Original Article

Application of a randomized response technique for the sensitive question survey in female sex workers in Xichang, China

Kejin Chen[1,2], Zhilei Mao[3], Zongda Jin[2,4], Ge Gao[2], Chaorong Bian[5]

[1]Group Health Division, Changzhou Maternal and Child Health Care Hospital, Changzhou, China; [2]Department of Epidemiology and Biostatistics, School of Public Health, Medical College of Soochow University, Suzhou, China; [3]Science and Education Department, Changzhou Maternal and Child Health Care Hospital, Changzhou, China; [4]Medical Service Division, The Third Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China; [5]Medical Service Division, Changzhou Children's Hospital Affiliated to Nantong University, Changzhou, China

Contributions: (I) Conception and design: G Gao, K Chen; (II) Administrative support: G Gao, C Bian; (III) Provision of study materials or patients: G Gao, Z Jin; (IV) Collection and assembly of data: Z Jin, Z Mao; (V) Data analysis and interpretation: K Chen, C Bian, Z Jin; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Chaorong Bian. Medical Service Division, Changzhou Children's Hospital Affiliated to Nantong University, 468 Yanling Road, Changzhou 213000, China. Email: 895215152@qq.com.

Background: The technique of sampling is a particularly important aspect of sampling surveys, and the method of determining sample size is the key link in sampling technique. However, a sample size determination method for complex sampling surveys of sensitive issues using a randomized response model is not yet available.

Methods: In this work, we drew from the theory of probability and mathematical statistics, and developed a series of formulas for sample size determination that are required to estimate the overall proportion and mean on dichotomous sensitive questions in 3-stage sampling using Simmons models.

Results: Shortly after generating the formulas, we successfully applied them to a pre-survey regarding sensitive issues among female sex workers (FSWs) in Xichang, and found their practical application to achieve satisfactory effects. By using a large number of Monte Carlo simulation sampling tests, the reliability of the survey methods and formulas for sensitive question surveys were found to be high, but the randomized response technique (RRT) cannot completely eliminate the possibility of social acceptability bias.

Conclusions: RRT is an acceptable method for sensitive question surveys within sensitive populations, and this new sampling method is liable to be used for sensitive question surveys of large populations. We believe that the results obtained in the field of public health can provide some data support for clinical practice.

Keywords: Randomized response technique (RRT); sensitive question survey; female sex worker (FSW)

Submitted Oct 28, 2020. Accepted for publication Jan 02, 2021.
doi: 10.21037/apm-20-2262
View this article at: http://dx.doi.org/10.21037/apm-20-2262

Introduction

Surveillance data from 2004 to 2013 in China showed that the incidence of human immunodeficiency virus (HIV) and syphilis infection had the most rapid increase, with an annual increase in prevalence of 16.3% [95% confidence interval (CI), 11.5–21.2], and 16.3% (95% CI, 13.8–18.8) respectively (1). The increased incidence of sexually transmitted diseases (STDs) was attributed to a variety of reasons, including internal labor migration, unsafe sexual practices, and lack of consulting healthcare services (2-4). Female sex workers (FSWs) play a very important role in the spread of STDs, and have been viewed as key links of such diseases between groups of STD infections and the general population (5). There has been much investigation into FSWs, but the majority of them have used the traditional questionnaire inquiry method only (6,7). As commercial sexual activities are illegal in China, and Chinese women are conservative about sex, this demographic is highly private and has a strong sense of self-protection. It is difficult to obtain data from the respondents encompassing sensitive topics such as sex services. Effective policies and measures to resolve these increasing problems can only be created using the results of an authentic survey of the STD epidemic in China.

The basic thinking of the randomized response technique (RRT), as used in this study, is still based on the idea of the Warner model (8), but we incorporated a non-sensitive and unrelated issue, as opposed to a control issue as, to directly oppose the sensitive issue as in the Warner model, with the aim of further alleviating the privacy concerns of the respondents. In our model, the respondents believed that their answers would not expose their profession, so they could more freely cooperate with the investigators. Due to the absence of the investigators during the inquiry, the conductors did not know which question the respondents answered, only the respondents themselves knew this detail (9). This enabled the respondents have a sense of security and therefore the answers are more authentic.

In this study, we applied the above method to sensitive issues and complex sampling, as mentioned in another research article (10). Beyond this, through a preliminary survey on the features of sensitive issues for FSWs in Xichang, Sichuan Province, China, we obtained the required relevant statistical values in the calculated formulae to estimate the sample size. This would be helpful in informing policy makers when developing prevention strategies and for guiding further research.

We present the following article in accordance with the SURGE reporting checklist (available at http://dx.doi.org/10.21037/apm-20-2262).

Methods

Study population

We surveyed FSWs from May to July 2018 in Xichang, Sichuan Province, China. According to the sixth national census performed in 2010, there was a population of 278,023 in downtown Xichang. The male to female ratio there was 1.0562:1, and the female population was 135,212 at the time. According to the HIV and AIDS (acquired immune deficiency syndrome) Data Hub for Asia-Pacific (https://www.aidsdatahub.org/), FSWs account for 3.4–3.6% of urban women in the southwest cities of China such as Guiyang and Xingyi. Xichang is located in the hinterland of Liangshan, in the southwest of the Sichuan Province. The Liangshan Prefecture has been regarded as one of the hardest-hit areas for HIV in China. Based on that proportion, we can deduce that the number of FSWs in Xichang, the target population of this research, is 4,876.

Ethics statement

All participants provided their written informed consent to participate in this study anonymously. The data were also collected and analyzed anonymously. This research, including the consent procedure, was approved by the Ethics Committee of Soochow University. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). We confirm that all methods were performed in accordance with the relevant guidelines and regulations.

Public and patient involvement

The survey in this study was an application of the new sampling method, and was a cross-sectional study. We recruited the FSWs at active locations in 3-stage sampling. The places of activity of providing sexual services could be divided into high, middle, and low grade, and the places we selected included all these grades. Our investigators explained RRT to each respondent, ensuring that everyone completed the questionnaires independently after they had signed the informed consent form. The authors of this article trained the investigators in Xichang, and did not directly participate in the investigation themselves.

Survey methods

In general, sensitive issues are related to the privacy or interests of some institutions, organizations, individuals, or what most people are reluctant to state publicly, such as attitudes toward ethics, economic income, and life conduct (11). In accordance with the general characteristics of sensitive questions, they can be divided into sensitive questions of quantitative characteristics and sensitive questions of attribute characteristics (12), which can be further divided into those of 2 attribute characteristics and those of multiple attribute characteristics, according to the number of attribute characteristics. The 2 answers to sensitive questions of 2 attribute characteristics must be mutually independent and exclusive. For example, when the question is "Did you use a condom the last time when you had sex?", the answers must be either "Yes" or "No".

A randomized device was conducted as follows. Red and white balls, which were indistinguishable in size, weight, and texture, were placed into a bag. The proportion of red balls in the bag is referred to as P. Every participant was instructed to draw a ball randomly from the bag. If the participant drew a red ball, they would be asked to answer "Yes" or "No" to the sensitive question A: "Do you have the characteristics A?" If the participant drew a white ball, then they would be asked to answer "Yes" or "No" to the non-sensitive question B: "Do you have the characteristics B?" The respondents participated in this exercise in isolation and put the ball back to the bag between 2 respondents to preserve the ratio P.

