



# The Randomized Response Technique Application in the Survey of Homosexual Commercial Sex among Men in Beijing

*Xiangyu CHEN, Qiaoqiao DU, Zongda JIN, Tian XU, Jiachen SHI, \*Ge GAO*

*Dept. of Epidemiology and Health Statistics, School of Public Health, Medical College of Soochow University, Suzhou, China*

**\*Corresponding Author:** Email: gaoge01@163.com

(Received 11 Nov 2013; accepted 21 Feb 2014)

## Abstract

**Background:** Traditional survey methods may cause refusals to respond or untruthful replies when encounter the sensitive questions. The Randomized Response Technique (RRT) is designed to decrease social desirability bias and obtain reliable estimates. This study aimed to apply these new methods on the sensitive questions survey.

**Methods:** Simmons model was applied in the survey of issues relevant to commercial sex among men who have sex with men (MSM) in Beijing, China. Stratified two-stage sampling and stratified random sampling were applied.

**Results:** During July to December 2010 in Beijing, the commercial sex proportion among MSM was 0.051(95% CI: 0.017, 0.085), its estimated variance was  $3.01 \times 10^{-4}$ . The last male condom use rate during commercial sex was 0.778(95% CI: 0.746, 0.810), its estimated variance was  $2.6 \times 10^{-4}$ .

**Conclusion:** We have introduced new approaches for the sensitive question survey among MSM. The new approaches are proved to be reliable and valid.

**Keywords:** Randomized response technique, Homosexual commercial sex, Stratified random sampling, Stratified two-stage sampling, China

## Introduction

Sampling survey is commonly used in numerous fields of statistics research. Sampling survey which related to sensitive information is inevitably encountered. Sensitive questions mean that the content of questions themselves can be considered as intrusive and the answers to questions are socially undesirable or the meanings of questions have threat of disclosure. Such as premarital sex, HIV-positive, commercial sex, homosexuality, and drug abuse. If respondents are faced with a question that they find embarrassing, they may refuse to answer, or choose a response which prevents them from having to continue with the questions. Potential threat which leads to the decline of accuracy is both nonresponse and reporting error (1).

A proposed solution is the RRT which designed for the purpose of decreasing social desirability bias, guaranteeing confidentiality, improving respondent cooperation and procuring reliable estimates (2).

The Chinese Ministry of Health reported that in China the percentage of HIV infection caused by homosexual behavior rose from 14.7% in 2009 to 17.4% in 2011(3). It is significant to acquire the character of the sexual behavior especially the commercial sex behavior among MSM when involved preventing the spread of AIDS. However, the current survey of commercial sex behavior among MSM is carried out by asking questions directly. As these questions are sensitive, it could

lead to bias. In this case, we use RRT to improve the reliability.

In this study, we focused on introducing new approaches for the sensitive question survey among MSM, for a better prevention of aids.

## Materials and Methods

### *Study population*

The study population was MSM ages from 15 to 49 in Beijing, China. According to the sixth national census, the resident population (referred to the population which often lived in Beijing for more than six months) in Beijing was 19.612 million (4), the male people ages from 15 to 49 was 6.775 million, and the MSM accounted for 1.0% of the male people ages from 15 to 49 (5), therefore, the total number of MSM ages from 15 to 49 in Beijing was 67750. We got the ID number of each MSM with the help of the local center for disease control and prevention. We used SPSS to generate a random number table. Then the MSM whose ID numbers in the table were selected representing the participants.

### *Stratified random sampling*

In this sampling, the stratum was formed based on the differences in the venues where the MSM seeking sexual partners, known as the actual venues and the internet. In December 2010, 372 respondents were selected in the actual venues, 545 respondents were selected via the internet, 917 respondents in total. The sample size was calculated based on the study of William. G. Cochran (6) so as to minimize the sampling errors for specified cost. The number of participants in each stratum was calculated based on optimum allocation stratified sampling method (6).

