

Assessment of non-response bias in a probability household survey of male same-gender sexual behavior

José Antonio Izazola-Licea, M.D., Sc. D.,⁽¹⁾ Steven L. Gortmaker, Ph. D.,⁽²⁾
Victor De Gruttola, Sc. D.,⁽²⁾ Kathryn Tolbert, Ph. D.,⁽³⁾ Jonathan Mann,[†] M.D., M.P.H.⁽⁴⁾

Izazola-Licea JA, Gortmaker SL,
De Gruttola V, Tolbert K, Mann J.
Assessment of non-response bias
in a probability household survey
of male same-gender
sexual behavior.
Salud Publica Mex 2000;42:90-98.

Izazola-Licea JA, Gortmaker SL,
De Gruttola V, Tolbert K, Mann J.
Determinación de sesgo de no respuesta
en una encuesta probabilística de hogares
de comportamiento sexual con personas
del mismo género.
Salud Publica Mex 2000;42:90-98.

Abstract

Objective. To assess non-participation bias in a survey of male sexual behavior. **Material and methods.** A household survey was carried out in 1992-1993 using a probability sampling frame in Mexico City. Demographic variables were available for all eligible men. The extent of non-participation bias was estimated using a version of the Heckman method, which utilizes two equations, one to predict participation and the other to predict reports of same-gender sexual behavior. **Results.** A total of 8 068 of the 13 713 eligible men completed a face-to-face questionnaire (response rate 59%); 173 men (2.1%) reported bisexual behavior in their lifetime, and 37 (0.4%) reported only male partners. Survey participation was predicted using demographic variables: 67% of the observations were correctly predicted by a probit regression model: 82% of participants and 53% of non-participants (pseudo- $r^2=0.13$). Same-gender sexual behavior was predicted by variables indicating attachment to gay/bisexual social networks, history of sexually transmitted diseases, positive attitudes towards gay and bisexual males, and lack of support from male relatives. Ninety-seven per cent of the cases was correctly predicted by the probit model (pseudo- $r^2=0.14$). The correlation between these two equations was not statistically significant. **Conclusions.** These results indicate that prevalence esti-

Resumen

Objetivo. Determinar y cuantificar el sesgo de no participación en una encuesta de comportamiento sexual masculino. **Material y métodos.** Una encuesta probabilística de hogares se llevó a cabo utilizando un marco muestral probabilístico (jurisdiccional) en la ciudad de México, en 1992-1993. Se obtuvieron las variables demográficas para todos los hombres elegibles de los hogares seleccionados. La magnitud del sesgo de no participación se estimó utilizando el método Heckman, el cual utiliza dos ecuaciones, una para predecir la participación en la encuesta y la segunda para predecir la variable sustantiva de interés, en este caso, comportamiento sexual con otros hombres. **Resultados.** De los 13 713 hombres elegibles en los hogares seleccionados en el marco, sólo 8 068 participaron brindando información completa a cuestionarios cara a cara (tasa de respuesta 59%); 173 hombres (2.1%) señalaron comportamiento bisexual a lo largo de toda su vida, y 37 (0.4%) informaron haber tenido parejas sexuales masculinas exclusivamente. La participación en la encuesta se predijo usando variables demográficas: 67% de las observaciones fueron correctamente predichas utilizando una regresión probit -82% de los participantes y 53% de los no participantes (seudo- $r^2=0.13$). El comportamiento sexual con otros hombres se predijo mediante variables que indicaban pertene-

- (1) The author was with Consejo Nacional para la Prevención del Sida (Conasida), when the study was conducted; Post-doctoral fellow, Harvard AIDS Institute, Harvard School of Public Health, when the manuscript was written; presently, he is with Iniciativa Regional sobre Sida para América Latina y el Caribe (Sidalac), Fundación Mexicana para la Salud, México, D.F., México.
- (2) Department of Health and Social Behavior, Harvard School of Public Health, Boston, Massachusetts (MA), United States of America (USA).
- (3) The Population Council, Office for Latin America and the Caribbean. Mexico City, Mexico.
- (4) Francois-Xavier Bagnoud Center for Health and Human Rights, Harvard School of Public Health, Boston, MA, USA.

Received on: April 8, 1999 • Accepted on: January 14, 2000

Reprint requests to: José Antonio Izazola Licea. Iniciativa Regional sobre Sida para América Latina y el Caribe (Sidalac),
Fundación Mexicana para la Salud (Funsalud). Avenida Periférico Sur 4809, colonia El Arenal, Tepepan, Delegación Tlalpan, 14510 México, D.F., México.
E-mail: jizazola@funsalud.org.mx

mates of same-gender sexual behavior among Mexico City men were not biased by selective survey participation. Careful selection and training of household interviewers may have assisted in minimizing potential bias.

