GTSS Global Youth Tobacco Survey (GYTS)

Sample Design and Weights

Global Youth Tobacco Survey (GYTS)
Sample Design and Weights

Version 1.1

August 2014

<div style="border: 2px solid gray; border-radius: 10px; text-align: center;">

Global Youth Tobacco Survey (GYTS)
Comprehensive Standard Protocol

....

GYTS Core Questionnaire with Optional Questions
GYTS Sample Design and Weights
GYTS Implementation Instructions
GYTS Analysis and Reporting Package
GYTS Data Dissemination Guidance
GYTS Data Release Policy

</div>

Acknowledgements

GYTS Collaborating Organizations

- Centers for Disease Control and Prevention

- CDC Foundation

- RTI International

- World Health Organization

Suggested Citation

Global Youth Tobacco Survey Collaborative Group. *Global Youth Tobacco Survey (GYTS): Sample Design and Weights, Version 1.1*. Atlanta, GA: Centers for Disease Control and Prevention, 2014.

Contents

Exhibits

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 eight 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[1]. An efficient and systematic surveillance mechanism is essential to monitor, manage, and help prevent the epidemic.

The Global Tobacco Surveillance System (GTSS) is a set of globally standardized surveys to monitor tobacco use and key tobacco control policies. It is comprised of both youth and adult surveys. Launched in 1999, the Global Youth Tobacco Survey (GYTS) is a global standard for systematically monitoring youth tobacco use (smoking and smokeless) and tracking key tobacco control indicators. GYTS assists countries to enhance their capacity to design, implement, and evaluate tobacco control interventions. It is an important tool to assist countries in supporting WHO MPOWER, a package of six evidence-based demand reduction measures contained in the WHO's Framework Convention on Tobacco Control (FCTC).

GYTS is one of the largest public health surveillance systems, and it has been active in more than 188 countries/sites. The survey is repeated every 4-5 years generating data that are comparable within and across countries.

1.1 Overview of the Global Youth Tobacco Survey

GYTS is a nationally representative school-based survey of students in grades associated with 13 to 15 years of age and is designed to produce cross-sectional estimates for each country. It uses a global standardized methodology that includes a two-stage sample design with schools selected with a probability proportional to enrollment size. The classes[2] within selected schools are chosen randomly and all students in selected classes are eligible to participate in the survey. The survey uses a standard core questionnaire with a set of optional questions that permits adaptation to meet the needs of the country on tobacco use and key tobacco control indicators. The questionnaire consists of the following topics: tobacco use (smoking and smokeless), cessation, secondhand smoke (SHS), pro- and anti-tobacco media and advertising, access and availability to obtain tobacco products, and knowledge and attitudes regarding tobacco. The questionnaire is self-administered; using scannable paper-based bubble sheets, it is anonymous to ensure confidentiality.

As a companion to GYTS, the seven-question School Policy questionnaire was developed to collect information about school policies related to tobacco use among staff and students. This self-administered questionnaire is completed by one of the school administrators.

To maximize the efficiency of the data collected from GYTS, a series of manuals has been created. These manuals are designed to provide countries with standard requirements as well as

---

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

[2] The term class is frequently used in this manual. A class is defined as a group of students who gather regularly to be taught a course, while a grade is defined as the body of students for whom the year of graduation is the same. The terms class and classroom are used interchangeably in this manual.

---

recommendations on the design and implementation of the survey at every step of the GYTS process. They are also designed to offer guidance on how a particular country might adjust features of the GYTS protocol to maximize the utility of the data within the country. Following the standard protocol and procedures in these manuals is strongly encouraged to maintain consistency and comparability across countries and for the successful implementation of GYTS.

1.2   Use of this Manual

This manual is designed to offer requirements and recommendations, as well as suggested guidelines for countries to follow as they develop an appropriate sample design for their implementation of GYTS. This chapter provides background information on GYTS. Subsequent chapters are summarized below:

- **Chapter 2** summarizes the survey objectives.

- **Chapter 3** presents a definition of the target population for the survey and provides a discussion of the school sample frame (enrollment list) that will be used. Clearly defining the target population and sample frame is imperative since they define the scope of the survey. A clear definition of the target population allows for better understanding of the extent of sample coverage. This definition is also used to determine survey eligibility for sampling and data collection.

- **Chapter 4** presents the basic features of the sample design that, when included, contribute to cross-country comparability of findings as well as to statistical validity of tobacco use estimates. The chapter also presents a brief overview of the recommended approach for choosing, recruiting, and collecting data from GYTS samples.

- **Chapter 5** illustrates in more detail the suggested selection process at each stage of the sample design.

- **Chapter 6** provides an overview of the process of producing a sample weight for each student. These weights are typically adjusted for nonresponse and calibrated to target population counts (school enrollments sizes) before they are used for data analysis.

- **Appendices:** Appendices A through G provide various templates including sampling parameters worksheet, school enrollment list, rationale for the recommended overall sample size for GYTS surveys and information and templates used for enrollment, sample selection, and forms for school-level and class-level information.

2.   Survey Design Objectives

GYTS is designed to produce a nationally representative sample of students in grades associated with 13-15 years of age from which cross-sectional estimates by gender can be generated.

The targeted sample of GYTS is grades that represent the 13-15 year age group of students within each sample school. The target sample is designed to cover a vast majority of students aged 13-15 years in the sampling process. In GYTS, both private and public schools are recommended to be included in the national sampling frame.

The sample for each GYTS survey should be designed to obtain completed questionnaires from at least 1,500 students with roughly half the sample respondents being boys and the other half being girls. These minimum sample sizes will generate nationally representative estimates that meet the expected precision requirements of the GYTS.

Recommendation:
The design of GYTS seeks to obtain nationally representative estimates for students aged 13-15 years.

In addition to gender, some countries may wish to report survey estimates by important demographic domains such as urbanicity, regions or provinces within a country. Generating estimates by these domains generally requires larger overall sample sizes than if no domain-specific estimates are required. For example, if a country is interested in producing estimates by urban and rural areas, it is recommended that the survey samples 1,500 students in each area. These additional requirements are recommended provided that estimates of sufficient statistical quality can be produced and do not compromise the quality and comparability of other key intra- and inter-country estimates.

3. Target Population and Sampling Frame

3.1 Definitions of the GYTS Target Population and Study Eligibility

In general, the target population of the GYTS should include all eligible schools (both public and private) in all geographic areas within a GYTS country. In rare instances, it may be necessary to exclude schools in some geographic areas. Areal exclusions will be considered, provided a significant proportion of the country's eligible student population does not reside in these schools. For example, extremely remote areas or areas that cannot be visited due to war, political unrest, etc. may be considered for exclusion from the GYTS target population. To propose exclusion, a country should (i) explicitly define the geographic areas of the country that include the schools to be excluded, and (ii) provide an estimate of the percentage of the target student population that resides in those areas being excluded. This percentage will provide an indication of the potential bias effect of any areal exclusion on estimates that are otherwise intended to be national in scope.

> Only in rare instances should countries exclude geographic portions of the country from the GYTS target population.

The target population for GYTS surveys should include those students enrolled in public and private schools located in the included areas of the country and who are in grades that typically include boys and girls 13-15 years of age. Thus, all public and private schools, regardless of level (e.g., primary, middle, secondary, high), that enroll 13-15 year olds should be included in the school sampling frame. The grades might also include students outside of this age group. However, these students are excluded at the time of analysis. Moreover, the schools/grades, though included at the time of preparing the enrollment list, are explicitly excluded from the survey if they do not include students 13-15 years of age.

> GYTS Target Population:
> Include those students enrolled in either public or private schools located in the included areas of the country and who are in grades that typically include boys and/or girls 13-15 years of age.

The target students are sampled as follows. After the selection of schools is completed, the survey administrator will create a roster of classes in all eligible grades in the school that meet the age criteria at the time the roster is completed for selection of classes. This is generally accomplished at the initial visit made by the survey administrator to each selected school for obtaining the class-level information and permission to conduct the survey. Moreover, all eligible classes should be included on the class roster, and all reasonable efforts should be made to obtain completed questionnaires from all the students in the selected classes.

A country implementing agency is required to provide information on target population and sample parameter specifications for the development of the sample design for a country to select the schools and subsequently the classes. **Appendix A** (Sampling Parameters Worksheet) provides details on information required for sample design development and to ensure representativeness and that the standards are followed.

The *GYTS Implementation Instructions* will provide country-specific information on which schools should be included in or excluded from the target population, as well as a discussion of when to follow-up with selected schools and the processes and standard guidelines for class selection and completion of questionnaires.

3.2   School Sampling Frame

The list from which the sample is to be chosen is called the "sampling frame". Entries on this frame are called "sampling units". Thus, to select the sample of schools for each GYTS survey, a school sampling frame is needed. The frame should consist of all sampling units (schools) in the survey population from which the sample is to be drawn. The sampling frame (list frame or cluster frame) used in most school-based surveys of students, including GYTS, is called a School Enrollment List or School Enrollment File. There are two types of frames used in most surveys of students: list of individual students or cluster of students (grades or sections of grades).

Based on past GYTS experience, most countries provide school enrollment lists of eligible students by grade and gender that will be later used as the size measure for Probability Proportional to Size (PPS) sampling. The most up-to-date or current list of eligible public and private schools is often available from the Ministry of Education. In some instances, obtaining a complete school list requires coordination among several ministries, departments, and organizations (e.g., Ministry of Secondary Education, Ministry of Higher Education, associations of private schools if public agencies lists of schools do not include private schools, etc.).

The key information needed for each entry on a school frame, at a minimum, should include the following: a unique school ID, school name, location or contact information (address), school type (private/public), and the student enrollment count by eligible grades and sex. Please refer to **Appendix B** for more information on the school enrollment list. This list will be adapted and used for sample selection at a later stage using a specially designed software. For the GYTS, as with most school-based surveys, a comprehensive list of all eligible schools in a country is necessary to ensure completeness of the sampling frame. The up-to-date sampling frame is essential to drawing a nationally representative sample. It is recommended that the school enrollment list be prepared or updated within the past one year of the proposed GYTS implementation to ensure the quality of the school sampling frame.

> A school enrollment list should be up-to-date and is prepared, preferably, within the past one year of the proposed GYTS implementation. A school enrollment list is a simple list of the number of students by each eligible grade in a school or a list of the number of students in a school by grade and gender.

3.3   Small School Exclusion

There may be some schools in the country's enrollment list with relatively few eligible students. While it is always best to sample from a list of all eligible schools, regardless of size, in order to maximize survey efficiency, some previous GYTS countries have decided to exclude schools with relatively small enrollments. For example, previous GYTS sampling frames have excluded schools with enrollments below 40 students. While the cut-off point may vary from country to country, any GYTS country that elects to exclude smaller schools should specify its size cutoff and report the percentage of the target population that will be excluded in applying this cutoff. This percentage would be computed as the count of 13-15 year old students in the excluded schools, divided by the count of 13-15 year old student in all eligible schools, including those that are excluded.

