

Methodology of the Global Adult Tobacco Survey in China, 2010

JASON HSIA^{*,1}, GONG-HUAN YANG[#], QIANG LI^{#+}, LIN XIAO[#],
YAN YANG[#], YAN-WEI WU[±], AND SAMIRA ASMA^{*}

^{*}Office On Smoking And Health, Centers For Disease Control And Prevention, Atlanta, Ga, USA;

[#]National Tobacco Control Office, Chinese Center For Disease Control And Prevention, Beijing 100050, China; ⁺University Of Waterloo, Waterloo, Ontario, Canada;

[±]World Health Organization Country Office, Beijing, China

INTRODUCTION

The Global Adult Tobacco Survey (GATS) is a component of Global Tobacco Surveillance System (GTSS) under auspices of the Bloomberg philanthropy and the Bill and Melinda Gates Foundation. GATS is a household survey with a standard protocol and its goals are to measure tobacco use, to assess changes due to policy and to facilitate cross country comparison. China is the largest consumer and producer of tobacco in the world. China was selected as one of 14 countries of high burden of tobacco use, large population, and mostly low income, to conduct the GATS.

It is essential to establish an effective monitoring, supervising and evaluation system in China to monitor tobacco use and obtain nationally representative periodic data on the key indicators of tobacco use among adolescents and adults. This is a responsibility clearly given by the WHO Framework Convention on Tobacco Control (FCTC) to all parties, as well as an important component of the six most effective strategies for tobacco control (the MPOWER package) vigorously advocated by WHO. GATS China, started in 2009 and completed in 2010, serves as an important measurement of the Chinese tobacco control monitoring system. In this paper, we describe the design features, data collection methods, and analytic methods to help tobacco control experts and researchers to understand GATS features and to use GATS China data.

QUESTIONNAIRE DEVELOPMENT

Since 2007, US Centers for Disease Control and Prevention and World Health Organization in collaboration of the worldwide experts have developed the GATS standard protocols, including questionnaire. Each country participated the GATS customized the questionnaire and sample design under the generic requirement to assure the periodical and cross country comparison.

The objectives of GATS China are to systematically monitor adult tobacco use (smoking and smokeless) and track key tobacco control indicators in a nationally representative sample of China and to track implementation of the FCTC recommended policies outlined in the MPOWER package. The China GATS Questionnaire added optional and country-specific questions to the GATS core questionnaire. The questionnaire included the following sections: background characteristics, tobacco smoking, smokeless tobacco, cessation, secondhand smoking, economics, media, and knowledge as well as attitudes and perceptions. The Chinese Center for Disease Control and Prevention (China CDC) invited subject matter experts for three meetings to adapt the GATS questionnaire. The adapted questionnaire was translated and back-translated to ensure the quality of the translation. The questionnaire was then pre-tested in the field in May 2009 and August 2009. Based on the pretest, the questionnaire was slightly modified and finalized for

¹Correspondence should be addressed to Dr. Jason HSIA, Global Tobacco Control Branch, Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, GA USA 30341. E-mail: JHsia@cdc.gov

full study implementation. The final version of the questionnaire was approved by the GATS Questionnaire Review Committee in August 2009.

SAMPLE DESIGN

The target population was all non-institutionalized men and women age 15 years and older stratified by urban/rural and male/female. Institutional people are not included in the survey population, such as those living in military base or group quarters, hospitals, prisons, or nursing homes.

According GATS sampling protocol (REF GATS Sampling Manual), to meet the desired precision with 95% confidence interval, a sample size of 2 000 is needed. Therefore, GATS China required a sample size of 2 000 individuals to meet the desired precision for a specific estimate of each interested stratum of urban/rural and gender. Based on 1996 national

tobacco survey and 2002 behavioral risk factor survey conducted by the Chinese Center for Disease Control and Prevention, household and individual eligible rate was assumed to be 90% and 98% respectively and household and individual response rate was assumed to be 95% and 85% respectively. In China, the male response rate is lower than the female's and the response rate in urban areas is lower than that in rural areas. In order to assure that there are at least 2 000 male respondents, a total sample size of 14 185 was needed. Therefore the designed sample size was 15 000 for all of China. To take overhead and other costs into consideration, 100 primary sampling units, districts at urban areas and counties or county-level cities at rural areas, were determined. These PSUs were proportionally allocated to the 6 regions and equally allocated to urban/rural areas (Table 1). Within each PSU, 2 next level units and 75 households within the next level units were sampled.

