screening and treatment of sexually transmitted diseases and drug detoxification and maintenance therapy.

Despite recognition of their basic health needs and official policies to meet them, the population of FSW in Iran still faces potentially severe legal action and a fragile situation persists. Sex that occurs between any persons who are not married to each other, including in prostitution, can be prosecuted as a capital offense. Periodic police sweeps target houses, hotels and other venues that gain a reputation for being places where people find commercial sex. The context of FSW in Iran therefore drives the sex work further underground and hinders the accurate reporting of behavioral data—a situation similar to other parts of the world albeit usually less intensely.

As in most studies worldwide, measuring risk behaviors in sero-behavioral surveys in Iran is mainly done by face-to-face interview (FTFI) where trained professionals systematically ask questions and record respondents' answers [12]. Studies have compared the different modalities of data collection, such as the audio computer-assisted survey instrument (ACASI) [13] and its derivatives and coital diary [14] against FTFI. Others have compared FTFI with in-depth interview (IDI) [15, 16]. Results indicate that questionnaire delivery mode affects reported sexual behaviors with a trend towards higher reporting (i.e., less under-reporting) of stigmatized behaviors by IDI [16] and by methods that secure more privacy and confidentiality. For example, in a randomized trial, 1,283 male and female Thai students aged 15–21 years in 2002 were allocated into four subgroups. One of four techniques (Palmtop-assisted self-interviewing (PASI), ACASI, self-administered questionnaire or FTFI) was used to collect behavioral data in each. Only 2.5 % reported ever having a genital ulcer in FTFI while by ACASI and PASI the level was 8.0 and 6.7 %, respectively [17]. With regard to the number of sexual partners, there has been significant heterogeneity between FTFI and other modalities [12]. Other studies corroborate large heterogeneity on observed differences by collection methods and context [12, 18], making it difficult to propose a true gold standard for collecting sensitive behaviors, although generally reporting overall appears higher when applying IDI, PASI or ACASI compared to FTFI.

Several sero-behavioral surveys among high risk populations including FSW have been recently conducted or are currently underway in Iran using FTFI for collecting behavioral data. The present study was implemented to assess the level of concordance (i.e., potential bias) in self-reported risk behaviors by FTFI through comparison to IDI (considering as gold standard). The present article aims to

quantitatively compare the two methods of data collection, FTFI and IDI, and further triangulate the amount of reporting bias using a panel of female clinical psychologists with many years of collective experience working with FSW.

## Methods

### Study Subjects

From May to October 2011, 63 FSW were recruited for a behavioral survey after obtaining verbal informed consent. During the study period, all FSW referred to one of several collaborating non-governmental organizations and public health facilities serving sex workers in two cities were contacted by the sites' female clinical psychologists and, if eligible, were invited to participate, and, if consenting, interviewed on the same day. The recruitment centers in Tehran (Iran's capital and largest city) were Hamyari Sabz, Shahriyar, Behroozan, Emami Health Center and the Rebirth Center; in Kerman (another large city) the Rezvan health center was included. These six clinics serve FSW from different socio-economic classes, but primarily the low- to middle-level. FSW age 18–65 years and selling sex in the last 6 months were eligible. The study protocol and procedures were reviewed and approved by the Research Review Board of the Kerman University of Medical Sciences.

### Measures

The female clinical psychologists of each facility conducted the FTFI on all participants, completing the questionnaires in Farsi including vernacular terminology used by the population. The FTFI began with the interviewer asking rapport-building questions on socio-demographic characteristics and the reasons for visiting the particular site. Subsequent questions were progressively more sensitive within the Iranian cultural context, including marriage, sexual history, condom use, drug use, sexually transmitted disease history, knowledge on HIV routes of transmission and HIV testing. The interviewer read the questions one-by-one ensuring that the respondent understood each. We created a classification we refer to as the "transparency probability" which was measured by asking the participant which of her family members and friends (whether involved in sex trade and not) knew that they sold sex. This measure of openness was hypothesized to provide a proxy for the degree of self-censorship and social desirability response bias that may be seen in other measures.

