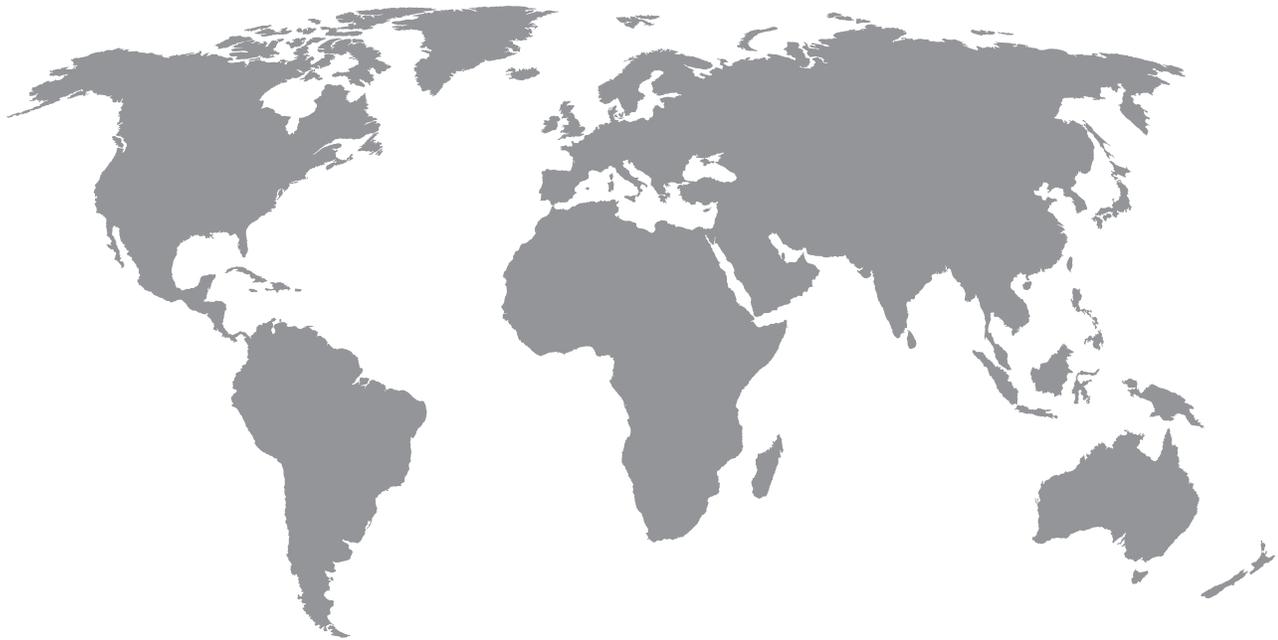




Global Adult Tobacco Survey (GATS)



Sample Weights Manual

Global Adult Tobacco Survey (GATS) Sample Weights Manual

September 2020

Global Adult Tobacco Survey (GATS) Comprehensive Standard Protocol

GATS Questionnaire

Core Questionnaire with Optional Questions
Question by Question Specifications

GATS Sample Design

Sample Design Manual
Sample Weights Manual

GATS Fieldwork Implementation

Field Interviewer Manual
Field Supervisor Manual
Mapping and Listing Manual

GATS Data Management

Programmer's Guide to General Survey System
Core Questionnaire Programming Specifications
Data Management Implementation Plan
Data Management Training Guide

GATS Quality Assurance: Guidelines and Documentation

GATS Analysis and Reporting Package

Fact Sheet Templates
Country Report: Tabulation Plan and Guidelines
Indicator Definitions

GATS Data Release and Dissemination

Data Release Policy
Data Dissemination: Guidance for the Initial Release of the Data

Suggested Citation

Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Sample Weights Manual*. Atlanta, GA: Centers for Disease Control and Prevention, 2020.

Acknowledgements

GATS Collaborating Organizations

- United States Centers for Disease Control and Prevention (CDC)
- CDC Foundation
- Johns Hopkins Bloomberg School of Public Health (JHSPH)
- RTI International
- World Health Organization (WHO)

Financial Support

Financial support was provided by the *Bloomberg Initiative to Reduce Tobacco Use* through the CDC Foundation with a grant from Bloomberg Philanthropies.

Disclaimer: The views expressed in this manual are not necessarily those of the GATS collaborating organizations.

Contents

Chapter	Page
1. Introduction	1-1
1.1 Overview of the Global Adult Tobacco Survey	1-1
1.2 Use of this Manual	1-2
2. Overview of Sample Weights in GATS.....	2-1
3. Recommended Approach.....	3-1
3.1 Background for the Illustrative Example	3-1
3.2 Base Weight.....	3-2
3.3 Adjustment for Unit Nonresponse	3-5
3.4 Post-Stratification Calibration Adjustment.....	3-11
4. Assuring the Quality of GATS Sample Weights.....	4-1
4.1 Careful Documentation of Sample Selection and Associated Probabilities.....	4-1
4.2 Control the Adverse Statistical Effect of Variable Weights	4-1
4.3 Carefully Follow the Recommended Computational Procedures and Quality Checks in this Manual	4-2
4.4 Prepare a Detailed Written Document Describing the Weights Computation Process that Is Actually Used	4-3
4.5 Consult with GATS Central Office Staff During the Weighting Process	4-3
5. Bibliography	5-1

1. Introduction

Tobacco use is a major preventable cause of premature death and disease worldwide, with approximately 1.4 billion people age 15 years or older using tobacco¹. Furthermore, more than 8 million people die each year due to tobacco-related illnesses². If current trends continue, tobacco use may kill a billion people by the end of this century, and it is estimated that more than three quarters of these deaths will be in low- and middle-income countries³. An efficient and systematic surveillance mechanism is essential to monitor and manage the epidemic.

The **Global Adult Tobacco Survey (GATS)**, a component of Global Tobacco Surveillance System (GTSS), is a global standard for systematically monitoring adult tobacco use and tracking key tobacco control indicators. GATS 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.

In order to maximize the efficiency of the data collected from GATS, a series of manuals has been created. These manuals are designed to provide countries with standard requirements as well as several recommendations on the design and implementation of the survey in every step of the GATS process. They are also designed to offer guidance on how a particular country might adjust features of the GATS protocol in order to maximize the utility of the data within the country. In order to maintain consistency and comparability across countries, following the standard protocol is strongly encouraged.

GATS manuals provide systematic guidance on the design and implementation of the survey.

1.1 Overview of the Global Adult Tobacco Survey

GATS is designed to produce national and sub-national estimates among adults across countries. The target population includes all non-institutionalized men and women 15 years of age or older who consider the country to be their usual place of residence. All members of the target population will be sampled from the household that is their usual place of residence.

GATS uses a geographically clustered multistage sampling methodology to identify the specific households that Field Interviewers will contact. First, a country is divided into Primary Sampling Units, segments within these Primary Sampling Units, and households within the segments. Then, a random sample of households is selected to participate in GATS.

The GATS interview is composed of two parts: *Household Questionnaire* and *Individual Questionnaire*. These questionnaires are administered using an electronic data collection device.

¹ World Health Organization. WHO report on the global tobacco epidemic, 2019: Offer help to quit tobacco use. Geneva, Switzerland: World Health Organization; 2019. <https://apps.who.int/iris/bitstream/handle/10665/326043/9789241516204-eng.pdf?ua=1>

² GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Seattle, WA: Institute for Health Metrics and Evaluation; 2018.

³ Mathers, C.D., and Loncar, D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine*, 2006, 3(11):e442.

The GATS interview consists of two parts: the *Household Questionnaire* and the *Individual Questionnaire*. The *Household Questionnaire* (household screening) and the *Individual Questionnaire* (individual interview) will be conducted using an electronic data collection device.

