The Global Adult Tobacco Survey (GATS): Sample Design and Related Methods

William D. Kalsbeek[1], J. Michael Bowling[1], Jason Hsia[2], Sara Mirza[2], Krishna M. Palipudi[2], Samira Asma[2,3]

[1] Survey Research Unit, Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA;

[2] Office of Smoking and Health, Centers for Disease Control and Prevention, Atlanta, GA, USA;

[3] On behalf of the Global Adult Tobacco Survey Collaborative Group, Atlanta, GA, USA;

Abstract:

Begun in 2007 as a component of the Global Tobacco Surveillance System (GTSS), the Global Adult Tobacco Survey (GATS) is a systematically and comparably designed set of international surveys on the tobacco use behavior of civilian, non-institutionalized persons 15 years of age and older in participating low-to-middle income countries around the world. These surveys are intended to provide an ongoing source of data to plan and evaluate interventions to prevent and reduce tobacco use in each country. This paper focuses on the design features and practical outcomes of sampling-related activities in GATS. The paper also presents the rationale for and the main design features for samples selected in 14 GATS countries to date, a summary of country adaptations, an overview of steps taken to compute sample weights and to assure quality in selected sample, and a tabular summary of the designs and related recruitment experience in eight of the 14 countries. We conclude by reflecting on several lessons learned and future issues to address in further developing this system of population surveys.

Key Words: sample design health surveys, survey nonresponse, sample weights

1. Introduction

Tobacco use is a major preventable cause of premature death and disease worldwide. Approximately 5.4 million people die each year due to tobacco-related illnesses – a figure expected to increase to more than 8 million a year by 2030. If current trends continue, tobacco use may kill a billion people by the end of this century. It is estimated that more than three quarters of these deaths will be in low- and middle-income countries (Mathers and Loncar, 2006). An efficient and systematic surveillance mechanism is essential to monitor and manage the epidemic.

The Global Adult Tobacco Survey (GATS), a component of the Global Tobacco Surveillance System (GTSS), is an international mechanism to systematically monitor adult tobacco use and to track key tobacco control indicators over time. GATS in each

participating country is a nationally representative household survey of adults 15 years of age or older using a standard core questionnaire, sample design, and data collection and management procedures that were reviewed and approved by international experts. GATS is intended to enhance the capacity of countries to design, implement and evaluate tobacco control interventions. GATS is funded through the Bloomberg Initiative to Reduce Tobacco Use, a program of Bloomberg Philanthropies.

GATS is a multi-partner initiative of the World Health Organization (WHO), US Centers for Disease Control and Prevention (CDC) and national governments. GATS surveys in individual countries exist within a system of surveys designed to ensure cross country comparisons and consistency over time. From a public health policy perspective, GATS will assist countries to monitor the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) and MPOWER policies.

A survey administered to a nationally representative sample is a useful strategy to provide both the cross-sectional and longitudinal information that are needed for tobacco use prevention and control. Several ongoing health surveys offer the advantage of providing high quality population data at relatively low cost by employing systematic methods of objective data gathering from scientifically representative samples of their targeted populations, such as the Global Youth Tobacco Survey (GYTS), World Fertility Survey (WFS), Behavioral Risk Factors Surveillance System (BRFSS), etc. Critical to the success and utility of these data systems is the use of scientifically rigorous sampling methods to assure the statistical validity and precision of survey findings. GATS has adopted the highest standards in this regard, by utilizing state-of-the-art methods of sample selection, recruitment, and data collection using electronic handheld devices to assure that the best possible sample estimates can be produced.

A well-conceived survey protocol is applied to each GATS household sample. The sample is first chosen following a stratified multi-stage cluster sampling approach in which probability proportional to size (PPS) random selection methods are used to successively choose clusters in one or more selection steps. Teams of thoroughly trained field interviewers then apply well-established sample household recruitment strategies to attempt to persuade a knowledgeable adult in each sample households to assist in creating a roster of all household residents that are eligible for GATS.[1] In addition to current status criteria defined by age, institutionalization status, and participation in the military, individual residents must meet country and household requirements of usual residence, as defined by the percent of the past year that was spent in the country and selected household, respectively.[2] Implementation of these criteria implied a *target population* for each survey consisting of civilian, non-institutionalized persons 15 years of age or old who are "usual" residents of the country. In some instances geographic portions of the country were excluded because of extreme inaccessibility and/or civil unrest. A

[1] Residency in the selected household was initially based on what individuals would consider to be their primary place of residence the night just prior to completing a roster of eligible residents and selecting one of them at random for the GATS interview. This notion of a "primary place of residence last night" was later clarified to become an individual's usual place of residence in the year just prior to rostering and selection.

[2] To meet the requirement of "usual" country residency, an individual must be: (i) a citizen of the country, or (ii) a non-citizen who has lived in the GATS country for at least half of the time during the previous year. To be considered a "usual" household resident, (i) the individual must have no other place of residence, or (ii) the selected household must be where he/she has lived at least half of the time during the previous year if he/she has multiple places of residence.

household-level response occurred when one of the eligible residents on a roster of household residents was chosen at random. Person-level participation happened once the selected resident fully completed the 30-45 minute Individual Questionnaire, which contains both demographic and tobacco use questions that are answered in an in-person interview administered by an interviewer with the assistance of a hand-held (iPAQ) computer that had been pre-programmed to complete the household rostering and randomly choose a resident to complete the survey interview.