For the present study, we focus on several variables we hypothesized to be vulnerable to varying degrees of under- or over-reporting. The selection of these key variables and a gauge of their likely directions and magnitudes of bias

were done in consultation with a panel of four health professionals conducting HIV behavioral surveillance, two clinical psychologists and two qualitative researchers. After introducing the objectives and methods of the study, we asked them to indicate which of the questionnaire behavioral measures would be more or less sensitive to FSW and prone to over- or under-reporting. This filtering phase led to a shorter list of key questions, which was subsequently finalized by discussion among the group. This process arrived at the following categorical variables thought to be under-reported by FSW: arrested and incarcerated in the last 12 months, ever use of drugs, history of genital ulcer or discharge in the last 12 months, non-condom use at last sex act with a client and being associated with a venue (e.g., home or shelter) where persons find commercial sex. The last category was hypothesized to be under-reported based on concern over drawing unwanted attention of the police to certain places. Testing for HIV and obtaining results of the test were hypothesized to be over-reported based on the social desirability of responding affirmatively to a clinical psychologist in the setting of the interview. Ever being married was hypothesized to be over-reported as sex occurring between two persons who are not husband and wife is highly stigmatized and illegal, and particularly stigmatized if a woman has never been married. Additional continuous variables (i.e., age at first sex for money or other needs and the number of sex acts, the number of non-condom use acts with clients, the number of clients, and the number of days exchanging sex in the last week) were also considered as potentially under-reported.

Following the FTFI, checking the internal consistency of the reported behaviors and discussing the respondents' general health and background, the clinical psychologists conducted an IDI with each of the 63 FSW as a cognitive cross-check of their answers. 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 follow the questionnaire as a step-by-step reading of the questions, rather followed a natural discussion leading to the 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 separated column after finishing the IDI. Interviewers then probed in greater detail the FSW's life directing the narrative towards any inconsistencies between their story and the behaviors reported in the FTFI questionnaire.

To further independently explore the likely amount of social desirability bias for the key behaviors listed above, ten female clinical psychologists were consulted in two focus group discussions (FGD). In another words, we used FGD to further quantitatively explore the amount of bias in reported risk behaviors. These included the six

interviewers, the two health professional panel members, and two additional clinical psychologists. The FGD participants had from 1.3 to 5.5 years of experiences with FSW in Tehran and Kerman, providing psychotherapy consultations for this marginalized group at both private and public health centers, including the recruitment study sites. The panel members and the instructor were blinded to the results of the comparison between FTFI and IDI. The FGD solicited opinions on the amount of potential bias in FSW reported behaviors. For each of the key behaviors mentioned above, the instructor explained the meaning in terms of the survey objectives and facilitated the group towards consensus in the amount of under- or over-reporting likely in the FTFI using two scenarios. First, each FGD participant was asked to imagine the FSW reporting not engaging in the behavior (e.g., they do not use drugs) in the FTFI, then to write the proportion of FSW they believe in reality do have the risky behavior (e.g., do use drugs) and discuss with the group to arrive at a consensus figure.

The resulting conditional probability is the negative predictive value (NPV). Thus, the NPV represents the accuracy with which participants in the FGD accurately estimate the percentage of behaviors that FSWs engage in when FSWs deny a certain behavior. The second scenario was that the FSW reported having the risky behavior (e.g., using drugs) in the FTFI, with the panel similarly arriving at proportion they believe to have the risk behavior (e.g., using drugs). This conditional probability is the positive predictive value (PPV). Thus, the PPV represents the accuracy of FTFI in estimating the percentage of risky behaviors among FSW engaging in such risky behaviors. In this way, consensus was built within the FGD to quantify the NPV and PPV for all selected variables.

#### Statistical Analysis

Stata v.10 was used for analysis. Answers arising from the IDI were considered as the comparative (“gold standard”) and the answers to the FTFI as the test measure. Standard test performance measures were calculated as follows:

- Sensitivity is defined as the conditional probability,  $P$  (classified as FSW reports having the risky behavior | truly having the risky behavior)
- Specificity as the conditional probability,  $P$  (classified as FSW reports not having the risky behavior | truly not having the risky behavior)
- PPV as the conditional probability,  $P$  (truly having the risky behavior | classified as FSW reports having the risky behavior)
- NPV as the conditional probability,  $P$  (truly not having the risky behavior | classified as FSW reports not having the risky behavior)

For each of the selected sensitive behaviors, point sensitivity and specificity measures as well as the exact binomial confidence intervals were calculated. Additionally, for the FGD-derived conditional probabilities, normal-approximation 95 % confidence intervals for the PPV and NPV of the sensitive behaviors were also calculated compared to IDI responses (thus, two sets of PPV and NPV are calculated). For the continuous measures of behaviors listed above, we calculated the absolute discrepancies in each FSW response and the mean difference between the FTFI and IDI results with 95 % CI. The paired  $t$  test was used to assess whether the two responses were significantly different (i.e., that the mean difference is not equal to zero).