Key words: non-response bias; selection bias/evaluation; health surveys; homosexuality, male; Mexico

cer a redes sociales de homosexuales o bisexuales, historia de infecciones transmitidas sexualmente, actitudes positivas hacia hombres homosexuales y bisexuales y falta de apoyo de familiares masculinos. En 97% de los casos se predijo correctamente utilizando el modelo probit (seudo- $r^2=0.14$). La correlación entre estas dos ecuaciones no fue estadísticamente significativa. **Conclusiones.** Estos resultados indican que los estimadores de prevalencia de comportamiento sexual con personas del mismo género en hombres de la ciudad de México no fueron sesgados por participación selectiva en la encuesta. La selección cuidadosa del personal de campo y la capacitación de los entrevistadores podría haber coadyuvado en minimizar el sesgo potencial.

Palabras clave: sesgo de no respuesta; sesgo de selección/evaluación; encuestas epidemiológicas; homosexualidad masculina; México

The first approaches to research sexuality were the work of biologists¹ or medical clinicians.² Later, several disciplines from the social sciences (e.g., demography or anthropology) added their perspective to the study of human sexuality. Studies on sexual behavior became more frequent due to the AIDS epidemic, and public health was added to the spectrum of disciplines that focus on the scientific study of sexuality.

Most of the scientific literature regarding same-gender sexual behavior (SGSB) has been based on convenience and other non-probability samples.³ There are relatively few surveys that have used probability samples for the study of sexual behavior, mainly in the United States,^{4,7} and in Western Europe.^{8,9} To our knowledge, there are no published reports of similar studies using probability sampling frames in developing countries, particularly in Latin America.

The use of probability sampling frames allows sample results to be generalized to a population. However, probability surveys may be biased because of selective participation. Responses may be biased due to a variety of reasons, including self-presentation bias at the time of the interview.¹⁰ Self-presentation or social desirability bias reflects the underlying values that a culture or specific subcultures place on revealing sexual experiences to others, and is considered to strongly influence responses in sex surveys, particularly for more stigmatized behaviors (e.g., same-gender sexual behavior).¹¹ This type of bias may also be present as misclassification of the outcome measures: e.g., individuals who have a history of SGSB may not participate in the survey, and if they do, they may not report SGSB.

Participation bias results in a systematic selection of the respondents in such a way that non-respondents

are different from respondents. Differences in participation may strongly affect survey estimates. For example, in the context of a sex survey, homosexual and bisexual men could be less willing to participate in a household survey compared to heterosexual men. Or, heterosexual individuals might not want to participate in a sex survey related to HIV transmission because they consider themselves not to be at risk. If non-participants are not different from participants, it can be assumed that the final sample is not a biased sample. If question non-response is random with respect to the sexual behavior under investigation, as well as with respect to other variables of interest, then no significant bias is expected. The direction of the bias of self-reported sexual behavior is usually assumed to be under-reporting. However, there are instances when the opposite may occur (e.g., male adolescents may report higher number of partners as an indication of sexual prowess). In general, however, self-reports of SGSB are considered to be lower-bound estimates of such behavior, because of the stigma attached to it.

The expected impact of non-participation bias is of such magnitude, that funds have been refused for the study of sexual behavior in population-based surveys by the governments of the United States and of Great Britain, using as a main argument the potential for participation-bias.^{4,7,8}

The representativeness of a sample is directly ascertainable only if data on the population from which the sample was drawn are available; however, data on observable differences between respondents and non-respondents may not suffice to determine selection bias. Furthermore, even if bias is detected, usually, there is little or no information about non-respondents. When information is available, the usual analytical approach is to test for differences in demographic cha-

racteristics between respondents and non-respondents, without correcting for differences if they are found. A 1970 survey of sexual behavior in the United States detected significant differences between survey participants when compared to the Population Census, but no correction or formal testing for participation bias was reported.¹²

One way of testing for non-response bias and for correcting estimates from a non-randomly selected sample is to treat sample selection bias as a specification error using simultaneous equations: one equation to model participation and the second one to model an empirical equation to predict the prevalence of the issue under study for non-respondents, based on the information provided by the participants (substantive equation). The best known method is a two-stage regression algorithm (Heckman method) for assessment and control of selection bias.¹³ Several extensions have widened its applicability.¹⁴ This method has been used with survey data in a study of medicine use by the elderly,¹⁵ and has also been applied to the analysis of non-response to sexual behavior questions of the General Social Survey in the United States.*