The sample in population surveys must be designed to meet the information needs set for it. Findings from GATS are used to quantify and to better understand circumstances surrounding public exposure to the harmful effects of tobacco in devising plans to limit this exposure. This requires that various types of tobacco use prevalence rates be sufficiently accurate for several population subgroups so that public health policy and other interventions to limit exposure will more effectively target those in need of reduced exposure. It was therefore important that the sample design in each GATS country produce prevalence estimates of acceptable statistical quality for important population reporting domains defined by the urbanicity (urban/rural), gender (male/female) and in some instances regional location of survey respondents.

There are two common standards of statistical quality in survey estimates that apply to GATS. One is validity which was met by requiring probability sampling methods, employing effective sample recruitment strategies that would lead to high levels of survey participation, correctly computing sample weights for each survey respondents, and properly accommodating key sample design features such as stratification, cluster sampling, and planned disproportionate sample allocation (e.g., by urbanicity) to control sample sizes. A second key standard set for GATS sample designs was the reliability of estimates as measured by the precision of cross-sectional estimates at each survey round and by the statistical power to detect both intra- and inter-country differences. Design features affecting reliability included minimum primary sampling unit (PSU) sample sizes for the sample as a whole, acceptable respondent sample sizes for important population domains, and planned sample overlap among rounds of data collection to enhance the quality of temporal comparisons and trends.

The purpose of this paper is to introduce and to share experience regarding the GATS research design, particularly as it relates to sampling adopted in eight of the 14 GATS countries. We begin by reviewing some of the key features of the design requirements for sample selection and recruitment. A summary of design results is then presented for completed GATS surveys, followed by a discussion of some lessons learned that may affect how future GATS samples are selected.

2. Sample Design and Results

2.1 Participating Countries

To date, GATS surveys have been completed in the following countries located on all five of the world's major continents: Bangladesh, Brazil, China, Egypt, India, Mexico, Philippines, Poland, Russian Federation, Thailand, Turkey, Ukraine, Uruguay, and Vietnam. Except for Uruguay, countries were selected based on the highest number of smokers in the world. These countries account for 61% of the world's adult population of persons 15 years of age and older and about 62% of the world's adult smokers. They vary in population size from very large (China and India) to relatively small (Uruguay). GATS

is currently expanding to other countries such as Argentina, Indonesia, Malaysia, Nigeria, Pakistan and Romania. Moreover, some countries that have completed a baseline round of GATS are expected to repeat the survey in 2011-2013.

Planning and implementing GATS in each participating country generally involved several agencies and organizations, including the country's national statistics office, the ministry of health (including its various research affiliates), and occasionally a private or university-affiliated survey research organization. The health ministry was usually the main coordinating agency for this effort. Also involved were representatives from CDC, the WHO, and various academic and non-academic consultant organizations. Procedurally, completing a GATS in each country typically involved the following steps: decide on the content and format of the survey questionnaire, pretest the questionnaire, design and select the household sample, recruit the sample and collect the GATS interview data from a randomly chosen eligible resident of each sampled household.

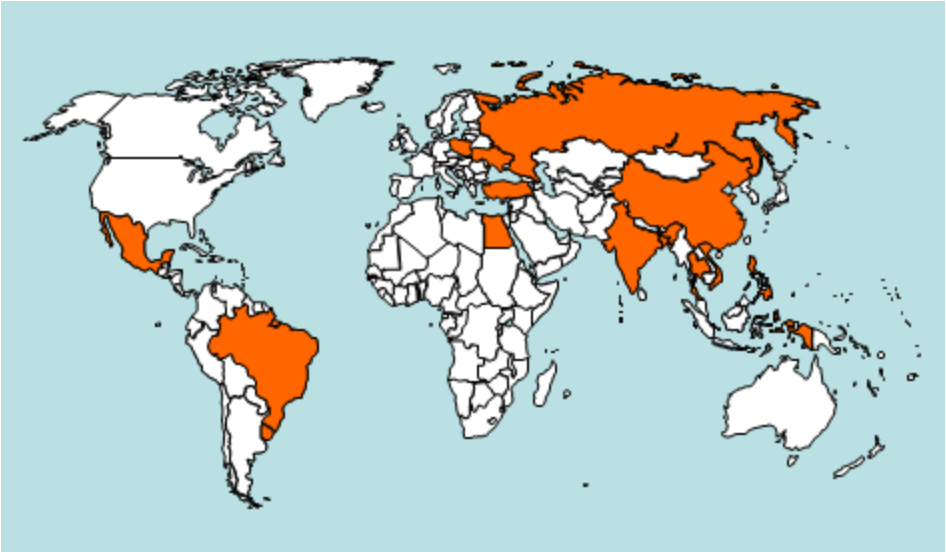

Figure 1: Existing GATS Countries**;** Bangladesh, Brazil, China, Egypt, India, Mexico, Philippines, Poland, Russian Federation, Thailand, Turkey, Ukraine, Uruguay, and Vietnam