Suppose that totality is composed of N_1 primary units, and the i^{th} primary unit is composed of N_{i2} second-stage units (i=1, 2…N_1). Then the j^{th} second-stage unit of the i^{th} primary unit is comprised of N_{ij3} third-stage units (i=1, 2…N_1; j=1, 2…N_{i2}). The population consists of N tertiary units. In other words, each primary unit is made up of $\overline{N}_1$ second-stage units, and each second-stage unit is made up of $\overline{N}_3$ tertiary units.

In the first phase, n_1 primary units were randomly selected. In the second phase, n_{ij3} second-stage units were randomly drawn from the i^{th} selected primary unit (i=1, 2…n_1). In the third phase, n_{ij3} third-stage units were randomly chosen from the j^{th} drawn second-stage unit of the i^{th} selected primary unit (i=1, 2…N_1; j=1, 2…N_{i2}). That is to say, $\overline{n}_2$ secondary units were drawn from each selected primary unit, and $\overline{n}_3$ tertiary units were chosen from each drawn secondary unit of each selected primary unit. Every selected tertiary unit was investigated with the Simmons model for dichotomous sensitive questions.

In this study, we combined the Simmons Model for dichotomous sensitive questions with a 3-stage sampling survey method to estimate sample size in each stage. Our previous study (13) had derived the formulas for the optimal sample size of 3-stage sampling; on this basis, we used RRT to minimize sampling error under the condition that the cost of the investigation was limited, and to minimize the cost under the condition that the sampling error was limited. After that, we used SAS® (Statistical Analysis System, SAS Institute Inc., Cary, NC, USA) software to evaluate the survey methods and formulas by Monte Carlo simulation, and found that they had both high reliability and validity.

Formulae applied in the survey

The formulas (13) (See Appendix 1) applied in the survey were listed as follows:

$$p = \frac{N_1}{Nn_1}\sum_{i=1}^{n_1}\frac{N_{i2}}{n_{i2}}\sum_{j=1}^{n_{i2}}N_{ij3}p_{ij} \quad [1]$$

$$V(p) = \frac{\sigma_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right)+\frac{\sigma_2^2}{n_1\overline{n}_2}\left(1-\frac{\overline{n}_2}{\overline{N}_2}\right)+\frac{\sigma_3^2}{n_1\overline{n}_2\overline{n}_3}\left(1-\frac{\overline{n}_3}{\overline{N}_3}\right) \quad [2]$$

$$\sigma_1^2 = \frac{1}{N_1-1}\sum_{i=1}^{N_1}(\pi_i-\pi)^2 \quad [3]$$

$$\sigma_2^2 = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_{i2}-1}\sum_{j=1}^{N_{i2}}(\pi_{ij}-\pi_i)^2 \quad [4]$$

$$\sigma_3^2 = \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_{i2}}\sum_{j=1}^{N_{i2}}\pi_{ij}(1-\pi_{ij}) \quad [5]$$

$$s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(p_i-p)^2 \quad [6]$$

$$s_2^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{n_{i2}-1}\sum_{j=1}^{n_{i2}}(p_{ij}-p_i)^2 \quad [7]$$

$$s_3^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{n_{i2}}\sum_{j=1}^{n_{i2}}\left\{\frac{p_{ij}(1-\pi_{ij})}{n_{ij3}}+\frac{p_{ij}\cdot g(Y,A)+(1-\pi_{ij})g(N,\overline{A})}{n_{ij3}\left[g(Y,A)-1\right]\left[g(N,\overline{A})-1\right]}\right\}$$

$$g(Y,A) = \frac{P+(1-P)R_{ij}}{(1-P)R_{ij}} \quad [8]$$

$$g(N,\overline{A}) = \frac{1-(1-P)R_{ij}}{(1-P)(1-R_{ij})}$$

$$p_{ij} = \frac{m_{ij}/n_{ij3}-(1-P)R_{ij}}{P} \quad [9]$$

$$C = C_0 + C_1n_1 + C_2n_1\overline{n}_2 + C_3n_1\overline{n}_2\overline{n}_3 \quad [10]$$

$$V(p) = -\frac{\sigma_1^2}{N_1}+\frac{1}{n_1}\left(\sigma_1^2-\sigma_2^2/\overline{N}_2\right)+\frac{1}{n_1\overline{n}_2}\left(\sigma_2^2-\sigma_3^2/\overline{N}_3\right)+\frac{\sigma_3^2}{n_1\overline{n}_2\overline{n}_3} \quad [11]$$

$$\overline{n}_2 = \sqrt{\frac{\sigma_2^2-\sigma_3^2/\overline{N}_3}{\sigma_1^2-\sigma_2^2/\overline{N}_2}\bullet\frac{C_1}{C_2}} \quad [12]$$

$$\overline{n}_3 = \sqrt{\frac{\sigma_3^2}{\sigma_2^2-\sigma_3^2/\overline{N}_3}\bullet\frac{C_2}{C_3}} \quad [13]$$

$$n_1 = \frac{\sigma_1^2-\sigma_2^2/\overline{N}_2+\left(\sigma_2^2-\sigma_3^2/\overline{N}_3\right)/\overline{n}_2+\sigma_3^2/\overline{n}_2\overline{n}_3}{V+\sigma_1^2/N_1} \quad [14]$$

$$n_1 = \frac{C-C_0}{C_1+C_2\overline{n}_2+C_3\overline{n}_2\overline{n}_3} \quad [15]$$

$$n_{i2} = N_{i2}\bullet\frac{\overline{n}_2}{\overline{N}_2} \quad [16]$$

$$n_{ij3} = N_{ij3}\bullet\frac{\overline{n}_3}{\overline{N}_3} \quad [17]$$

Investigation target

In our survey, we asked the FSWs to answer some sensitive questions with 2 attribute characteristics, and the answers had to be mutually independent and exclusive. In order to shorten the experiments and improve the enthusiasm of the respondents simultaneously, we found the following 3 sensitive issues to be relatively useful for our goals: whether you possess a regular sex partner (RSP) or not? (Issue C01), whether you stop sexual services after suffering from venereal diseases or not? (Issue C02), whether you support the legalization of prostitution or not? (Issue C03) (See Appendix 2).

RRT model

The Simmons model for dichotomous sensitive questions was applied in the survey. The randomized device was a semi-closed bag, including 6 red balls (P=0.6) and 4 white balls (P=0.4). The balls were identical in size, weight, and texture. The respondents who retrieved the red ball answered the sensitive question. Otherwise, they answered the non-sensitive question. In the survey about the proportion of FSWs possessing an RSP, the sensitive question was: "Apart from your chargeable clients, do you have a spouse or other regular sexual partner?" The non-sensitive question was: "Is your date of birth an odd number?" Depending on the ball they picked out, the participants should answer "Yes" or "No" to the corresponding question in accordance with their actual situation. There were 3 independent sensitive issues in our survey, and we designed 3 sets of randomized devices accordingly, which were identical except that the sensitive issues were different.