A previous report described a 2.6% lifetime prevalence of SGSB among Mexico City men 15 to 60 years of age.^{16,17} Exclusive lifetime homosexual behavior was 0.5% and bisexual behavior was 2.1%. It was estimated that 0.8% of this population had sex with other males in the year previous to the interview. These results appear to be comparable to results from other probability studies in the United States, Britain and France. The present study was conducted to test for non-participation bias in the report of SGSB in this household probability survey of Mexico City men. Our main hypothesis is that we expect significant evidence for non-participation bias of SGSB men in Mexico City. Because of the great stigma attached to SGSB, we expected that a bias against participation of homosexual and bisexual men will occur in the population survey, and hence prevalence estimates of SGSB may be underestimated.

Material and methods

A household probability survey was carried out in the Mexico City Metropolitan Area (MCMA) from June

1992 to March 1993. A sampling framework of Sanitary Jurisdictions from the National Health Surveys System of the Mexican Ministry of Health was used.

A household listing was performed as the first step of the survey in order to obtain a census of all persons living in each household. Respondents were asked to participate in this health survey without mentioning AIDS or sexual behavior as contents of the questionnaire, since these topics were not included in the household questionnaire. Interviewers were instructed to obtain the household information from the first knowledgeable adult found in the selected households. Households were revisited up to 10 times for each potential respondent. Face-to-face interviews were carried out among eligible individuals using a 25-minute structured questionnaire. An intensive two-week training course on sexuality was mandatory for field personnel. Individuals who showed negative attitudes toward sexual diversity, or who showed uneasiness while conducting the interviews in the pilot test were not allowed to conduct interviews in the main study.

Since inquiring about sexual behavior is socially sensitive, interviewers were instructed to question the respondent in private. If the conditions of the interview did not guarantee confidentiality, the interviewers were instructed to postpone or terminate the interview. A more detailed description of the survey has been reported elsewhere.¹⁸

Sociodemographic information was collected for all household members: age, gender, relationship to the head of the household, schooling and occupation. For eligible men, information about having a stable female partner or having children was also obtained. Other variables regarding the structure of the households were also taken into account: position of the respondent in the household (head, son or other relationship to the head of the household), number of eligible men within the household, and number of visits by the interviewers. The gender of the interviewers, the field team to which they belonged to, and geographic areas of the households were also coded.

At the beginning of the individual questionnaires, information about social networks and social support was gathered. As part of this section, respondents were asked if any of three of their closest male friends had ever had sex with other men.

Respondents were asked if they were aware of the presence of homosexual or bisexual men in their workplace or school, in their household, or in bars, social or sport clubs that they attend. In addition, respondents who reported to be aware of the presence of homosexual or bisexual men in these social environments were

* Michaels R, Veldman C. Behind doors: An analysis of non-response to the GSS sexual behavior supplements. Chicago: National Opinion Research Center, 1992. Typescript, page 17. Cited in reference 4:558.

asked if they had conviviality or socialized with the men they recognized as homosexuals or bisexuals.

Sources of social and emotional support were classified according to the relationship of the respondents with the providers of support; in case there were males who provided support, respondents were questioned if they were homosexual or bisexual.

A history of sexually transmitted diseases was obtained. A history of urethritis was obtained by asking about symptoms and recalling medical diagnosis. In addition, reports of having had syphilis, genital warts, genital herpes, and lice were sought.

The operational definition of homosexual behavior included having engaged in oral or anal intercourse with a male, and having had active physical contact with other men during masturbation.

After the sexual behavior questions, positive and negative attitudes regarding male homosexuality and bisexuality were asked: respondents were asked if they considered that a man who had sex with other men was a sinner, a sexual pervert, had a mental or a hormonal disease, or was a normal person.

The Heckman method is a two-stage estimation method that provides consistent estimates for interdependent truncated variables. This two-stage method for assessment and control of sample selection bias was originally suggested by J. Heckman,^{13,19} and later extended to a wider class of models by Lee and others.²⁰

The two-stage Heckman procedure first estimates regression parameters from a maximum likelihood estimate (MLE) probit model (the selection equation), and then estimates a substantive equation by ordinary least squares (OLS) regression. These two regression equations are assumed to have correlated disturbances (e_1 and e_2). The general model would be:

$$Y_{1i} = \beta_1' X_{1i} + \varepsilon_{1i} \quad \text{Substantive equation} \quad (1a)$$

$$Y_{2i} = \beta_2' X_{2i} + \varepsilon_{2i} \quad \text{Selection equation} \quad (2a)$$

i = Observation from $i=1$ to $i=n$

Y_{1i} = Reports of male same-gender sexual behavior
(1=Yes; 0=No)

Y_{2i} = Participation (1=Yes; 0=No)

β_1, β_2 = Vectors of exogenous explanatory variables for equations 1 or 2.