### *Stratified Two-stage sampling*

As the condom use rate varied across ages (7), in order to reduce the sampling bias, the MSM population in Beijing was divided into two strata according to the age. The first stratum consisted of the MSM ages from 15-29 yr and the second stratum consisted of the MSM ages from 30-49

yr. The first stratum accounted for 58.24% while the second is 41.76%. In each stratum, the 16 districts in Beijing were taken as the first-stage units ( $N_1 = N_2 = 16$ ), and the MSM population were taken as the second-stage units (67750 in total). Each district contained 4234 MSM ( $\bar{N}_2 = 4234$ ), therefore, in each district, the first stratum contained 2466 MSM ( $\bar{N}_{21} = 2466$ ) and the second stratum contained 1768 MSM ( $\bar{N}_{22} = 1768$ ). According to the formulae for the optimum sample sizes for stratified two-stage sampling carried out by Jianfeng Wang (8), so as to minimize the sampling errors for specified cost, 13 districts were randomly selected from each stratum in the first stage, 1523 MSM were randomly selected from the selected districts in the second stage. According to the sample allocation study carried out by Marcin Kozak (9), in the first stratum, 68 MSM were selected from each selected district ( $\bar{n}_{21} = 68$ ), in the second stratum, 49 MSM were selected from each selected district ( $\bar{n}_{22} = 49$ ).

### *Simmons model*

Simmons model which based on Warner's RRT was put forward by Simmons in 1967(10). In this design, respondents were given two unrelated questions. One was a sensitive question and the other was a non-sensitive question. The kind of question for choosing depended on the outcome of a randomization set. The probability ( $P$ ,  $P \neq 0.5$ ) of selecting the sensitive question was determined during the randomization set design. Depending on the outcome of the randomization process, the respondents needed to answer "Yes" or "No" to the selected questions. In the survey of the commercial sex proportion, the sensitive question was: In the last 6 months, have you had commercial sex with men? The non-sensitive question was: Is your domicile Beijing? In the survey of condom use rate during commercial sex, the sensitive question was: Did you use condom during last male commercial sex? The non-sensitive was: Is your birth date an odd number? The randomization set included a semi-closed box, 6 red balls and 4 white balls. The balls were same in size, weight and tactility. The respondents who

selected the red ball answered the sensitive question. Otherwise, the respondents answered the non-sensitive question.

**Formulas applied in the survey**

The formulas that applied in the commercial sex proportion survey are listed as follows (11):

$$p_h = \frac{m_h / n_h - (1 - P)R_h}{P} \quad [1]$$

$$v(p_h) = \frac{m_h / n_h (1 - m_h / n_h)}{n_h P^2} \quad [2]$$

$$p = \frac{\sum_{h=1}^L N_h P_h}{N} = \sum_{h=1}^L W_h P_h \quad (W_h = \frac{N_h}{N}) \quad [3]$$

$$v(p) = \sum_{h=1}^L W_h^2 v(p_h) \quad [4]$$

The formulas that applied in the condom use rate survey are listed as follows:

$$p = \sum_{h=1}^L W_h P_h \quad (W_h = \frac{N_h}{N}) \quad [5]$$

$$p_h = \frac{\sum_{i=1}^{n_{1h}} N_{i2h} P_{ih}}{\sum_{i=1}^{n_{1h}} N_{i2h}}, \quad h = 1, 2, \dots, L \quad [6]$$

$$P_{ih} = \frac{\hat{\lambda}_{ih} - (1 - P)R_{ih}}{P} = \frac{m_{ih} / n_{i2h} - (1 - P)R_{ih}}{P} \quad [7]$$

$$V(p) = \sum_{h=1}^L W_h^2 \left( \frac{\sigma_{1h}^2}{n_{1h}} \left( 1 - \frac{n_{1h}}{N_{1h}} \right) + \frac{\sigma_{2h}^2}{n_{1h} \bar{n}_{2h}} \left( 1 - \frac{\bar{n}_{2h}}{N_{2h}} \right) \right), \quad [8]$$

$$S_{1h}^2 = \frac{1}{n_{1h} - 1} \sum_{i=1}^{n_{1h}} \left( \frac{N_{i2h}}{N_{2h}} \right)^2 (P_{ih} - p_h)^2 \quad [9]$$

$$S_{2h}^2 = \frac{1}{\sum_{i=1}^{n_{1h}} N_{i2h}} \sum_{i=1}^{n_{1h}} N_{i2h} P_{ih} (1 - P_{ih}) / n_{i2h} \quad [10]$$

**Data management**

All the participants had signed the informed consents. The questionnaires were completed by respondents independently and all of the questionnaires were checked carefully. The response rate reached 100% and all questionnaires were qualified. The collected data was checked twice and used to set up the database with EpiData 3.1. SAS 9.2 was used to analyze the data.