We took some measures in order to better play the role of RRT. Our investigators had a deep understanding of the principles and practices of RRT. They explained RRT to each respondent, ensuring everyone realized that RRT was a scientific survey method and could cleverly protect their privacy. If necessary, we could allow the respondents to check the randomized device to dispel their misgivings completely. In our survey, we were cautious to choose the non-sensitive issues since they may involve general characteristics of individual. Our investigators were strangers to the FSWs, otherwise the unrelated question would not be able to achieve its proper protection. Moreover, the unrelated question we chose was very easy to answer for the respondents, and it did not require contemplation. Our survey was of an adequate sample size to ensure the authenticity of findings.

Three-stage sampling

Considering that the population of each administrative region in Xichang may not be the same, we divided Xichang into 10 districts of similar population, here referred to as first-stage units (N_1=10). Places known to provide paid sexual services, such as nightclubs, saunas, karaoke bars, and bars, were taken as survey spots, here referred to as second-stage units ($\overline{N}_2$=12). A total of 4876 FSWs were taken as third-stage units ($\overline{N}_3$=41). According to the formulas for optimum sample size of 3-stage sampling derived by Jianfeng Wang, we could minimize the sampling error for a specified cost. In our 3-stage sampling, we randomly chose 5 districts from those 10 first stage districts (n_1=5), and chose 6 second stage survey spots in each district randomly and equally ($\overline{n}_2$=6). At each survey spot, we randomly selected an average of 27 people in the third stage unit ($\overline{n}_3$=27). Finally, a total of 800 FSWs were recruited. So, relative to the total number, the sampling fractions for the first, second, and third stages were 50%, 25%, 16.4%, respectively.

We recruited the FSWs at sites of activity. The staff of the State Key Laboratory of Infectious Disease Prevention and Control in China also used this approach to recruit FSWs and conducted 3 cross-sectional surveys of HIV infection and behavior from 2010 to 2015 (14). The places activity for providing sexual services could be divided into high, middle, and low grade, and the places we selected included all the grade divisions. Therefore, the population we selected was a very good representative sample; however, we cannot rule out that very few FSWs provide commercial services in the name of an individual.

Monte Carlo simulation

In line with the answers to the dichotomous sensitive issues among FSWs in the actual survey in Xichang, the sample proportions and their standard deviations were taken as simulative summary measures, and we then set up an analog population through SAS® (Statistical Analysis System, SAS Institute Inc., Cary, NC, USA) (See Appendix 3).

In this work, for the investigation method on dichotomous sensitive questions in 3-stage sampling under Simmons models, we used SAS® to simulate 100 pre-survey samples, each of which equally contained 2,000 simulated respondents. According to the calculation formulas of optimal sample size deduced in this paper, we could calculate the sample sizes at each stage required for the 100 sampling surveys respectively (See https://cdn.amegroups.cn/static/application/239522e014618b88688e70af39d65930/apm-20-2262-1.pdf). After that, we conducted 100 formal sampling surveys on the simulated population, and obtained 100-point estimates and their 95% CIs of the population proportion (See https://cdn.amegroups.cn/static/application/db099841f298c0defc3b617c569c01fd/apm-20-2262-2.pdf). If almost all of the 100 95% CIs contained the overall proportion of the simulated population and almost all of the 100 sample proportions were close to the same value, the survey method would be considered to have good reliability and validity.

Statistical analysis

All participants provided written informed consent. The questionnaires were completed by the respondents independently, and all questionnaires were checked carefully. The response rate was 100%, and all returned questionnaires qualified for inclusion. The collected data were checked twice and entered into a database with EpiData (version 3.1, EpiData Association, Odense, Denmark). The software SAS® (version 9.3, SAS Institute Inc., Cary, NC, USA) was used to analyze the data.

Results

Survey and quantitative results

Taking Issue C01 as an example, in investigation site 1 in district A, 23 FSWs were randomly selected (n_{ij3}=n_{113}=23), of which 12 answered "yes" (m_{ij}=m_{11}=12) (See https://cdn.amegroups.cn/static/application/3144ad73360446f10a7ad469ad08e447/apm-20-2262-3.pdf). Based on $\hat{\lambda}_{ij} = m_{ij}/n_{ij3}$, the ratio of answering "yes" was $\hat{\lambda}_{11} = m_{11}/n_{113} = 12/23$. The true proportion of respondents having a date of birth being an odd number was 0.5 (R_{ij}=R_{11}=0.5), and the proportion of respondents being asked to answer the sensitive question was 0.6 (P=0.6).

According to Eq. [9], p_{11} (i=1; j=1), the sample proportion of FSWs possessing an RSP in site 1 in district A was:

$$p_{11} = \frac{\hat{\lambda}_{11} - (1-P)R_{11}}{P} = \frac{m_{11}/n_{113} - (1-P)R_{11}}{P} = \frac{12/23 - (1-0.6)\times 0.5}{0.6} = 0.5362 \quad [18]$$

The proportion of FSWs possessing RSP in other investigation sites in district A or in other districts could be determined in the same way (*Table 1*).

Analytical results of sensitive issues

According to Eq. [1], the estimate of the overall rate of FSWs possessing an RSP in Xichang, here referred to as p, could be determined as follows:

$$p = \frac{10}{4876\times 5}\sum_{i=1}^{5}\frac{N_{i2}}{n_{i2}}\sum_{j=1}^{n_{i2}} N_{ij3}p_{ij} \doteq 0.5594 \quad [19]$$

Concurrently, based on Eqs. [6], [7], and [8], the estimate based on sample of σ_1^2, σ_2^2, and σ_3^2, here referred to as S_1^2, S_2^2, and S_3^2, respectively, could be determined:

$$s_1^2 = \frac{1}{5-1}\sum_{i=1}^{5}(p_i - p)^2 \doteq 0.0098 \quad [20]$$

$$s_2^2 = \frac{1}{5}\sum_{i=1}^{5}\frac{1}{n_{i2}-1}\sum_{j=1}^{n_{i2}}(p_{ij} - p_i)^2 \doteq 0.0251 \quad [21]$$

$$s_3^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{n_{i2}}\sum_{j=1}^{n_{i2}}\left\{\frac{p_{ij}(1-\pi_{ij})}{n_{ij3}} + \frac{p_{ij}\cdot g(Y,A) + (1-\pi_{ij})g(N,\bar{A})}{n_{ij3}[g(Y,A)-1][g(N,\bar{A})-1]}\right\} \doteq 0.0251 \quad [22]$$

Then, through Eq. [2], we could obtain the sample variance of p:

$$v(p) = \frac{s_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right) + \frac{s_2^2}{n_1\bar{n}_2}\left(1-\frac{\bar{n}_2}{\bar{N}_2}\right) + \frac{s_3^2}{n_1\bar{n}_2\bar{n}_3}\left(1-\frac{\bar{n}_3}{\bar{N}_3}\right) \doteq 0.0014 \quad [23]$$

From this, we could obtain the 95% CI of the proportion of FSWs possessing an RSP in Xichang:

$$p \pm 1.96\times\sqrt{v(p)} \doteq 0.4858 \sim 0.6330 \quad [24]$$

By the same token (See https://cdn.amegroups.cn/static/application/341debe83b0f0575fc0b83c88a70a335/apm-20-2262-4.pdf, https://cdn.amegroups.cn/static/application/cc61a978cda5cefbaf21f1b175ebcd7c/apm-20-2262-5.pdf), the analytical results and their 95% CIs of 3 sensitive issues, could be determined as shown in *Table 2*.