X_{1i}, X_{2i} = Parameter vectors of equations 1 or 2.

$\varepsilon_j \sim N(0,1)$ with correlation $[\varepsilon_1, \varepsilon_2] = \rho$

$Y_j = 1$ if $Y_j^* > 0$ and 0 else

(Y_{1i}, X_{1i}) is observed only when $y_{2i} = 1$

The endogenous variable of the substantive equation (reports of homosexual behavior or Y_{1i}) could only

be observed if the endogenous variable of the selection equation exceeded a threshold and participation occurred. If there is selection bias, there would be a correlation between the error or disturbance terms in both equations, therefore, the correlation $-\rho$ would be significantly different from zero.

The substantive equation to be estimated can be rewritten:

$$Y_{1i}^* = E(Y_{1i} | X_{1i}, Y_{2i} \geq 0) = \beta_1' X_{1i} + [\sigma_{12}/(\sigma_{22})^{1/2}] \lambda_i \quad (1b)$$

For equation (1b) the conditional expectation of the endogenous variable is equal to the expected value of the right hand side of the original substantive equation (1a) plus a new term. The new term of the substantive equation (1b) is the product of two terms. The first is the ratio of the covariance between the disturbances in equations 1a and 2a ($\varepsilon_1, \varepsilon_2$) to the standard deviation of the disturbances in equation 2a. The ratio serves as a regression coefficient, and if the covariance of the two disturbances are uncorrelated, this term disappears. The second term, λ , is a hazard rate. This rate represents for each observation, the instantaneous probability of being excluded from the sample. The hazard rate also captures the expected values of the disturbances in the substantive equation after the nonrandom selection has occurred. By including the hazard rate as an additional variable, one is controlling for the nonzero expectations of the disturbance term.¹⁰

Both ρ and λ are used to characterize selectivity effect. Some researchers, especially economists, are used to the selectivity effect summarized not by ρ but by λ . Another way of defining the hazard rate (λ)²¹ is as the product of the correlation between the selection and the substantive equations (ρ) multiplied by the standard error of the residual of the substantive equation ($\sigma = \sigma$): $\lambda = \rho \sigma$.

If ρ or σ are equal to zero, then λ is zero. Therefore, the effect of sample selection can be assessed by testing the null hypothesis that any of λ , ρ or σ are equal to zero.

For continuous dependent variables in the substantive equations, the Heckman method provides consistent although not fully efficient estimates. When the dependent variable of the substantive equation (Y_{1i}) is dichotomous, this method does not provide consistent estimates, and a modification to the general model is needed. An extension of the Heckman method was developed to improve efficiency and to obtain consistent estimates, using the final values from the Heck-

man procedure (using OLS) as the starting values for a maximum likelihood estimation using two probit equations.^{13,14,22}

This two-stage equations system is very sensitive to correct specification of the model in order to provide consistent estimates. Therefore, tests for multicollinearity were performed to improve the specification of the models.

Probit regressions were used to model participation and reports of same-gender sexual behavior. The most parsimonious models for each of the regression equations were selected by deleting non-statistically significant predictors in the Heckman system of simultaneous equations.

The independent probit equations (SGSB and participation) and the modified Heckman two-stage simultaneous regressions* were estimated using LIMDEP version 5.1.²³ All estimates assumed simple random sampling, even though some clustering of the sample occurs by design. Clustered samples will not bias estimates, but will likely result in conservative standard errors and estimated confidence intervals.²⁴

Results

Participation

An initial total of 13 057 addresses were included in the sampling frame. One-third of the households were excluded for several reasons: vacancy at the time of the survey (13%); not being a household (6%); non-response to the household interview (1.6%); and having no eligible men between 15 and 60 years old in the household (13%). Consequently, 8 759 households were considered eligible, and 13 713 men between 15 and 60 years old lived in these households. Of these, 8 600 (63%) were able to be contacted on a person-to-person basis and asked to respond to the individual questionnaire; 532 (6%) refused to answer the questionnaire and 8 068 were successfully interviewed. The overall response rate of the survey was thus 59%.