**Results**

**Baseline results of the commercial sex proportion survey**

This survey was conducted with 917 respondents during July to December 2010 in Beijing. In which 15.8% of respondents' domiciles are Beijing, 47 in actual venues whereas 98 in the internet. The MSM who were older than 25 accounted for 68% in actual venues while the proportion via the internet was 51.4%. The MSM who seek sexual partners in actual venues were younger than those who seek sexual partners via the internet ( $\chi^2 = 25.13, P = 5.35 \times 10^{-7}$ ). The proportion of the MSM who received college education in actual venues is 23.4% while this proportion via the internet is 64.8%. The education level of the MSM differ between the two venues ( $\chi^2 = 151.70, P = 7.36 \times 10^{-35}$ ) (Table 1).

**Table 1:** Demographic characteristics of the MSM

Variables	Actual venues	Internet	$\chi^2$	P
Age(years)				
≤25	119	265	25.13	5.35*10 <sup>-7</sup>
>25	253	280		
Education level				
Lower than Tertiary	285	192	151.70	7.36*10 <sup>-35</sup>
Tertiary and higher	87	353		
Domicile				
Beijing	47	98	4.75	0.0293
Outside Beijing	325	447		

**The commercial sex proportion among MSM**

The commercial sex proportion among MSM who seek sexual partners in the actual venues in Beijing during July to December 2010 was 11.94%, its estimated variance was  $8.09 \times 10^{-4}$ . This proportion among MSM who seek sexual partners via the in-

ternet was 3.29%, the estimated variance was  $4.25 \times 10^{-4}$ . The total commercial sex proportion among MSM in Beijing during July to December 2010 was 5.11% and its variance was  $3.01 \times 10^{-4}$ , the 95% CI was 0.0171-0.0851 (Table 2).

**Table 2:** The proportion of commercial sex among MSM

Venues	$m_h$	$n_h$	P	$R_h$	W	P	$v(p)$
Actual venues	46	372	0.6	0.13	0.21	0.1194	$8.09 \times 10^{-4}$
Internet	50	545	0.6	0.18	0.79	0.0329	$4.25 \times 10^{-4}$
Total	-	-	-	-	-	0.0511	$3.01 \times 10^{-4}$

- $m_h$ : the number of respondents who answered "Yes" in corresponding stratum
- $n_h$ : the total number of respondents in corresponding stratum
- $R_h$ : the proportion of respondents with non-sensitive character in corresponding stratum
- $p$ : the commercial sex proportion in corresponding stratum
- $v(p)$ : the estimated variance of commercial sex proportion in corresponding stratum

**The comparison of the commercial sex proportion among MSM**

As the sample size of these two stratum is large and  $p_1, p_2$  are not too close to 0 and 1,  $p_1$  and  $p_2$  obey approximate normal distribution. Thus,  $p_1 - p_2$  obey approximate normal distribution. As  $p_1, p_2$  were mutually independent,  $v(p_2 - p_1) = 0.001234$ .

The test statistic Z is -2.46.

The commercial sex proportion among MSM who seek sexual partners in the actual venues is higher than the proportion via the internet ( $Z = -2.46, P = 6.90 \times 10^{-3}$ ).

**Results of the condom use rate survey**

In the first stratum (ages 15-29,  $h=1$ ), district 1,66 MSM were selected as respondents ( $n_{12h} = n_{121} = 66$ ), 38 MSM answered "Yes". Thus, the proportion is:  $\sigma_{2h}^2$ . The proportion of MSM whose

birth date was the odd number accounted for 0.5 ( $R_1 = 0.5$ ). The red ball percentage accounted for 0.6 ( $P = 0.6$ ). According to the formula [7], in the primary stratum district 1, the sample ratio of last male condom use during commercial sex is 0.6263. The last male condom use sample ratio in the second stratum district 1 and the ratios in these two stratum in other districts were calculated in the same way.

The last male condom use sample ratio in the two stratum and the sample estimators of  $\sigma_{1h}^2$  and  $\sigma_{2h}^2$  were calculated according to the formula [6, 9, 10] (Table 3). According to the formula [5], the last male condom use behavior during anal sex ratio in Beijing is 0.7780. According to the formula [8], the estimated variance of P is 0.00026. The 95% CI of the last male condom use rate is 0.7463-0.8097.