> Small schools (e.g., those with fewer than 40 students) may be excluded from the GYTS school frame.

4. Sample Design: Specifications and Stages of Selection

The quality and usefulness of results from the GYTS depend largely on the procedures used to select schools and classes (as clusters of students). Since surveying each student in a country is not practical, a random probability sample is selected from the target population. The results from a probability sample can be generalized to the entire student population from which the sample was drawn since random selection is used in each sampling step, or stage (see below).

In general, the GYTS sample is a multi-stage geographically clustered design. A multi-stage design is defined as a sample design that entails progressively selecting subsamples from a previously selected sample until a sample of population members is identified via the planned selection process. In the GYTS, a two-stage design is utilized with schools selected at the first stage with probability proportional to the enrollment size (PPS)[2] followed by classes[3] chosen randomly within selected schools. All students in the selected class are eligible to participate in the survey. The sampling frame usually includes students from both public and private schools. Inclusion of students from private schools depends on the availability of the school sampling frame.

PC Sample, a sample selection software specifically written for CDC youth surveys (e.g., YRBS and school health surveys) by Westat, is used for school and class selection. The software selects a probability sample of schools based on the school sampling frame (i.e., School Enrollment List) provided by a country (see **Appendix B**). PC Sample also provides instruction sheets used in class selection within each school.

There are certain requirements and recommendations that should be followed in order to maximize the comparability of the results between countries that are conducting GYTS. However, each country has the option of introducing design enhancements that would allow them to increase the usability of the results from this survey (e.g., selecting the sample to ensure precise estimates by a region). In this chapter, we present some of the basic survey design requirements. Any design enhancement that a country wishes to introduce will generally be acceptable provided the modified design yields a valid probability sample of students and the quality of all key estimates produced from the data will be satisfactory to the country and consistent with recommended design features included in this section. In summary, some of the required basic survey design features of the GYTS are presented in Sections 4.1 to 4.3.

---

[2] Please refer to Karlton, G. Introduction to Survey Sampling pp 38-46 for additional details about probability proportional to size (PPS) sampling.

[3] Most importantly, 'class' is operationalized as a listing of all eligible students in a school such that each eligible student is in EXACTLY ONE class. For example, in US school-based surveys where schools have students in classes for several "periods" of the day, "class" has been operationalized as the classroom where the student spends his/her first period of the school day. This way, in principle, each student is associated with one and only one class, and thus avoids the probably of multiplicity (arising when population members have multiple chances to be selected in a sample). Based on prior GYTS experience, some examples of "classes" that have been used for prior GYTS surveys include 'classrooms' in the countries of South East Asian Region, 'grades' in the countries of European Region, and 'forms' in the countries of African Region.

4.1   Sample Design Specifications

Requirements related to the sample design include the following:

1. Random selection must be used in each sampling stage so that every member of the target population has a non-zero chance of being selected into the sample.

2. The probability of selection for every unit (school, class, and student) selected at each stage of the design must be computable and retained on the final analytic files for the survey. For example, if the sample is selected in two stages (i.e., schools and then classes within schools to include all students in selected classes for the survey sample), then the final analytic files must contain the corresponding probabilities of selection.

> REQUIREMENT:
> At each stage of selection, every unit must be selected randomly and probabilities of selection at each stage must be recorded.

3. Sample Sizes and Expected Precision

    Requirements and recommendations related to sample size are based on the following indicators of statistical quality that were established for GYTS findings:

    a. GYTS should be designed to produce nationally representative estimates, by gender, with a minimum sample size of 1,500 student respondents from a minimum of 20 schools. Additional detail on how this respondent sample size was computed is provided in **Appendix C**. Further guidance and technical assistance can be provided to understand the precision and power implications of proposed sample sizes for students and schools.

        If any country is interested in obtaining sub-national (e.g., regional/provincial) estimates and would like to produce estimates in each region/province, then the survey should be designed to obtain 1,500 student respondents in each region/province.

    > Recommendation:
    > A student sample size of 1,500 from a minimum of 20 schools is recommended for a country. Larger numbers of students and schools, particularly schools, is preferred.

    b. Note that estimates produced with 1,500 students from 20 schools could sometimes produce a margin of error up to 15%, and the power to detect differences may be much less than the usual benchmark level of 80%. The larger the sample sizes in terms of schools and students, the higher the precision of estimates and the lower the margin of error. The fewer the number of schools and students in a sample, the lower the precision of estimates and the higher the margin of error.

    > Recommendation:
    > A student sample size of at least 1,500 is recommended for each domain (e.g., region/province or school type) when estimates are to be reported for these domains by gender.

c. The design of the survey should correctly reflect anticipated levels of nonresponse and ineligibility in determining how many students must be selected in order to yield the recommended number of respondents indicated above. Nonresponse and ineligibility may be observed at the school, the class, and the student level. For example, a selected school asked to participate in the survey or a class selected for participation may refuse to participate (nonresponse). Similarly, a student from a selected class may be absent.

> REQUIREMENT:
> GYTS sample sizes should be adjusted upward to take into account the attrition effects of nonresponse and ineligibility at each stage in deciding how many schools and students to select.

If, for example, a country's survey is designed to sample 20 schools with a target to achieve 1,500 students and it expects to observe the following:

Rate	Comment	Assumption
School Participation Rate	Accounts for eligible school acceptance to participate in the survey.	80%
Student Attendance Rate	Accounts for those eligible students from a selected class to be present at the time of questionnaire administration or data collection.	80%

Then, the survey must be designed to achieve a selected sample of

Number of schools to be selected: $\dfrac{20}{0.80} = 25$

Number of students to be selected: $\dfrac{1,500}{0.80} = 1,875$

4.2 Overview of the Sampling Approach

As described in earlier sections, the sample for GYTS should be selected using a multi-stage, geographically clustered design to ensure adequate coverage of the entire target population while simultaneously minimizing data collection costs. Costs are reduced because the geographic clustering of the sample design will minimize the travel needed for a research coordinator to visit each selected school. Since no national listing of students is available in many countries, it would be impractical to sample from such a list. This implies the need for sampling clusters of students defined by the schools they attend to control survey costs; however, the cost benefit of sampling school clusters also implies a loss in statistical precision for GYTS estimates of tobacco use.

Careful documentation of selection probabilities at each stage of sampling is needed later for producing sample weights. Separate identifiers for sample members at the school stratum, school, and student levels are needed for analysis later. Details on stratum identifiers (both explicit and implicit strata) that are needed to properly measure the precision of survey estimates are described in **Chapter 6**. An overview of the GYTS sample design is outlined in **Exhibit 4-1**. In summary, this sample selection process should proceed as follows.

Exhibit 4-1. Overview of GYTS Sample Design

```
┌─────────────────────────────────────────────────────────────────────┐
│                         Target Population                            │
│  ▪ Students in grades associated with 13-15 years of age             │
└─────────────────────────────────────────────────────────────────────┘
                                  ∨
┌─────────────────────────────────────────────────────────────────────┐
│                           Select schools                             │
│  ▪ Schools should be selected with probability proportional to       │
│    size (PPS)                                                         │
│  ▪ Size measure is the school enrollment size                        │
│  ▪ Approximately 20 schools are selected for a target sample of      │
│    1,500 students                                                    │
└─────────────────────────────────────────────────────────────────────┘
                                  ∨
┌─────────────────────────────────────────────────────────────────────┐
│                           Select classes                             │
│  ▪ Classes should be selected systematically from each selected      │
│    school                                                            │
│  ▪ Class selection is made at school by research coordinator         │
│  ▪ Double draw of classes is considered based on GYTS coordination   │
│    with other surveys, as needed                                     │
└─────────────────────────────────────────────────────────────────────┘
                                  ∨
┌─────────────────────────────────────────────────────────────────────┐
│                 Include students in selected classes                 │
│  ▪ All the students in a selected class are eligible to participate  │
│    in the survey                                                     │
└─────────────────────────────────────────────────────────────────────┘
```

Stage 1: Selection of Schools

Most countries, as a standard practice, extend their surveys to students in both private and public schools. However, countries may limit which types of schools will be included in the survey (such as the exclusion of private schools) and must specify this on the Sampling Parameters Worksheet (see **Appendix A**) with a justification for such decision provided separately. Schools that contain grades associated with students 13-15 years of age are selected to produce a sample of schools. Twenty schools and 1,500 students are targeted using the probability proportional to size technique. In practice, 25 schools and 1,875 students are sampled to compensate for school and student attrition caused by ineligibility, nonresponse and refusal during the survey, with the assumed school response rate and student response rate both at 80%. There will be NO replacement or substitution for schools that do not agree to participate. When schools refuse to participate, the representativeness of the sample is compromised. Thus, it is important that every effort is made to enlist the participation of the selected schools. The process in this stage includes the following features: (i) explicit (e.g., by region) and/or implicit stratification in selecting schools, (ii) selection of certainty schools, and (iii) selection of non-certainty schools via PPS systematic selection within explicit strata. The details on stratification and selection of certainty and non-certainty schools are described in detail in **Chapter 6**. More information about the selection process and suggested ways to increase participation is described in the *GYTS Implementation Instructions*.

Stage 2: Selection of Classes

The selection of classes (to identify students) is the most complex part of the selection process since construction of the sampling frame of classes and identification of the set of selected classes within each

school must be carried out by the individuals administering the survey (e.g., University students, or Ministry of Health staff or research coordinators or survey administrators). The process in this stage includes the following features: (i) selection of whole clusters (classes) of students, (ii) defining "class" in a given school, constructing the frame listing of classes for a given school, and (iii) determining the systematic selection interval so that the overall selection probabilities are the same for all students in the target population. Eligible students *cannot* be enrolled in more than one of these classes simultaneously, and each student in these classes must have an equal opportunity of selection.

Worksheets to identify classes within participating schools are produced by PC Sample. Classes are selected systematically from a sequentially numbered list containing every eligible class in a selected school. PC Sample calculates the sampling interval for classes assuming a fixed class size and specifies the classes to be chosen (see the two right-most columns in **Appendix D**). Each eligible student within a selected class is approached for participation in the study, thus making the conditional selection probability for all students in a school equal to 1 divided by the sampling interval for classes. The selection of classes should be based on the school-level information forms produced by PC Sample (see **Appendix D** (School Selection List), **Appendix E** (School and Class-Level Sampling Information Forms for GYTS), and **Appendix F** (School and Class-Level Sampling Information Forms for Double Draw of GYTS and Survey X)).

School and Class Eligibility and Recruitment Status

A school may be considered in-scope or out of scope as per the information provided in the enrollment list. For example, if a school is selected to participate, and at the time of data collection the research coordinator finds that the school no longer has grades that correspond to 13-15 year old students, the school would be classified as out of scope and dropped from the study due to its ineligibility. It would not be replaced by an in-scope school.