Taking into account only those households in which there was at least one eligible man, the median number of individuals living in each household was 5 and the median number of eligible respondents per household was 2. In 42% of the households, all eligible men were successfully interviewed, conversely, in 26% of the households no interviews were obtained.

In the remaining sample, varying degrees of response were obtained: in 24% of the sample half or more responded (but not all eligible men); in 8% of the households, less than half of the eligible men in the household were interviewed. The median number of visits per individual in each household was two: 28% of the interviews were gathered during the first visit to the household, 23% in the second visit.

Participation rates were modeled using a probit regression with individuals' sociodemographic characteristics and with variables regarding the context of the interview. The multivariate parameter estimates and their statistical significance are shown in Table I.

There was a significant trend in the response rate according to the number of eligible men per household. For example, in the households in which there was only one eligible man the response rate was 61%, and in those with more than five eligible men, the response rate was 53% ($p < 0.0001$). The participation rate was negatively associated with the number of visits ($p = 0.0001$).

Age was a significant predictor of participation, with a higher response rate among younger individuals. Men with more than primary school but less than 3 years of college and those identified as students were more likely to participate than others.

Participation rate was significantly associated with residence in one of four Sanitary Jurisdictions in Mexico City. There was also an association of the response rate with the field teams in charge of different geographic areas. Less than one-third of the households were visited by female interviewers (29%). There was no significant difference in the response rate according to the gender of the interviewer. There was no significant association between participation and the relationship of potential respondents with the "head of the household."

This probit model correctly predicted 67% of the observations;* 82% of participants and 53% of non-participants. The pseudo- r^2 of this regression model was 0.13.²³

Predictors of same-gender sexual behavior

Almost 16% of participants had not initiated their "active sexual life", e.g. never had anal, oral or vaginal intercourse. Thirty-seven men (0.4%, 95% CI: 0.3-0.6%) reported exclusive lifetime homosexual behavior and

* Known also as "bivariate probit model with selectivity" by W. Greene, in reference to the bivariate normal distribution of the two dependent variables of the simultaneous equations system.

* Correct predictions were those with values of 0.5 or higher for the predicted probabilities.

Table I
**PARTICIPATION RATE ACCORDING TO INDIVIDUAL'S
 CHARACTERISTICS AND INTERVIEW CONTEXT,
 AND COEFFICIENT ESTIMATES FROM PROBIT REGRESSION.
 MEXICO CITY, 1992-1993**

Characteristics	n	Participation	Coefficient estimates	p
	13 480			
No. of eligible men in the household			-0.0701	0.00001
No. of visits to the household			-0.2487	0.00001
Age			-0.0701	0.00001
Education				
Primary or less	3 668	50.2%	Reference	
High school	7 060	61.6%	0.1501	0.00001
≤3 years of college	1 007	61.7%	0.1642	0.001
≥4 years of college	2 179	53.2%	0.1461	0.0002
Occupation				
Students	2 115	67.5%	0.2638	0.0001
Blue collar	4 282	60.0%	0.0721	0.005
White collar	3 222	55.8%	-0.0124	0.7
Other	3 845	59.3%	Reference	
Female interviewer				
Yes	3 972	62.3%	0.0326	0.3
No	9 942	56.3%	Reference	
Geographic region				
Coyoacan	1 028	50.6	-0.3311	0.00001
Cuauhtemoc	653	69.5	0.3905	0.00001
Iztapalapa	2 474	56.6	-0.1602	0.00001
Xochimilco	498	54.4	-0.3031	0.00001
Others	8 827	60.2	Reference	
Have children				
Yes	7 408	57.1	0.1029	0.001
No	6 506	59.0	Reference	
Field team				
5	3 666	60.7	-0.1110	0.003
6	3 145	49.0	-0.4839	0.00001
7	3 191	58.6	-0.2365	0.00001
50	1 034	57.5	-0.2534	0.00001
Others	2 878	63.8	Reference	

Pseudo- $r^2=0.13$; % of "correctly" predicted cases=67%

The coefficient estimates of this "probit" regression equation show the strength of the association—and its statistical significance in multivariable equations—of the independent variables and the probability of participating in the survey (dependent variable: participation rate). The strength of the association between independent variables can be made directly since they have been transformed to have normal distribution (mean=0 and variance=1). The sign of the coefficient implies direct or inverse association between the dependent and the independent variables. When the independent variables are continuous, the coefficient estimates are interpreted as the increase in the probability that the dependent variable occurs for each unit increase of the independent variable, after being standardized $N(0,1)$. When the independent variables are categorical, the reference value of the variable does not show a coefficient estimate because it is fixed to zero, i.e. non statistical significance as coefficient to the slope of the equation line

173 (2.1%, 95% CI: 1.7-2.4%) reported having had sexual intercourse with both men and women in their lifetime. The remaining 81.6% reported exclusive heterosexual behavior.