The estimation of sample size

As mentioned above, we calculated the variances of the sample proportion of FSWs possessing an RSP in Xichang: S_1^2=0.0098, S_2^2=0.0251, S_3^2=0.0251. According to Eq. [12], we could then calculate the quantity of survey spots which should be selected randomly from each chosen district ($\sigma_i^2 \doteq s_i^2$):

Table 1 The proportion of FSWs possessing an RSP in each site of each district in Xichang using Simmons model for qualitative characteristic questions under three-stage sampling

i	j	m_{ij}	p_{ij}
A	1	12	0.5362
A	2	18	0.6667
A	3	15	0.5595
A	4	17	0.7564
A	5	20	0.7778
A	6	13	0.6515
B	1	21	0.8333
B	2	12	0.5362
B	3	14	0.5000
B	4	11	0.4306
B	5	12	0.5758
B	6	13	0.5000
C	1	13	0.3889
C	2	17	0.7160
C	3	10	0.3333
C	4	11	0.3214
C	5	18	0.7011
C	6	12	0.3810
D	1	10	0.3077
D	2	16	0.8788
D	3	14	0.5641
D	4	16	0.5862
D	5	19	0.7976
D	6	18	0.7778
E	1	9	0.2222
E	2	11	0.4000
E	3	16	0.5862
E	4	12	0.5362
E	5	13	0.3889
E	6	15	0.5287

i, the number of district; j, the number of site; m_{ij}, the number of FSWs who answered "Yes" in the corresponding site; p_{ij}, the sample proportion of FSWs possessing an RSP in the corresponding site. FSW, female sex worker; RSP, regular sexual partner.

$$\bar{n}_2 = \sqrt{\frac{s_2^2 - s_3^2/\bar{N}_3}{s_1^2 - s_2^2/\bar{N}_2} \bullet \frac{C_1}{C_2}} \doteq 6 \qquad [25]$$

According to Eq. [13], we could compute the number of FSWs to be recruited equally and randomly from every selected spot ($\sigma_i^2 \doteq s_i^2$):

$$\bar{n}_3 = \sqrt{\frac{s_3^2}{s_2^2 - s_3^2/\bar{N}_3} \bullet \frac{C_2}{C_3}} \doteq 21 \qquad [26]$$

In our survey, under the circumstance of minimizing the cost for the fixed sampling error of V=0.0015, with the help of Eq. [14], the required minimum number of districts in Xichang to be investigated could be determined ($\sigma_i^2 \doteq s_i^2$):

$$n_1 = \frac{s_1^2 - s_2^2/\bar{N}_2 + \left(s_2^2 - s_3^2/\bar{N}_3\right)/\bar{n}_2 + s_3^2/\bar{n}_2\bar{n}_3}{V + s_1^2/N_1} \doteq 6 \qquad [27]$$

According to the budget, the basic outlay of our entire survey was C_0=100,000 yuan, including cooperation fees for institutions of the project, travel expenses, training expenses of staff, and so on. The average cost of investigating each district was C_1=100,000 yuan, including cooperation costs for agencies of the project, car fares, communication expenses, and so on. The average cost of looking into each place of activity was C_2=10,000 yuan, including cooperation costs for places providing paid sexual services, rewards to female volunteers for intervention in HIV, and so on. The direct cost of interviewing each FSW on average was C_3=25 yuan, which was paid as a reward to the FSW for undergoing interview.

In our survey, under the circumstance of minimizing sampling error for the fixed cost of C=100000 yuan, with the help of Eq. [15], the required minimum number of districts in Xichang to be investigated could be determined ($\sigma_i^2 \doteq s_i^2$):

$$n_1 = \frac{C - C_0}{C_1 + C_2\bar{n}_2 + C_3\bar{n}_2\bar{n}_3} \doteq 6 \qquad [28]$$

Then, we could estimate the required sample size for each dichotomous sensitive question as listed below (*Table 3*).

After deducing $\bar{n}_2$, according to Eq. [16], the number of survey spots randomly chosen from the i^{th} drawn district could be determined. For example, if there were a total of 12 survey spots in a certain district, the quantity of spots randomly selected from this district should have been:

Table 2 Analytical results of dichotomous sensitive questions

Dichotomous sensitive questions	Proportion	95% CI
Proportion who possesses an RSP	55.94%	(48.58–63.30%)
Ratio who stops sexual services after suffering from venereal diseases	75.85%	(70.07–81.63%)
Percentage who supports legalization of prostitution	56.77%	(48.82–64.72%)

CI, confidence interval; RSP, regular sexual partner.

Table 3 Estimation of required sample size for each dichotomous sensitive question

Dichotomous sensitive questions	$\overline{n}_2$	$\overline{n}_3$	n_1
Proportion who possesses an RSP	6	21	6
Ratio who stops sexual services after suffering from venereal diseases	7	24	6
Percentage who supports legalization of prostitution	6	19	6

$\overline{n}_2$, the quantity of spots selected from each chosen district randomly; $\overline{n}_3$, the number of FSWs drawn from each selected venue randomly; n_1, the required sample size to maintain a fixed cost for a district. RSP, regular sexual partner; FSW, female sex worker.

$$n_{i2} = N_{i2} \bullet \frac{\overline{n}_2}{\overline{N}_2} \doteq 6 \quad [29]$$

After working out $\overline{n}_3$, according to Eq. [17], the number of FSWs randomly selected from j^{th} drawn site of the i^{th} chosen district could be calculated. For instance, if there were 36 FSWs in a certain site of a particular district, the number of FSWs randomly selected from there should have been:

$$n_{ij3} = N_{ij3} \bullet \frac{\overline{n}_3}{\overline{N}_3} \doteq 19 \quad [30]$$

Determination of the final sample size

When minimizing the sampling error for a fixed survey cost, we needed to randomly select 24 FSWs from each of the 7 sites in each of the 6 districts in each stage.

Evaluation of reliability and validity

Regarding the dichotomous sensitive issue C01, "Whether there is a spouse or any regular sexual partner other than those paying for sexual services?", the overall proportion calculated from the simulated population was 0.5594. The Simmons model was used to carry out analogue investigations on the dichotomous sensitive issue for all the respondents of 100 simulated samples respectively, and the results are shown in *Table 4*.

According to *Table 4*, we can see 96 (except the sample no's 30, 37, 55, 76) of the 100 95% CIs contain the overall proportion of the simulated population, that is 0.5594. Meanwhile, 96 of the 100 sample proportions are close to 0.5594. These 2 results indicate that the survey method and the calculation formulas both had good reliability and good validity.