In other situations, an eligible school may be allowed to participate by the Ministry of Education, but the school administrator may refuse to participate due to various reasons (e.g., school level IRB approval is not obtained). In this instance, the school will be considered eligible but treated as a non-respondent. Non-respondents of this sort are not to be replaced with other schools.

A class may be considered out of scope or ineligible if the selected class in a school does not consist of or teach the eligible grades. A class may be later determined to be ineligible if the class includes primarily ineligible students (e.g., a class of special education such as mental disabilities).

Number of Classes to Be Selected

The number of classes selected in a given school partially depends upon the number of schools that are selected to participate, the total number of completed questionnaires desired, the average size of the classes, and the school enrollment size in each school. The number of schools selected is dependent upon the specific sample design that was agreed upon by the country and CDC. Generally speaking:

1. As described previously, 25 schools and 1,875 students are typically sampled for the survey with the expectation that 20 schools will participate with a school response rate of 80%. Thus, an average of 94 (1,875/20 = 94) students are to be selected per school. This means that, on average, 3 classes will be selected per participating school, provided the class sizes across schools are all approximately equal to 30 students.

2. As the number of selected schools goes below/above 20 and the total number of students remains 1,500, the average number of classes selected per school will increase/decrease. For example, if 40 schools (twice as many as the example above) are selected to participate but the total number of students remains 1,500 then, on average, 1 to 2 classes (half as many as the example above) will be selected per participating school.

3. Various circumstances may dictate a sample design that selects fewer or more schools and classes than those in either example above.

4. Variations in class sizes will affect the number of classes chosen in individual schools, increasing the number of classes to be selected within school as class size falls and decreasing the selected number of classes as class size rises. However, PC sample software assumes a constant class size while selecting the classes.

5. Timing (season or month of the year) of the survey administration can affect school and student response rates. Please refer to general guidelines on recommended dates for survey administration in the ***GYTS Implementation Instructions***.

Practices to Avoid when Selecting Classes

1. Replacing a selected class. Do not replace a selected class with any other class from the same school if the teacher of the class refused to participate.

2. Arbitrary selection of classes within schools. For example, do not accept the principal's choice of classes if they are different from the randomly selected classes.

Double Draw of Classes

Certain countries coordinate GYTS with other surveys to reduce the burden on schools and students and to save resources. This coordination is typically done by sampling double or triple the number of classes based on the number of coordinating surveys. For example, if GYTS is coordinated with another school-based health survey ("Survey X"), doubling the number of classes sampled produces two non-overlapping samples of classes. This means that selected classes are either assigned to GYTS or to Survey X, but not both. The samples can be used during the same or separate semesters, and classes can be assured that they will be asked to participate in only one survey. Doubling the number of classes increases the burden within sampled schools but minimizes the number of schools asked to participate. When the number of classes is doubled, both surveys are administered simultaneously. This approach is useful in countries with few schools and in countries where both surveys have been conducted successfully. Regardless of the type of coordination selected, the CDC's Global Tobacco Control Branch in the Office on Smoking and Health, the country Ministry of Health or implementing agency, and WHO work together to plan and implement the coordination. It is important to have a common understanding between GYTS and Survey X to preserve trends and comparability within the two surveys without compromising the standard protocol of the two surveys.

Double draw is a kind of sampling selection approach for selecting classes within each selected school when GYTS is coordinated its data collection with any other national-level school-based survey in a country. Upon completion of school selection, a sequence of random numbers for class selections will be generated by PC Sample software in a school-level sample information form. At this stage, GYTS and the other school-based national survey select alternate classes and distribute them to each survey. This process allows GYTS to select half of the eligible classes in a school. Therefore, the within-school interval

for each selected school will be doubled to adjust for the imbalance and stored for weighting purposes. A description of double draw is illustrated in **Appendix F** with an example of how classes are selected in this situation.

Stage 3: Identification of all Students in Selected Classes

Identifying students by class is generally less disruptive to the school than selecting students randomly from a list of all students in the school. Thus in GYTS, all the students in a selected class are included in the survey. The number of students completing the questionnaire, in most GYTS surveys, ranges between 1,500 and 10,000 based on sample design requirements. The table in **Appendix C** shows that a sample of approximately 1,500 students will yield key estimates at a precision level of ± 5% for country-wide estimates when the number of students in the country's target population is at least 30,000. Required sample sizes are smaller in countries with a smaller target population, but sample sizes must be larger than 1,500 if the countries wish to produce estimates at the +/- 5% precision level for important population subgroups (e.g., defined by gender, region, etc.).

4.3   Data Collection Procedures

Requirements related to data collection procedures include the following:

1. Each country must collect the data via a self-administered data collection method using bubbled answered sheets specially designed for GYTS for scanning and processing the data at later stages.

2. Response Rates: A major component of assessing the quality and usability of the GYTS data that are collected is the overall response rate. Each country must develop data collection procedures to achieve a combined

REQUIREMENT:
Obtain a combined response rate of at least 80%.

response rate of 80% or greater. At each point of the selection process where nonresponse can occur, the response rate is defined as the total number of respondents divided by the total number of sample members eligible to respond. The combined response rate is defined as the product of the point-specific response rates. In GYTS, the combined response rate is defined as the product of the school-, class-, and student-level response rates.

There are three sequential points in the sampling process where non-participation can occur. There is a response rate associated with each of these points:

The <u>school response rate</u> is calculated by dividing the number of participating schools by the number of selected schools that turned out to be eligible. This rate may be computed for all schools or for subgroups of schools (e.g., by region). For example, if 95 out of 100 selected eligible schools participated in the survey, the school response rate would be 95%.

The <u>class response rate</u> is calculated by dividing the number of participating classes by the number of selected classes. For example, if 98 out of 100 selected classes participated in the survey, the class response rate would be 98%.

The <u>student response rate</u> is calculated by dividing the number of participating students by the number of eligible students. For example, if 1,750 out of 1,875 selected students (those students enrolled in the selected classes within participating schools) participated in the survey, the student response rate would be 93.3%.

The overall response rate is the product of these three rates. Using the examples above, the overall response rate would be 0.95 x 0.98 x 0.933 = 0.869, or 86.9%.

3. A more detailed description on nonresponse rates is provided in **Chapter 6**.

4. In some studies, sample substitutions are allowed. This refers to the practice of replacing eligible non-respondents or ineligibles with another randomly selected unit from the sampling frame. Sample substitution can alter the representativeness of a sample and thus lead to biased estimates from that sample, and therefore are NOT allowed in GYTS.

5. Sample supplementation (e.g., of schools) refers to the practice of introducing an additional sample into the survey to compensate for sample attrition due to ineligibility and/or nonresponse. This is NOT allowed in GYTS. The anticipated school response rates are considered in the initial stages of the sample design and accordingly, the number of selected schools is adjusted upward to hopefully achieve the intended number of survey respondents (selecting 25 schools or more). Similar compensation for sample attrition is considered for within-school sampling of students (selecting 1,875 students or more).

> REQUIREMENT:
> No substitution for eligible non-respondents or ineligibles is permitted in GYTS.

6. Timing of survey: Among the practical considerations, determining the "best" time for conducting the GYTS to increase the participation and availability of eligible students is critical and will vary by country. However, it is suggested not to engage in data collection during the examination period (standardized testing days), and immediately before or after the holidays. It is also recommended to conduct the survey in the first half of the academic year (preferably during the first semester) to avoid busy schedules that decrease attendance. It is recommended to conduct the survey during the middle of the morning in each school, and lunch time should be avoided. Later times of the day become poorer choices as students may leave the building early.

If any country implementing the GYTS has questions regarding the sample design requirements and recommendations presented in this section, or if a country would like to request a deviation from them, it should contact CDC or WHO.

5.   GYTS Sample Selection: An Illustrative Example

This chapter provides an illustrative example for a fictitious GYTS country sample design and selection.

5.1   Information Request for Sample Selection

As a first step of the sample design review and finalization process, CDC requests that the country implementing agency and research coordinator provide information required for sample selection prior to the training workshop in the GYTS process. The template used for obtaining such information should be based on the school enrollment list (see **Appendix B**). Additional requirements and selection procedures are discussed during the training workshop and filled out in the sampling parameters worksheet (**Appendix A**). After the review and jointly agreeing on the proposed sample design, the eligible schools are sampled using PC Sample software.

As indicated in the earlier chapters, the sample is drawn using a two-stage cluster design with schools selected at the first stage with a probability proportional to the enrollment size (PPS systematic selection) followed by classes chosen randomly (systematically) within selected schools. All the students within a selected class are eligible to participate in the survey. An important feature of the GYTS sample is that sample substitution is not allowed anywhere in the sampling process (i.e., in choosing schools, classes, or students).

Below is an example of sample selection for a typical GYTS survey (e.g., GYTS Finland, 2012) involving the selection of a nationally representative sample of 20 schools and 1,500 students with an anticipated overall school response rate of 80% and a student response rate of 80%. This implies a selection of 25 (20 divided by 0.80) schools and 1,875 (1,500 divided by 0.80) students from the most current and up-to-date school enrollment list of both private and public schools. The target sample includes grades 7-9 in both private and public schools that include students aged 13-15 years in Finland with an overall enrollment size of 48,677 students.

Before the sample selection process, the final school enrollment list is prepared and reviewed for accuracy. If provided separately, all the regional/provincial specific school enrollment lists are combined to make a final master enrollment list. All the schools are sorted in descending order by enrollment size. Small schools (e.g., with enrollment size less than 40, approximately 1.7% of the total eligible student population) are excluded from the enrollment list file since they are costly to reach and sample. The final school enrollment list is created for PC Sample to select the schools as shown in **Exhibit 5-1**.

5.2   Selection of Schools Using PC Sample

PC Sample software sorts the enrollment list by listing the schools from largest to smallest enrollment and assigns a continuous sequence of unique numeric identifiers starting at one. This is labeled as 'GYTS School ID'. This ID, in addition to the country-provided school ID, is useful to uniquely identify a school from the list along with other school information. This GYTS School ID is of particular importance when a master school enrollment list is pooled together from various regional-specific enrollment lists with duplicate school IDs.

Measure of size (MOS) in the GYTS sample is the number of eligible students in each school. The Sampling Interval (SI) for school sampling by PPS systematic selection is calculated by dividing the total student enrollment (sum of all eligible students in the sampling frame from all schools or sum of all MOS)

by the target number of schools. If the number of eligible students is not available in a country, the total enrollment size of the school can be considered approximations for MOS.

In this example, the sampling interval is 48,677 divided by 25 = 1,947.

The software then scans the list to see if there are any schools with an enrollment size greater than the sampling interval. All the schools with enrollment size greater than 1,947 are called 'Certainty Schools' or 'self-representing schools'. These schools are treated as individual strata when analyzing the data. These schools are excluded from the school sampling frame for the selection of the remaining schools. In the current example, assume there are no such schools in the entire list so that the software selects 25 schools. Otherwise, the software would re-compute the sampling interval for the selection of the remaining schools from the list after excluding the self-representing schools.