TABLE 1

Number of Districts or Counties/County-level Cities in Target population and in Sample by Region and Urban/rural

Region	Number of Districts or Counties/County-level Cities (Sampled)		
	Total	Urban	Rural
North	428 (16)	110 (8)	318 (8)
Northeast	288 (10)	138 (5)	150 (5)
East	643 (22)	242 (11)	401 (11)
Central and South	633 (22)	214 (11)	419 (11)
Southwest	511 (18)	85 (9)	426 (9)
Northwest	355 (12)	64 (6)	291 (6)
Total	2858 (100)	853 (50)	853 (50)

A multi-stage stratified cluster sampling design was implemented in the survey. The country was divided into six regions: North, Northeast, East, Central, and South, Southwest, and Northwest. Each of the six regions was further divided into urban and rural areas, making 12 strata in total. At the first stage, urban districts or rural counties/county-level cities were selected, using the probability proportionate to size (PPS) sampling method from the 2007 household registration database administrated by the Ministry of Public Security of China. The measure of size (MOS) was the number of households. The selection probability of i^{th} district or county was given by

$$P_{hi} = \frac{a_h M_{hi}}{M_{h..}}$$

where M_{hi} is number of households of i^{th} district or county in h^{th} stratum, $M_{h..}$ is number of households in h^{th} stratum, a_h is the number of districts or counties selected ($i=1,2,...m_h$ and $(h=1,2,...12)$). A SAS procedure SURVEYSELECT (REF: SAS) was used for PPS sampling.

At the second stage, two urban neighborhood committees or rural villages were selected from each of the selected primary sampling units using the PPS method. The selected secondary sampling unit was partitioned into segments of around 1 000 households (using mapping and listing to determine the number) and one segment was randomly selected. The selection probability of j^{th} neighborhood committee or village was given by

$$P_{hij} = \frac{2M_{hij}}{M_{hi}}$$

where M_{hij} is number of households of j^{th} neighborhood committee or village in i^{th} district or county of h^{th} stratum, M_{hi} is number of households in i^{th} district or county of h^{th} stratum ($j=1,2; i=1,2,...m_h$ and $h=1,2,...12$).

At the third stage, 75 households were selected using simple random sampling from each selected segment, using systematic random sampling. The selection probability at stage 2 was given by

$$P_{hijk} = \frac{75}{N_{hij}}$$

Where N_{hij} is number of households were counted from the listing of j^{th} neighborhood committee or village of i^{th} district or county in h^{th} stratum ($j=1,2$; $i=1,2,\dots,m_h$ and $h=1,2,\dots,12$). Within the selected household, one eligible person was randomly selected at the last stage with selection probability of one over eligible persons from the selected k^{th} household ($k=1,\dots,75$).

DATA COLLECTION

The implementing agency responsible for GATS-China data collection was China's Centers for Disease Control and Prevention (CDC). Eight full survey implementation training sessions were conducted. The training sessions took place in small groups and were given by the same trainers to ensure quality. All field interviewers and county CDC staff at the primary sample unit level attended the training workshops. Fieldwork took place from December 1, 2009 to March 15, 2010.

There were 200 interviewers and 100 supervisors in the secondary sampling units. Two interviewers and one supervisor were assigned to each secondary sampling unit. The supervisor went with the interviewer to some of the households to ensure that the interviewer followed the survey guidelines. The supervisor also re-interviewed about 10% of households with a shorter version of the questionnaire as a validation check. All problems were reported to the provincial supervisors. GATS was conducted in 28 of 32 provinces in China. Each province had one provincial-level supervisor, who reported to the national office if a problem could not be solved at the provincial level. IT staff at the county CDC transferred the survey data from iPAQs to laptops and sent them to the national office weekly. The national office summarized the data and monitored data collection. The final data cleaning was conducted at the national office of the China CDC.

During the process of data collection, an interview would be terminated if selected respondents aged less than 17 years did not have formal permission from either their parents or guardians or if respondents aged 18 and over did not give their verbal permission. All collected data were kept strictly confidential.

ESTIMATION ISSUES

The weighting process for the GATS included three main steps: 1) the base weight or design weight,

calculated from all steps of random selection in the sample design; 2) an adjustment for non-response by sample households and samples individuals eligible for the survey; and 3) a post-stratification calibration adjustment of sample totals to projection of population 15 years old and above by region, area, gender, and age group.