Discussion

Culturally sensitive issues appear to be particularly prominent in the field of public health, and have a large negative impact on economic development and social stability. It is difficult for investigators to draw a sample directly due to the wide variance and enormous numbers of the sample, and the dispersed distribution of the hidden population. Multi-stage sampling is a scientific and wise choice to achieve the goals of a survey and improve sampling efficiency and representativeness. Multi-stage sampling refers to a sampling process that is carried out in stages, and the sampling method used in each stage is usually different. It has many characteristics of organizing the participants easily and sampling flexibility. By contrast, proportional sampling does not take the sample variability into account, and only uses a uniform proportion for sampling. It is absolutely simple and convenient, but could lead to a low representative level of the sample. The extraction of the basic unit to be investigated is not a step in place, furthermore, sampling frames for the second or lower

Table 4 The result of computer simulation regarding the issue C01

Sample number	Sample proportion P	Sample size	95% LCL	95% UCL
1	0.5255	2,261	0.4911	0.5599
2	0.5664	2,127	0.5324	0.6005
3	0.5410	2,254	0.5048	0.5773
4	0.5860	2,268	0.5469	0.6251
5	0.5597	1,994	0.5223	0.5971
6	0.5789	2,168	0.5454	0.6124
7	0.5275	1,781	0.4915	0.5634
8	0.5881	2,198	0.5538	0.6224
9	0.5630	2,493	0.5279	0.5981
10	0.5616	2,952	0.5312	0.5921
11	0.5442	2,127	0.5060	0.5823
12	0.5806	1,620	0.5339	0.6273
13	0.5327	2,094	0.5036	0.5618
14	0.5416	2,659	0.5087	0.5745
15	0.5661	1,749	0.5267	0.6055
16	0.5594	2,040	0.5232	0.5956
17	0.5830	2,134	0.5485	0.6175
18	0.5307	1,968	0.4894	0.5719
19	0.5730	2,166	0.5343	0.6117
20	0.5453	2,123	0.5112	0.5794
21	0.5878	2,017	0.5579	0.6177
22	0.5734	1,726	0.5346	0.6121
23	0.5462	2,422	0.5182	0.5743
24	0.5359	2,395	0.5036	0.5681
25	0.5502	1,948	0.5022	0.5982
26	0.5237	1,549	0.4855	0.5618
27	0.5569	2,600	0.5229	0.5908
28	0.5491	1,786	0.4999	0.5984
29	0.5562	2,115	0.5158	0.5965
30	0.5907	2,368	0.5608	0.6207
31	0.5422	2,712	0.5098	0.5745
32	0.5757	2,537	0.5435	0.6078
33	0.5617	2,531	0.5358	0.5875
34	0.6017	1,343	0.5572	0.6462
35	0.5395	1,900	0.5007	0.5784
36	0.5571	2,398	0.5236	0.5905
37	0.5156	1,853	0.4727	0.5585
38	0.5823	1,519	0.5389	0.6256
39	0.5467	2,477	0.5153	0.5780
40	0.5328	1,772	0.4813	0.5843
41	0.5771	2,219	0.5428	0.6114
42	0.5547	1,870	0.5166	0.5927
43	0.5600	2,282	0.5222	0.5978
44	0.5711	1,688	0.5206	0.6215
45	0.5300	2,185	0.4866	0.5733
46	0.5658	2,183	0.5187	0.6129
47	0.5723	2,075	0.5312	0.6135
48	0.5751	2,117	0.5486	0.6016
49	0.5538	2,054	0.5093	0.5983
50	0.5408	1,684	0.4949	0.5867
51	0.5654	2,347	0.5388	0.5919
52	0.5253	2,866	0.4884	0.5622
53	0.5283	1,833	0.4805	0.5762
54	0.5757	1,936	0.5213	0.6301
55	0.5099	1,394	0.4693	0.5504
56	0.5567	2,474	0.5261	0.5874
57	0.5391	1,536	0.4888	0.5894
58	0.5506	2,484	0.5187	0.5825
59	0.5693	1,643	0.5320	0.6066
60	0.5385	1,775	0.4995	0.5776
61	0.5915	1,510	0.5459	0.6371
62	0.5530	1,981	0.5168	0.5892
63	0.5395	2,024	0.4856	0.5933
64	0.5807	1,969	0.5386	0.6228
65	0.5703	2,442	0.5233	0.6173
66	0.5648	2,195	0.5281	0.6015
67	0.5858	1,591	0.5322	0.6394
68	0.5520	1,514	0.5034	0.6006

Table 4 (continued)

Table 4 (*continued*)

Sample number	Sample proportion P	Sample size	95% LCL	95% UCL
69	0.5410	1,634	0.5092	0.5728
70	0.5811	2,490	0.5506	0.6115
71	0.5369	1,747	0.4937	0.5800
72	0.5484	2,184	0.5157	0.5810
73	0.5593	2,293	0.5333	0.5853
74	0.5695	1,842	0.5375	0.6016
75	0.5539	1,947	0.5121	0.5957
76	0.5032	2,108	0.4626	0.5439
77	0.5851	1,569	0.5317	0.6386
78	0.5626	1,977	0.5176	0.6075
79	0.5474	1,995	0.5078	0.5870
80	0.5545	2,053	0.5160	0.5929
81	0.5677	2,061	0.5243	0.6111
82	0.5654	1,831	0.5194	0.6114
83	0.5399	2,296	0.5029	0.5769
84	0.5544	2,428	0.5077	0.6011
85	0.5394	1,879	0.4988	0.5801
86	0.5589	2,713	0.5182	0.5997
87	0.5799	1,872	0.5479	0.6119
88	0.5340	1,764	0.5023	0.5658
89	0.5651	2,006	0.5186	0.6117
90	0.5759	2,319	0.5462	0.6056
91	0.5351	2,285	0.5002	0.5700
92	0.5919	2,503	0.5492	0.6346
93	0.5534	1,880	0.5169	0.5899
94	0.5522	2,481	0.5209	0.5835
95	0.5579	2,397	0.5240	0.5918
96	0.5543	2,192	0.5123	0.5963
97	0.5825	2,275	0.5411	0.6239
98	0.5423	1,433	0.4915	0.5931
99	0.5433	1,956	0.5054	0.5812
100	0.5632	1,985	0.5133	0.6130

LCL, lower confidence limit; UCL, upper confidence limit.

order unit at the time of sampling are not required.

Over the past 40 years, RRT has been widely applied to different domains, such as tax evasion, drug abuse, academic deception and so on. Bias caused by social expectations could be reduced by RRT. Furthermore, when compared with traditional investigation methods, such as self-administered questionnaires (SAQ), face-to-face interviews, telephone surveys, and computer-assisted interviews, RRT has significant advantages in terms of accuracy, reliability, and parameter estimation (15). The existence of social acceptability bias can't be ruled out when using RRT, and we cannot guarantee that all FSWs were inclined to admit behaviors that may have a negative impact on them. A slight increase in variance may be due to RRT; however, if other investigation methods were used, this would have resulted in an increase in the errors that could not be estimated and a decrease in the credibility of the results. Therefore, with the aims of obtaining relatively reliable data and better understanding behavioral characteristics, we used RRT to desensitize FSWs to the sensitive questions.