#### Results

A total of 63 FSW age 18–44 years (mean 28.5) were recruited to the validation study (Table 1), with 23.8 % under 25 years. Most (81.0 %) reported having ever been married with 28.5 % in a marital union. Slightly over half completed high school. History of arrest and incarceration were reported by 49.2 and 23.8 %, respectively, in the last year. The vast majority (96.8 %) had ever used drugs while injection was reported by 14.5 %. Only 52.4 % reported being tested for HIV in the last year, with 18.2 % reporting not receiving their results. Unprotected sex with a client in the last act was reported by 39.3 % of FSW and 33.3 % acknowledged that they were associated with a venue for finding sex partners.

On average, the women reported 3.4 sexual contacts with their clients in last 7 days, with a mean of 1.9 times without condom use. The first sex act for money, drugs or shelter was at the age of 21 years. In the last 7 days, FSW reported an average of just over 3 clients over an average of approximately 3 days. They earned \$277.80 USD in a month by selling sex (Table 1), with the last act averaging \$21.90. The transparency probability (likelihood of disclosing sex work) with friends who were also FSW was 0.6 (95 % CI 0.5–0.7). The probability decreased significantly to 0.4 considering other friends or family members ( $t$  test 4.17,  $P = 0.001$ ).

Holding the IDI responses as the standard, sensitivities and specificities of the FTFI measures are shown in Fig. 1. Sensitivities were high (>90 %) for ever being married, ever use of drugs, and never testing for HIV. The lowest sensitivities were noted for being associated with a venue for commercial sex (52.4 %), having symptoms of STI (63.9 %), being incarcerated (66.7 %) and not receiving their HIV test result (66.7 %). Specificities for the selected variables were generally high, with only never being tested for HIV falling below 90 %.

**Table 1** Characteristics of FSW included in a validation study of reported behavior, Iran, 2011 ( $N = 63$  FTFI)

Variable	Point estimate mean or % (95 % CI)
Age in years (mean)	28.5 (27.0–29.9)
Ever married (%)	81.0 (69.0–89.0)
Current marital status (%)	
Single	20.6 (12.2–32.7)
Married	28.6 (18.6–41.2)
Divorced	38.1 (26.7–50.9)
Other	12.7 (6.4–23.8)
Education level achieved (%)	
Illiterate	1.6 (0.2–11.0)
Primary school	12.7 (6.4–23.8)
Middle school	34.9 (24.0–47.8)
High school	49.2 (36.8–61.7)
University degree	1.6 (0.2–11.0)
Reported risk behaviors	
Categorical (%)	
Arrested in last 12 months	49.2 (36.8–61.7)
In prison in last 12 months	23.8 (14.7–36.2)
Ever used drugs	96.8 (87.8–99.2)
Ever injected drugs	14.5 (7.6–26.0)
Did not test for HIV in last 12 months	47.6 (35.4–60.2)
Did not receive results of the HIV test in last 12 months	18.2 (8.0–36.1)
Had genital ulcer or discharge in last 12 months	57.1 (44.4–69.0)
Non-condom use at last sex act	39.3 (27.7–52.4)
Ever associated with a place, home or shelter in order to find commercial and other sex partners	33.3 (22.6–46.1)
Continuous (mean)	
Number of sexual contacts, last 7 days	3.4 (2.6–4.3)
Number of non-condom use acts, last 7 days	1.9 (1.3–2.4)
Age at first sex act for money, drugs or shelter	21.0 (19.5–22.4)
Number of clients, last 7 days	3.2 (2.4–4.0)
Number of days in last week with sex with a client	2.9 (2.3–3.5)
Average monthly income through commercial sex (US \$)	277.8 (111.3–444.3)
Income through last commercial sex act (US Dollar)	21.9 (17.1–26.8)
Transparency probability <sup>1</sup>	
Sex work activity known to friends who themselves are FSW	0.6 (0.5–0.7)
Sex work activity known to friends not FSW	0.4 (0.3–0.6)
Sex work activity known to first degree family members	0.4 (0.2–0.5)

<sup>1</sup> The proportion of FSW family members and friends (involved in sex trade and not) who already knew they had sold sex

Figures 2 and 3 show PPV and NPV derived from the FTFI (again using IDI as the standard) and the FGD. Most measures had >90 % PPV using the FTFI or FGD outcomes. In effect, FTFI and FGD members concurred that when FSW acknowledged the risk behavior, their response was correct. The PPV for not testing for HIV was lower for both FTFI and FGD measures (84.8 and 74.0 %, respectively), and for the FTFI measure of non-condom use (87.5 %), and the FGD measure of not receiving the HIV test result (84.0 %). NPV of FTFI and FGD measures were lower and more variable (Fig. 3) than PPV, indicating more discrepancy when FSW denied the risk behavior. Moreover, with the exception of drug use, FTFI measures were more sanguine than the FGD measures with respect to NPV. Ever use of drugs, ever married, history of STI symptoms, and non-condom use at last sex with a client had particularly low NPV.