Base weights equal to the reciprocal of the probability of selection were assigned to each sample person. The probability of selection is the product of the probabilities of selection: 1) the primary sampling unit (PSU); 2) the household within the PSU; and 3) the eligible sample person within the household. Base weight was trimmed at the 99th percentile to reduce sampling variance without compromising much of unbiased estimation. After trimming, the weight was adjusted to the same total. The trimmed base weights were adjusted for non-response on two factors: household level non-response adjustments, and person level non-response adjustments. Household level non-response adjustments were made within each PSU. The corresponding household level weighting class adjustment were computed as one divided by the weighted household response rate for each sample PSU. The person level response rate was computed within strategically formed subgroups: region, urban/rural, gender, age-grouped, and smoking status. This produced non-response-adjusted base weights for sample persons who responded. Population projections of persons by region, urban/rural, gender and age-grouped were used for post-Stratification calibration adjustment. The final weights assigned to each responding unit were computed as the product of the base weights, the non-response adjustment and post-stratification calibration adjustment. The final weights were used in all analyses to produce estimates of population parameters.

QUALITY ASSURANCE MEASURES

As a part of GATS protocol, quality assurance measures have been carefully examined. The GATS China designated sample size was 15 000 households. After excluding ineligible households, 13 562 households completed a household interview. The household response rate was 97.5% (96.0% urban and 98.8% rural). The individual response rate was 98.5% (98.3% urban and 98.6% rural). The overall response rate was 96.0% (94.4% urban and 97.4% rural). Because of this high response rate, the adjustment for nonresponse was small.

Summation of base weights prior to any adjustments is only 4 million apart from the actual total of target population. However, like most

household surveys, age distribution in the sample is imbalanced with fewer people between 15 and 24 years old and more adults older than 65 years than the actual distribution. Poststratification calibration was performed to overcome the imbalance problem.

Multiplicative effect of variable sample weight,

$$Meff_w = 1 + \frac{s_w^2}{\bar{w}^2},$$

where $s_w^2 = \sum_{i=1}^n (w_i - \bar{w})^2 / (n-1)$, was calculated

from the final data. The measure is below 2 for most of strata. Three of six regions where the rural stratum has the measure of higher than 2. Kish's design effect (REF: Kish), defined by $Deff_{CS} = 1 + \rho_o(\bar{m}-1)$, which is a measure of variances caused by clustering was checked. Since GATS sample involves both sampling clusters and unequal sample weights, the overall design effect on a survey estimate, denoted by $Deff(\hat{\theta})$, shown by Gabler, *et al.* (REF: Gabler *et al.*) to be the product of the multiplicative effect of variable sample weights ($Meff_w$) and Kish's design effect was also examined. For all key variables, the overall design effects are around 3.5. The relative standard errors for the key variables are small.

STATISTICAL ANALYSIS

To take the complex survey sample design into account, SUDAAN 10.0.1 (REF: SUDAAN) was used to compute the estimates and proper standard errors of population parameters. SPSS 18.0 (REF: SPSS) was used for sample weighting and data quality assurance. Sample weights were developed by the U.S. Centers for Disease Control and Prevention. Each responding unit was assigned a unique survey weight to be used in the calculation of survey estimates.

The final weight assigned to each respondent was computed as the product of the base weight, the non-response adjustment and post-stratification calibration adjustment. The final weight was used in all analyses to produce estimates of population parameters. All computations were performed using the SUDAAN complex survey data analysis procedure.

SAMPLE AND POPULATION CHARACTERISTICS

Overall, 13 354 surveyed individuals represented a population of 1 068 752 451 men and women aged 15 years and older in China. Table 2 shows selected demographic characteristics of the weighted

respondent data. There were 6 603 men and 6 751 women respondents, representing a target population of 50.9% men and 49.1% women. For age distribution, 5 000 and 5 001 individuals were interviewed in the 25 to 44 and 45 to 64 age groups, respectively. The number of respondents aged 15 to 24 was lower and the number of respondents aged 65 and above was higher. This was because many young people, particularly in rural areas, did not live at home during the time period of the survey, and also due to the smaller age range between 15 and 24. The imbalance was adjusted in post-stratification.

By residence, 5 832 and 7 522 individuals were interviewed from urban and rural areas, respectively. The weighted percentage estimate of people in urban areas is 46.1%, and in rural areas 53.9%. This indicated a tremendous change in urban/rural distribution since the year 2000. For education level, 36.4% had attended secondary school, 33.3% had attended primary school or less, 18.8% were high school graduates, and 11.6% were college graduates or above. Occupation was based on definitions provided by the National Bureau of Statistics. 31.6% were agricultural workers, 15.7% were business or service employees, 10.8% were machine operators. Worthy of note is the fact that 11.0% were unemployed, and 9.6% were retired. Also, only 1.7% were medical/health personnel, and 1.6% were teaching staff representing close to 18 million and 17 million people, respectively.