At each address in the sample, Field Interviewers will administer the *Household Questionnaire* to one adult who resides in the household. The purposes of the *Household Questionnaire* are to determine if the selected household meets GATS eligibility requirements and to make a list, or roster, of all eligible members of the household. Once a roster of eligible residents of the household is completed, one individual will be randomly selected to complete the *Individual Questionnaire*. The *Individual Questionnaire* asks questions about background characteristics; tobacco smoking; electronic cigarettes; smokeless tobacco; cessation; secondhand smoke; economics; media; and knowledge, attitudes, and perceptions about tobacco.

1.2 Use of this Manual

This manual is designed to offer the requirements for countries to follow as they compute the sample weights for GATS. This first chapter provides background information on GATS. The subsequent chapters are summarized below:

- *Chapter 2* defines a sample weight, indicates when in the survey process weights are produced, gives an overview of the rationale behind the computational process, and describes how weights are used by the data analyst.
- *Chapter 3* describes in some detail each of the steps of the recommended approach to computing weights. Each of these steps is illustrated with an example based on the type of respondent one might find in a GATS sample.
- *Chapter 4* recommends several steps to assure that high quality weights are produced.

It should be noted that while this manual offers a step-by-step template for the computation of survey weights, modifications to this computational template may be needed due to country-specific circumstances regarding the sample design, analysis needs, and the availability of ancillary data to compute adjustments. Collaboration between in-country statisticians, CDC country focal points, and the Sample Review Committee (SRC) is essential to ensuring the quality of survey sample weighting and adjustment.

2. Overview of Sample Weights in GATS

A *sample weight* is a statistical measurement linked to a data record for any survey respondent in population samples fully utilizing random selection methods to choose the sample. In general terms, an individual sample weight is usually simply the inverse of the adjusted probability of obtaining the data for the respondent. In most cases, this probability is simply the respondent's original selection probability based on the sample design. The inverse probability, or base weight (B_j), is usually adjusted to account for unintended sample imbalance arising during the process of conducting the survey⁴. More than one weight adjustment may be applied. All are multiplicative. Unless a weight is rescaled for analytic purposes (e.g., "calibrated" to sum to the total population covered), its value can be interpreted as an indication of the number of population members represented by the respondent. Separate sets of weights may be necessary when data to be analyzed are gathered for different types of data items or units of analysis associated with the respondent. For example, if data in a household survey are gathered for the selected households, and for one resident who is chosen at random in each of those households, separate sets of weights would be produced for the household and for the resident data. However, since only person-level data are to be processed in GATS, only person-level weights will be required.

Whereas the general statistical rationale behind the use of weights for estimation from population samples is well established (Horvitz and Thompson, 1952), no universally held protocol exists for computing them. This is partly because of variation in circumstances from sample to sample regarding its design, the quality of documentation of the sample selection and recruitment processes, and the availability of ancillary information about the sample and the population to identify and deal with imbalance in the sample due to differential frame coverage and nonresponse. Thus, the actual computational steps for a sample weights can vary among surveys. However, some combination of the steps listed below is typically followed in producing a weight for each (i.e., the *j*-th) individual respondent data record, with the final adjusted weight (W_j) being the product of the value generated in each step, as described in more detail below:

1. Determine the base weight (B_j) to account for all steps of random selection that led to the sample of population members,
2. Adjust for nonresponse (A_j^{nr}) to compensate for sample imbalance due to differential success in sample recruitment,
3. Further adjust for incomplete frame coverage (A_j^{cov}) to adjust for imbalance due to a sampling frame that does not fully cover the population targeted for study, and
4. Further adjust to calibrate (A_j^{cal}) the final set of adjusted weights to the distribution of the population by characteristics that are highly correlated with the key study outcome measures (i.e., tobacco use behavior in GATS).

⁴ The concept of "imbalance" here is simply to indicate that the demographic representation of the sampled population is skewed somewhat by forces related to the sample selection and recruitment phases of the study. In other words, the otherwise representative sample that random selection produces become somewhat less representative of the population.

The actual sequence of steps followed in producing the set of statistical weights for the population sample is important since the weights produced through any given step depend on the computational outcomes of each prior step. The final weight for any respondent is the product of the computational outcomes for the respondent from all the steps in order:

$$W_j = B_j A_j^{(nr)} A_j^{(cov)} A_j^{(cal)} \quad (1)$$

To understand the logic behind the process of producing sample weights for survey samples, where residents of selected households are members of the target population, first consider that a population member in GATS can only provide data if all three of the following events (i.e., F, S, and R) occur:

- **Event F:** The member's household and the member are included on the household and person-within-household sampling frames used for GATS sampling;
- **Event S:** Given Event F, the member's household and the member must be randomly selected for participation in GATS; and
- **Event R:** Given Event S, the member's household and the member must, in turn, agree to become a GATS participant and complete a valid questionnaire.

The probability that the member's data are used for sample estimation is the product of the probabilities of observing these three events. We will see later that completing Steps (1), (2), and (3) requires that one determine or estimate the probabilities of Events F, S, and R. Sample calibration (or "post-stratification") serves to control loss in the sample's external validity due to forces not specifically accommodated by the nonresponse and coverage adjustments in Steps (2) and (3). These forces could be random variations in the sample's demographic composition with respect to variables not used to define sampling strata, as well as differential nonresponse and under-coverage associated with variables other than those that define the adjustment cells used to produce the adjustments in Steps (2) and (3). Since it is often impractical to do Step (3), adjustment for incomplete frame coverage is usually accomplished as a part of calibration.

Sample weights are important for many analysis situations. For instance, they are used in producing point estimates of population characteristics (e.g., a current smoking prevalence rate) as well as in estimating the variance of those characteristics, although other design features of the sample (i.e., stratification, cluster sampling, without-replacement selection) are also important for the latter. Thus, weights are important for many data uses, including descriptive analysis and comparative testing (e.g., for significant differences in prevalence rates among regions of a country). They are also sometimes used for regression modeling (e.g., to identify predictors of smoking behavior). Weights, therefore, are often a necessary but not always sufficient design feature to use in analysis of survey data. For example, if one only uses weights, but ignores stratification and cluster sampling in estimating the variance of survey estimates, reported variances and conclusions from tests of hypothesis may very well be incorrect.

One set of weights is produced for each set of observation units in the study. Studies often have multiple sets of these observational units. For example, if the study's data collection plan calls for gathering information for households (e.g., income, socioeconomic status, distance from health providers) and individual members of those households (e.g., smoking behavior, demographics, etc.), a separate set of weights may be needed for each observational unit (i.e., households and persons). Moreover, the weights may differ for households and their resident respondents if sampling is done within household to choose

the individuals providing the person-level data. This kind of difference would be true in GATS if both household-and person-level data are produced since one eligible household resident is selected within each participating household.

Where available, master samples provide countries with preselected samples of geopolitical clusters to meet the survey needs of different organizations (see *Section 3.3* of the *GATS Sample Design Manual*). Master samples of geographical units are then subsampled to provide GATS with clusters of households to conduct household listing and selection in accordance with the *GATS Mapping and Listing Manual*. The GATS sample design must correspond with that of each country's master sample and probabilities of selection of master sample units must be readily available at the time of the GATS selection. The probability of selecting the GATS cluster is then the product of the initial probability of selecting the master sample unit and the probability of selecting the GATS cluster of households. As the design of subsampling from the master sample varies across countries, close coordination between country statistical offices, country GATS coordinators, CDC country focal points and statisticians, and members of the SRC is essential in order to ensure that probabilities of selection and corresponding weights are computed accurately. The following computations in this manual will assume that master samples were not used but rather geopolitical clusters were sampled directly from a frame of all such units in the country, often provided by country census statistical offices. Those countries relying on master samples for first stage selections must incorporate probabilities of selecting master sample units into selection probabilities of GATS clusters.