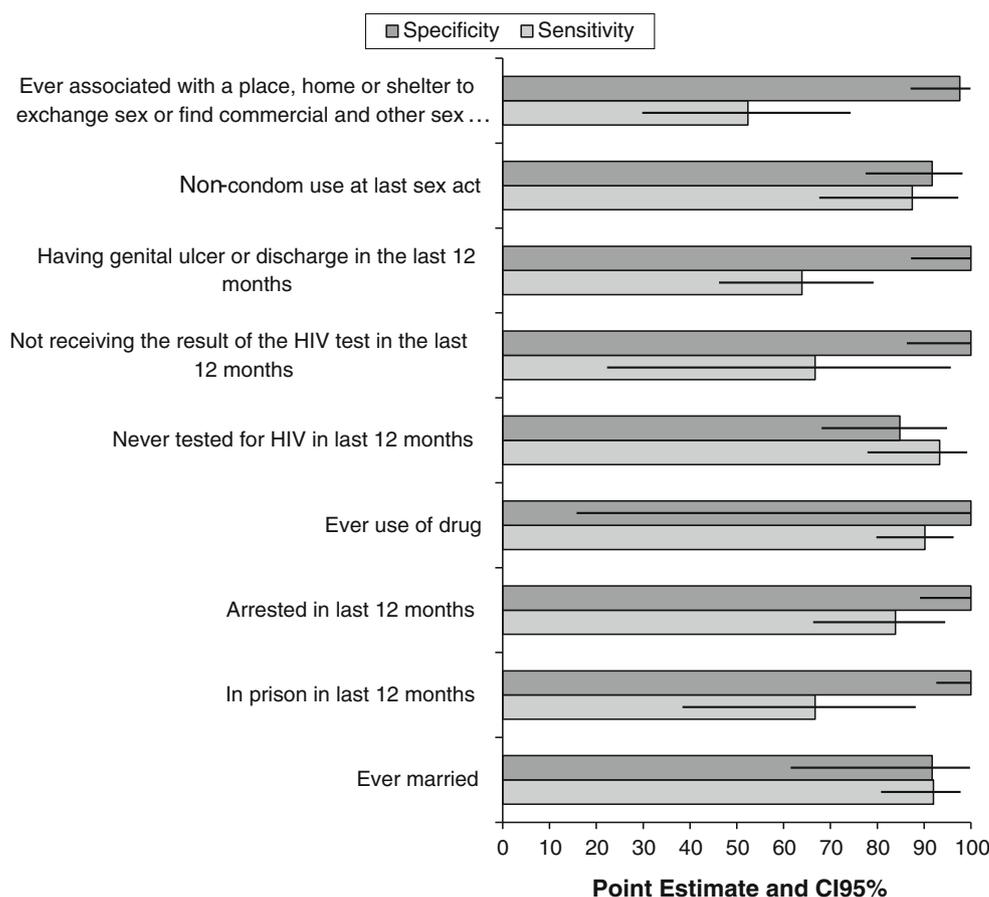
Table 2 shows the magnitude and tests of significance for discrepancies between the IDI and FTFI for key continuous variables. Comparing the FTFI to the IDI, the FSW reported 1.5 fewer sexual contacts ( $t = 3.69$ ,  $P < 0.001$ ), 0.4 fewer non-condom use sexual acts ( $t = 2.03$ ,  $P = 0.04$ ), 0.8 fewer clients ( $t = 2.68$ ,  $P = 0.01$ ), and 0.9 fewer days exchanging sex ( $t = 2.80$ ,  $P = 0.01$ ) in the last week.