The software then calculates the running cumulative enrollment size, which is the sum of the measures of size among all remaining schools on the frame. PC Sample randomly chooses a number between 1 and the sampling interval. This is the Random Start (RS). For example, in this fictitious example, the RS is 1,321.

Calculate the following series: RS; RS + SI; RS + 2SI; RS + 3SI; RS + 4SI; RS + 5SI; RS + 6SI; RS + 7SI; RS + 8SI; RS + 9SI.

Example: RS + 2SI is calculated as 2 times the sampling interval added to the random start. In this case, 1,321 + 2(1,947) = 5,215.

Following the above sequence, all the 25 schools are selected by PC Sample as shown in **Exhibit 5-2**.

5.3   Selection of Classes and Students

When selecting the schools, the software also produces school-level forms and class-level forms for selection of classes (see **Appendix E**). The school-level sample information form includes a sequence of systematic random numbers that are generated by PC Sample for class selection within each school, based on the target student sample and fixed class size values that are supplied as sampling parameters. The school selection list and the school- and class-level sample information forms are provided to the implementing agency or research coordinator for class selections. Based on the class list in each eligible school, classes are randomly selected using the random numbers on the school-level form.

For example, the school with GYTS School ID 24 and enrollment size of 488 has 13 classes associated with grades 7-10 as shown in **Exhibit 5-3.** In the school-level sample information form, the random numbers generated for this school are 2, 6, 11, 15, 20, 24, 29, 33, 38, and 42. Using these random numbers, three classes are selected in this school (i.e. 2, 6 and 11) as shown in **Exhibit 5-4** with student enrollments of 27, 30 and 28, respectively. All the students in these three classes are eligible to participate in completing the GYTS questionnaire. After the selection process of classes is complete, the survey administrator or research coordinator needs to complete the school-level (2-page form) and class-level sample information (1-page form) forms as shown in **Exhibit 5-5** and **Exhibit 5-6**. It is necessary to fill the forms completely and accurately for data management, sample weighting and data analysis. As shown in these exhibits, in the first page of school level sample information form, all the second period classes are used for survey administration to gain the maximum response rate in the selected school.

The specific information that is needed to be completed in this form (see **Exhibit 5-5**) includes: grades taught in the school, school participation status, school eligibility status, and the total number of classes in the selected school. In addition, completion status of the school policy questionnaire needs to be included. In the second page, the survey administrator completes the information on class ID, teacher's name and information on the receipt of completed answer sheets and the class-level sample information form. The information that needs to be completed in the class-level sample information form (see **Exhibit 5-6**) includes: class ID, number of students enrolled in a specific grade, grade level in the class, and number of students that participated in the sampled class. Completed school- and class-level sample information forms should be sent along with completed answer sheets of the GYTS questionnaire and the school policy questionnaire answer sheet. During the process of sample weighting and data analysis, the information provided in these forms will be utilized and a sampling description (see **Appendix G** for a template) will be developed for the purpose of creating a factsheet and country report.

Exhibit 5-1. School Enrollment List Prepared for Sample Selection

| | | | | | | | | Enrollment by Grade and Sex | | | | | | | | | Total School Enrollment (No. of students from all eligible grades) |
| | | | | | | | | Grade 7 | | | Grade 8 | | | Grade 9 | | | |
School ID*	School Name	Address	Province/State	Region	Type of School (Public/Private)	Grades Taught	Level of School (Primary/Middle/Secondary/High)	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	
03219	JONGUHLANGA JS SCHOOL	ROAD,UMTATA,MTH	MTHATHA	MTHATA	Public	1 - 10	Primary/Secondary	100	70	170	59	68	127	50	54	104	401
08587	GUGWINI JS SCHOOL	A/A,LAMBASI,MOUN	MOUNT AYLIFF	LIBODE	Public	6 - 10	Secondary	81	60	141	48	32	80	32	22	54	275
03212	SWANE JS SCHOOL	LOCATION,FLAGSTA	FLAGSTAFF	MTHATA	Public	6 - 10	Secondary	12	5	17	2	5	7	7	4	11	35
03577	MAHLUBINI JS SCHOOL	MAHLUBINI	COFIMVABA	MT FRERE	Public	6 - 10	Secondary	12	6	18	8	10	18	6	3	9	45
03642	GWELANE HIGH SCHOOL	TEKO C,KENTANI,BUTTER WORTH,4961	BUTTERWORTH	MBIZANA	Public	6 - 12	Secondary/High	10	3	13	4	5	9	5	3	8	30
03209	NYANGA SS SCHOOL	KALINYANGA A/A,ENGCOBO,NGC OBO,	NGCOBO	COFIMVABA	Public	6 - 10	Secondary	204	149	353	156	147	303	87	105	192	848
08048	NOZUKO SS SCHOOL	15 IS KALANE STREET,UMTATA,M THATHA,5099	MTHATHA	BUTTERWORTH	Public	6 - 10	Secondary	13	12	25	9	7	16	5	7	12	53
05536	MALUTI SS SCHOOL	RAMOHLAKOANA A/A,MALUTI,MATATI ELE,4740	MATATIELE	TO BE UPDATED	Public	6 - 10	Secondary	14	11	25	14	6	20	13	5	18	63
03176	PAKAMANI SS SCHOOL	MAQWATINI,CALA,5 455	CALA	MTHATA	Public	6 - 10	Secondary	12	10	22	13	6	19	7	8	15	56
03526	MTETUVUMILE SS SCHOOL	MCAMBALALA A/A,COFIMVABA,53 80	COFIMVABA	MALUTI	Public	6 - 10	Secondary	103	70	173	59	56	115	63	68	131	419
03190	GELVANDALE SS SCHOOL	MARTIN,GELVANDA LE,	GELVANDALE	LADY FRERE	Public	6 - 10	Secondary	92	72	164	77	68	145	34	29	63	372

Global Youth Tobacco Survey (GYTS)
School Enrollment List - [Country / Region, Country], [Year]

Exhibit 5-2.   School Selection List

Global Youth Tobacco Survey (GYTS)
[Country, Region] 2012
Selected School List

GYTS School ID*	School ID	School Name	Municipal ID	Name of City/Municipali	Enrollment (Grade 7-9)	School Weight	Within School Weight	Classes
3	08621	Kirkkoharjun koulu	257	Kirkkonummi	577	3.401170026	6.106384912	1, 7, 13, 19, 25, 31, 37, 43, 49, 56
7	03404	Oulunkylän yhteiskoulu	91	Helsinki	503	3.901540965	5.323243693	1, 6, 12, 17, 22, 28, 33, 38, 44, 49
11	08101	Kartanon koulu	186	Järvenpää	474	4.140242838	5.016336999	2, 7, 12, 17, 22, 27, 32, 37, 42, 47
15	08102	Järvenpään Yhteiskoulu	186	Järvenpää	456	4.303673476	4.825843189	4, 9, 14, 19, 23, 28, 33, 38, 43
20	03219	Simonkylän koulu	92	Vantaa	434	4.52183204	4.593017421	2, 6, 11, 16, 20, 25, 29, 34, 39, 43
24	08587	Hyrylän yläaste	858	Tuusula	426	4.606749073	4.508353506	2, 6, 11, 15, 20, 24, 29, 33, 38, 42
29	03212	Kilterin koulu	92	Vantaa	397	4.94326223	4.201446811	3, 7, 12, 16, 20, 24, 28, 33, 37
33	03577	Puistolan peruskoulu	91	Helsinki	381	5.150853294	4.03211898	4, 8, 12, 16, 20, 24, 28, 32, 36
40	03642	Harjun koulu	444	Lohja	371	5.28969031	3.926289086	4, 8, 12, 16, 20, 24, 28, 32, 36
45	03209	Hakunilan koulu	92	Vantaa	363	5.406267507	3.84162517	4, 7, 11, 15, 19, 23, 27, 30, 34
50	08048	Klaukkalan yläaste	543	Nurmijärvi	354	5.543714986	3.746378265	3, 7, 11, 15, 18, 22, 26, 30, 33
56	05536	Sipoonjoen koulu	753	Sipoo	345	5.688333638	3.65113136	2, 5, 9, 13, 16, 20, 24, 27, 31, 35
62	03176	Kasavuoren koulu	235	Kauniainen	333	5.893318634	3.524135487	2, 5, 9, 12, 16, 19, 23, 26, 30, 33
68	03526	Sepän koulu	257	Kirkkonummi	322	6.094643184	3.407722603	3, 7, 10, 14, 17, 20, 24, 27, 31
74	03190	Mikkolan koulu	92	Vantaa	303	6.476815529	3.206645803	3, 6, 9, 13, 16, 19, 22, 25, 29
81	03768	Isoniitun koulu	543	Nurmijärvi	283	6.934541008	2.994986014	1, 4, 7, 10, 13, 16, 19, 22, 25, 28
88	03104	Botby grundskola	91	Helsingfors	273	7.188553499	2.88915612	2, 5, 8, 11, 14, 16, 19, 22, 25
95	07251	Askolan koulu	18	Askola	268	7.322668303	2.836241172	3, 6, 8, 11, 14, 17, 20, 23, 25
103	08084	Härkävehmaan yhtenäisk	106	Hyvinkää	249	7.881426125	2.635164373	3, 6, 8, 11, 13, 16, 19, 21, 24
111	03724	Aurinkolahden peruskoul	91	Helsinki	235	8.350957894	2.487002521	3, 5, 8, 10, 12, 15, 17, 20, 22
119	08607	Otalammen koulu	927	Vihti	220	8.920341387	2.328257679	2, 4, 7, 9, 11, 14, 16, 18, 20
128	03633	Sökövikens skola	49	Esbo	206	9.526578181	2.180095827	1, 3, 5, 7, 9, 12, 14, 16, 18, 20
139	08847	Albert Edelfeltin koulu	638	Porvoo	182	10.78283025	1.92610408	2, 4, 6, 8, 10, 12, 14, 15, 17
151	03001	Ala-Malmin peruskoulu	91	Helsinki	144	13.62829934	1.523950481	1, 3, 4, 6, 7, 9, 10, 12, 13, 15
169	03171	Koulumäen koulu	49	Espoo	44	20.76885333	1	1, 2, 3, 4, 5, 6, 7, 8, 9

*Note: Use the GYTS School ID when filling in the Header Sheet - please see Appendix H

Exhibit 5-3.   Sequential Listing of Classes

School Name: Hyrylän yläaste	GYTS School ID: 24
Classes	No. of students
7A	49
7B	27
7C	39
7D	36
8A	40
8B	39
8C	38
9A	45
9B	37
9C	35
9D	41
10A	29
10B	33

Exhibit 5-4.   Selection of Classes Using Random Numbers on School-Level Sample Information Form

For example, in a school-level sample information form, the following random numbers are generated:

2   6   11   15   20   24   29   33   38   42

School Name: Hyrylän yläaste			GYTS School ID: 24
Classes	No. of students	Random numbers	Selection status
7A	49	1	
7B	27	2	Selected class
7C	39	3	
7D	36	4	
8A	40	5	
8B	39	6	Selected class
8C	38	7	
9A	45	8	
9B	37	9	
9C	35	10	
9D	41	11	Selected class
10A	29	12	
10B	33	13	
		14	
		15	There is no 15th class.
In this school, only three classes (7B, 8B and 9D) are selected to participate in the survey since there are only 13 classes in total.			