The proportion asked the sensitive question or the level of confidentiality could affect the efficiency of the Simmons model. In this paper, we have already mentioned the proportion of red balls (P=0.6) in the randomized devices, which has been proven to maximize the efficiency of the Simmons model (16). We cannot provide the accurate proportion of those who were asked the sensitive questions during the practical investigation, on account of the investigators having been absent at that time, but the formulas we had derived took into account the variance produced by using RRT, which made the result more reliable.

Previous studies on sampling techniques of culturally sensitive issues have been confined to improving the model in simple random sampling. This application is only useful in special populations or at small scales, although studies on sensitive issues have been conducted extensively in China and abroad. Few studies have been conducted regarding RRT under complex sampling methods, such as multi-stage sampling and stratified multi-stage sampling (10). In addition, the estimation of sample size under complex sampling is rarely considered. Therefore, it is important to estimate the sample size for RRT model under multi-stage sampling and to evaluate its reliability and validity.

In this paper, we presented an investigative method for culturally sensitive issues using the Simmons model under 3-stage sampling. And have provided formulae concerning the optimal sample size in 2 conditions of satisfying minimum sampling error and lowest investigation costs. The application in this study was the pre-survey of the Chinese National Natural Science Fund project conducted by our team in 2019–2020. Based on the formulas for sample size derived from this study, we can scientifically estimate the sample size required for each sampling stage. The prevalence of STDs among FSWs in Xichang was obtained preliminarily, which provided evidence for the study of behavioral characteristics of groups at high risk of STDs in Xichang. Effective methods and scientific statistical formulas are vital for successful investigation of sensitive issues, but sample size must be estimated before a sampling survey begins.

Reliability and validity are 2 important indicators for evaluating a method measurement. A high-quality measuring method should have both high reliability and validity (17). Few evaluations concerning the reliability and validity of investigation methods for culturally sensitive issues have been conducted. In the last decade, our team has carried out much research on the methods for investigating various sensitive issues, which has combined various random response models with different sampling methods, such as cluster sampling, stratified cluster sampling, stratified random sampling, 2-stage sampling, stratified 2-stage sampling, and stratified 3-stage sampling methods (18-23). Formulas for the overall proportion, the population mean, and its variance were deduced. The performances of the sampling methods and formulas in the survey of premarital sex and cheating on exams at Soochow University, as well as sensitive issues among men who have sex with men (MSM), were also provided. The reliability of the survey methods and formulas for sensitive questions was found to be high.

This survey showed that 55.94% of the FSWs have a regular sexual partner apart from clients, and 75.85% stop providing sexual services after contracting venereal diseases. Their working situations easily lead to their infection with STDs. We also feel that the local FSWs lacked traditional workplace skills and relied on the provision of sexual services as their primary source of income. The local government needs to commit to assisting the re-employment of the FSWs, including the establishment and improvement of the training of laid-off female workers and attracting investment to increase local employment opportunities. Of the FSWs interviewed, 56.77% support the legalization of prostitution. Legalization might protect the rights and interests of FSWs and reduce crime, which has already become the case in some developed countries. In China, it is difficult to implement legalization of prostitution, but the local relevant departments can strengthen visibility to enhance the awareness and self-protection capabilities of FSWs.

Conclusions

In this paper, we introduced a new sampling method to draw a random sample that is representative for such a hidden population as FSWs who are illegal and stigmatized in China. We derived the formulae to estimate the optimal sample size in 2 conditions of satisfying minimum sampling error and lowest investigation costs. Moreover, the sample sizes required for each stage in the survey were ascertained, which provided critical data for the technique of sampling. We applied the method to the pre-survey of FSWs in Xichang and obtained important data concerning characteristics of sexual behaviour, which could be helpful in informing policy makers in developing prevention strategies and guiding further research. In this study, a large number of Monte Carlo simulation sampling tests were used to validate reliability and validity, which were found to be high. This new sampling method is liable to be used for sensitive question surveys of large populations.

Acknowledgments

We are most grateful to the female sex workers in Xichang who made it possible for us to perform this study. We extend our thanks to the Chinese Center for Disease Control and Prevention and Xichang Dermatology and Venereal Disease Prevention and Control Station for working with us. Finally, Kejin Chen wants to thank, in particular, the invaluable support received by Qihua Chen and Yang Cao over the years.

Funding: This work was supported by the National Natural Science Foundation of China (81273188), China Association for the Prevention and Control of STDs and AIDS· AbbVie Fund for Maternal and Child Care and Prevention of Mother-to-Child Transmission (Project No.PMTCT201904), Changzhou Health Young Cultivation Project (CZQM2020101).

Footnote

Reporting Checklist: The authors have completed the

SURGE reporting checklist. Available at http://dx.doi.org/10.21037/apm-20-2262

Data Sharing Statement: Available at http://dx.doi.org/10.21037/apm-20-2262

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/apm-20-2262). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All participants provided their written informed consent to participate in this study anonymously. The data were also collected and analyzed anonymously. This research, including the consent procedure, was approved by the Ethics Committee of Soochow University (No. 2017084). All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Yang S, Wu J, Ding C, et al. Epidemiological features of and changes in incidence of infectious diseases in China in the first decade after the SARS outbreak: an observational trend study. Lancet Infect Dis 2017;17:716-25.
2. Lin CC, Gao X, Chen XS, et al. China's syphilis epidemic: a systematic review of seroprevalence studies. Sex Transm Dis 2006;33:726-36.
3. Xu JJ, Yu YQ, Hu QH, et al. Treatment-seeking behaviour and barriers to service access for sexually transmitted diseases among men who have sex with men in China: a multicentre cross-sectional survey. Infect Dis Poverty 2017;6:15.
4. Zou X, Chow EP, Zhao P, et al. Rural-to-urban migrants are at high risk of sexually transmitted and viral hepatitis infections in China: a systematic review and meta-analysis. BMC Infect Dis 2014;14:490.
5. Shah NS, Shiraishi RW, Subhachaturas W, et al. Bridging populations-sexual risk behaviors and HIV prevalence in clients and partners of female sex workers, Bangkok, Thailand 2007. J Urban Health 2011;88:533-44.
6. Jeal N, Patel R, Redmond NM, et al. Drug use in street sex workers (DUSSK) study protocol: a feasibility and acceptability study of a complex intervention to reduce illicit drug use in drug-dependent female street sex workers. BMJ Open 2018;8:e022728.
7. Prangnell A, Shannon K, Nosova E, et al. Workplace violence among female sex workers who use drugs in Vancouver, Canada: does client-targeted policing increase safety? J Public Health Policy 2018;39:86-99.
8. Warner HR, Cox A. A mathematical model of heart rate control by sympathetic and vagus efferent information. J Appl Physiol 1962;17:349-55.
9. Arnab R, Singh S. Randomized response techniques: An application to the Botswana AIDS impact survey. Journal of Statistical Planning and Inference 2010;140:941-53.
10. Pu X, Gao G, Fan Y, et al. Parameter Estimation in Stratified Cluster Sampling under Randomized Response Models for Sensitive Question Survey. PLoS One 2016;11:e0148267.
11. Moshagen M, Musch J, Ostapczuk M, et al. Reducing socially desirable responses in epidemiologic surveys: an extension of the randomized-response technique. Epidemiology 2010;21:379-82.
12. Chaudhuri A, Dihidar K. Estimating Means of Stigmatizing Qualitative and Quantitative Variables from Discretionary Responses Randomized or Direct. The Indian Journal of Statistics 2009;71:123-36.
13. Wang J, Ge G, Fan Y, et al. The estimation of sample size in multi-stage sampling and its application in medical survey. Applied Mathematics and Computation 2006;178:239-49.
14. Xiang Z, Xiaoyi Y, Wanli Z. Investigation of HIV infection and related behaviors among female sex workers in Xichang, 2010-2015. China Tropical Medicine 2017;17:569-72.
15. Lensvelt-Mulders GJLM, Hox JJ, Heijden PGMVD, et al. Meta-analysis of randomized response research thirty-five years of validation. Sociological Methods & Research 2004;33:319-48.
16. Giordano S, Perri PF. Efficiency comparison of unrelated