## Discussion

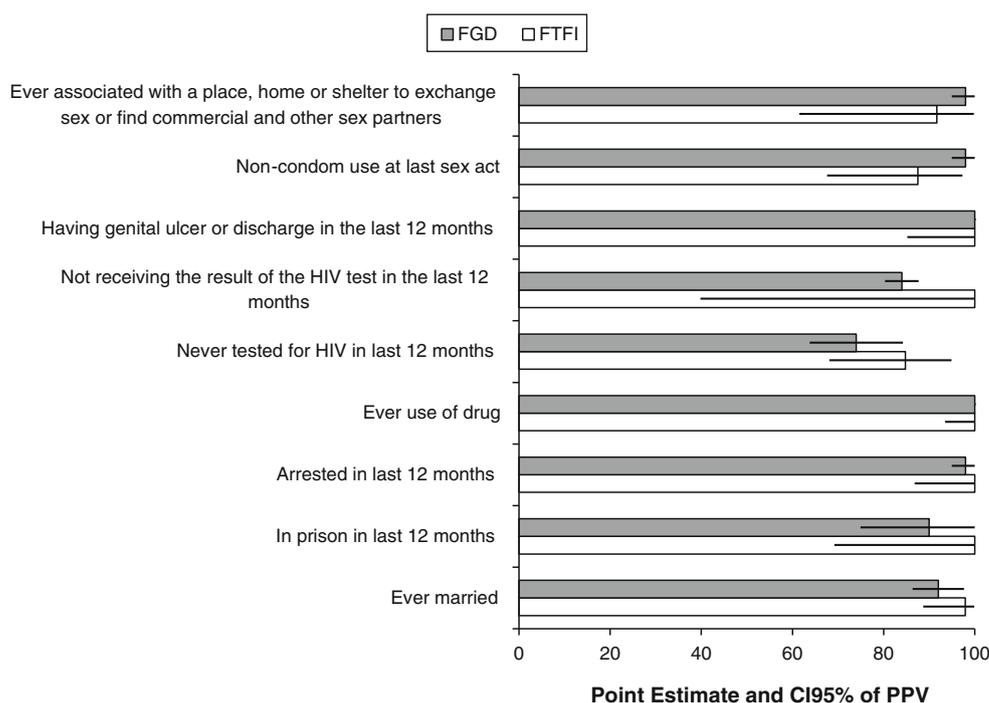
Our results confirm that face-to-face interviewing (FTFI), which remains the most common questionnaire delivery mode worldwide, is prone to under-reporting of stigmatized, risky behaviors [13, 19–21]. Even for women acknowledging engaging in commercial sexual acts and receiving services from organizations serving sex workers, reporting being in prison, never tested for HIV, having STI symptoms and being associated with venues for commercial sex is likely self-censored in behavioral surveys. FSW also under-report the number of sexual partners, sexual acts and non-condom sexual acts with clients. While these biases and their directions are noted in other studies [16, 20, 22] and reviews [12, 18] in other contexts, in the present study we developed a technique to quantify the amount of potential bias using multiple data sources and mixed methods.

As in the literature, our respondents tended to under-report sensitive behaviors with FTFI compared with other questionnaire delivery modes [19], such as ACASI elsewhere and by IDI in the present study. We also corroborate and quantify that the amount of under-reporting is heterogeneous according to the question. In our evaluation of sensitivity and specificity of FTFI against IDI, we found,

**Fig. 1** Sensitivity and specificity point estimates and 95 % CI for behaviors reported by FSW in FTFI compared to IDI, Iran, 2011 (N = 63). Sensitivity is the conditional probability, P(classified as FSW having the risky behavior for HIV | truly having the risky behavior); specificity is the conditional probability, P(classified as FSW without the risky behavior for HIV | truly don't have the risky behavior)



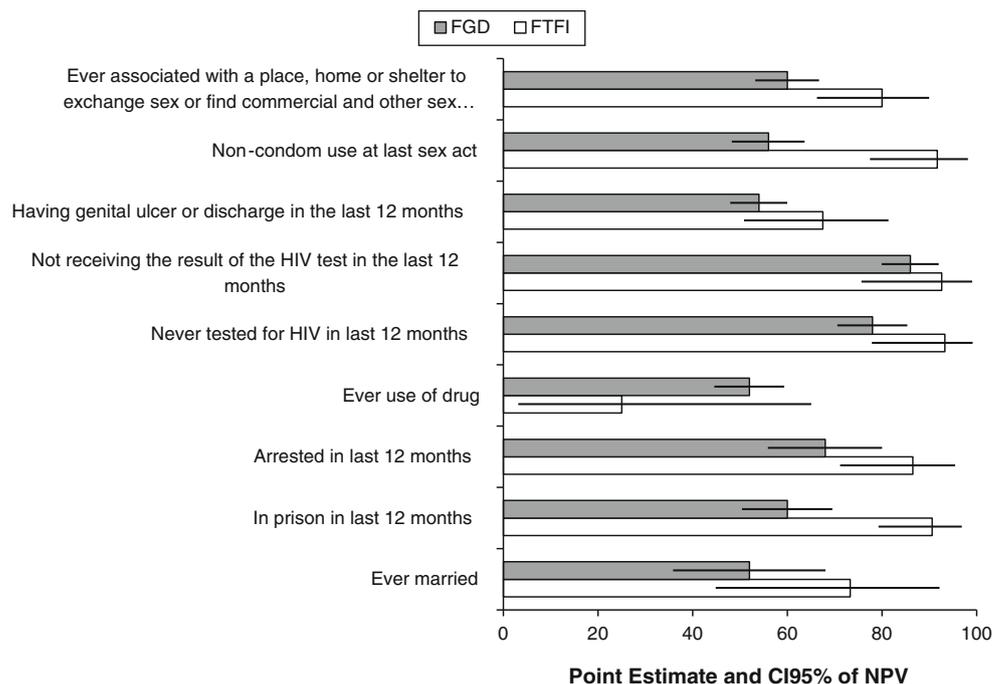
**Fig. 2** PPV point and 95 % CI for behaviors reported by FSW in FTFI compared to IDI (FTFI, bottom, light bar) and health professional focus group discussion consensus opinion (FGD, top, dark bar), Iran, 2011 (N = 63). PPV is the conditional probability, P(truly having the risky behavior | classified as FSW having the risky behavior for HIV)