Exhibit 5-5.   Properly Completed School-Level Sample Information Form

Global Youth Tobacco Survey (GYTS)
Finland, 2012
School-Level Sample Information Form

PAGE 1 OF 2

GYTS Coordinating Agency: **National Institute for Health and Welfare, Finland**

School: **Hyrylän yläaste**

School ID: **8587**

GYTS School ID: **24**

What grades are taught in this school?   **7 – 9**
(Examples: K-12, 9-12)

School participation status: **PARTICIPATING**   REFUSING   INELIGIBLE
(Circle one)

If the school is ineligible, please explain: **Not Applicable**

If this school refused to participate or is ineligible, you do not need to complete the rest of this form.

Obtain a list of classes in this school to use for sampling. Number the classes on the list beginning with 1. The list of classes must be such that each student in the surveyed grades must be represented on the list ONE AND ONLY ONE time. Refer to the GYTS Implementation Instructions for more information.

Enter the TOTAL NUMBER of eligible classes on the list: **13**

For each number below, select the corresponding class on your list. Keep selecting classes until you reach the END of your class list. If the first number listed below is greater than the number of classes on your list OR if you run out of numbers before you reach the end of your list of classes, contact [Name of the research coordinator and contact information with telephone and email address].

2,   6,   11,   15,   20,   24,   29,   33,   38,   42

Please keep your class list until you receive your report from CDC in case there are any questions concerning these classes. If you prefer not to keep this list, you may send it to [Name of the research coordinator] with these forms. On the next page is a class tracking form. For each class selected, enter the class ID (use the number above that was used to select the class) and the teacher's name. As you receive completed questionnaires and class-level sample information forms, put a check mark in the space provided.

School policy questionnaire completed   Yes   √   No, please explain: _____

PAGE 2 OF 2

GYTS Coordinating Agency: **National Institute for Health and Welfare, Finland**

School: **Hyrylän yläaste**

School ID: **8587**

GYTS School ID: **24**

CLASS TRACKING FORM

Class ID	Teachers Name	Completed answer sheets received	Completed Class-Level Sample Information Form Received
2	Mark	√	√
6	Peter	√	√
11	Lisa	√	√

If you need more lines, make a copy of this page.

Note: Send this form in with your Header Sheet and Answer Sheets/Booklets for this class. After
      completing this form, please send it to Global Tobacco Control Branch in the Office on Smoking
      and Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.
.

------------------------------ FOR OFFICE USE ------------------------------
School sampling interval: 13.6282993416307
Within-School sampling interval: 1.52395048073902
Random number: 0.4076209068

Exhibit 5-6.   Properly Completed Class-Level Sample Information Form

Global Youth Tobacco Survey (GYTS)
Finland, 2012
Class-Level Sample Information Form

Note: Make enough copies of this form so there is one for EACH selected class in each school.

A Class-Level Sample Information Form MUST be completed for each of the classes selected for your survey, whether or not the class participated.

GYTS Coordinating Agency: **National Institute for Health and Welfare, Finland**

School: **Hyrylän yläaste**

School ID: **8587**

GYTS School ID: **24**

Class ID: (**2**)

        (Use the random class number shown on the School-Level Sample Information Form.)

How many students are ENROLLED in this class?   (**27**)

What is the GRADE in this class?   **7**   (8)   9   Other
                                                                (Circle one)

How many students in this class participated in this survey?   (**24**)

If the CLASS did not participate in this survey, please explain:

Note: Send this form in with your Header Sheet and Answer Sheets/Booklets for this class. After completing this form, please send it to Global Tobacco Control Branch in the Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

6.   Sample Weights for GYTS Data

This chapter describes the sample weighting procedure for GYTS data. The weighting procedures described herein are based on the weighting procedures developed for the Youth Risk Behavior Survey (YRBS). Weighting procedures may differ slightly for "atypical" samples which involve additional stratification, oversampling, or coordinated sampling with other school-based surveys.

A *sample weight* is a statistical measurement linked to a data record for any survey respondent in population samples that fully utilize random selection methods to choose the sample. In general terms, an individual sample weight is the inverse of the adjusted probability of obtaining the data for the respondent. In most cases, this probability is the respondent's original selection probability based on the sample design. The inverse probability, or base weight, is adjusted to account for unintended sample imbalance arising during the process of conducting the survey[4]. More than one weight adjustment may be applied, and all weight adjustments are multiplicative. Unless a weight is rescaled for analytic purposes (e.g., "calibrated" to sum to the total respondent sample size), its value cannot 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 school-based student survey like GYTS are gathered for the selected schools and/or classes and if one or more classes are chosen at random in each of those selected schools, separate sets of weights should be produced both for the school/class and for the student data. However, since only student-level data are to be processed in GYTS, only student-level weights will be required.

6.1   General description of weights process

No universally held protocol exists for computing the weights. This is partly because of the variation in circumstances from sample to sample regarding its design, the quality of documentation of the sample selection and recruitment processes, the availability of ancillary information about the sample, and the population to identify and deal with imbalance in the sample due to differential nonresponse. Thus, the actual computational steps for sample weights can vary among surveys. However, some combination of the steps listed below is typically followed in producing a weight for each individual student data record, with the final adjusted weight being the product of the values generated in each step. The weighting process for GYTS data involves a three-step process:

1.   Determine the base weight to account for all steps of random selection that led to the student sample of population members,

2.   Adjust for nonresponse to compensate for sample imbalance due to differential success in sample recruitment, and generally adjust for school-, class- and student-level nonresponse.

3.   And, further adjust to calibrate the final set of adjusted weights to the distribution of the eligible student population by characteristics that are correlated with the key study outcome measures (i.e., grade and gender in GYTS data for the known student population totals), if they are available from the participating country.

---

[4] The concept of "imbalance" here indicates 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 becomes somewhat less representative of the population.

The sequence of steps followed in producing the set of statistical weights for the student sample is important since the weights produced at 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.

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 the 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.

RECOMMENDATION:
The following three-step approach is recommended for each GYTS country sample: (1) compute a base weight for each student sample, (2) adjust the base weights for nonresponse, and (3) calibrate the adjusted weights to known eligible student population totals, if they are available for the country.

One set of weights is produced for each student data record.
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. Weights must also be completed before analysis can begin since the analysts using the survey's data will need them.

The final weight produced for a student record in GYTS data is given by:

$$W = W1 \times W2 \times f1 \times f2 \times f3 \times f4$$

W1 = a school selection weight defined as the inverse of the probability of selecting the school

W2 = a class selection weight defined as the inverse of the probability of selecting a class within a selected school

f1 = a school-level nonresponse adjustment factor calculated by school size category (small, medium, large)

f2 = a class-level nonresponse adjustment factor calculated by school

f3 = a student-level nonresponse adjustment factor calculated by class

f4 = a post-stratification adjustment factor calculated by gender and grade

6.2 Calculating the Base Weight

The base weight for any student respondent in any class within a particular sample school is the product of the school-selection weight (W1) and the class-selection weight (W2) for that school (see for example, the third and fourth columns from the right in **Appendix D**). The W1 and W2 come from the probabilities of selection of schools and then classes within each school, as calculated by PC Sample.

6.2.1 Overall Sampling Fraction

The overall sampling fraction is the targeted probability of selecting a student as specified in the sample design. For example, suppose that a country wants to complete 1,500 questionnaires with secondary school students in grades 7 to 9. If the total school enrollment is 150,000 students in 200 schools, then the overall sampling fraction will be 1,500/150,000 or 1/100 (0.01). However, since we know we are not likely to get exactly 1,500 completed questionnaires due to survey nonresponse at both the school and student levels, the actual formula includes an adjustment to the total school enrollment in grades 7 to 9. The denominator of the overall sampling fraction is multiplied by the product of the expected student response rate and the expected school response rate. Thus in our hypothetical example, suppose that the school response rate is expected to be 0.95 and the student response rate is thought to be 0.85 (i.e., due to absences on the survey day or refusal from student to participate). Thus instead of 1/100, the overall sampling fraction will be 1/80.75 (0.012384), where 80.75 = (100)(0.95)(0.85).

$$\text{Overall sampling fraction (f)} = \frac{\text{Target sample size}}{\text{Total school enrollment in sampling frame}}$$

or

$$\text{Overall sampling fraction (f)} = \frac{\text{Target Sample Size}}{(\text{Student RR} \times \text{School RR} \times \text{Total school enrollment in sampling frame})}$$

Target Sample Size = Target Number of Students to Sample
Student RR = Expected Student Response Rate
School RR = Expected School Response Rate

This overall sampling fraction can be broken into component probabilities--the probability of selecting a school and the probability of selecting a class within the selected school.

In the resulting sample of students selected by PC Sample, each student has the same chance of being selected for the survey.

6.2.2 First Stage Sampling Fraction (Probability of Selecting a School)

The next step involves calculating the probabilities of selection for schools (first stage) and then classes (second stage). Schools are selected using systematic sampling with a random start and Probabilities Proportional to Size (PPS). Probabilities of school selection are proportional to a measure of size (MOS) that is based on the enrollment for each school. Except for very large and very small schools, the measure of size is exactly equal to the enrollment in the target grades. Prior to sampling, schools on the frame are sorted in descending order of size. Depending on the number of schools that are to be sampled, some very large schools may be selected with certainty. As each "certainty" school is selected, it is removed from the frame of eligible schools.

Determining Certainty Schools

An initial sampling interval is calculated by dividing the sum of the enrollments of all the schools in the sampling frame by the number of schools desired in the sample. GYTS usually selects 25 schools for a sample of 1,875 students. If the total enrollment for all schools in the sampling frame equals 75,000, then the sampling interval will be 3,000 (75,000/25=3,000). Schools that have an enrollment greater than or equal to 3,000 are treated as certainty schools and are removed from the sampling frame. Certainty schools are always selected for the sample. Each time a school is selected with certainty and removed from the frame, the sampling interval is recomputed based on the enrollment of the schools remaining on the sampling frame and on the number of schools remaining to be selected.

$$\text{Revised Sampling Interval} = \frac{\text{Total school enrollment in revised sampling frame}}{\text{Adjusted number of schools}}$$

Sampling of certainty schools continues until the enrollment of the largest school remaining on the frame is less than the revised sampling interval. At this point, sampling of certainty schools is complete.

Sampling fraction for certainty schools: $\quad f_{\text{school}} = 1$

The school sampling weight (W1) is simply the inverse of these sampling intervals, depending on whether the school is certainty or not.