Same-gender sex was predicted by variables indicating belonging to gay/bisexual social networks, history of sexually transmitted diseases, positive attitudes towards gay and bisexual males and lack of support from male relatives. Ninety-seven per cent of the cases were correctly predicted in the probit model (pseudo- $r^2=0.14$). See Table II.

There was no evidence to indicate that there was a higher prevalence of SGSB in Mexico City because of migration, i.e., the prevalence of SGSB was not different according to place of birth or time of residence in Mexico City. SGSB estimates were not different according to religion. Age was not significantly associated with reports of same gender sexual behavior in the multivariate model.

As could be expected, men who had a stable female partner or had children, had a lower frequency of reported SGSB. Since these two variables were essentially measuring the same construct, only the variable indicating children was used in the model to avoid multicollinearity. Two per cent of those who had children reported SGSB at least once in their lives. Occupation was not associated with SGSB, except for reporting being a student. Students showed a lower frequency of sexual activity in general and of SGSB in particular.

A history of syphilis, genital herpes, or hepatitis, was not a significant predictor of SGSB. Having had urethritis and the presence of genital warts (condilomas) significantly predicted SGSB. Among those who reported a history of urethritis or condilomas, the prevalence of SGSB was almost five times higher than among those who did not report a history of these sexually transmitted diseases.

Most males reported not receiving social support from any source. However, the most preferred source of support were male-friends. Twenty-two per cent of the sample reported receiving social or emotional support from male relatives, especially the father. Among those who did not receive support from their male relatives the prevalence of SGSB was almost twice that of those who reported receiving support from male relatives. Few reported receiving support or advice from women or from homosexual and bisexual males (0.6%). Fifty-three per cent of those who reported having received support from homosexual and bisexual men also reported SGSB.

Eighty per cent of the respondents stated that they were not aware of the presence of homosexual or

Table II
COEFFICIENT ESTIMATES FROM AN INDEPENDENT PROBIT REGRESSION AND A MODIFIED HECKMAN MODEL PREDICTING SAME-GENDER SEXUAL BEHAVIOR IN 15-60 YEARS OLD MEN IN MEXICO CITY, 1992-1993

Characteristics	n 8 053	% SGSB	Independent Probit	p	Modified Heckman	p
Demographics						
Have children						
Yes	4 222	2.0%	-0.3979	0.00001	0.3724	0.6
No	3 831	3.3				
Student						
Yes	1 426	1.2	-0.4776	0.0001	-0.1453	0.8
No	6 627	2.9				
History of sexually transmitted diseases						
Ever had medical diagnosis of urethritis						
Yes	369	10.6	0.5049	0.00001	0.5060	0.01
No	7 684	2.2				
Ever had condilomas						
Yes	103	11.7	0.4634	0.01	0.5338	0.03
No	7 950	2.5				
Social and support networks						
Have homosexual or bisexual friends						
Yes	109	39.4	1.1450	0.00001	1.1422	0.0006
No	7 944	2.1				
Socialize with homosexual or bisexual men in social environments						
Yes	756	9.6	0.2548	<0.00001	0.1319	0.06
No	7 297	1.8				
Receive social support from gay or bisexual men						
Yes	47	53.2	0.6143	0.0002	0.7976	0.01
No	8 006	2.3				
Receive social support from male relatives						
Yes	1 744	1.5	-0.1981	0.01	-0.1949	0.05
No	6 309	2.9				
Other behavioral variable						
Masturbation in the presence of other men (lifetime)						
Yes	627	11.6	0.6156	<0.00001	0.5793	0.003
No	7 426	1.8				
Attitudes towards SGSB						
Consider a man who has sex with other men as a normal person						
Yes	2 347	4.4	0.2365	0.001	0.2146	0.02
No	5 706	1.9				

Pseudo- $r^2=0.14$; % of "correctly" predicted cases=97%

SGSB= same-gender sexual behavior

The coefficient estimates of the probit regression equation show the strength of the association (simultaneously) of the independent variables and the probability of reporting same-gender sexual relationships (dependent variable); allowing direct comparisons of the strength of the associations between independent variables since they have been transformed to the normal distribution (mean=0 and variance=1). The sign of the coefficient determines the direct or inverse association between the dependent and the independent variables