for example, that asking about association with places for finding clients is a very sensitive issue in the context of police activity in Iran. For similarly reasons, having outside

or before being married is also subject to under-reporting as police and society also targeting this behavior. Also similar to the literature, we found STI symptoms and

**Fig. 3** NPV point and 95 % CI for behaviors reported by FSW in FTFI compared to IDI (FTFI, bottom, light bar) and health professional focus group discussion consensus opinion (FGD, top, dark bar), Iran, 2011 ( $N = 63$ ). NPV is the conditional probability,  $P$ (truly don't have the risky behavior | classified as FSW without the risky behavior for HIV)



**Table 2** Differences in behaviors between IDI and FTFI, FSW in Iran, 2011 ( $N = 63$ )

Outcome/Behaviors	$N$	Mean difference between IDI and FTFI (95 % CI)	Paired $t$ test ( $DF^1$ )	$P$ value
Age	63	-0.02 (-0.1-0.1)	-0.38 (62)	0.71
Sexual contacts, last 7 days	58	1.5 (0.7-2.3)	3.69 (57)	<0.001
Non-condom use acts, last 7 days	52	0.4 (0-0.7)	2.03 (51)	0.04
Age at first sex for money, drugs or shelter	55	0.05 (-1.1-1.2)	0.09 (54)	0.92
Clients, last 7 days	50	0.8 (0.2-1.5)	2.68 (49)	0.01
Days in last week with sex with client	53	0.9 (0.2-1.5)	2.80 (52)	0.01
Monthly income through commercial sex (US \$)	46	5.2 (-22.9-33.2)	0.37 (45)	0.71
Income through last commercial sex act (US \$)	49	1.4 (-0.6-3.4)	1.43 (48)	0.16

Positive "mean difference" represent under-reporting

<sup>1</sup> Degrees of freedom

disclosure of sex work are highly socially stigmatized [23] and therefore prone to under-reporting in FTFI questionnaires. Validation of reports in FTFI against IDI has been previously assessed by Konings et al. [16], with a notable under-reporting of casual sex partners in FTFI. In another survey in Switzerland, condom use in last intercourse was reported at about 40 % in a FTFI interview compared to 46 % in a second interview by telephone [20]. In our study, we found that 87 % of those who later disclosed a non-condom use sexual act with a client had reported it in the initial FTFI. We concur with the Konings study that IDI provides a more accurate reflection of reality than FTFI because of the extensive rapport building between the interviewer and the interviewees or because it gives more time for the respondents to recall their behaviors [16].

Considering the process and findings from the IDI, standard FTFI could be improved in ways to have less measurement bias. First, interviewers must acknowledge how difficult it is to discuss the stigmatized behaviors with participants. Secondly, they need to ensure that participants are confident their information is confidential and whenever possible anonymous. As mentioned above, rapport building is a crucial step and its importance in behavioral surveys must not be underestimated. Reducing the number of questions to those behaviors really needed and giving more time to participants for better recall will also help in reducing under-reporting. However, it should be emphasized these techniques might reduce the bias in FTFI but not resolve it.

Generally, we observed that for FSW who acknowledge risky behaviors in the FTFI, their response was correct.

This is clear from the PPV, in both FTFI and FGD, presented in Fig. 2. The minimum PPV was reported for “never tested for HIV”. In this case, we perceive a potential conflict in that disclosing HIV status is a stigmatized issue. If at a health facility, an FSW may prefer to deny they have tested for HIV to avoid being asked their HIV status, even if actually they did test. This was affirmed by the FGD. The same interpretation would apply for “receiving back the HIV test results” outcome.