Determining Non-certainty Schools

If more schools are needed in the sample after the certainty schools have been selected, they are sampled from the schools remaining on the frame. The sampling procedure for these "non-certainty schools" includes adjustments to the measure of size for schools that have very small enrollments. This procedure ensures that each student has the same probability of selection for the sample and that this probability is equal to the overall sampling fraction. Except for certainty schools and schools for which the minimum measure of size was used in sampling, although the actual number may vary, the expected number of students selected per school is the same in all schools.

In our example, if there was one school with an enrollment of 3,500, the new sampling interval would be 2,979 (71,500/24=2,979) rather than 3,000. If there are no schools with an enrollment as large as 2,979, non-certainty schools are now selected using systematic sampling and the recalculated sampling interval. Probabilities for these schools are proportional to school enrollment. Special adjustments are made to MOS for very small schools as described below.

6.2.3   Adjusting for Schools with Very Small Enrollments

For non-certainty schools with very small enrollments, it is possible that the probability of selection based on enrollment is so small that the overall probability of selection for students in these schools may be less than the required overall rate. Specifically, this happens when the school enrollment is so small that, even if all students were selected with certainty, their probability of selection would not be equal to "f", the overall probability. Therefore, an adjustment is made to the measure of size for small schools so that students from these schools will have the required overall probability of selection. Essentially, the

probability of selection for the small schools is increased so that selecting students with certainty from these schools will match the overall probability of selection for all students. This has the effect of slightly decreasing the probabilities of selection for the larger, non-certainty schools. The number of students sampled from the small schools is thus slightly larger than it would be ordinarily.

After the certainty schools have been removed from the sampling frame, the remaining schools are ordered according to decreasing enrollment. It is then necessary to determine a minimum measure of size MINMOS for each school. When a school's enrollment falls below $MINMOS_i$, the actual enrollment is replaced by $MINMOS_i$. The procedure that PC Sample uses for determining $MINMOS_i$ is relatively simple. For the first and largest school remaining on the frame $MINMOS_i = 0$ (remember the schools are sorted in descending order by size). Thereafter, each succeeding school will have:

$$MINMOS_i =$$

$$\frac{\text{Overall sampling fraction f} \times \text{Sum of the enrollments of all schools preceding school}_i}{\left[\text{Number of schools remaining to be selected} - \left(f \times \left(\text{Number of schools in the adjusted frame} - \text{Number of schools preceding school}_i + 1\right)\right)\right]}$$

In a real world example, a country had 737 schools with a total enrollment of 583,396 students on their sampling frame. The country asked that schools with enrollments of 40 or less be eliminated, and so the frame finally contained 723 schools with a total enrollment of 583,313. The country wanted to select 60 schools to ensure a minimum of 2,500 completed questionnaires from students. The overall sampling interval was 9,722 (583,313/60) students, and no school exceeded this enrollment meaning that no certainty schools were selected from the frame. The overall sampling fraction was 0.0043 (2,500/583,313). Even after eliminating schools with enrollments of 40 or less, there still remained 14 schools with enrollments of less than 50 students. $MINMOS_i$ increases as the enrollment decreases, and it reached a maximum of 52.10641 before the actual enrollment of the remaining 14 schools fell below this number. Thus, the revised MOS for these 14 schools became equal to the maximum MINMOS of 52.10641.

Selection of non-certainty schools is carried out using systematic sampling with a random start and an adjusted school sampling interval that uses a total enrollment based on the revised MOS_i.

$$\text{Adjusted school sampling interval} = \frac{\text{Total enrollment based on revised MOS}_i}{\text{Number of non} - \text{certainty schools required}}$$

The default selection procedure uses implicit stratification based on school enrollment. This procedure helps to ensure that schools of varying sizes are selected and helps to control the precision of estimates.

Thus in our real world example, the adjusted school sampling interval is 9,726 (583,571/60) instead of 9,722.

6.2.4   Second Stage Sampling Fraction (Probability of selection within school)

Certainty Schools

Returning to our original example for simplicity, we see that all students in certainty schools have an overall probability of selection equal to the overall sampling rate (0.01 in our original example).

$$f_{class} = f$$

$$\text{Within school sampling inteval } = \frac{1}{\text{Overall sampling fraction}} = \frac{1}{0.01} = 100$$

Non-certainty Schools

For non-certainty schools, the within-school sampling probabilities are based on the adjusted school sampling interval and on the adjusted school probability of selection so that

$$\text{Adjusted school probability } (f_{school}) = \frac{\text{Adjusted school measure of size}}{\text{Adjusted school sampling interval}} \text{ , and}$$

$$\text{Within school interval} = \frac{\text{Adjusted school probability}}{\text{Overall sampling fraction}}$$

Assuming in our original example that no schools had an enrollment less than MINMOS$_i$, a school with 987 students would have a within-school probability of selection of:

$$f_{school} = \frac{987}{2,979} = 0.33$$

For a school with an enrollment of 600, the comparable calculation would be:

$$f_{school} = \frac{600}{2,979} = 0.20$$

The within-school intervals in each case are 33 (0.33/0.01) for the school with 987 students and 20 (0.20/0.01) for the school with 600 students.

The second stage probability for selecting classes within a school is calculated by dividing the overall sampling fraction by the school probability of selection.

$$f_{class} = \frac{0.01}{0.33} = 0.0303, \text{ if the school has 987 students and}$$

$$f_{class} = \frac{0.01}{0.20} = 0.05, \text{ with an enrollment of 600 students}$$

The product of the first- and second- stage sampling fractions should equal the overall sampling fraction. Thus, for the school with 987 students

$$f = f_{school} \times f_{class} = 0.33 \times 0.0303 = 0.01$$

The corresponding product for the school with 600 students would be:

$$f = f_{school} \times f_{class} = 0.20 \times 0.05 = 0.01$$

This two-stage sampling procedure yields an overall sample with probability of selection for each student equal to the overall sampling fraction. Within each selected school, the within-school sampling interval is applied to a random start. Except for certainty schools and schools for which the minimum measure of size was used in sampling, the expected number of students selected per school is the same in all schools. The actual number of students selected depends on several other variables, including the accuracy of the reported enrollments, the variability of class size, and the number of classes selected in each school. For the school with 987 students the random start is, say, 22 (the random start must be within the sampling interval) and so the following classes are selected for the survey: 22, 55, 88, 121, and 154, for example. It is expected that only the first one to four classes will actually be selected. However, the sampling interval is provided with the sample documentation so that more classes can be chosen if necessary.

6.2.5   School Selection Weight

The school selection weight (W1) is the inverse of the school (first stage) sampling fraction.

Thus for the school with 987 students, the weight is: $\dfrac{1}{\text{Probability of selecting a school}} = \dfrac{1}{0.33} = 3.03$

For the school with 600 students, the weight is: $\dfrac{1}{\text{Probability of selecting a school}} = \dfrac{1}{0.20} = 5.0$

6.2.6   Class Selection Weight

The class selection weight (W2) is the inverse of the class (second stage) sampling fraction.

The class selection weight for the school with 987 students is:

$$\frac{1}{\text{Probability of selecting a class (student) within a school}} = \frac{1}{0.0303} = 33.0$$

The class selection weight for the school with 600 students is:

$$\frac{1}{\text{Probability of selecting a class (student) within a school}} = \frac{1}{0.05} = 20.0$$

6.3 Non-Response Adjustments

An adjustment is made for school non-response, class non-response, and student non-response. The purpose of the non-response adjustments is to refine the weights to adjust for bias from non-response.

f1 - School Non-Response Adjustment

Schools are placed into 3 categories – large, medium, and small – using tertiles of enrollment. For each group, a school non-response adjustment factor is calculated as follows:

$$\text{School nonresponse adjustment factor} =$$

$$\frac{\sum_{\text{selected schools in tertile}}(\text{School selection weight} \times \text{School enrollment})}{\sum_{\text{participating schools in tertile}}(\text{School selection weight} \times \text{School enrollment})}$$

f2 - Class Non-Response Adjustment

For situations in which entire classes within a participating school do not respond, a class adjustment factor is computed for each school in the following manner:

$$\text{Class nonresponse adjustment factor} = \frac{\text{Number of classes selected in the school}}{\text{Number of classes participating in the school}}$$

f3 - Student Non-Response Adjustment

Within each class, a student-level adjustment is made for students who are non-respondents. The adjustment is computed by doing the following:

$$\text{Student nonresponse adjustment factor} = \frac{\sum_{\text{eligible students in the class}}(\text{Student weight})}{\sum_{\text{completed questionnaire in the survey}}(\text{Student weight})}$$

Overall Non-Response Adjustment

Of the components in the overall non-response adjustment, the school adjustment factor will be the same for all units regardless of school or class. The class adjustment will be the same for all students in a school. The student adjustment factor will be the same for students within the same class. If all of the selected classes participate, the class adjustment factor drops out of the calculation.

$$\text{Overall nonresponse adjustment} =$$

$$\text{School adjustment factor} \times \text{Class adjustment factor} \times \text{Student adjustment factor}$$

6.4 Post-Stratification Adjustment

For the post-stratification adjustment (f4), the sample data is adjusted to match the school population data using known frequencies. The post-stratification is done by gender and grade. Students in grades that are "out-of-scope" are taken out of the data. For students who have missing data on grade or gender, the values are imputed as described below. If these frequencies are not available for any GYTS samples, then post-stratification adjustment is not generally done. The nonresponse-adjusted weight might be sufficient if the population frequencies used for post-stratification adjustment are unavailable.

For respondents missing gender:

The respondent is randomly assigned as a 'male' or 'female' value based on a uniform distribution of random numbers.

For respondents missing grade:

The respondent is assigned a grade value based on the average grade (rounded) of the student's school.

(Note: The values are only imputed for weighting purposes. The imputed values are not retained in the final data set).

$$\text{Post stratification adjustment} = \frac{\text{Population counts (grade–gender category)}}{\text{Sum of adjusted weights (grade–gender category)}}$$

For GYTS data sets, the intention is to post-stratify by grade and gender whenever possible.

6.5   Final Weight

The final weight is the product of the school selection weight (W1), class selection weight (W2), overall nonresponse adjustment (f1×f2×f3), and post stratification adjustment (f4).

$$\text{Final weight} =$$

$$\text{School selection weight} \times \text{Class selection weight} \times \text{Overall nonresponse adjustment} \times$$
$$\text{Post stratification adjustment}$$

6.6   Stratum and PSU Assignment for Complex GYTS Data Analysis

The assignment of strata and PSUs is as follows.

Certainty Schools

If the school was selected with certainty (probability =1), then the school forms its own stratum and the classes within the schools are identified as the PSUs.

Non-Certainty Schools

The non-certainty strata are sorted from largest to smallest. The schools are then paired together to form strata for variance estimation during analysis. The schools within each stratum are identified as the PSUs. In the case of an odd number of non-certainty schools, the last 3 smallest schools are placed into the same stratum.

7. Bibliography

Barnett V. Sample Survey Principles and Methods, Oxford University Press; 2002.