question models based on same privacy protection degree. Statistical Papers 2011;53:987-99.
17. Dare OO, Cleland JG. Reliability and validity of survey data on sexual behaviour. Health Transit Rev 1994;4 Suppl:93-110.
18. Bo Y, Ge G, Zhilong H. The Survey among MSM Population in Beijing Using Two-Stage Sampling Method on RRT Model. Chinese Journal of Health Statistics 2012;29:351-3.
19. Chen X, Du Q, Jin Z, et al. The Randomized Response Technique Application in the Survey of Homosexual Commercial Sex among Men in Beijing. Iran J Public Health 2014;43:416-22.
20. Lei W, Ge G, Mingrun Y. Stratified Two-stage Cluster Sampling and Its Application on the Greenberg Model for Sensitive Question Survey. Chinese Journal of Health Statistics 2011;28:37-8, 46.
21. Mian W, Ge G. Cluster Sampling and its Application on Quantitative Sensitive Questions. Chinese Journal of Health Statistics 2008;25:586-9.
22. Yunhua Z, Ge G, Xiangke P. The Method to Estimate the Sample Size for Quantitative Sensitive Questions in Stratified Two-Stage Sampling and its Application. Chinese Journal of Health Statistics 2014;31:45-8.
23. Zhilong H, Ge G, Mian W. Statistical Methods of Two-stage Sampling on the Randomized Response Technique for Sensitive Question Survey with Multiple Choices and its Application. Chinese Journal of Health Statistics 2009;26:580-2.

(English Language Editor: J. Jones)

Cite this article as: Chen K, Mao Z, Jin Z, Gao G, Bian C. Application of a randomized response technique for the sensitive question survey in female sex workers in Xichang, China. Ann Palliat Med 2021;10(1):392-403. doi: 10.21037/apm-20-2262

Appendix 1

Formulas Derivation for Simmons Model under Three-stage Sampling

Let the overall survey cost C be

$$C = C_0 + C_1 n_1 + C_2 n_1 \bar{n}_2 + C_3 n_1 \bar{n}_2 \bar{n}_3. (10)$$

The variance of p can also be written in the following alternative form:

$$V(p) = -\frac{\sigma_1^2}{N_1} + \frac{1}{n_1}\left(\sigma_1^2 - \sigma_2^2 / \bar{N}_2\right) + \frac{1}{n_1 \bar{n}_2}\left(\sigma_2^2 - \sigma_3^2 / \bar{N}_3\right) + \frac{\sigma_3^2}{n_1 \bar{n}_2 \bar{n}_3}. (11)$$

To minimize the sampling cost C under a given variance ($V(p) = V$), the optimum sampling size can be considered as the minimal values of function (10) subject to the constraint (11). The Lagrange function L is defined as

$$L(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda) = C + \lambda\left[V - V(P)\right],$$

that is,

$$L(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda) = C_0 + C_1 n_1 + C_2 n_1 \bar{n}_2 + C_3 n_1 \bar{n}_2 \bar{n}_3 + \lambda\left[V - V(P)\right],$$

where λ is a Lagrange multiplier.

The necessary conditions for the solution of the problem are

$$\begin{cases} \dfrac{\partial L}{\partial n_1} = C_1 + \dfrac{1}{n_1^2}\left(\sigma_1^2 - \sigma_2^2 / \bar{N}_2\right) \times \lambda = 0 \\ \dfrac{\partial L}{\partial (n_1 \bar{n}_2)} = C_2 + \dfrac{1}{(n_1 \bar{n}_2)^2}\left(\sigma_2^2 - \sigma_3^2 / \bar{N}_3\right) \times \lambda = 0 \\ \dfrac{\partial L}{\partial (n_1 \bar{n}_2 \bar{n}_3)} = C_3 + \dfrac{1}{(n_1 \bar{n}_2 \bar{n}_3)^2}\sigma_3^2 \times \lambda = 0 \end{cases}.$$

Equations above-mentioned gives

$$\bar{n}_2 = \sqrt{\frac{\sigma_2^2 - \sigma_3^2 / \bar{N}_3}{\sigma_1^2 - \sigma_2^2 / \bar{N}_2} \bullet \frac{C_1}{C_2}}, (12)$$

$$\bar{n}_3 = \sqrt{\frac{\sigma_3^2}{\sigma_2^2 - \sigma_3^2 / \bar{N}_3} \bullet \frac{C_2}{C_3}}, (13)$$

$$n_1 = \frac{\sigma_1^2 - \sigma_2^2 / \bar{N}_2 + \left(\sigma_2^2 - \sigma_3^2 / \bar{N}_3\right) / \bar{n}_2 + \sigma_3^2 / \bar{n}_2 \bar{n}_3}{V + \sigma_1^2 / N_1}. (14)$$

 http://dx.doi.org/10.21037/apm-20-2262

The minimum value of $V(p)$ under a cost function (fixed survey cost C), the optimum sampling size is obtained as the minimum values of function (11) subject to the constraint (10). Consider the following Lagrange function :

$$L\left(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda\right) = V\left(P\right) + \lambda\left[C - C_0 - C_1 n_1 - C_2 n_1\bar{n}_2 - C_3 n_1\bar{n}_2\bar{n}_3\right],$$

and the detailed form is as follows,

$$\begin{aligned} & L\left(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda\right) = \\ & -\frac{\sigma_1^2}{N_1} + \frac{1}{n_1}\left(\sigma_1^2 - \sigma_2^2/\bar{N}_2\right) + \frac{1}{n_1\bar{n}_2}\left(\sigma_2^2 - \sigma_3^2/\bar{N}_3\right) + \frac{\sigma_3^2}{n_1\bar{n}_2\bar{n}_3} \\ & +\lambda\left[C - C_0 - C_1 n_1 - C_2 n_1\bar{n}_2 - C_3 n_1\bar{n}_2\bar{n}_3\right], \end{aligned}$$

where λ is a Lagrange multiplier.