2.2 Design Standardization

Inter-country comparisons are facilitated by standardizing the design in each GATS country. Design standardization requires that countries adopt key features of a carefully developed survey design protocol that has been thoroughly spelled out in a series of procedural manuals and guidelines available in country-specific languages, provided to partner organizations in each country. Topics of these documents include: (i) how to identify and assemble the appropriate consortium of organizational partners to plan and conduct the survey, (ii) how to design and compute weights for GATS sample s, (iii) questionnaire development manuals that describe each of the core and optional GATS survey questions, and provide instructions on how to administer each question, (iv) field instructions on how to complete household mapping and listing, (v) operational instructions for field interviewers and supervisors, and (vi) analysis and reporting manuals that spell out how the initial findings ("fact sheets") and a more thorough set of tobacco use findings (the "country report") are to be produced. Copies of these manual are available at the GATS website (*http://www.cdc.gov/tobacco/global/gats/*).

GATS procedural documents are rigid yet flexible. On the one hand, they collectively reflect the need for conformity by each country to high professional standards and inter-country comparability in survey design. However, they also offer sufficient flexibility to enable each GATS survey design to adapt to the country's socio-cultural norms (e.g., unacceptability for a respondent to interviewed by a person of the opposite gender or a different ethnic group) and the specific information needs of agency and organizational partners (e.g., the need for regional estimates). For example, countries were offered the option to randomize their GATS household sample by gender within the area cluster from which the household samples were chosen. One male resident was randomly chosen from a list of eligible male residents in households randomly assigned to be "male households;" and likewise for "female households."

2.3 Specifications and Requirements

The statistical purpose of the sample and gathered data from survey respondents is to produce several dozen *prevalence rates* variously indicative of tobacco use by members of the target population in each country. These rates are derived from several groups of interview questions including: (i) tobacco smoking behavior (past, current, frequency, attempts to quit), (ii) smokeless tobacco behavior (past, current, frequency, attempts to quit), (iii) secondhand smoke (exposure in the home, at work and at other common public places and venues), (iv) economics of personal tobacco use (what was purchased, where, and what did it cost), (v) recent exposure to media (encouraging the use of tobacco products, health warnings about tobacco use), (vi) knowledge and attitudes (about tobacco use and its impact on health), and (vii) respondent demographic characteristics.

Moreover, tobacco use prevalence rates of interest to a country were produced for the nation as a whole, as well as for various other *key reporting domains*. These domains have usually been defined by some combination of the following individual demographic characteristics: (i) urbanicity, (ii) gender , and (iii) the geographic region of the country in which the respondent resides.

Recommended sample sizes were based on two specific sets of precision requirements for estimated rates ($\hat{p}$) computed by reporting domain, with the goal of achieving a combined design effect ($Deff_c$) averaging about 2.00 for these estimates. One requirement was that a 95% margin of error (i.e., $MOE(\hat{p}) = (1.96)SE(\hat{p})$) on a current tobacco use rate of less than 40% would be no greater than 3 percentage points. Another was the need for at least 80% power to detect a drop from 40% to 34% in tobacco use prevalence between rounds of GATS, based on a two-sided alternative with type I error of $\alpha = 0.05$ and independently chosen samples in the two rounds.

To assure sufficient validity of the respondent sample in each country, the GATS sample design manual (GATS Collaborative Group, 2010a) requires that countries collect data by face-to-face interviews in a national sample of households chosen following current accepted practices for multi-stage area household sampling as described by Kish (1965) and other more recent sampling texts. This implied that the sample of households be chosen in two or more stages, where sampling units in the first and/or second selection stages were well-defined geo-political areas in the country and the final household selection stage involved choosing a sample from a relatively current and acceptably complete household listing for a relatively small area unit consisting of no more than about 250 households. While it was recommended that countries construct a dedicated

household frame for GATS by using standard mapping and listing techniques, use of existing household lists (e.g., from local administrative sources or updated household lists from the most recent census) was more likely. Spin-the-bottle random pathway and other similar quasi-random household selection techniques were not allowed.

In countries where for cultural reasons it was important to match interviewer and respondent on gender, countries were given the option to randomize their household sample into male or female subgroups. Households in the male group listed and selected only eligible male residents (who were assigned to male interviewers), while only eligible females residents were selected, and a female interviewer assigned, in households in the female subgroup. Finally, within-household sampling involved choosing one eligible resident to interview in each participating household, as previously noted. Thus, a general requirement was that fully random selection methods be used in each sampling stages so that the final sample of respondents would be as close to a valid probability sample as practically possible. Use of random or non-random substitution methods to deal with non-participation at any stage of sampling was also prohibited.