Centers for Disease Control and Prevention (CDC). 2009 Youth Risk Behavior Survey (YRBS).National YRBS Data Users Manual; 2009. http://www.cdc.gov/HealthyYouth/yrbs/pdf/national_usersmanual_yrbs.pdf

Centers for Disease Control and Prevention (CDC). Methodology of the Youth Risk Behavior Surveillance System. MMWR Recommendations and Reports. 53, RR12, 1-13; 2004. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5312a1.htm

Centers for Disease Control and Prevention (CDC). Youth Tobacco Survey –Survey Administrator Handbook. Available at: http://www.cdc.gov/tobacco/data_statistics/surveys/yts/pdfs/ytsHandbook.pdf

Chaudhuri A. Survey Sampling: Theory and Methods, M. Dekker, New York; 2005.

Cochran WG. Sampling Techniques, 3rd Ed., New York: Wiley and Sons; 1977.

Deville JC, Sarndal CE. Calibration Estimators in Survey Sampling. Journal of American Statistics Association 1992; 87, 376-382.

Global Adult Tobacco Survey Collaborative Group. Global Adult Tobacco Survey (GATS): Sample Design Manual, Version 2.0. Atlanta, GA: Centers for Disease Control and Prevention; 2010.

Global Adult Tobacco Survey Collaborative Group. Global Adult Tobacco Survey (GATS): Sample Weights Manual, Version 2.0. Atlanta, GA: Centers for Disease Control and Prevention; 2010.

Global Youth Tobacco Survey Collaborative Group. Global Youth Tobacco Survey (GYTS): Core Questionnaire with Optional Questions, Version 1.0.; July 2012.

Global Youth Tobacco Survey Collaborative Group. Global Youth Tobacco Survey (GYTS): Implementation Instructions, Version 1.0.; August 2012.

Grosh M, Munoz J. A Manual for Planning and Implementing the Living Standard Measurement Survey Study. LSMS Working Paper No. 126. World Bank, Washington, DC; 1996.

Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. Survey Methodology, Wiley and Sons, New York; 2004.

Henry GT. Practical Sampling, Sage Publications, Newbury Park; 1990.

Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 1952; 47, 663-685.

Iannacchione VG, Milne JG, Folsom RE. Response Probability Weight Adjustments Using Logistic Regression. Proceedings of the Section on Survey Research Methods, American Statistical Association, 637-642; 1991.

Kalton G. Compensating for Missing Survey Data. Ann Arbor, MI: University of Michigan; 1983.

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

Kalton G. Introduction to Survey Sampling, Sage Publications; 1983.

Kish L. Survey Sampling, Second Printing, New York: Wiley and Sons; 1965.

Krewski D. Current Topics in Survey Sampling, Academic Press, New York; 1981.

Lepkowski JM, Kalton G, Kasprzyk D. Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel. Proceedings of the Section on Survey Research Methods, American Statistical Association, 296-301; 1989.

Lessler JT, Kalsbeek WD. Nonsampling Error in Surveys. New York: Wiley and Sons; 1992.

Levy PS, Lemeshow S. Sampling of Populations — Methods and Applications, Wiley, New York; 1999.

Lohr S. Sampling: Design and Analysis, Pacific Grove: Duxbury Press; 1999.

Macro International. Sampling Manual. DHS Basic Documentation No. 6. Calverton, MD; 1996.

Murthy MN. Sampling Theory and Methods, Statistical Publishing Society, Calcutta; 1967.

Namboodiri NK. Survey Sampling and Measurement, Academic Press, New York; 1978.

PC Sample Manual and Users Guide 2004.

Potter FJ. The Effect of Weight Trimming on Nonlinear Survey Estimates. Proceedings of the Section on Survey Research Methods, American Statistical Association, 758-763; 1993.

Raj, D. (1968). Sampling Theory, New York: McGraw-Hill.

Sarndal, Carl-Erik, Swensson, Bengt, and Wretman, Jan H'akan, Model Assisted Survey Sampling, Springer-Verlag, New York; 1992.

Scheaffer RL, Mendenhall W, and Ott L. Elementary Survey Sampling (6th Edition), Duxbury Press; 2006.

Stuart A. Basic Ideas of Scientific Sampling, 2nd Ed., Griffin's Statistical Monographs and Courses, No. 4, Hafner Press, New York; 1976.

Sukhatme PV, et al. Sampling Theory of Surveys with Applications, Ames: Iowa State University Press; 1984.

Thompson SK. Sampling, Wiley, New York; 2002.

TIMSS (1998a). Survey Operations Manual–Main Survey (Doc. Ref.: TIMSS 1999 98-0026). Prepared by the International Study Center. Chestnut Hill, MA: Boston College.

Trends in International Mathematics and Sciences Study (TIMSS, 1997). TIMSS 1999 School Sampling Manual–Version 2 (Doc. Ref.: TIMSS 1999 97-0012). Prepared by Pierre Foy, Statistics Canada. Chestnut Hill, MA: Boston College. http://timss.bc.edu/timss1999i/pdf/T99_TR_Chap02.pdf

United Nations. Household Sample Surveys in Developing and Transition Countries. New York, NY; 2005.

Warren CW, Asma S, Lee J, Lea V, Mackay J. Global Tobacco Surveillance System: The GTSS Atlas. CDC Foundation, Myriad Editions, UK; 2009.

Warren CW, Jones NR, Peruga A, Chauvin J, Baptiste JP, Costa de Silva V, el Awa F, Tsouro A, Rahman K, Fishburn B, Bettcher DW, Asma S. Global youth tobacco surveillance, 2000-2007. MMWR Surveill Summ. 2008 Jan 25;57(1):1-28.

Wolter KM. Introduction to Variance Estimation, Springer-Verlag, New York; 1985.

Yates Frank. Sampling Methods for Censuses and Surveys, Macmillan, New York; 1981.

Appendix A: GYTS Sampling Parameters Worksheet

Date:___/____/_____

Country:

Implementing Agency: _____

Contact Name: _____

Mailing Address: _____

Telephone Number: _____

Email Address: _____

1. At what level is the GYTS to be implemented? (check all that apply):
    [ ] *National level*
    [ ] *Regional level*
    [ ] *Both*
    [ ] *Other:_____*

2. Domain/strata for which estimates are required:
    a. *Number: _____*
    b. *List (e.g., region, type of school), please specify:_____*

3. What type of school is eligible for selection?
    [ ] *Public*
    [ ] *Private*
    [ ] *Both*
    [ ] *Other, please specify:_____*

4. Type of sampling:
    [ ] *Sample*
    [ ] *Census*

5. Grades and number of students to be sampled:
    a. Grades:_____
    b. Number of students to be sampled (Target sample size):_____

6. Which grades and school levels are associated with the ages below in your country?

Age	Grades	School level (e.g., primary, secondary, middle, high)
11		
12		
13		
14		
15		
16		
17		

7. Requirement of sample selection:

Domain/Strata Name/Level	No. of Schools (without Inflation)	School Response Rate	No. of Students (without Inflation)	Student Attendance Rate	Average Class Size
		%		*%*	
		%		*%*	
		%		*%*	
		%		*%*	

8. Small School Exclusion:
There may be some schools in the country's enrollment list with very few eligible students. In order to maximize survey efficiency and school response rate, some countries decide to exclude schools with enrollment of eligible students less than a pre-specified cut-off point (e.g., 40 students). This cut-off point may vary from country to country.

Exclude from the enrollment list any schools with enrollments below …
   [ ] *Thirty students*
   [ ] *Thirty five students*
   [ ] *Forty students*
   [ ] *Other : _____*
   [ ] *Do not eliminate any schools*

9. Does the GYTS need to be coordinated with other surveys:
   [ ] *Yes …………………If yes,*
                   a. Please specify survey:_____
                   b. Sample double draw [ ]
   [ ] *No*

10. Materials needed:
   Number of answer sheets: _____
   Number of header sheets: _____
   Type of courier: _____
   [ ] *FedEx*
   [ ] *DHL*
   [ ] *Other:_____*

11. Expected date to receive school enrollment list: _____

12. Expected fieldwork/data collection period: _____ Month(s) _____ Year(s)

13. Other Notes (oversampling requirements, etc.):

Appendix B: Sample GYTS School Enrollment List

SAMPLE GYTS SCHOOL ENROLLMENT LIST

This school enrollment list shows the information that CDC/OSH/GTCB needs in order to draw a sample for the GYTS. The file must formatted into MS-Excel and the data layout below illustrates the minimum information needed in order to draw your sample. Separate files should be submitted for your MS and HS frames. School enrollment counts are used as the sampling frame for GYTS eligible school samples. Please make sure that your enrollment numbers are listed as whole numbers without decimals, fractions or alpha-numeric designators. Enrollment list that do not meet these minimum specifications will be returned for correction and resubmission.

Please note that student demographic information (enrollment information by grade and sex) is not needed in order to draw a GYTS sample. However, participating countries will need to provide this information in order to weight the final GYTS datasets. Note that weighting, and thus data analysis, cannot begin until such information is received.

NOTES: (1) Enrollment list do not have to follow this exact format and can be adapted as per country requirements (e.g., adding eligible additional grades) as long as they still clearly provide the same information displayed below. (2) Ensure that the enrollment figures only include your eligible student population (target grades representing students aged 13-15 years); (3) In the event your sample design includes stratification or regional sampling, it will be necessary to include some type of regional identifiers in your enrollment list.

EXAMPLE: This example provides the suggested enrollment information needed to select your sample and weight the data. As noted above, the enrollment counts broken out by grade and sex are necessary to weight the sample, and thus data analysis cannot be conducted without receiving complete information as per enrollment list. Information in the below table is an example only and does not necessarily represent the actual school enrollments or identifiers.