The coefficient estimates of the Heckman equation show the adjusted estimates controlling for non-response bias by means of maximum likelihood estimates for the censored cases. These coefficient estimates are interpreted in the same way as the probit multivariable equations, but corrected for bias

bisexual men in their workplace, school-place, or in the social and sport clubs that they attend. Of the people who knew of the presence of men who had SGSB in these social situations, 48% reported not socializing with them. Of the people who did socialize with homosexual and bisexual men in these environments, 9.6% reported SGSB. A small proportion of the sample (1.4%) declared that they knew that at least one of their three closest friends had had sex with other men. Among these respondents, 39% also reported SGSB.

A question concerning attitudes towards men who had sex with other men was also associated with SGSB. Those who considered "a man who has sex with other men" as a normal person, had higher rates of SGSB (4.4%).

Almost 8% of the sample reported having practiced masturbation in the presence of other males. However, they were not considered as having SGSB because there was no physical contact with these other men. However, among those who reported this practice, 11.6% acknowledged having SGSB.

Heckman method modeling of same-gender sexual behavior

The sociodemographic variables were used in the equation to predict participation, taking advantage of the fact that they were available for the total sample.* The modeling of SGSB was performed using variables regarding social and support networks (with emphasis in the interaction with homosexual or bisexual men), history of sexually transmitted diseases and attitudes towards men who had sex with other men.

A modified MLE Heckman model for dichotomous dependent variables was estimated to assess non-participation bias. After the probit regression equations of participation and same-gender sexual behavior were entered simultaneously in the model, there were only two significant changes in the estimated coefficients of the substantive equation (see Table II).

* The use of the Heckman model can also be advisable when information about the population from which the sample was drawn is not available (as in the case of convenience samples) and the study design only allows the observation of a selected subsample. In addition to selectivity, data for some outcomes may only be observed given the result of a previous or joint outcome (partial observability). The proper way to deal with a model with sample selection and partial observability was developed by Poirier (simultaneous determination) and by Abowd and Farber in the case of sequential determination. When there exists partial observability, instead of observing Y_1 and Y_2 for all observations, only the product ($Y = Y_1 Y_2$) is observed, and the disturbance terms of both equations are correlated.

The variables measuring having children and occupation as a student lost their statistical significance after introducing the new term as described in equation 2a. It was expected that these variables would experience the greatest changes since they were the only statistically significant variables in both the substantive and the selection equations. One way to interpret these changes is that, after controlling for the probability of not participating in the survey by introducing the hazard rate in the model, the independent relationships between having children and being a student are no longer associated with reports of SGSB.

Since there is a negative correlation between these two equations ($\rho = -0.67$) under-representation of the individuals who were more likely to report SGSB would be expected. However, this correlation did not reach statistical significance ($p = 0.10$). Therefore, there is no significant evidence of a selection effect.

Discussion

In this probability survey, regression models which described participation and same-gender sexual behavior were estimated. The results indicate that the prevalence estimates of same-gender sexual behavior among Mexico City men are not biased by selective survey participation, as was also found in another study of a probability sample of the US population.*

With the experience of this and other surveys, it may be concluded that individuals are as willing to participate in surveys regarding sexual behavior as they are to participate in any other household survey.⁴

For example, the overall response rate of this survey (59%) was lower than the national response rate found in a 1993 multipurpose national survey of chronic diseases in Mexico, that used the same sampling framework (response rate=67%). However, in the chronic disease survey the lowest response rate was for males living in Mexico City (64%). Therefore, the response rate found in these surveys may be more related to the difficulties of conducting a probability sample in a large city, rather than to the topic of the survey.²⁵

The careful selection and training of the interviewers may have assisted in minimizing potential selection bias by avoiding hostile attitudes towards potential respondents, by guaranteeing confidentiality, and by providing an adequate environment for the develop-

* Michaels R, Veldman C. Behind doors: An analysis of non-response to the GSS sexual behavior supplements. Chicago: National Opinion Research Center, 1992. Typescript, page 17. Citado en referencia 4:558.

ment of the interview. However, it still remains to be tested if the use of self-administered questionnaires performs better than a face-to-face approach, in an attempt to minimize inaccurate reporting of SGSB. In the pilot studies for this survey, more people complained about seeing cards that portrayed explicit sexual acts than hearing the description of sexual behaviors.