A minimum of 2,000 respondents has been required for each key reporting domain in order to meet GATS requirements for the precision of cross-sectional prevalence estimates and for the power to detect differences estimates between rounds. The required overall respondent sample size is thereby based on the number of key reporting domains that are needed. One common set of required domains in participating countries has called for producing national estimates overall and by the four population subgroups formed by the cross-classification of gender by urbanicity, thereby leading to a recommended overall sample size of 8,000 = 4 x 2,000. When national estimates have been needed for the marginals of these two domain variables, the recommended overall respondent sample size has been reduced to 4,000 = 2 x 2,000. Finally, if domain estimates involving gender and urbanicity were required for each of R regions to accommodate more localize health planning or policy development, recommended overall sample sizes have expanded to R x 8,000 for domains involving the joint classification of gender and urbanicity and R x 4,000 for corresponding marginal domains.

Stratification in the first stage of sampling helped to control domain sample sizes at recommended levels, the goal being approximately 2,000 respondents in each domain regardless of the population percent for the domain. With approximately equal percentages of males and females in the population and a goal of equal sample sizes by gender in most countries, stratification by gender was unnecessary. However, the urban:rural percent mix in the population in most countries departed somewhat from 50:50. Therefore it was essential to explicitly stratify the first stage sample by urbanicity so that roughly equal urban and rural sample sizes could be achieved. This equal allocation by urbanicity, especially when the urban:rural mix departed substantially from 50:50, created disproportionality in the final overall sample.

2.4 Country Adaptations

Stratification was one of several design features where modification to the standard design protocol was needed to address particular information needs and circumstances in GATS countries. For example, Egypt required estimates marginally by gender for the 5 administrative regions (partly defined by urbanicity) that became first stage strata: Urban Cosmopolitan Governorates (e.g., Cairo), Urban Lower Egypt, Rural Lower Egypt,

Urban Upper Egypt, and Rural Upper Egypt. This implied a target respondent sample size of about 20,000 = 5 x 4,000.

The actual survey protocol also varied by the type of in-country organization that did the sampling and data collection work for GATS. In some countries the national statistical office (NSO) completed these survey activities since this agency typically has the most prior experience with conducting national surveys. Even when the NSO was not the main implementing agency of the survey,, it was frequently the main source of the data, maps and household listings that the lead survey organization would need to successfully complete a GATS survey. In some instances another governmental ministry, such as the one responsible for health promotion, did the sampling and data gathering, since it may have been conducting its own surveys and thus have access to the required expertise and field operations capacity to mount a face-to-face national sample survey. When governmental agencies were unable or unavailable to conduct this survey work, country staff would look to academic or private survey research firms with staff possessing the needed skills to satisfactorily complete a national survey effort.

A third variation in the eventual survey design was determined by the role other samples played in selecting the sample for GATS. In some countries other samples played no role at all, thus making the GATS sample the result of a "stand-alone design" developed solely to meet the information needs of GATS. In other countries the GATS sample was at least partially dependent on the sample chosen via another sample design, thus integrating the two designs somehow. In some instances the other source was a "master sample" selected for more general use in various national survey efforts, while in others it was the sample selected for another national survey. Regardless of source, the GATS sample was chosen from the other design and design integration occurred in one or more sampling stages, with the first stage being the most common point of integration. Designs could only be considered to be fully integrated when the GATS and the other design included the same households, thus implying complete overlap in all sampling stages. Any other integration scheme was partial, and the absence of any planned sample overlap was the result of samples that had been independently chosen.

2.5 Sample Weights

The GATS sample weights manual (GATS Collaborative Group, 2010b) offers a detailed description of the usual three-step approach for computing weights in area probability samples for each country sample (Lessler and Kalsbeek, 1992; Kalton and Flores-Cervantes, 2003):

2.5.1 Compute a Base Weight

This is the inverse of the respondent's selection probability, determined as the product of all stage-specific conditional selection probabilities. These component probabilities were often readily available since statisticians in charge of sample selection were encouraged to compile selection probabilities as each sampling step was completed and selection probabilities noted.

2.5.2 Adjust the Base Weights for Nonresponse

The adjustment for each respondent was the product of separate weighting class adjustments that were produced for the two levels of survey nonresponse affecting the respondent. The household-level adjustment was the inverse of the response rate in a cluster from which the respondent's household was chosen, while the respondent's person-level adjustment was the inverse of the response rate in the respondent's

adjustment cells defined by the categorical cross-classification of age, gender, smoking status, and (when applicable) region. Except for region, the latter set of adjustment cell variables were obtained from items recorded in the household roster.

2.5.3 Calibrate the Adjusted Weight

A calibration factor was multiplied times the adjusted weight for each respondent so that the final weighted distribution of the sample would match the corresponding distribution of population counts. The distribution of population counts to which the sample was calibrated was strategically defined by individual population characteristics that are known to predict tobacco use behavior (i.e., age, gender, and education), or region and urbanicity which define key reporting domains. Weights were typically calibrated to original or updated national findings from the last census, or to estimated counts from a larger, highly credible survey sample.

Variation among nonresponse and calibration adjustment cells, along with disproportionality in the sample arising mainly out of the goal of equal allocation of urban and rural respondents regardless of the urban:rural percent mix in the population, created weights that varied. This variation in sample weights among respondents used to produce survey estimates can reduce the precision of these estimates, as discussed below (Section 3.4).