The process of producing weights typically occurs after data collection and once the data have been processed and cleaned for the analyst. They cannot be generated until after field work is completed since they are applied to the final sample of respondents and computing them relies on final outcome information from data collection. They must also be completed before analysis can begin since the analysts using the survey's data will need them. Technical staff involved in choosing the sample and supporting the field work are usually best equipped to produce the weights because of their knowledge about the target population and how the sample was drawn and recruited. As needed, GATS central staff can provide assistance and support to the weights computation process. Finally, to assure the quality of weights, the weights computation process is fully reviewed by external experts to ensure that appropriate procedures are followed (See *Chapter 5* of the *GATS Quality Assurance: Guidelines and Documentation*).

3. Recommended Approach

This chapter of this manual sets out in some detail the three-step approach corresponding to the major components that are recommended for in-country technical staff follow in producing weights for their GATS sample. These steps are: compute a base weight for each sample respondent, adjust the base weights for differential nonresponse in the sample, and calibrate the adjusted weights to known population totals. More general discussion of each of these steps is found in reviews by Lessler and Kalsbeek (1992) and Kalton and Flores-Cervantes (2003). Our approach in presenting these steps here is to integrate a realistic GATS-like example into the discussion. The discussion of each of the three steps begins by listing what is needed to complete the step and then presents the specific formulae that are recommended to compute that component of the weight as part of the illustrative example. Alternatives to some recommendations are offered for consideration by GATS country staff, although choosing these alternatives must be fully justified and made with full collaboration of CDC country focal points and SRC before proceeding with them.

RECOMMENDATION:

The following three-step approach is recommended for each GATS country sample: (1) compute a base weight for each sample respondent, (2) adjust the base weights for nonresponse, and (3) calibrate the adjusted weights to known population totals

3.1 Background for the Illustrative Example

The example used to illustrate the computational process for GATS weights focuses on the problem of determining the final adjusted sample weight for a fictitious individual GATS respondent (referred to in the sequel as, \mathcal{R}) in a country sample where gender randomization was not required. We presume that the sample design that led to selecting this person is a stratified three-stage household sample with some recognized geo-political area unit of variable size as the primary sampling unit (PSU), an approximately equal-sized area “segment” consisting of roughly 200 households as the secondary sampling unit (SSU), and individual households selected within sampled segments. In this example, we assume that there were no missed households in constructing lists of households within \mathcal{R} 's segment and that gender allocation was not part of the selection process of \mathcal{R} ; but in actual practice these issues might need to be addressed (The appropriate methods for addressing those issues are included later in this manual)⁵. Finally, we assume that \mathcal{R} is chosen from a household roster that was entered into a handheld computer, which is also used to complete the GATS interview with \mathcal{R} . Other assumed details surrounding the selection and recruitment of \mathcal{R} will be mentioned in describing each step.

ILLUSTRATIVE EXAMPLE:

Background on fictitious GATS respondent (\mathcal{R})

⁵ Subselection of missed households is needed when households not included on the household list frame are discovered through a missed HU procedure (e.g., the “half-open interval”) and the number of discovered households is sufficiently large to require random subselection of unlisted households. Gender allocation is to be done in GATS samples when for cultural reasons countries wish to match interviewers with the gender of selected respondents and will use same sex interviewer teams, or when a gender group must be oversampled for statistical reasons. Each sampled household is randomly assigned to be a “male household” (only eligible male residents rostered for respondent selection) or a “female household” (only eligible female residents rostered).

3.2 Base Weight

3.2.1 Needed to Complete This Computational Step

- Well-documented selection probabilities for each stage of the process of sample selection if, as preferred, they have already been computed.

OR

- Specific knowledge of how to compute the probabilities corresponding to the random selection methods used in each step of the sampling process. Selection steps include: for sampling stages, gender allocation of households (if required), and sub-selection of households discovered through a missed household procedure if applicable.
- Selection worksheets and/or computer code for selection software for each step of the process of sample selection, *if* selection probabilities must be determined at the time that weights are computed.

3.2.2 Computing Base Weights with Illustrative Example

The base weight of a respondent in any probability sample is simply 1 divided by the overall selection probability for the respondent given the steps completed in selecting the respondent. Calculating the base weight for a GATS respondent thus requires that one answers the question: what was the statistical probability that the sample design would lead to the selection of the respondent?

The *GATS Sample Design Manual (Chapter 11)* describes the following relevant components of the overall selection probability when the country follows the recommended multi-stage sampling approach. The subscripts α and k (jointly for the α -th PSU and k -th SSU, respectively) in this description jointly corresponding to the “area” (segment) g , which is chosen in two sampling stages in selecting \mathcal{R} :

- $p_{\alpha k}^{(1)}$ = Unconditional probability of selecting the α -th PSU (geo-political area unit in which \mathcal{R} lives) and k -th SSU (segment in which \mathcal{R} lives);
- $p_{\alpha k}^{(2)}$ = Conditional probability (given PSU and SSU selections) of selecting the household in which \mathcal{R} lives;
- $p_{\alpha k i}^{(3)}$ = For a fictitious GATS respondent (\mathcal{R}), this probability=1, However, conditional probability (given PSU, SSU, and household selections) of randomly assigning \mathcal{R} 's household to be a “female” household, is required if gender randomization used;
- $p_{\alpha k i}^{(4)}$ = Conditional probability (given PSU, SSU, household selections and gender allocation) of randomly selecting \mathcal{R} 's household, if it had not been on the original household frame for the αk -th segment and was discovered as part of a missed household procedure, otherwise this probability=1; and
- $p_{\alpha k i j}^{(5)}$ = Conditional probability (given PSU, SSU, household selections, gender allocation, and missed household selection) of randomly selecting \mathcal{R} from a roster of eligible residents of \mathcal{R} 's household.

Notice that each of the selection events corresponding to these probabilities must occur in order for \mathcal{R} to be selected in the GATS sample. Then the unconditional joint probability of selecting \mathcal{R} (the $\alpha k i j$ -th person) into the GATS sample is:

$$p_{\alpha k i j} = p_{\alpha k}^{(1)} \cdot p_{\alpha k}^{(2)} \cdot p_{\alpha k i}^{(3)} \cdot p_{\alpha k i}^{(4)} \cdot p_{\alpha k i j}^{(5)} \quad (2)$$

since the probability of joint sequential events is the unconditional probability of the first event in the sequence times the conditional probabilities of each subsequent event given the outcome of the prior events in the sequence. Thus, the associated base weight for \mathcal{R} is:

$$B_{\alpha k i j} = \frac{1}{p_{\alpha k i j}} = \frac{1}{p_{\alpha k}^{(1)} \cdot p_{\alpha k}^{(2)} \cdot p_{\alpha k i}^{(3)} \cdot p_{\alpha k i}^{(4)} \cdot p_{\alpha k i j}^{(5)}} \quad (3)$$

As seen in Eq. (3), we must determine each of the components of unconditional joint probability of selecting \mathcal{R} in order to compute \mathcal{R} 's base weight. Assuming that some form of without-replacement probability proportional to size (PPS) sampling is used to select PSUs within first stage sampling strata, then if $N_i = 2,462$ is the size measure (in number of households as of the last census) for \mathcal{R} 's PSU, $I = 2$ is the number of PSUs chosen in the sampling stratum from which \mathcal{R} 's PSU was chosen, and the sum of size measures for all PSUs in that stratum is $\sum_{\alpha} N_{\alpha} = 338,754$, the unconditional selection probability for \mathcal{R} 's PSU will be,

$$p_{\alpha}^{(1)} = \frac{I \cdot N_{\alpha}}{\sum_{\alpha} N_{\alpha}} = \frac{(2) \cdot (2,462)}{338,754} = 1.4536 \times 10^{-2} \quad (4)$$