A test for misclassification of respondents' reports of SGSB was not presented in this study. However, in a previous communication, results of HIV testing in a subsample of 1 700 participants of this survey did not suggest misclassification: 4.2% of men reporting SGSB, and 0.08% of men reporting lifetime exclusive heterosexual behavior were HIV reactive in blood and/or saliva ($p < 0.001$).¹⁸

A potential limitation of the results of this study is that even though the correlation between the error terms in the selection and the substantive equation was relatively large (-0.67), the fact that it was not statistically significant ($p = 0.10$) could be explained by lack of statistical power due to the relatively small number of individuals who reported SGSB, and particularly to the expected low prevalence of SGSB among non-participants. The fact that this sample size was obtained after making a probability selection of 13 057 addresses exemplifies the difficulties in assessing and correcting for selection bias when the variable under study has such a low frequency.

References

1. Kinsey A, Pomeroy W, Martin C. Sexual behavior in the human male. Filadelfia: WB Saunders, 1948.
2. Masters WH, Johnson VE. Human sexual response. Boston: Little Brown, 1966.
3. Boulton M. Review of the literature on bisexuality and HIV transmission. En: Tielman RAP, Carballo M, Hendriks AC, ed. Bisexuality and AIDS. A global perspective. Buffalo (NY): Prometheus Books, 1991.
4. Lauman EO, Gagnon JH, Michael RT, Michaels S. The social organization of sexuality. Sexual practices in the United States. Chicago (ILL): The University of Chicago Press, 1994.
5. Michael RT, Laumann EO, Gagnon JH, Smith TW. Number of sex partners and potential risk of sexual exposure to HIV. MMWR Morb Mortal Wkly Rep 1988;37:565-584.
6. Rogers SM, Turner CF. Male-male sexual contact in the USA: Findings from five sample surveys, 1970-1990. J Sex Res 1991;28:491-519.
7. Reiss IL. Is this the definitive sexual survey? [Book Review of "The Social Organization of Sexuality"]. J Sex Res 1995;32:77-91.
8. ACSF investigators. AIDS and sexual behavior in France. Nature 1992;360:407-409.
9. Johnson AM, Wadsworth J, Wellings K, Bradshaw S, Field J. Sexual lifestyles and HIV risk. Nature 1992;360:410-412.
10. Berk RA. An introduction to sample selection bias in sociological data. Am Sociol Rev 1983;48:386-398.
11. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. Psychol Bull 1990;108:339-362.
12. Fay RF, Turner CF, Klassen AD, Gagnon JH. Prevalence and patterns of same-gender sexual contact among men. [Published erratum appears in Science 1989;244:1531]. Science 1989;243:338-348.
13. Heckman JJ. Sample selection bias as a specification error. Econometrica 1979;47:153-161.
14. Greene, W. Econometric Analysis. 2nd ed. Nueva York: Macmillan, 1993.
15. Grotzinger KM, Stuart BC, Ahern F. Assessment and control of nonresponse bias in a survey of medicine use by the elderly. Med Care 1994;32:989-1003.
16. Izazola-Licea JA. Same-gender Sexual behavior and HIV/AIDS risk among adult males in Mexico. Doctoral thesis. Boston (MA): Harvard School of Public Health, 1995.
17. Izazola-Licea JA, Gortmaker S, De Gruttola V, Tolbert K, Mann J. Prevalence of same-gender sexual behavior and HIV in probability household survey in Mexican men. J Sex Res 2000;37(1):37-43.
18. Nieto-Andrade B, Izazola-Licea JA. Uso del condón en hombres con parejas no estables en la ciudad de México. Salud Publica Mex 1999;41:85-94.
19. Maddala GS. Limited-dependent and qualitative variables in econometrics. Londres: Cambridge University Press, 1983.
20. Lee LF. Some approaches to the correction of selectivity bias. Rev Econom Stud 1982;XLIX:355-372.
21. Stata Corp. Stata Statistical Software: Release 4.0. College Station, Texas: Stata Corporation, 1993; vol. 2:441.
22. Lee, LF. Identification and estimation in binary choice models with limited (censored) dependent variables. Econometrica 1979;47:977-996.
23. Greene WH. Manual for LIMDEP Version 5.1. Nueva York: Econometric Software, 1989.
24. Sudman S. Applied sampling. Quantitative studies in social relations. Nueva York: Academic Press, 1976.
25. Secretaría de Salud. Encuesta Nacional de Enfermedades Crónicas. México, D.F.: SSA, 1994.