DISCUSSION

GATS China was conducted under collaboration between US CDC and China CDC, following Global Adult Tobacco Survey standard protocol and serving as a component Global Tobacco Surveillance. GATS China 2010 has laid the foundation for systematic China adult tobacco surveillance with well implemented and documented standards. One of difficulties in comparison of tobacco use indicators and policy implementation is source of national data. Tobacco use data often comes from national representative surveys and the different design, questionnaire, and field operation standard of these surveys provides methodological challenges for monitoring tobacco use. GATS China has made up for this gap in the world, including the largest tobacco consumption and production country, China. China has conducted several national health related surveys, such as the National Health Household Survey and a national nutrition survey. Lack of quality control both in sample design and data collection introduces bias and variation of estimates.

TABLE 2
Distribution of Adults ≥ 15 Years Old by Selected Demographic Characteristics

Demographic Characteristics	Weighted		Number of Adults (in Thousands)	Un-weighted Number of Adults
	Percentage (95% CI)			
Overall	100		1 068 752	13 354
Gender				
Male	50.9	(49.0, 52.9)	544 452	6 603
Female	49.1	(47.1, 51.0)	524 300	6 751
Age (years)				
15-24	21.5	(19.3, 23.9)	229 512	1 146
25-44	39.5	(37.5, 41.5)	422 039	5 000
45-64	29.2	(27.7, 30.8)	312 131	5 001
65+	9.8	(9.1, 10.7)	105 071	2 207
Residence				
Urban	46.1	(35.9, 56.6)	492 389	5 832
Rural	53.9	(43.4, 64.1)	576 363	7 522
Education Level				
Primary School or Less	33.3	(30.1, 36.7)	279 339	4 966
Attended Secondary School	36.4	(34.1, 38.7)	305 039	4 111
High School Graduate	18.8	(16.7, 21.0)	157 424	1 887
College Graduate or Above	11.6	(9.2, 14.5)	97 218	1 239
Occupation				
Agriculture Worker	31.6	(25.3, 38.7)	338 008	5 377
Machine Operator	10.8	(8.7, 13.4)	115 740	1 253
Business or Service Employee	15.7	(13.8, 17.7)	167 240	1 694
Leaders of Organizations	5.2	(4.0, 6.8)	55 610	641
Clerks	1.7	(1.3, 2.3)	18 624	242
Specialized Technicians	4.0	(3.3, 4.9)	42 513	476
Medical/Health Personnel	1.7	(1.2, 2.3)	17 863	169
Teaching Staff	1.6	(1.2, 2.1)	16 603	210
Students	6.4	(5.3, 7.7)	68 340	350
No Jobs	11.0	(9.0, 13.4)	117 457	1 370
Retired	9.6	(7.8, 11.9)	103 009	1 451
Others	0.5	(0.3, 1.1)	5 715	96
Region				
East	38.9	(30.0, 48.6)	415 285	4 402
Central	27.3	(19.4, 37.0)	291 964	3 940
West	33.8	(27.2, 41.1)	361 503	5 012

Note. Education level is reported only among respondents 25+ years old.

GATS China has much more robust design and strict data collection procedure, increasing precision and minimizing nonsampling errors.

The main challenge in designing GATS China was the size of the population, geographic distribution, and survey cost. There is a large

variation in socioeconomic development and culture background. From the precision and sampling error perspective, larger number of PSUs is more desirable. On the other hand, a larger number of PSUs could significantly increase survey cost, in particular overhead cost, and nonsampling error as quality of

training of interviewers and supervisors and the number of interviewers and supervisors can be quite different from the same survey in a much smaller county. Our design tried to balance sampling errors and nonsampling errors.

A limitation in 2010 GATS China design is in the study of China's migrating population, also called it "floating population". Most of them move from their home in rural areas to urban areas to work. There are currently over 200 thousand adults belonging to this special population, many of whom do high risk job for little pay. It is difficult to locate them by conventional mapping and listing methods. The smoking prevalence of the floating population is believed to be higher in the urban areas where they work than in their rural areas of origin. In spite of this

limitation, the design of GATS China is effective overall.

REFERENCES

1. Kish L (1965). Survey Sampling, Wiley and Sons, New York.
2. Gabler S Haeder S and Lahiri P (1999). A model based justification of Kish's formula for design effects for weighting and clustering. *Survey Methodology*, **25**(1), 105-106.
3. Global Tobacco Surveillance System (GTSS) (2009). Global Adult Tobacco Survey (GATS): Sample Design Manual.
4. SAS Institute Inc, SAS 9.1.3 Help and documentation, Cary, NC: SAS Institute Inc, 2000-2004.
5. SPSS Inc, PASW Statistics 18 Command Syntax Reference, Chicago, IL: SPSS Inc, 1993-2007.
6. Research Triangle Institute, SUDAAN Language Manual, Release 10.0 Research Triangle Park, NC: Research Triangle Institute, 2008.

(Received December 2, 2010 Accepted December 17, 2010)