If \mathcal{R} 's SSU is one of $K_{\alpha} = 2$ segments chosen by (without-replacement) simple random sampling from $S_{\alpha} = 12$ segments in \mathcal{R} 's PSU, then the conditional probability (given PSU selection) of selecting \mathcal{R} 's SSU is,

$$p_{k(\alpha)}^{(1)} = \frac{K_{\alpha}}{S_{\alpha}} = \frac{2}{12} = \frac{1}{6} \quad (5)$$

and the unconditional joint probability of selecting \mathcal{R} 's PSU and \mathcal{R} 's SSU is,

$$p_{\alpha k}^{(1)} = p_{\alpha}^{(1)} p_{k(\alpha)}^{(1)} = \left[\frac{I \cdot N_{\alpha}}{\sum_{\alpha} N_{\alpha}} \right] \cdot \left[\frac{K_{\alpha}}{S_{\alpha}} \right] = (1.4536 \times 10^{-2}) \cdot (1/6) = 2.4226 \times 10^{-3} \quad (6)$$

The *GATS Sample Design Manual* describes two methods of selecting households from selected segments or PSUs. If systematic sampling is used to select households, the conditional probability of selecting each household is $1/K$ where K is the selection interval. If simple random sampling is used to select households, this probability is the ratio of household sample size in each segment to the total number of households on the frame list for each segment. Assuming that $H_{\alpha k} = 28$ households are selected by without-replacement simple random sampling from $L_{\alpha k} = 212$ listed households in the αk -th segment in which \mathcal{R} 's household is located, then the conditional probability of selecting \mathcal{R} 's household is,

$$p_{\alpha k}^{(2)} = \frac{H_{\alpha k}}{L_{\alpha k}} = \frac{28}{212} = 0.13208 . \quad (7)$$

Since the fictitious respondent is found in a country where no gender randomization was needed, the respondent selection probability,

$$p_{\alpha k i}^{(3)} = 1 \quad (8)$$

Note that in general, when gender assignment of households is required, and $M_{\alpha k}$ and $F_{\alpha k}$ are, respectively, the numbers of $H_{\alpha k} = M_{\alpha k} + F_{\alpha k}$, selected households in the αk -th “area” (segment) assigned to be “male” and “female,” then for male respondents,

$$p_{\alpha k i}^{(3)} = \frac{M_{\alpha k}}{H_{\alpha k}} \quad (9)$$

and for female respondents,

$$p_{\alpha k i}^{(3)} = \frac{F_{\alpha k}}{H_{\alpha k}} \quad (10)$$

In general if a respondent is selected in a household that was discovered by means of a missed household procedure (e.g., half-open interval) and a sampling rate ($f_{\alpha k i}$) is applied in sub-selecting the household by simple random sampling, then the sub-selection probability for any (i.e., the $\alpha k i$ -th household chosen this way will be,

$$p_{\alpha k i}^{(4)} = f_{\alpha k i} . \quad (11)$$

Since we assume that \mathcal{R} 's household was not selected as part of a missed household procedure,

$$p_{\alpha k i}^{(4)} = 1 . \quad (12)$$

For the specific case of determining \mathcal{R} 's within-household selection probability, we note that the respondent was chosen at random among the $R_{\alpha k i} = 4$ members of the respondent's household that were included on the household's roster. Thus, for the household,

$$p_{\alpha k i j}^{(5)} = \frac{1}{R_{\alpha k i}} = \frac{1}{4} = 0.25 \quad (13)$$

Combining all of the probabilities for the selection process that led to choosing \mathcal{R} , we have in summary then that \mathcal{R} 's unconditional overall selection probability is

$$p_{\alpha k i j} = p_{\alpha k}^{(1)} \cdot p_{\alpha k}^{(2)} \cdot p_{\alpha k i}^{(3)} \cdot p_{\alpha k i}^{(4)} \cdot p_{\alpha k i j}^{(5)} = (2.4226 \times 10^{-3}) \cdot (0.13208) \cdot (1) \cdot (1) \cdot (0.25) = 7.9992 \times 10^{-5} , \quad (14)$$

and the respondent base weight is,

$$B_{\alpha k i j} = \frac{1}{p_{\alpha k i j}} = \frac{1}{7.9992 \times 10^{-5}} = 12501.3081 , \quad (15)$$

and that the base weight for the household is

$$B_{aki} = p_{aki}^{(5)} / p_{akij} = (0.25) \cdot (125013081) = 3125.2370 \quad (16)$$

Note that $B_{\alpha k i j}$ is denoted as B_j in the later sections of the manual.

3.3 Adjustment for Unit Nonresponse

3.3.1 Needed to Complete This Computational Step

- The base weight for each GATS respondent.
- The final recruitment disposition (e.g., responding, refusal, not at home, etc.) **for all selected households** to use in calculating household response rates among eligible sample households in each sample “segment.” This means that counts of the number of selected, study eligible and participating households will be needed for each sample segment. For those sample designs that select households directly in first stage sampling units, final recruitment dispositions in each PSU will be used for household nonresponse calibration.
- The final recruitment disposition as well as gender, age, and smoking status (current smoker or not) information from the household rosters **for all selected residents** of all participating households. Additional information available from PSU selection such as rural/urban and region can also be used for nonresponse adjustment. This ancillary information about the selected sample will be used to produce response rates by selected weighting class variables. It is essential that all variables used for response rate calculations be non-missing. If smoking status is missing or unknown it should be imputed for that roster member to the status of “nonsmoker” for person-level nonresponse weighting adjustment.

3.3.2 Computing Nonresponse Adjustments with an Illustrative Example

The nonresponse bias of estimates of simple population characteristics like means, totals, and proportions is based solely on respondent data and the base weight is partly determined by the member-level covariance in the population ($\sigma_{\pi y}$) between the propensity (i.e., probability in a stochastic sense) of the j -th individual member to respond (π_j) and the survey measurements (y_j) for what is being estimated from the survey data (Lessler and Kalsbeek, 1992)⁶. Notice that we temporarily consider all individuals to be members of the population at large, and thus drop the label for the household (i) of which the individual is a member.

For (target) populations of size N (assumed known) we can write the bias due to nonresponse of an

unadjusted estimate ($\hat{y}_r = \frac{\hat{t}_r}{N} = \frac{\sum_{j=1}^N s_j r_j y_j / p_j}{v}$) of the population mean

($\bar{Y} = t / N$) as,

⁶ The term “propensity” rather than “probability” is used in connection with survey nonresponse, since presumed stochastic behavior, rather than explicit randomization, determines the outcome of the process determining whether or not a member of a survey sample responds.

$$Bias(\hat{y}_r) = \frac{\sigma_{\pi y}}{\bar{\pi}}, \quad (17)$$

where for the j -th member of the population, s_j is the 0/1 indicator for Event S, p_j is the overall selection probability (i.e., $Pr\{s_j = 1\}$), r_j is the 0/1 indicator for Event R, and $\bar{\pi}$ is the mean of all response propensities (π_j) in the population. Adjusting the base weight of a survey respondent for nonresponse requires an estimate of the response propensity for the respondent. The inverse of (i.e., 1 divided by) this estimated propensity becomes the adjustment for nonresponse, which is multiplied times the base weight from the prior step to produce a nonresponse-adjusted weight for the respondent.

More precisely, adjusting sample weights for nonresponse requires that each value of π_j must be estimated empirically based on the nonresponse experience in the sample. The member-level multiplicative adjustment for nonresponse is then simply the reciprocal of the estimated response propensity for the respondent:

$$A_j^{(nr)} = \hat{\pi}_j^{-1} \quad (18)$$

and the nonresponse-adjusted weight is,

$$W_j^{(2)} = B_j A_j^{(nr)} \quad (19)$$

A key issue in producing this adjustment then is how to estimate individual response propensities. Weighting class response rates and predicted response propensities from a fitted logistic model are two approaches that have been used to adjust for unit nonresponse. Since the weighting class approach is simpler to implement and the predicted propensity approach has not generally been found to be superior in its ability to control (**not** eliminate) nonresponse bias, the recommended approach for GATS is the weighting class adjustment.