The optimum n_1, $\bar{n}_2$ and $\bar{n}_3$ are the solutions of the following numerical problem:

$$\begin{cases} \dfrac{\partial L}{\partial n_1} = -\dfrac{1}{n_1^2}\left(\sigma_1^2 - \sigma_2^2 / \bar{N}_2\right) - \lambda C_1 = 0 \\ \dfrac{\partial L}{\partial\left(n_1\bar{n}_2\right)} = -\dfrac{1}{\left(n_1\bar{n}_2\right)^2}\left(\sigma_2^2 - \sigma_3^2 / \bar{N}_3\right) - \lambda C_2 = 0. \\ \dfrac{\partial L}{\partial\left(n_1\bar{n}_2\bar{n}_3\right)} = -\dfrac{1}{\left(n_1\bar{n}_2\bar{n}_3\right)^2}\sigma_3^2 - \lambda C_3 = 0 \end{cases}$$

Results are presented as below:

$$\bar{n}_2 = \sqrt{\frac{\sigma_2^2 - \sigma_3^2/\bar{N}_3}{\sigma_1^2 - \sigma_2^2/\bar{N}_2} \bullet \frac{C_1}{C_2}}, (12)$$

$$\bar{n}_3 = \sqrt{\frac{\sigma_3^2}{\sigma_2^2 - \sigma_3^2/\bar{N}_3} \bullet \frac{C_2}{C_3}}, (13)$$

$$n_1 = \frac{C - C_0}{C_1 + C_2\bar{n}_2 + C_3\bar{n}_2\bar{n}_3}. (15)$$

 http://dx.doi.org/10.21037/apm-20-2262

Appendix 2

Questionnaire of Random Response Technique Method

Hello! This survey was conducted on the sensitive issues concerning personal sexual behaviors, which was jointly carried out by Chinese Center for Disease Control and Prevention and Xichang Dermatology and Venereal Disease Prevention and Control Station.

In order to better protect your privacy, we have adopted random response technique (RRT) method for some more sensitive issues, that is, fill in the corresponding survey according to the figure on the ball.

Your participation is very important for us to carry out AIDS prevention and research work in the future. We hope to get your understanding and cooperation. We will keep the results strictly confidential. The survey will take about 10 minutes. Thank you! Please write down the corresponding characters or values on________ and the relevant values or order numbers of answers in □.

I agree to participate and truly answer（fill in □ with √）:
1、□ Yes　　2、□ No （The investigation is over.）

B、General conditions

B01 Your age: ________ one full year of life

B02 Your marital status: □

1、Married cohabitation　2、Married separation　3、Unmarried cohabitation
4、Unmarried single　5、Divorced / widowed cohabitation　6、Divorced / widowed single

B03 Your place of domicile: □

1、Xichang　2、Liangshan Prefecture barring Xichang
3、Sichuan Province barring Liangshan Prefecture（Please specify: __________city/ Prefecture）
4、other provinces / Municipality（Please specify: _______________province/ Municipality）

B04 Your living time in Xichang: □

1、≤3 months　2、3～6 months　3、6～12 months　4、1～2 years　5、≥2 years

B05 Your nationality: □

1、Han nationality　2、Yi nationality　3、other（Please specify: _______________ nationality）

B06 Your degree of education: □

1、illiteracy　2、Primary school　3、middle school
4、High school or special secondary school　5、Junior College
6、Undergraduate and above

http://dx.doi.org/10.21037/apm-20-2262

B07 Are you still engaged in other work: ☐

1、No 2、Yes（Please specify:_______________）

B08 Where is the main place of your sexual services in Xichang now: ___________ Street / Road

B09 Your current place of providing sexual services:_______________

（Please answer the specific type of venue. The following categories are for reference only.）

1、Separate sauna/bath center/massage center

2、Separate nightclub/ Karaoke Hall / discotheque /bar

3、Star-rated hotel（sauna/bath center/ hair salon / nightclub / Karaoke Hall / discotheque /bar etc.）

4、Non-star hotel（sauna/bath center/ hair salon / nightclub / Karaoke Hall / discotheque /bar etc.）

5、Small hotel or rest house 6、Own home / private home

7、Roadside/streetside: hair salon/beauty salon/hairdressing salon/shampoo room/footbath room

8、Station pile 9、Other（Please specify:_______________）

C、Dichotomous sensitive questions

C01：

- **Draw No. 1 ball：**Please answer the question: "Whether there is a spouse or any regular sex partner other than the sexual serving object that charges?"
- **Draw No. 2 ball：** Please answer the question: "Is your date of birth an odd number?"

☐1、Yes 2、No

（Note：No. 1 ball：No. 2 ball= 6:4）

C02：

- **Draw No. 1 ball：** Please answer the question: "When you were diagnosed with a sexually transmitted disease by a doctor, did you stop sexual services?"
- **Draw No. 2 ball：** Please answer the question: "Is your date of birth an odd number?"

☐1、Yes 2、No

（Note：No. 1 ball：No. 2 ball= 6:4）

 http://dx.doi.org/10.21037/apm-20-2262

C03：

- **Draw No. 1 ball：** Please answer the question: "Someone suggested that the state should allow prostitution to be legalized. Do you agree?"
- **Draw No. 2 ball：** Please answer the question: "Is your date of birth an odd number?"

☐ 1、Yes 2、No

（Note：No. 1 ball：No. 2 ball= 6:4）

The investigation is over. Thank you for your participation!

Questionnaire quality controller： __________________ **Date：** _______________

Appendix 3

```
%macro simulated_population(N=,
                            N1=,
                            N2=,
                            seed=);
data xc1;
do id=1 to &N;
    seed=&seed;
    h=rantbl(&seed,0.5525,0.4475);
    birthdate=rantbl(&seed,0.5,0.5);
    primary_unit=ceil(0+(&N1-0)*ranuni(&seed));
    second_unit=ceil(0+(&N2-0)*ranuni(&seed));
output;
end;
run;
proc sort data=xc1 out=xc2;
    by h primary_unit second_unit;
run;
data xc_population;
    set xc2;
    district=put(primary_unit,2.);
    site=catx('-',district,put(second_unit,2.));
    if h=l then do;
        c3_real=233.3122+256.23*rannor(&seed);
        d1_real=rantbl(&seed,0.5188,0.4812);
        e4_real=rantbl(&seed,0.1153,0.1504,0.7343);
    end;
    else do;
        c3_real=189.4191+201.3559*rannor(&seed);
        d1_real=rantbl(&seed,0.6095,0.3950);
        e4_real=rantbl(&seed,0.1123,0.1318,0.7559);
```

 http://dx.doi.org/10.21037/apm-20-2262

```
    end;
    drop primary_unit second_unit;
run;
proc sort data=xc_population;
    by h district site;
run;
%mend simulated_population;
%simulated_population (N=60000,
                       N1=10,
                       N2=12,
                       seed=1);
run;
```


http://dx.doi.org/10.21037/apm-20-2262