2.6 Quality Assurance

While the development of several sets of written guidelines and manuals helped establish a scientific standard for sampling in GATS, it was believed that review of the sampling process was also important to assure that these standards would be met. Formal reviews of sampling activity occurred at two points in the survey timeline for each country. An external review committee of sampling experts was formed to offer a constructive review of sample design documents prepared by in-country staff. This review occurred just prior to sample selection and sought to assure that the proposed sampling plan was in compliance with the protocol set out in the sample design manual, and that the design appropriately accommodated the country's particular tobacco use information needs. A second external review was completed by members of the sampling review committee soon after the completion of data collection but prior to the initial release of survey findings. The purpose of this review was to confirm that sample weights had been computed in accordance with the sample weights manual, and that other statistical indicators of sampling quality (e.g., response rates, *MOE* and $Deff_c$ for survey estimates, etc.) were as expected based on the final survey design.

3. Summary of Country Sampling Results

Table 1 includes several indicators of the circumstances, design, and statistical outcomes we have summarized for eight of the 14 GATS countries where initial survey findings have been released to the public. These countries included: Bangladesh, Brazil, Egypt, Mexico, Philippines, Thailand, Turkey, and Uruguay. Each of the reported indicators is presented in turn below.

3.1 Design Considerations

We examined two facets of the circumstances surrounding the GATS surveys in these countries. One was the key reporting domains for findings, for which we observed that five (5) of the countries gave priority to producing national estimates jointly by

Table 1: Summary of Sampling Results for Eight GATS Countries

	Bangladesh	**Brazil**	**Egypt**	**Mexico**	**Philippines**	**Thailand**	**Turkey**	**Uruguay**
Key reporting subgroups	National; Urban / Rural; Male / Female	Regional (5); Urban / Rural; Male / Female	Regional (5); Urban / Rural; Male / Female	National; Urban / Rural; Male / Female	National; Urban / Rural; Male / Female	Regional (4) and Bangkok metropolis; Male / Female	National; Urban / Rural; Male / Female	National; Urban / Rural; Male / Female
Type of organization that did sampling and conducted data collection:	Academic Survey Research Center	National Statistical Office	National Statistical Office	Other Govern-mental Agency	National Statistical Office	National Statistical Office	National Statistical Office	National Statistical Office
Role of other in-country samples	Stand alone	Integrated	Integrated	Stand alone	Integrated	Stand alone	Stand alone	Stand alone
Number of stages to sample households	3	3	2	3	3	2	2	3
Overall number of selected first stage sampling units (PSUs)	400	851	880	181	794	1,088	400	150
Overall respondent sample size	9,629	39,425	20,946	13,627	9,705	20,566	9,030	5,581
Average sample cluster size $(\bar{m})$	24	46	24	75	12	19	23	37
Overall household-level response rate (AAPOR: RR1)	97.7%	95%	98.9%	89.6%	97.4%	97.9%	93.7%	97.0%
Overall person-level response rate (AAPOR: RR1)	95.8%	98.9%	98.4%	92.1%	97.4%	96.2%	97.0 %	98.5%
Combined overall response rate	93.6%	94%	97.3%	82.5%	94.8%	94.2%	90.8%	95.6%
Household-level refusal rate (AAPOR: REF1)	0.2%	NA	0.4%	6.4%	0.1%	1.5%	0.2%	1.0%
Personal-level refusal rate (AAPOR: REF1)	0.2%	NA	0.2%	2.2%	0.1%	1.7%	0.1%	0. 6%
Multiplicative effect of variable weights for the overall sample $(Meff_w)$	1.95	NA	1.81	1.76	1.59	2.49	1.51	1.81
Urban:Rural Percentage Split[a]	25:75	84:16	43:57	76:24	63:37	32:68	67:33	92:8
Estimated combined design effect $(Deff_c)$ **for the overall-population estimate of the current tobacco use prevalence rate**[b]	2.78	NA	1.67	3.68[c]	1.75[c]	3.21	1.82[c]	2.10

NA Not available

a Source:*The World Factbook 2005*. Washington, DC: Central Intelligence Agency, 2005. *<<https://www.cia.gov/library/publications/the-world-factbook/index.html>>*

b Current tobacco use includes both smokers and smokeless tobacco users

c Indicates the $Deff_c$ calculated for current tobacco smokers

urbanicity and gender, while three (3) required estimates by urbanicity and gender separately by region. A slight variation on national estimates occurred in Uruguay where gender comparisons were only required in urban areas since such a small percent of the population live in rural parts of the country, thus implying three reporting domains (i.e., urban male, urban female, and rural). Regions were variously defined in the second grouping of three countries. For instance estimates for the five regions in Brazil and Thailand were based solely on geographic location within the country while, as noted previously, Egypt's five regions were defined both by location and urbanicity.

Another aspect of the countries' circumstances we observed was the type of survey organization that was responsible for sampling and data collection. In six of the eight countries profiled in Table 1, GATS was implemented by the national statistics office. Another government agency, an institute of public health, did this work in Mexico, while a university survey research center led these tasks in Bangladesh.