Estimating propensities is made a bit more complicated in GATS since nonresponse can occur at both the household and person level of sampling in respondent recruitment. The combined response propensity for the j -th respondent (π_j) therefore has two multiplicative components to be estimated separately, one to accommodate the respondent's household's propensity to respond by completing the household roster ($\pi_j^{(HH)}$) and the other to reflect the respondent's propensity to respond by completing the GATS interview, once the respondent is chosen from the household roster ($\pi_j^{(person)}$).

Under the recommended approach to adjust for nonresponse, each propensity component is estimated for the j -th respondent as the response rate for members of the subgroup (i.e., "class") of selected sample members with similar characteristics and response tendencies as the respondent. Computation of the rate of "response" for the household and person-level components should follow the guidelines for weighted or unweighted versions of response rate RR1 as defined by AAPOR (American Association of Public Opinion Research, 2009) and presented in the *GATS Sample Design Manual*. Weighted response rates using base weights are preferable to unweighted response rates. Ultimately, it is the prerogative of

countries to decide which to use. Using household disposition codes, the household response rate computed separately for each sample segment (see description below) is computed as

$$\text{Household-level response rate} = \frac{[1]}{[1] + [3] + [4] + [5] + [6] + [9]} \quad (20)$$

where

- 1 = Completed Household Questionnaire, One Person Selected
- 2 = Completed Household Questionnaire, No One Selected
- 3 = Completed Part of Household Questionnaire, Could Not Finish Roster (Incomplete Interview)
- 4 = Household Questionnaire Not Complete, Could Not Identify Appropriate Screening Respondent
- 5 = Nobody Home
- 6 = Household Refusal
- 9 = Other Household Nonresponse.

Note that the household final disposition code 2 is excluded from both numerator and denominator of the household-level response rate since these households are considered ineligible, whether or not gender randomization of households is done.

The household-level response rate above excludes ineligible households from the denominator, and it assumes that all of the selected households with unknown eligibility (final disposition codes 3, 4, 5, 6, 9) are eligible to participate in GATS. This may lead to underestimates of household-level response rates if households of unknown eligibility are really ineligible. It is recommended that countries estimate this proportion (e) by calculating the known eligibility rate, which is the known eligibles (disposition code 1) divided by the known eligibles plus the known ineligibles (disposition code 2):

$$e^{\text{household-level}} = \frac{[1]}{[1] + [2]} \quad (21)$$

If this proportion for the sample as a whole is less than 0.90, countries should adjust the unknown component of the household-level response rate by multiplying unknowns (final disposition codes 3, 4, 5, 6, 9) by this proportion (e). The following formula conforms to the AAPOR (2009) response rate RR3:

$$\text{Household-level response rate} = \frac{[1]}{[1] + e^{\text{household-level}} [3 + 4 + 5 + 6 + 9]} \quad (22)$$

The corresponding household-level weighting class adjustment ($A_{hi}^{(nr, HH)}$) would be computed as one divided by the weighted household response rate for each sample segment.

The person-level response rate is computed within strategically formed subgroups (see description below) as:

$$\text{Person - Level } RR = \frac{[11]}{[11] + [12] + [14] + [15] + [16] + [17]} \quad (23)$$

where

- 11 = Completed Individual Questionnaire
- 12 = Incomplete Interview
- 13 = Selected Individual Was Later Determined to Be Ineligible for GATS
- 14 = Selected Respondent Not Home
- 15 = Selected Respondent Refusal
- 16 = Selected Respondent Incompetent
- 17 = Other Individual Nonresponse.

The person-level response rate above excludes ineligible individuals (final disposition code 13) from the denominator, and it assumes that all of the selected individuals with unknown eligibility (final disposition code 14) are eligible to participate in GATS. This may lead to underestimates of person-level response rates if interviewers often select respondents who are found to be ineligible for the survey once the interview begins. It is recommended that countries estimate the proportion of those respondents selected from the roster who are truly eligible to respond to the GATS survey (e) using weighted dispositional code frequencies as:

$$e^{\text{person-level}} = \frac{[11] + [12] + [15] + [16] + [17]}{[11] + [12] + [13] + [15] + [16] + [17]} \quad (24)$$

If this proportion for the sample as a whole is less than 0.90, countries should adjust the unknown component of the person-level response rate by multiplying unknowns (final disposition code 14) by this proportion (e). The following formula conforms to the AAPOR (2009) response rate RR3:

$$\text{Person-level response rate} = \frac{[11]}{[11] + [12] + [e^{\text{person-level}} [14]] + [15] + [16] + [17]} \quad (25)$$

The person-level adjustment component ($A_{hj}^{(nr, \text{person})}$) of the weighting class adjustment for the j -th person in the i -th household would be computed as one divided by the weighted person-level response rate.

Adjustments based upon small subgroup or cluster sizes may suffer from considerable variation and yield excessively large adjustment values. Therefore, an upper bound of 3.00 will be set on all computed household and person-level weighting class adjustments. Values larger than 3.00 in either component of the adjustment weight will be capped at 3.00.

RECOMMENDATION:
Any household- or person-level nonresponse adjustment components that exceed 3.00 should be set to 3.00.

Finally, the corresponding household-level components ($A_{hi}^{(nr, HH)}$) and person-level component ($A_{hj}^{(nr, person)}$) of the adjustment for the j -th person in the i -th household should be multiplied to produce the combined nonresponse adjustment for that respondent ($A_j^{(nr)}$).

The choice of characteristics to use in defining the subgroupings (i.e., the “weighting classes”) for each component is strategically important, since bias reduction associated with this approach is directly related to the extent of correlation between the response rate and the parameter of interest for these classes (Kalton, 1983). Weighting classes in most situations are defined by the cross-classification of several categorical variables (e.g., or continuous variables that have been categorized, such as age in individual years being divided into several categories defined by age groupings). These subgroupings might be sampling strata, sample clusters, or defined by other information known for all selected members of the sample. For GATS samples, we recommend that the weighting classes for the household-level component of the nonresponse adjustment be defined by the set of selected households within sample “segments.” For those countries that have selected geographical clusters for direct household listing in one stage, the sample PSUs will be the weighting classes for the household component. For the person-level component of the nonresponse adjustment we recommend that weighting classes be separately defined for each region for which GATS estimates and 8,000 (4000 as applicable) respondents are required. Furthermore, region by a combination or all of the following person-level weighting class variables should be used for adjustment: urban/rural and the roster-reported age (15-24, 25-34, 35-44, 45-54, 55+), gender (male, female) and current smoking status (smoking, not smoking). In the event that no regional estimates are to be made, weighting classes for the person-level components can be formed by a combination of urban/rural, gender, age, and smoking status variables.

RECOMMENDATION:

Form weighting class cells for person-level component of the nonresponse adjustment by roster-reported gender, age, and current smoking status; also by region if quality regional estimates are needed.

Once weighting classes have been formed for each component, the two multiplicative contributions to the weighting class adjustment for GATS would be computed for \mathcal{R} as follows. Numerators and denominators in the examples below are sums of base weights. For the household component, if all of the $H_{\alpha} = 56$ selected households in the α -th PSU in which \mathcal{R} 's household is located are eligible for the study and the sum of the base weights among these households is 170,013.3, and $r_{\alpha}^{(HH)} = 50$ of them agree to participate in the study (weighted total among them = 156,261.9), then the household component of the weighted nonresponse adjustment would be computed from the weighted response experience for the selected household sample in the PSU or segment in which \mathcal{R} 's household is located. For this example, suppose we have found that $e^{household-level}$ exceeded 90% enabling us to use formula 20 above. The household-level weighted response rate of the weighting class for \mathcal{R} (residing in the i -th household) is

$$RR_h^{(weighted, HH)} = \hat{\pi}_{hi}^{(wca, HH)} = \frac{\sum_{i=1}^{r_{\alpha}^{(HH)}} B_{hi}}{\sum_{i=1}^{H_{\alpha}} B_{hi}} = \frac{156,261.9}{170,013.3} = 0.9191 \quad (26)$$

The household-level component of the weighting class adjustment for \mathcal{R} (residing in the i -th household) would then be computed as

$$A_{hi}^{(nr,HH)} = \frac{1}{[\hat{\pi}_{hi}^{(wca,HH)}]} = \frac{1}{\sum_{i=1}^{H_{hi}} B_{hi} / \sum_{i=1}^{H_{hi}} B_{hi}} = \frac{1}{0.9191} = 1.088 \quad (27)$$

where the B_{hi} are variable base weights for the households within the h -th weighting class.