The story for those denying or not disclosing risky behaviors is different. Their responses are affected by the level of stigma around each risky behaviors and the stigma level is translated into the variability of the NPV (observed in Fig. 3). An interesting finding is that the FGD participants apparently underestimated the stigma around drug use and this is reflected in the NPV. For other risky behaviors, FGD participants have a more pessimistic view on the accuracy of reporting when FSW deny it. Nonetheless, FGD participants also believed that accuracy varies regarding different risk behaviors.

We recognize several limitations of our study. While a strength is our use of mixed methods to assess the likely amount and directions of biases of multiple measures with different contexts, the primary limitation is that there is no true gold standard. As such, self-reported sexual behavior in an interview is difficult or even impossible to be externally validated. In our study, we have considered the IDI as a proxy “gold standard”. New methods, further investigation and triangulation of data continue to be needed to validate sexual behavior reporting. A promising area of research is the use of biological markers of behaviors [24, 25], for example, as demonstrated in a clinical trial in Zimbabwe [24]. Reported sexual behavior was validated by measuring prostate-specific antigen (PSA) by vaginal swab and comparing it to the FTFI results on sexual behaviors. The authors found that only 52 % of PSA-positive women reported unprotected sex during the previous 2 days. STI laboratory tests have also been used to cross validate reported behaviors [25], but unfortunately they have limited routine applicability because of the cost and different exposure periods captured by the biological markers and the interview.

Another limitation of our study is its generalizability. Iran may represent a particularly severe context in which denial of sexual behavior is high due to legal and social consequences. Our validation study was conducted with this very high potential for under-reporting in mind. Nonetheless, the stigma associated with sex work and other sexual behaviors does apply to most contexts around the world. Our setting helps illuminate the relative amounts of over- and under-reporting of behaviors that can be expected. Our results may also generalize to similar contexts in the wider region of the Middle East. Internally, we

recruited FSW from the health facilities serving them in two metropolitan areas, Tehran and Kerman, and recognize that these women may not be typical of FSW in other Iranian cities or of those not accessing services. The sites were selected because they matched the recruitment of the larger sero-behavioral surveys under way. We believe that further investigation in ways to improve community-based sampling for FSW and other hidden populations is urgently needed to address multiple potential biases with facility and convenience sampling.

## Conclusions

In conclusion, the findings of this study have indicated that strongly stigmatized behaviors like non-condom use, symptoms of STI, and venue-based sexual acts are less likely to be reported in a routine face-to-face interview. Despite limitations, our study makes an attempt to quantify the level of reporting bias for different sensitive behaviors in two cities using multiple methods. Our bias parameters could be used in correcting the estimates of the larger sero-behavioral surveys and the approach may be locally applied to behavioral surveillance efforts in other countries. Considering the fact that most countries use FTFI as the main mode of behavioral data collection in ongoing surveillance activities, such calibrations are needed over multiple measures, places, populations and time periods.

**Acknowledgments** The Regional Knowledge Hub for HIV/AIDS Surveillance (Kerman University of Medical Sciences) and Tehran University of Medical Sciences have jointly supported this project as a PhD thesis (for A.M. author). The authors would like to thank all the clinical psychologists for their contribution in interviewing the FSWs: Ms Abedpour, Ms Vafaie, Ms Amir-Sayafi, Ms Goodarzi, Ms Salami, and Ms Nekoie and the group of experts attending in FGD. We express our gratitude to Soodeh Arabnejad, the wonderful program assistant who helps in data extraction and entry.

## References

1. Garnett GP, Garcia-Calleja JM, Rehle T, Gregson S. Behavioural data as an adjunct to HIV surveillance data. *Sex Transm Infect.* 2006;82(Suppl 1):i57–62.
2. World Health Organization. UNAIDS, Guidelines on surveillance among populations most at risk for HIV. Geneva: WHO; 2011.
3. MOHME, Center for Disease Management. Statistics of HIV/AIDS in Islamic Republic of Iran: Quarterly, Oct 2011: Center for disease management. Ministry of Health and Medical Education, Islamic Republic of Iran. 2011.
4. Haghdoost AA, Mostafavi E, Mirzazadeh A, Navadeh S, Feiz-zadeh A, Fahimfar N, et al. Modelling of HIV/AIDS in Iran up to 2014. *J AIDS HIV Res.* 2011;3(12):231–9.
5. MOH, UNAIDS. Modeling of new HIV infections based on exposure groups in Iran 2010.
6. Mohebbi MR. Female sex workers and fear of stigmatisation. *Sex Transm Infect.* 2005;81(2):180–1.