3.2 Design Features

Several key features of the sample design for area household samples were also profiled for the eight countries. One was the connection between the GATS sample and other in-country national samples. In five (5) of these countries no other sample design played a role in selecting the sample for GATS. The "stand-alone" sample in these countries was designed exclusively to meet the tobacco use data needs of the country, whereas in the other three (3) countries another sample design was involved in producing the GATS sample. For example, in Brazil a 1/3 random subsample of households in all selected sampling units through the penultimate sampling stage of an ongoing national household sample survey, called Pesquisa Nacional por Amostra de Domicilios (PNAD), were assigned for use in GATS. In Egypt a sample of households was independently chosen from a recently updated household listing in each member of a random subset of a national master sample of PSUs. Similarly, one of four replicate subsamples of PSUs from a two-stage master sample, including an updated census list of households, was used to create the GATS household sample in the Philippines.

Another comparative design feature, the number of sampling stages in which the GATS sample of households was chosen, varied somewhat among the eight countries as well. Two stages were required for household selection in three (3) countries, and three stages were needed in the other five (5) countries. The number of stages here is likely to have been inversely related to quality of the country's road system, and directly related to the country's geographic size and the availability of census data for levels of the country's hierarchy of geo-political area units.

PSU and respondent sample sizes also varied somewhat, with the largest sizes occurring in countries that also required regional estimates. Countries requiring national estimates chose between 150 and 794 PSUs, while PSU sample sizes for countries requiring regional estimates ranged from 851 to 1,088. Similarly, overall respondent sample sizes for countries that required only national estimates varied from 5,581 to 20,946, while those also needing regional estimates were between 20,566 and 39,425. One also notes that these two sample sizes have a high positive correlation, thus implying for these GATS samples the common practice of increasing PSU sample size when overall samples increase so that one can control the average respondent sample size per PSU, which directly affects the variance of survey estimates from clustered samples.

3.3 Sample Recruitment

Success in the recruitment of selected survey samples is measured by the ability to limit survey non-participation and thus maximize survey response rates. Separate household- and person-level response and refusal rates based on standard AAPOR formulae were computed and are presented (Table 1) for the eight countries. We used AAPOR's most conservative response rate formulae (RR1), which excludes partial completes from the numerator and includes nonrespondents with an unknown eligibility status in the denominator. Using the same denominator as RR1, the AAPOR refusal rate (REF1) was used to measure willful non-participation in GATS.

The combined response rate for the overall sample in each GATS country was computed as the product of the RR1 response rates at the household- and person-level. Overall rates in the eight countries ranged from a low of 82.5% (Mexico) to a high of 97.3% (Egypt), with a median value of 94.1%. These combined rates are the result of household-level rates that varied from 89.6% (Mexico) to 98.9% (Egypt), with a median value of 97.2%. Person-level rates ranged from 92.1% (Mexico) to 98.5% (Uruguay), with a median of 97.2%.

In general, non-participation was very low in these GATS samples. Part of the reason for this finding is the relative rarity of refusals at both the household- and person-levels. We note from Table 1 that household-level refusal rates varied from 0.1% (Philippines) to 6.4% (Mexico), with a median of 0.4%; and that person-level refusal rates varied from 0.1% (Philippines) to 2.2% (Mexico), with a median of 0.2%.

3.4 Quality Indicators

The multiplicative effect of variable weights on the variance of a survey estimate is commonly modeled as, $Meff_w = 1+(CV_w)^2$, where CV_w is the coefficient of variation of the sample weights for those sample members used to produce the estimate (Kish, 1965; Section 11.7B). Kish's widely used multiplicative effect of cluster sampling is represented as $1+ICC(\bar{m}-1)$, where *ICC* is a measure of intra-class correlation (or, perhaps more appropriately, intra-PSU homogeneity) and $\bar{m}$ is the average sample cluster size, computed for overall population estimates as the overall respondent sample size divided by the PSU sample size (Kish, 1965; Section 8.2). The product of these two effects for an estimate is its combined design effect (Kish, 1987), justified by Gabler, et al., (1999), as,

$$Deff_c = \left[1+(CV_w)^2\right]\left[1+ICC(\bar{m}-1)\right] \quad (1)$$

An estimate of $Deff_c$ is often available on user request for each survey estimate from most survey analysis software (e.g., in SUDAAN, SPSS, etc.), and $Meff_w$ can be easily estimated for each reporting domain as a descriptive statistical measure of the survey respondents in that subgroup.

Table 1 presents values of $Meff_w$ and $Deff_c$ for overall population estimates of the current tobacco use prevalence rate. Values of $Meff_w$ for overall population estimates ranged from 1.51 (Turkey) to 2.49 (Thailand), with a median value of 1.81. Although variation in weights can arise from poor quality measures of size for PPS sampling and from differences in nonresponse and calibration adjustments among adjustment cells, a key factor in determining the size of $Meff_w$ for overall population estimates from GATS is the urban:rural mix in the population. Except for Uruguay, each of the eight countries

targeted an equal number of urban and rural respondents. Thus, oversampling associated with urbanicity contributed to disproportionaltity (and variable weight) in the sample to the extent that the country's urban:rural mix departed from 50:50.