If less than 90% of households providing eligibility information were found eligible to participate in GATS countrywide, the unknown component of the household eligibility would have been multiplied by (e) yielding slightly higher response rates and correspondingly lower household adjustments.

The household adjustment component (1.088) is less than 3.00 so there is no need to limit the value of the adjustment.

The weighted person-level response rate should be computed using (23) above if the proportion (e) of selected individuals who are eligible to complete the GATS questionnaire is 0.90 or more. If (e) is less than 0.90, formula (25) should be used to compute the person-level response rate. As with the household adjustment component, the person-level adjustment component for \mathcal{R} is one divided by the weighted response rate for \mathcal{R} 's weighting class. Now suppose that \mathcal{R} is an urban, female smoker in the Region X, and she is one of $r_h = 680$ respondents (weighted number=119,009,025) among the $n_h = 771$ household residents (weighted value= 134,935,233) in the person-level weighting class consisting of all selected household residents of Region X who are urban, female and current smokers. Suppose, furthermore, that the weighted response rate for \mathcal{R} 's weighting class is

$$RR_h^{(weighted, person)} = \hat{\pi}_{hj}^{(wca, person)} = \sum_{j=1}^{680} B_{hj} / \sum_{j=1}^{771} B_{hj} = \frac{119,009,025}{134,935,233} = 0.8820 \quad (28)$$

The person-level adjustment component for \mathcal{R} is thus computed as

$$A_{hj}^{(nr, person)} = \frac{1}{[\hat{\pi}_{hj}^{(wca, person)}]} = \frac{1}{\sum_{j=1}^{680} B_{hj} / \sum_{j=1}^{771} B_{hj}} = \frac{1}{0.8820} = 1.1338 \quad (29)$$

The person-level adjustment component is also less than 3.00, therefore no trimming is necessary to cap the value. The final nonresponse adjustment for \mathcal{R} is

$$A_j^{(nr)} = \frac{1}{\pi_j^{(wca)}} = \frac{1}{[\hat{\pi}_i^{(wca, HH)}][\hat{\pi}_j^{(wca, person)}]} = [A_i^{(nr, HH)}][A_j^{(nr, person)}] = [1.0880] \cdot [1.1338] = 1.2336 \quad (30)$$

Recalling the value of the base weight and the final nonresponse adjustment for \mathcal{R} , the nonresponse-adjusted sample weight for \mathcal{R} is computed as

$$W_j^{(nr)} = [B_j] \cdot [A_j^{(nr)}] = [12501.3081] \cdot [1.2336] = 15421.61 \quad (31)$$

3.4 Post-Stratification Calibration Adjustment

3.4.1 Needed to Complete This Computational Step

- Population frequency counts, from census counts conducted within 5 years of the interview or other reputable source of current population data, of persons 15 years of age or older jointly by categorical variables related to smoking behavior and remaining sample imbalance (e.g., coverage). Possible calibration variables include gender, education, age, urban/rural and region, if regional population counts are available and respondent sample sizes are large enough at the regional level (8,000 recommended) to produce quality regional estimates.
- Comparably worded questions on and question response categories for respondent gender, and education questions in the GATS questionnaire.
- The nonresponse adjusted weights ($W_j^{(nr)}$) for all sample respondents

Computing Post-Stratification Adjustments with Illustrative Example:

While the two types of nonresponse adjustments just described are effective in offsetting sample imbalance due to the variables that are used to define weighting classes and model response propensities, there may be other important sample characteristics for which no adjustment has been made. For example, there may be additional characteristics of the chosen sample for which differential response rates have occurred. There may also be population characteristics for which differential frame coverage rates exist, and there may be variation in the selected sample size on characteristics other than those on which the sample selection process was stratified. A common solution to dealing with this remaining imbalance is to further calibrate the sample, but this time to the population from which the sample was drawn.

Deville and Sarndal (1992) were the first to coin the term, “calibration,” in conjunction with sample weighting, but approaches which in effect constrain the behavior of weights have existed for more than 60 years. In principle, the goal of a calibration weight adjustment is to bring weighted sums of the sample data into line with the corresponding counts in the target population. Post-stratification and raking were important early forms of weight calibration, and can be shown to be a special applications of the generalized calibration framework discussed by Deville and Sarndal (1992). Both are still in common use today.

The role of calibration depends on which, if any, of the other adjustments are made, and the order in which they are made. For instance, when the order of the only three adjustments is nonresponse and then calibration, the calibration adjustment corrects for any sample imbalances not specifically addressed by the nonresponse adjustment. On the other hand, if only a calibration adjustment is practical, it becomes the sole accommodation for all sources of sample imbalance.

The final set of weights may be calibrated to the population distribution based on population data from a statistically superior external source (e.g., the most recent census or findings from another contemporary national survey with population size estimates of equal or greater quality). Reputable and generally accepted population projections can also be used as the object of calibration. In the event that the most recent census was five or more years prior to the date of GATS data collection, consideration should be made for other sources of adjustment data. Countries with no existing, or outdated, sources of calibration data may not be able to complete this adjustment step. CDC country focal points and statisticians from

the Data Coordinating Center and SRC should be consulted if this is the case. This step essentially involves adjusting the weighted sample (based on $W_j^{(nr)}$ from the GATS sample) to the population distribution of a set of categorical calibration variables in either of two ways: (1) through *post-stratification* (or cell weighting) to the joint or cross-classified population distribution of these variables or (2) through *raking* (or iterative proportional fitting) so that the margins of the joint population distributions of these variables match those in the population. Although variation in final adjusted weights is likely to be somewhat lower with raking, the relatively larger expected sizes of GATS samples will lend themselves more readily to post-stratification, which more precisely calibrates the sample to the population and is therefore the recommended calibration approach for GATS. Detailed instructions in the use of post-stratification are provided below. Country statisticians interested in using raking procedures should contact CDC country focal points and the SRC before proceeding.

RECOMMENDATION:

Form adjustment cells for post-stratification by respondent-reported gender, age, and education; also by region if quality regional estimates are needed. Rural/urban residence should replace education in defining calibration adjustment cells whenever urban-rural comparisons are thought to be of greater importance than the benefits to calibration of education as a predictor of tobacco use.

As with other adjustments, calibration is most effective when the variables used to define the control distributions are highly correlated with key study variables. Although the best set of predictors often varies among study variables in health-related surveys, gender and education are generally good predictors of tobacco use behavior and are thus a good choice for GATS samples. Age, rural/urban residence, and region, if sample sizes are sufficient for regional estimation, are also potential calibration variables. Rural/urban residence should replace education in defining calibration adjustment cells whenever urban-rural comparisons are thought to be of greater importance than the benefits to calibration of education as a predictor of tobacco use. Ultimately, the final analysis weight (W_j) for the j -th respondent data record is obtained from Equation (1) as, $W_j = B_j A_j^{(nr)} A_j^{(cov)} A_j^{(cal)}$, where $A_j^{(cal)}$ is computed by some calibration strategy. Note that $A_j^{(cov)} = 1$ since no adjustment is recommended specifically for frame coverage.