**Table 3:** The last condom use sample ratio and sample estimators of  $\sigma_{1h}^2$  and  $\sigma_{2h}^2$  in the two stratum

Stratum (h)	$p_h$	$S_{1h}^2$	$S_{2h}^2$
primary (ages 15-29, $h=1$ )	0.7689	0.0262	0.0033
second (ages 30-49, $h=2$ )	0.7908	0.0526	0.0027

## Discussion

In medical and epidemiological research, the MSM population are frequently targeted as a “high risk group” for HIV acquisition and transmission because most of the MSM engage in anal sex, which carries the highest risk of HIV transmission among the various sexual practices (12). The commercial sex among MSM is more significant in the AIDS spreading due to the multiple sex partners, improper sexual behavior and large fluidity. The correct use of condoms during sex can reduce the risk of HIV transmission. Therefore, to detect the commercial sex proportion among MSM and the condom use rate during commercial sex is important for AIDS prevention and control. However, the current survey of commercial sex behavior among MSM was carried out by asking questions directly. As the questions about commercial sex behavior are sensitive to these respondents, they often decline to take part or reply untruthfully and are more willing to give socially desirable responses rather than accurate ones. In this case, we use RRT to guarantee confidentiality, improve respondents’ cooperation and obtain reliable estimates and decrease social desirability bias. It was the first time that we use Simmons model combined with these two sampling methods to investigate the MSM in Beijing.

Reliability and validity are two important indicators which can reflect the reliable and true character of objects in the study. While establishing good quality studies through reliability and validity in qualitative research, states that the “trustworthiness of a research report lies at the heart of issues conventionally discussed as validity and reliability” (13). This study is a part of the project supported by the National Natural Science Foundation. Before this paper, corresponding methods were studied which combined several RRT models with complicated sampling methods. Monte Carlo methods were used to evaluate the reliability and validity of the methods. All these methods were proved to be reliable and valid (14, 15).

In order to get the accurate data, two points should be paid attention in conducting the sam-

pling. Training the investigators strictly for the overall understanding about the sampling methods is the first point; the other is to ensure that the sample is representative for the total population. In order to reduce response bias and get reliable answers, the investigators informed the participants that the RRT can protect their personal privacy and also provide cash rewards to them.

In previous studies, the researchers only investigated the MSM in the actual venues, ignoring the massive online MSM population. The internet has become a major tool in MSM socializing, especially for younger generations (16-18). In this study, the online MSM population was taken into account. The actual venues in the study include bathhouses, gay clubs, parks etc. In the commercial sex proportion survey, the estimated proportions of commercial sex among MSM who seek sexual partners in actual venues and via the internet in Beijing during July to December 2010 were 11.94% and 3.29% respectively, 5.11% for overall. The commercial sex proportion among MSM who seek sexual partners in actual venues was higher than that via the internet, the result was close to the survey made by Han Delin et al (19). The MSM who seek sexual partners via the internet had a higher education level when compared with the MSM who seek sexual partners in the actual venues ( $\chi^2=151.70$ ,  $P=7.36*10^{-35}$ ); the MSM who seek sexual partners in the internet mostly were young men ( $\chi^2=25.13$ ,  $P=5.35*10^{-7}$ ), many of them had college degree, some were even students in school. A research showed that 92% of the HIV infections in school are transmitted through the sexual contact among men (20). In summary, on the one hand, the propaganda and intervention of AIDS in the actual venues is more urgent than in the internet. On the other hand, although the commercial sex proportion among MSM who seek sexual partners via the internet was lower than that in actual venues, however, the online MSM population was larger than that in the actual venues. Therefore, the disseminated of the AIDS prevention knowledge in the internet especially in the homosexuality websites should be enhanced. Besides, the educational administrative depart-

ments need to recognize that the school education plays an important role in the youth AIDS education and they must provide corresponding policy support. Students and teachers in the schools should avoid the discrimination against the MSM population.

In the condom use rate survey, as the survey results show, the condom use rate is 77.80%, there is still 22.20% unprotected sex behavior occurred. The situation of condom use rate among MSM is still need to be improved. Thus, measurements on safe sexual behavior towards MSM should be enhanced; the future prevention efforts should consider contextual components of condom use, including partner type, location of the sexual event, and semen exposure, to more accurately develop individualized risk reduction strategies (21-22).

## Limitations

The participants were selected only in Beijing. We can get the information of the MSM easily because of a well-established aids surveillance system in Beijing. However, it is difficult to obtain the information of the MSM in remote areas of China. Thus, larger scale surveys among MSM based on RRT needed to be further carried out.