Values of $Deff_c$ for overall tobacco use prevalence varied from a low of 1.67 (Egypt) to a high of 3.68 (Mexico), with a median value of 2.10, or slightly above the target value of 2.00. Since values of $Deff_c$ are specific to individual rate estimates, we also looked at the pattern of reported $Deff_c$ from SPSS (Version 18) for a fairly diverse set of 36 tobacco use rates (e.g., measuring things tobacco use attempts to quit, exposure to media, knowledge-attitudes about smoking, etc.) that were computed from Egypt's GATS sample (GATS- Egypt, 2010). . Values of $Deff_c$ ranged from 1.67 for current tobacco users to 4.08 for the rate of those believing that tobacco use causes stroke, with a median of 2.59, which is substantially higher than the target of 2.00. Finally, by solving for *ICC* in Eq. (1) we were also able to compute *ICC* for these use rate estimates. We found that ICC varied from a low of -0.0036 to a high of 0.0547 corresponding to the same low- and high-value tobacco use rates as reported for $Deff_c$, with a median value of 0.0189.

4. Discussion

We present evidence that efforts to implement a comprehensive design protocol for GATS country samples that following established principles of probability sampling have generally succeeded based on the experience summarized for eight of these countries. While country adherence to several key provisions of this protocol has been necessary, its implementation has been sufficiently flexible to adapt to the needs of individual countries, thus leading to some variation in the final design details and sample outcomes among countries. Furthermore, various statistical measures associated with these sampling outcomes suggest that following this design protocol has produced sample data whose estimates will be of acceptable quality though somewhat lower in precision than anticipated.

4.1 Some Lessons Learned

A number of important lessons have been learned in conducting GATS surveys to date. One is that success in using handheld technology for within-household sampling and respondent interviewing far exceeded expectations. Credit here is largely due to excellent coordinated efforts by project and in-country staff to install this substantial operational change. CDC staff and consultants from RTI International found the appropriate hardware, developed workable process code, and then effectively trained and supported in-country staff in using it, often for the first time. Moreover, country collaborators were highly receptive to this process change and worked diligently in transitioning from the pencil-and-paper mode of data collection that the majority of countries had been using previously in their surveys. In addition to streamlining questionnaire administration, data entry, and data editing, use of this technology also facilitated the creation of household rosters, randomly choosing the interview respondent, and storing selection probabilities for weights computation. In the end all GATS countries embraced this form of computer-assisted survey implementation.

Another insight gained from experience in the first set of GATS countries is that implications of integrating GATS sampling with another existing master or survey sample are mixed. On the one hand, integration offers resource efficiencies that can

translate into lower data collection costs. For example, since design integration typically implies the use of existing sampling frames and samples, some portion of the resources need for frame construction and sample selection are saved, thus reducing this component of sampling costs. Other cost reductions may arises in design integration with an existing survey sample if field interviewers in the other survey can also be used for GATS data collection, thus leading to savings in interviewer recruitment, training, and supervision. These types of savings were realized in Brazil.

Along with the potential cost advantages of design integration are potential statistical disadvantages. One is the dependence of the GATS sample on the design of the master sample or other survey sample with which the former is integrated. Not only is the composition of sampling strata determined by how strata are formed in the other design, but selection probabilities for the GATS sample are the product of the selection probabilities for the other design times the probabilities associated with whatever subsampling methods are used to select the GATS sample from the other sample. These types of dependencies can limit the final GATS sample if stratification for the other sample is not statistically helpful, or if subsampling complicates achieving desirable features for the GATS sample. For example, achieving selection probabilities indicative of PPS selection of PSUs may be difficult if selection of a master sample of PSUs does not lend itself easily to achieving the PPS probabilities.

A third lesson learned from the GATS experience thus far is that the coverage and general quality of the sampling frames, particularly the household frame in the penultimate stage of sampling, was difficult to gauge. The listings used for household selection were virtually all adapted from prior listings (e.g., from the most recent census or some other administrative source). None of the GATS countries to date have developed their household frames using conventional mapping and listing methods (see Kish, 1965; Section 9.6). Reliance on existing lists, often with little documentation as to their origins, made it difficult to assess coverage and other indicators of frame quality. These concerns about quality were at least partially offset in some countries where extensive efforts were made for GATS explicitly to update existing household lists in sampled areas (e.g., Bangladesh, Egypt). Due to time and cost constraints, no GATS participating country conducted a rigorous validation study of household listing by, for example, completing an independently adjudicated re-listing of a random subset of the selected area segments for which listings were used to choose its GATS household sample.

A fourth important insight from surveys done to date is that gender differences (typically higher for males than for females) are sufficiently great in some countries as to question the wisdom of the current GATS sampling protocol which calls for targeting equal numbers of male and female respondents. This lesson is especially important in countries where overall rates for key prevalence indicators are low, and much higher for males than for females, in which case somewhat larger respondent sample sizes would be needed for females to achieve the same relative precision (e.g., relative standard error) of prevalence estimates for both gender groups.