Post stratification adjustment (PSA) is produced by first dividing the complete set of sample respondents who completed the survey into calibration adjustment cells defined by a multi-way categorization of several demographic calibration variables that are known to be predictive of key tobacco use measurements in the target population (i.e., gender x education, and age x gender x education are two possibilities). Some calibration variables are more predictive than others. One measurable indication of predictive strength is the level of statistical significance of the logistic regression coefficient in predicting important dichotomous (0/1) tobacco use variables. Next, a calibration adjustment factor is computed within each adjustment cell as the ratio of two values. The value in the numerator of the ratio is a reputable count (e.g., from a recent census) of the number of the target population members in the cell. The denominator number is the sum of nonresponse adjusted weights for all respondents in the cell, which is an estimate of the target population count (numerator). The resulting ratio is likely to hover around 1.00. A PSA <1.00 suggests that the sample over-represents the population members in that cell, while a PSA >1.00 indicates that the population members are under-represented by the sample in that cell. Multiplying the PSA times the nonresponse-adjusted weight compensates for any over- or under-representation by the cell so that the final weighted distribution of the sample among the set of

adjustment cells will precisely match the population distribution among cells, thereby calibrating the final weighted respondent sample to the target population.

As previously indicated, the predictor variables we recommend (at minimum) for GATS sample calibration are the respondent's gender (male or female) and four categories of the respondent's level of completed formal education. The education categories should be defined so that the marginal percent distribution among categories is as close to uniform as possible (i.e., approximately 25% of the population in each group based upon census or other statistically superior external source).

In some instances, the number of respondents in an adjustment cells falls below some minimum standard (e.g., less than 50). To avoid extreme cell-specific values of the post-stratification adjustment due to the small respondent sample size in the cell, it is advisable to consider combining or merging each small-sample cell with one or more other cells. This merging process is called cell "collapsing". The question is: "With which cell(s) should a small cell be collapsed?" Since post-stratification is conceptually tied to sampling stratification, Kish (1965) counsels one to (as with sampling strata) combine cells that are as similar (i.e., have comparable means for key tobacco use measurements in GATS). This translates to collapsing categories of calibration variables that are least predictive (i.e., have the smallest logistic regression coefficients in predicting important dichotomous (0/1) tobacco use variables). So, for example, if there is a small-sample cell for the oldest males with the least formal education, and age is known to be the least predictive calibration variable, this cell and the cell for the next oldest males with the least formal education should be collapsed, provided the minimum cell count is achieved in the collapsed cells. If more than one cell must be collapsed with the small-sample cell, one could either go to the next cell up or down for the least predictive variable, or consider collapsing on the second most predictive calibration variable as described above. The ultimate goal is to have all cells exceed the minimum sample size standard.

Returning to our weights calculation example for the fictitious GATS respondent, \mathcal{R} , who would be assigned to the calibration adjustment cell including those in Region X who are female and in the same education category as \mathcal{R} . If the population count of those with these characteristics based upon the last census is found to be $N_h = 2,724,182$, and the weighted sum of the sample with these characteristics is,

$\sum_{j=1}^{r_h} W_{hj}^{(nc)} = 2,919,669.3442$, then the post-stratification adjustment ($A_j^{(cal)}$) is computed for \mathcal{R} and all other

GATS respondents in \mathcal{R} 's calibration adjustment cell as

$$A_j^{(cal)} = \frac{N_h}{\sum_{j=1}^{r_h} W_{hj}^{(nc)}} = \frac{2,724,182}{2,919,669.3442} = 0.9330, \quad (32)$$

where N_h is the population count according to the external source of population count data, such that

$N = \sum_{h=1}^H N_h$ is the total population size as indicated by the external calibration source, and r_h is the

respondent sample size in the adjustment cell. Note that $A_j^{(cal)}$ is same as for all the respondents in the adjustment cell 'h'. When checking the set of adjustments ($A_j^{(cal)}$) among all adjustment cells at the conclusion of this step, one should observe that most are either slightly greater or less than 1, with those segments of the population that were still underrepresented in the sample after adjustment for

nonresponse having values greater than one and those that were overrepresented have values less than one.

By multiplying the existing weight for \mathcal{R} times the adjustment for \mathcal{R} 's adjustment cell ($A_j^{(cal)}$), the final adjusted weight for \mathcal{R} (that is added to \mathcal{R} 's data record on the analysis file) will be,

$$W_j = W_j^{(nr)} \cdot A_j^{(cal)} = [15421.61] \cdot [0.9339] = 14402.24 \quad (33)$$

and the weighted relative frequency distribution of the sample using this final weight now matches the corresponding population distribution with respect to the calibration variables; i.e.,

$$\sum_{j=1}^{r_h} W_{hj} / \sum_{h=1}^H \sum_{j=1}^{r_h} W_{hj} = N_h / N \quad (34)$$

This equality can be used as a check of the calculations for this adjustment.

Alternative Approach

In countries with smaller respondent sample sizes and greater disproportionate sampling by region, raking may be a more practical alternative to post-stratification. Raking to force sums of final adjusted weights to match marginal categorical totals for a set of calibration variables (but not for cross-class cells) was first proposed by Deming and Stephan (1940). Starting with the same type of multi-way configuration of categorical calibration variables as post-stratification, each iteration of the raking involves separately forcing category sums for each variable to equal corresponding population categorical totals (see Kalton and Flores-Cervantes, 2003 for a computational example for this case where there are two calibration variables). Those countries interested in using raking rather than post-stratification for calibration of final GATS weights should discuss this alternative with each CDC's country focal point and the SRC.

4. Assuring the Quality of GATS Sample Weights

To assure that sample weights for GATS are produced consistently and accurately in each country, we offer several suggestions on ways to seek this assurance in concluding this manual.

4.1 Careful Documentation of Sample Selection and Associated Probabilities

The importance of carefully documenting the selection process followed in each sampling step of the GATS sample selection process cannot be overemphasized. It is critically important therefore that technical staff in each GATS country make a strong commitment to this documentation. There are several ways to assure that this documentation will be complete and fully useful for weights calculation. First, once the selection algorithm for each step is set, staff should consult a reputable sampling text to find the correct computational formula for the selection strategy that is used. Some commonly used selection strategies in survey samples are simple random sampling, systematic sampling, one-at-random sampling, PPS systematic sampling, PPS without replacement sampling, and PPS with replacement sampling. Texts that might be consulted are by Hansen, Hurwitz, and Madow (1953), Kish (1965), Raj (1968), Cochran (1977), Sukhatme, et al., (1984), and Lohr (1999). Second, the means to compute these selection probabilities should be in place to make the correct computations as selection occurs. In a practical sense, this means that computerized selection is preferred to manual selection and that the computer code to do sample selection should also include steps to compute and store the selection probabilities at the same time that individual sample members are chosen. It also means that if some sampling steps must be completed manually, the forms used for sample selection should provide for a way to assure that all information needed to compute the selection probability for that step can be produced at that time, or later if preferable. Third, a technical member of the in-country staff should be given the overall responsibility to assure that selection probabilities are computed correctly, in a timely manner, and properly stored for use when it comes time to produce the weights. Fourth, the procedures and formulae used to compute all components of the sampling weights should be planned and documented in sufficient detail in advance of the weights computation process so that another competent member of the technical staff not working with GATS could correctly complete weights computation process. This country-specific set of planning specifications for computing GATS weights should be modified at the conclusion of the weights computation process to serve as the final documentation for GATS weights. And finally, the documentation for the weighting process should be made generally available to GATS partners and the GATS sampling committee.