## Conclusion

The RRT model which combined with two sampling methods in this study was reliable and valid. The new approach is highly available and valuable in the future survey on MSM population.

## Ethical considerations

Ethical issues (Including plagiarism, Informed Consent, misconduct, data fabrication, double publication and/or submission, redundancy, etc.) have been completely checked by the authors.

## Acknowledgement

This study was supported by National Natural Science Foundation of China (No. 81273188). We are grateful to Tian Xu, Qiaoqiao Du, Wei Li, Zongda

Jin, for their invaluable help in field investigations and calculations. The authors declare that there is no conflict of interest.

## References

1. Tourangeau R and Yan T (2007). Sensitive questions in surveys. *Psy Bul*, 133(5): 859-63.
2. Cruyff M, Hout A, Heijden P and Böckenholt U (2007). Log-linear Randomized-Response models taking self-protective response behavior into account. *Methods & Research*, 36:266-82.
3. Ministry of Health of the People's Republic of China, UNAIDS, WHO (2011). *ES&PP of AIDS in China in 2011*.
4. Beijing Municipal Bureau of Statistics (2011). *Beijing Statistical Yearbook 2010*.
5. Wang LY, Xia DY, Wu YH (2004). Application of a multiplier method to estimate the population size of men who have sex with men. *SCJPM*, 32:9-11.
6. William GC (1977). Sampling techniques, 3rd edition. *Wiley*:89-111.
7. Rosenberger JG, Reece M, Schick V (2012). Condom use during most recent anal intercourse event among a U.S. sample of men who have sex with men. *J Sex Med*, 9(4):1037-47.
8. Wang JF, Gao G, Fan YB, et al.(2006). The estimation of sampling size in multi-stage sampling and its application in medical survey. *AMC*, 178:239-49.
9. Marcin K (2007). Stratified two-stage sampling in domains: sample allocation between domains, strata, and sampling stages. *Statistics & Probability Letters*, 57:1-8.
10. Simmons WR, Horvitz DG, Shah BV, et al. (1967). The Unrelated Question Randomized Response Model Proceedings in the Social Statistics Section. *ASA*, 64:520-39.
11. Liu W, Li XD (2010). Stratified Random Sampling on Simmons Model for Sensitive Question Survey. *Soochow University J Med*, 30(4): 759-62.
12. Varga CA (1997). The condom conundrum: barriers to condom use among commercial sex workers in Durban, South Africa. *AJRH*, 1(1):74-88.
13. Nahid G (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, 8(4):597-607.

14. Fan YB (2009). Stratified Cluster Sampling and Its Application on the Simmons RRT Model for Sensitive Question Survey. *JCHS*, 26: 562-65.
15. Du QQ (2012). Application of Monte Carlo Simulation in Reliability and Validity Evaluation of Two-Stage Cluster Sampling on Multinomial Sensitive Question. *LNCS ICICA2012*, 261-68.
16. Choi KH, Ning Z, Gregorich SE, et al. (2007). The influence of social and sexual networks in the spread of HIV and syphilis among men who have sex with men in Shanghai, China. *J AIDS*, 45:77-84.
17. Bolding G, Davis M, Hart G, et al. (2005). Gay men who look for sex on the internet: is there more HIV/STI risk with online partners? *AIDS*, 19:961-68.
18. Zhang D, Bi P, Lv F, et al. (2007). Internet use and risk behaviors: an online survey of visitors to three gay websites in China. *Sex Transm Infect*, 83:571-76.
19. Han DL, Wang XD, Liang X (2009). Comparative Study on AIDS related knowledge, attitude and behavior among MSM in different places of Chengdu. *JMPM*, 36:2910-15.
20. Hightow L, Macdonald P, Pilcher C (2005). The unexpected movement of the HIV epidemic in the Southeastern United States: transmission among college students. *AIDS*, 38:531-37
21. Mukandavire Z, Bowa K, Garira W (2007). Modelling circumcision and condom use as HIV/AIDS preventive control strategies. *Mathematical and Computer Modelling*, 1353-1372.
22. Rosenberger JG, Reece M, Schick V et al. (2012). Condom use during most recent anal intercourse event among a U.S. sample of men who have sex with men. *J Sex Med*, 9(4):1037-47.

Archive of SID