4.2 Future Directions

Lessons learned from prior experience in GATS will inform and can improve the quality of future efforts for two types of surveys that will emerge in coming years. For future GATS countries (six are currently are currently being considered) a key issue will be the

allocation of the sample by gender and urbanicity. As just noted, countries with substantial gender differences and relatively low overall prevalence rates may require that females be oversampled somewhat to produce similarly precise estimates by gender. This oversampling by gender may be best handled by modifying the randomization of households that is currently done to gender match interviewers and respondents. Similarly, if urban-rural differences become a lower priority for subgroup comparisons, future GATS samples might be made more nearly proportionate in allocation by urbanicity. GATS samples that are more proportionately allocated by urbanicity would have relatively smaller values of $Meff_w$, and thus of $Deff_c$, for most survey estimates.

For countries already having completed a round of GATS and considering another, priorities may shift from producing cross-sectional subgroup estimates for each round to estimating change in tobacco use between rounds as a way to evaluate the temporal effect of new ideas for controlling tobacco use. Since planned (as compared to random) overlap in samples over time can be an effective way to improve the statistical quality of estimated temporal change, another focus of design planning for followup rounds of GATS will be to examine the plausibility of ways to insert planned overlap into followup rounds.[3] One issue to resolve in considering the use of planned overlap is deciding which sampling stage, or stages, to create overlap. Generally speaking, the later the stage in which overlap is planned the greater the precision benefit for estimated temporal differences, since components of the variance of prevalence attributes tend to be greater in later stages (e.g., there is more variation among prevalence measurements for households within segments than for segments within PSUs, or among PSUs). Another practical consideration in the use of planned overlap is the ability to keep track of those sampling units used in the two rounds. So, for instance, if overlap is planned at the household level of sampling for the two rounds, one must be able to find the overlapping household from one round to another. This would require that information be obtained in the earlier round to improve the chances of finding the household later on, especially if they move between rounds.

It is anticipated that GATS will remain an international standard for consistently monitoring adult tobacco use and key tobacco control indicators that will in turn enable countries to develop effective public health interventions. Experience gained from developing a comprehensive standardized protocol for broadly quantifying adult tobacco use in populations and implementing it in14 countries (all within 3 years, an achievement in itself), together with the lessons learned from this experience, will inform further refinement of the protocol and processes to make it more efficient for countries that are interested to participate or repeat the survey in the future. Beyond the system of survey, a core subset of GATS questions have been identified for addition to other health surveys. While it is clear that there remain substantial challenges to providing high quality objective population data for these efforts, those who have created the GATS system of surveys have built a solid foundation to providing the quantitative information that will be required to control tobacco use around the world.

[3] "Planned overlap" occurs when a random portion of the sample in the earlier round is designated for membership in the later round, while "random overlap" happens when, by chance alone, some members of the sample in the earlier round are also chosen for the sample in the later round. Planned overlap can improve the precision of round-to-round difference, but random overlap has no such beneficial effect on precision.

Acknowledgements

The authors wish to recognize with sincere thanks our many esteemed colleagues, on staff at or serving as consultants to CDC, as well as numerous in-country colleagues from the ministries of health, national statistical offices, and their organizational partners in Bangladesh, Brazil, Egypt, Mexico, Philippines, Thailand, Turkey, and Uruguay, all of whom played pivotal roles in the development and implementation of GATS in those countries whose survey outcomes we have presented in this paper

References

Gabler, S., S. Haeder, and P. Lahiri. 1999. "A Model-Based Justification of Kish's Formula for Design Effects for Weighting and Clustering." *Survey Methodology,* 25(1):105-06.

Global Adult Tobacco Survey Collaborative Group. 2009. *Global Adult Tobacco Survey (GATS):* Egypt Country Report. <<http://www.emro.who.int/tfi/GATS_2009 /Annex_c.htm>> accessed on 18 Sept 2010.

Global Adult Tobacco Survey Collaborative Group. 2010a. *Global Adult Tobacco Surveys (GATS): Sample Design Manual.* <<www.cdc.gov/tobacco/global/gats>>.

Global Adult Tobacco Survey Collaborative Group. 2010b. *Global Adult Tobacco Surveys (GATS): Sample Weights Manual.* <<www.cdc.gov/tobacco/global/gats>>.

Global Tobacco Surveillance System (GTSS). 2010. Tobacco Questions for Surveys: A Subset of Key Questions from the Global Adult Tobacco Survey (GATS). CDC.

Kalton, G, and I. Flores-Cervantes. 2003. "Weighting Methods." *Journal of Official Statistics, 19*(2), 81-97.

Kish L., 1965. *Survey Sampling*. New York: Wiley

Kish, L., 1987. "Weighting in Deft^2", *The Survey Statistician*, June 1987.

Lessler, J.T. and W.D. Kalsbeek. 1992. *Nonsampling Error in Surveys*, Wiley and Sons, New York.

Mathers C.D, and D. Loncar D. 2006. "Projections of global mortality and burden of disease from 2002 to 2030". PLoS Medicine, 3(11):e442.