4.2 Control the Adverse Statistical Effect of Variable Weights

Kish (1965) has shown that a variation in sample weights can increase the variance of survey estimates by a factor of multiplicative effect (*Meff*). For instance, if $\bar{w}=12394.0091$ is the sample mean of the final adjusted weights, W_j , and $s_w^2=(5463.3760)^2$ is the variance of these same weights among the members of the GATS sample of which the fictitious respondent, \mathcal{R} , is a part, then

$$Meff_w = 1 + \frac{s_w^2}{\bar{w}^2} = 1 + \left(\frac{5463.3760}{12394.0091} \right)^2 = 1.1943 \quad (35)$$

Some of the factors contributing to this variation are features of the sample design intended to reduce the variance of estimates (e.g., sample allocation among strata), while others are the result of strategies to control bias in these estimates (e.g., adjustments for nonresponse and sample calibration). Controlling this variation effectively requires walking a fine line between reducing the adverse effect of $Meff_w$ on the precision of estimates while minimizing change in the bias reduction benefits of weight adjustment.

The recommended approach to dealing with the effect of variable weights for GATS samples will be to monitor $Meff_w$ for the final adjusted weights (W_j), and to consult with GATS central staff if $Meff_w > 2.00$. This consultation may lead to the use of various existing strategies (e.g. truncation of weights) for restricting the size of weight adjustments or to the use of various weight “trimming” strategies. For example, Kalton and Flores-Cervantes (2003) describe the process of further collapsing adjustment cells to restrict the size of adjustments produced by the final set of cells. This preventive approach might be used to control the size of adjustments for nonresponse, but it might also be used to restrict the size of calibration adjustments as a middle ground between post-stratification and raking (see Deville and Sarndal, 1992). Potter (1988; 1990; 1993), on the other hand, reviews several trimming approaches that create multiple sets of trimmed weights based on various trimming levels and then choose the set of trimmed weights with the lowest estimated mean squared error for key survey estimates. A disadvantage is that trimmed weights producing the lowest mean squared error for one survey estimate may not produce the same result for another estimate.

RECOMMENDATION:

Control the effect of variable weights by monitoring its multiplicative effect on the variance of survey estimates; consult with GATS staff as needed

4.3 Carefully Follow the Recommended Computational Procedures and Quality Checks in this Manual

The goal of this manual is to provide guidance in developing a strategy for weights computation in each GATS sample and to assure that the strategy can be completed successfully so that analysts of GATS data can be confident that the weights they are using will contribute to producing the best possible population estimates of tobacco use. To aid in meeting this goal, we have provided a set of specific procedures that can produce sample weights for GATS data that meet the highest standards for survey research and if followed uniformly among all GATS countries will contribute to the quality of between-country comparisons of findings. In describing the computational steps for weights, we have also suggested or implied ways to assure the quality of the sample weights. They are to:

RECOMMENDATION:

Follow the recommended steps for weighting and perform the suggested checks during computation

1. Have someone not involved in GATS weights computation but familiar with the contents of the recommended protocol as laid out in this manual check behind the work of those producing the weights during the computation process. This check might best be done after each step (i.e., compute base weights, compute nonresponse adjustment, and compute calibration adjustment) is completed, rather than once at the conclusion of the entire process.

2. Determine if the average size of the base weights (B_j), divided by the average size of the nonresponse-adjusted weights ($W_j^{(nr)}$) is roughly equal to the final overall RR1 response rate as described in *Section 3.2* of this manual.
3. Use Equation (34) to confirm that the weighted distribution using the final calibrated weights (W_j) matches (for each calibration adjustment cell) the population distribution to which the weights were calibrated.
4. Note that the sum of the final adjusted weights (W_j) over the full sample should be the size of the overall population count (of those 15 years of age or older) from the source used for calibration. If that source was the last census, then this sum of weights should be the population count from the last census.
5. Confirm that the post-stratification adjustments ($A_j^{(cal)}$) produced by the set of adjustment cells are mostly either slightly greater or less than 1.

4.4 Prepare a Detailed Written Document Describing the Weights Computation Process that Is Actually Used

As with most major survey activities that are planned and then implemented, the plan that is laid out (i.e., essentially the contents of this manual if the recommended weights computation process is followed) often differs somewhat from what is actually done. Departures from planning specifications can occur in producing weights at each of the three weights computational steps. For example, if another reputable method for within-household selection of the survey respondent is used, then the computation of the conditional probability of selecting the respondent within a participating GATS household would be computed in accordance with the replacement method. To avoid loss of details on this important pre-analysis activity in GATS, detailed written documentation of the actual weighting process following the same basic outline of this manual should be completed immediately after the weights computation, and should then become a permanent part of the documentation for all survey activities for GATS. This document will be used to conduct a formal evaluation of the weighting and calibration process by SRU biostatisticians and ultimately be included in each country's final country report as a permanent record of methods used in GATS. For details of format and specific computations needed for review see the *GATS Quality Assurance: Guidelines and Documentation* manual. The public release of weighted findings from the country data should not occur until this documentation is reviewed and formally certified as having met the quality standards for weights as set by this manual.

RECOMMENDATION:

Prepare a written document that gives details of how weights were computed for each GATS country

4.5 Consult with GATS Central Office Staff During the Weighting Process

Finally, GATS central staff should support the sample weights computation process in each GATS country. This support will include clarification of procedural methods laid out in this manual as well as assistance with troubleshooting for unanticipated process in completing this important set of survey tasks.

5. Bibliography

- American Association for Public Opinion Research. 2009. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 6th edition. AAPOR.
- Berry, C. C., Flatt, S. W., and Pierce, J. P. (1996). Correcting Unit Nonresponse via Response Modeling and Raking in the California Tobacco Survey. *Journal of Official Statistics*, 12(4), 349-363.
- Cochran, W.G. (1977). *Sampling Techniques*, 3rd Ed., New York: Wiley and Sons.
- Deming, W. E. and Stephan, F. F. (1940). On the Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals are Known. *Annals of Mathematical Statistics*, 11(4), 427-444.
- Deville, J. C. and Sarndal, C. E. (1992). Calibration Estimators in Survey Sampling. *Journal of American Statistics Association*, 87, 376-382.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Sample Design Manual*. Atlanta, GA: Centers for Disease Control and Prevention, 2020.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Field Interviewer Manual*. Atlanta, GA: Centers for Disease Control and Prevention, 2020.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Field Supervisor Manual*. Atlanta, GA: Centers for Disease Control and Prevention, 2020.
- Global Adult Tobacco Survey Collaborative Group. *Global Adult Tobacco Survey (GATS): Quality Assurance: Guidelines and Documentation*. Atlanta, GA: Centers for Disease Control and Prevention, 2020.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory*, Vols. I and II, 1953.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.
- Iannacchione, V. G., Milne, J. G., and Folsom, R. E. (1991). Response Probability Weight Adjustments Using Logistic Regression. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 637-642.
- Kalsbeek, W. D., Morris, C., and Vaughn, B. (2001). Effects of Nonresponse on the Mean Squared Error of Estimates from a Longitudinal Study. *Proceedings of the Section on Survey Research Methods, American Statistical Association*
- Kalton, G. (1983). *Compensating for Missing Survey Data*. Ann Arbor, MI: University of Michigan.
- Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.
- Kish, L. (1965). *Survey Sampling*, Second Printing, New York: Wiley and Sons.
- Lepkowski, J. M., Kalton, G., and Kasprzyk, D. (1989). Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 296-301.
- Lessler, J. T. and Kalsbeek, W. D. (1992). *Nonsampling Error in Surveys*. New York: Wiley and Sons.
- Lohr, S. (1999). *Sampling: Design and Analysis*, Pacific Grove: Duxbury Press.
- Potter, F. J. (1988). Survey of Procedures to Control Extreme Sampling Weights. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 453-458.
- Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 225-230.
- Potter, F. J. (1993). The Effect of Weight Trimming on Nonlinear Survey Estimates. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 758-763.
- Raj, D. (1968). *Sampling Theory*, New York: McGraw-Hill.
- Särndal, C.-E., Swensson, B., and Wretman, J. H. a. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.
- Sukhatme, P. V. et al, (1984). *Sampling Theory of Surveys with Applications*, Ames: Iowa State University Press.

Global Adult Tobacco Survey (GATS)