7. Mostashari G, Darabi M. Summary of the Iranian situation on HIV epidemic. Tehran: NSP Situation Analysis; 2006.
8. Ramezani Tehrani F, Malek-Afzali H. Knowledge, attitudes and practices concerning HIV/AIDS among Iranian at-risk sub-populations. *East Mediterr Health J.* 2008;14(1):142–56.
9. Kassaian N, Ataei B, Yaran M, Babak A, Shoaie P. Hepatitis B and C among women with illegal social behavior in Isfahan, Iran: seroprevalence and associated factors. *Hepat Mon.* 2011;11(5):368–71.
10. Jahani MR, Alavian SM, Shirzad H, Kabir A, Hajarizadeh B. Distribution and risk factors of hepatitis B, hepatitis C, and HIV infection in a female population with “illegal social behaviour”. *Sex Transm Infect.* 2005;81(2):185.
11. Hajiabdolbaghi M, Razani N, Karami N, Kheirandish P, Mohraz M, Rasoolinejad M, et al. Insights from a survey of sexual behavior among a group of at-risk women in Tehran, Iran, 2006. *AIDS Educ Prev.* 2007;19(6):519–30.
12. Phillips AE, Gomez GB, Boily MC, Garnett GP. A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries. *Int J Epidemiol.* 2010;39(6):1541–55.
13. Mensch BS, Hewett PC, Gregory R, Helleringer S. Sexual behavior and STI/HIV status among adolescents in rural Malawi: an evaluation of the effect of interview mode on reporting. *Stud Fam Plann.* 2008;39(4):321–34.
14. Hays MA, Irsula B, McMullen SL, Feldblum PJ. A comparison of three daily coital diary designs and a phone-in regimen. *Contraception.* 2001;63(3):159–66.
15. Mavhu W, Langhaug L, Manyonga B, Power R, Cowan F. What is ‘sex’ exactly? Using cognitive interviewing to improve the validity of sexual behaviour reporting among young people in rural Zimbabwe. *Cult Health Sex.* 2008;10(6):563–72.
16. Konings E, Bantebya G, Carael M, Bagenda D, Mertens T. Validating population surveys for the measurement of HIV/STD prevention indicators. *AIDS.* 1995;9(4):375–82.
17. van Griensven F, Naorat S, Kilmarx PH, Jeeyapant S, Manop-aiboon C, Chaikummao S, et al. Palmtop-assisted self-interviewing for the collection of sensitive behavioral data: randomized trial with drug use urine testing. *Am J Epidemiol.* 2006;163(3):271–8.
18. Langhaug LF, Sherr L, Cowan FM. How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries. *Trop Med Int Health.* 2010;15(3):362–81.
19. Langhaug LF, Cheung YB, Pascoe SJ, Chirawu P, Woelk G, Hayes RJ, et al. How you ask really matters: randomised comparison of four sexual behaviour questionnaire delivery modes in Zimbabwean youth. *Sex Transm Infect.* 2011;87(2):165–73.
20. Jeannin A, Konings E, Dubois-Arber F, Landert C, Van Melle G. Validity and reliability in reporting sexual partners and condom use in a Swiss population survey. *Eur J Epidemiol.* 1998;14(2):139–46.
21. Crosby R, Salazar LF, DiClemente RJ, Yarber WL, Caliendo AM, Staples-Horne M. Accounting for failures may improve precision: evidence supporting improved validity of self-reported condom use. *Sex Transm Dis.* 2005;32(8):513–5.
22. Ramjee G, Weber AE, Morar NS. Recording sexual behavior: comparison of recall questionnaires with a coital diary. *Sex Transm Dis.* 1999;26(7):374–80.
23. Hood JE, Friedman AL. Unveiling the hidden epidemic: a review of stigma associated with sexually transmissible infections. *Sex Health.* 2011;8(2):159–70.
24. Minnis AM, Steiner MJ, Gallo MF, Warner L, Hobbs MM, van der Straten A, et al. Biomarker validation of reports of recent sexual activity: results of a randomized controlled study in Zimbabwe. *Am J Epidemiol.* 2009;170(7):918–24.
25. Hewett PC, Mensch BS, Ribeiro MC, Jones HE, Lippman SA, Montgomery MR, et al. Using sexually transmitted infection biomarkers to validate reporting of sexual behavior within a randomized, experimental evaluation of interviewing methods. *Am J Epidemiol.* 2008;168(2):202–11.