Global Youth Tobacco Survey (GYTS)
School Enrollment List - [Country / Region, Country], [Year]

School ID*	School Name	Address	Province/State	Region	Type of School (Public/Private)	Grades Taught	Level of School (Primary/Middle/Secondary/High)	Grade 7 Boys	Grade 7 Girls	Grade 7 Total	Grade 8 Boys	Grade 8 Girls	Grade 8 Total	Grade 9 Boys	Grade 9 Girls	Grade 9 Total	Total School Enrollment (No. of students from all eligible grades)
03219	JONGUHLANGA JS SCHOOL	ROAD UMTATA,MTH	MTHATHA	MTHATHA	Public	1 - 10	Primary/Secondary	100	70	170	59	68	127	50	54	104	401
06587	GUGWINI JS SCHOOL	AVA LAMBASI IMOUN	MOUNT AYLIFF	LIBODE	Public	6 - 10	Secondary	81	60	141	48	32	80	32	22	54	275
03212	SWANE JS SCHOOL	LOCATION FLAGSTA	FLAGSTAFF	MTHATA	Public	6 - 10	Secondary	12	5	17	2	5	7	7	4	11	35
03577	MAHLUBINI JS SCHOOL	MAHLUBINI	COFIMVABA	MT FRERE	Public	6 - 10	Secondary	12	6	18	8	10	18	6	3	9	45
03842	GWELANE HIGH SCHOOL	TEKO C.KENTANI BUTTER WORTH 4961	BUTTERWORTH	MBIZANA	Public	6 - 12	Secondary/High	10	3	13	4	5	9	5	3	8	30
03209	NYANGA SS SCHOOL	KALINYANGA AVA ENGCOBO,NGC OBO	NGCOBO	COFIMVABA	Public	6 - 10	Secondary	204	149	353	156	147	303	87	105	192	848
06048	NQZUKO SS SCHOOL	15 IS KALANE STREET UMTATA,MT HATHA,5099	MTHATHA	BUTTERWORTH	Public	6 - 10	Secondary	13	12	25	9	7	16	5	7	12	53
05536	MALUTI SS SCHOOL	RAMOHLAKOANA AVA MALUTI,MATATI ELE,4740	MATATIELE	TO BE UPDATED	Public	6 - 10	Secondary	14	11	25	14	6	20	13	5	18	63
03176	PAKAMANI SS SCHOOL	MAQWATINI CALA.5 455	CALA	MTHATA	Public	6 - 10	Secondary	12	10	22	13	6	19	7	8	15	56
03526	MTETUVUMILE SS SCHOOL	MCAMBALALA AVA COFIMVABA,538 0	COFIMVABA	MALUTI	Public	6 - 10	Secondary	103	70	173	59	56	115	63	68	131	419
03190	GELVANDALE SS SCHOOL	MARTIN GELVANDAL E	GELVANDALE	LADY FRERE	Public	6 - 10	Secondary	92	72	164	77	68	145	34	29	63	372

*The School ID may be unique to the school. If no unique School ID exists then number the schools sequentially.

Appendix C: Rationale for Target Respondent Sample Size

CDC works with each country to determine the target sample size depending on the level of precision required by the country and the enrollment size of the population of interest.

GYTS Sample Size

Eligible Student Population	Margin of Error				
	±1%	±2%	±3%	±4%	±5%
500	494	475	448	414	377
1,000	975	906	810	706	606
2,000	1,901	1,655	1,362	1,091	869
3,000	2,783	2,286	1,762	1,334	1,016
4,000	3,623	2,824	2,065	1,501	1,110
5,000	4,424	3,288	2,303	1,623	1,176
6,000	5,190	3,693	2,494	1,715	1,224
7,000	5,921	4,049	2,652	1,788	1,260
8,000	6,621	4,365	2,783	1,847	1,289
9,000	7,292	4,646	2,895	1,896	1,313
10,000	7,935	4,899	2,992	1,936	1,332
20,000	13,153	6,488	3,518	2,144	1,427
30,000	16,845	7,275	3,737	2,223	1,462
40,000	19,596	7,745	3,857	2,265	1,480
50,000	21,725	8,059	3,933	2,291	1,491
60,000	23,421	8,279	3,985	2,309	1,498
70,000	24,804	8,445	4,023	2,321	1,504
80,000	25,953	8,575	4,052	2,331	1,508
90,000	26,924	8,678	4,075	2,339	1,511
100,000	27,754	8,762	4,094	2,345	1,513
200,000	32,226	9,164	4,179	2,373	1,525
300,000	34,055	9,306	4,209	2,382	1,529
400,000	35,050	9,379	4,223	2,387	1,531
500,000	35,675	9,423	4,232	2,390	1,532
1,000,000	36,995	9,513	4,250	2,395	1,534

Note: The design effect in the GYTS countries may be considerably higher than that was assumed in preparing this table. As a result, required sample sizes to obtain the indicated precision levels could much higher than that were shown in this table.

Appendix D: School Selection List Template

Global Youth Tobacco Survey (GYTS)
[Country, Region] 2012
Selected School List

GYTS School ID*	School ID	School Name	Municipal ID	Name of City/Municipality	Enrollment (Grade 7-9)	School Weight	Within School Weight	Classes
3	08621	Kirkkoharjun koulu	257	Kirkkonummi	577	3.401170026	6.10384912	1, 7, 13, 19, 25, 31, 37, 43, 49, 56
7	03404	Oulunkylän yhteiskoulu	91	Helsinki	503	3.901540965	5.323243693	1, 6, 12, 17, 22, 28, 33, 38, 44, 49
11	08101	Kartanon koulu	186	Järvenpää	474	4.140242838	5.016336999	2, 7, 12, 17, 22, 27, 32, 37, 42, 47
15	08102	Järvenpään Yhteiskoulu	186	Järvenpää	456	4.303673476	4.825843189	4, 9, 14, 19, 23, 28, 33, 38, 43
20	03219	Simonkylän koulu	92	Vantaa	434	4.52183204	4.593017421	2, 6, 11, 16, 20, 25, 29, 34, 39, 43
24	08587	Hyrylän yläaste	858	Tuusula	426	4.606749073	4.508353506	2, 6, 11, 15, 20, 24, 29, 33, 38, 42
29	03212	Kilterin koulu	92	Vantaa	397	4.94326223	4.201446811	3, 7, 12, 16, 20, 24, 28, 33, 37
33	03577	Puistolan peruskoulu	91	Helsinki	381	5.150853294	4.03211898	4, 8, 12, 16, 20, 24, 28, 32, 36
40	03642	Harjun koulu	444	Lohja	371	5.28969031	3.926289086	4, 8, 12, 16, 20, 24, 28, 32, 36
45	03209	Hakunilan koulu	92	Vantaa	363	5.406267507	3.84162517	4, 7, 11, 15, 19, 23, 27, 30, 34
50	08048	Klaukkalan yläaste	543	Nurmijärvi	354	5.543714986	3.746378265	3, 7, 11, 15, 18, 22, 26, 30, 33
56	05536	Sipoonjoen koulu	753	Sipoo	345	5.688333638	3.65113136	2, 5, 9, 13, 16, 20, 24, 27, 31, 35
62	03176	Kasavuoren koulu	235	Kauniainen	333	5.893318634	3.524135487	2, 5, 9, 12, 16, 19, 23, 26, 30, 33
68	03526	Sepän koulu	257	Kirkkonummi	322	6.094643184	3.407722603	3, 7, 10, 14, 17, 20, 24, 27, 31
74	03190	Mikkolan koulu	92	Vantaa	303	6.476815529	3.206645803	3, 6, 9, 13, 16, 19, 22, 25, 29
81	03768	Isoniitun koulu	543	Nurmijärvi	283	6.934541008	2.994986014	1, 4, 7, 10, 13, 16, 19, 22, 25, 28
88	03104	Botby grundskola	91	Helsingfors	273	7.188553499	2.88915612	2, 5, 8, 11, 14, 16, 19, 22, 25
95	07251	Askolan koulu	18	Askola	268	7.322668303	2.836241172	3, 6, 8, 11, 14, 17, 20, 23, 25
103	08084	Härkävehmaan yhtenäisk	106	Hyvinkää	249	7.881426125	2.635164373	3, 6, 8, 11, 13, 16, 19, 21, 24
111	03724	Aurinkolahden peruskoul	91	Helsinki	235	8.350957894	2.487002521	3, 5, 8, 10, 12, 15, 17, 20, 22
119	08607	Otalammen koulu	927	Vihti	220	8.920341387	2.328257679	2, 4, 7, 9, 11, 14, 16, 18, 20
128	03633	Sökvikens skola	49	Esbo	206	9.526578181	2.180095827	1, 3, 5, 7, 9, 12, 14, 16, 18, 20
139	08847	Albert Edelfeltin koulu	638	Porvoo	182	10.78283025	1.92610408	2, 4, 6, 8, 10, 12, 14, 15, 17
151	03001	Ala-Malmin peruskoulu	91	Helsinki	144	13.62829934	1.523950481	1, 3, 4, 6, 7, 9, 10, 12, 13, 15
169	03171	Koulumäen koulu	49	Espoo	44	20.76885333	1	1, 2, 3, 4, 5, 6, 7, 8, 9

*Note: Use the GYTS School ID when filling in the Header Sheet

Appendix E: School-Level and Class-Level Sample Information Forms - Template for GYTS

<div align="center">

Global Youth Tobacco Survey (GYTS)
[Country] – [Region], [Year]
School-Level Sample Information Form

</div>

PAGE 1 OF 2

GYTS Coordinating Agency: [Name of the coordinating agency]

School:　　　　[Name of the school]

School ID:　　　[Unique school identifier]

GYTS School ID:　　　[Unique identifier for sampling at CDC]

What grades are taught in this school? _____
<div align="center">(Examples: K-12, 9-12)</div>

School participation status:　　　PARTICIPATING　　REFUSING　　INELIGIBLE
<div align="center">(Circle one)</div>

If the school is ineligible, please explain: _____

If this school refused to participate or is ineligible, you do not need to complete the rest of this form.

Obtain a list of classes in this school to use for sampling. Number the classes on the list beginning with 1. The list of classes must be such that each student in the surveyed grades must be represented on the list ONE AND ONLY ONE time. Refer to the GYTS Implementation Instructions for more information.

Enter the TOTAL NUMBER of eligible classes on the list: _____

For each number below, select the corresponding class on your list. Keep selecting classes until you reach the END of your class list. If the first number listed below is greater than the number of classes on your list OR if you run out of numbers before you reach the end of your list of classes, contact [Name of the research coordinator and contact information with telephone and email address].

<div align="center">2　6　11　15　20　24　29　33　38　42</div>

Please keep your class list until you receive your report from CDC in case there are any questions concerning these classes. If you prefer not to keep this list, you may send it to [Name of the research coordinator] with these forms. On the next page is a class tracking form. For each class selected, enter the class ID (use the number above that was used to select the class) and the teacher's name. As you receive completed questionnaires and class-level sample information forms, put a check mark in the space provided.

School policy questionnaire completed　　　Yes　[ ]　　No, please explain:_____

Global Youth Tobacco Survey (GYTS)
[Country] – [Region], [Year]
School-Level Sample Information Form

GYTS Coordinating Agency: [Name of the coordinating agency]

School: [Name of the school]

School ID: [Unique school identifier]

GYTS School ID: [Unique identifier for sampling at CDC]

CLASS TRACKING FORM

Class ID	Teachers Name	Completed answer sheets received	Completed Class-Level Sample Information Form Received

If you need more lines, make a copy of this page.

Note: Send this form in with your Header Sheet and Answer Sheets/Booklets for this class. After completing this form, please send it to Global Tobacco Control Branch in the Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.
.

------------------------------- FOR OFFICE USE -------------------------------
School sampling interval: 13.6282993416307
Within-School sampling interval: 1.52395048073902
Random number: 0.4076209068

Global Youth Tobacco Survey (GYTS)
[Country] – [Region], [Year]
Class-Level Sample Information Form

Note: Make enough copies of this form so there is one for EACH selected class in each school.

A Class-Level Sample Information Form MUST be completed for each of the classes selected for your survey, whether or not the class participated.

GYTS Coordinating Agency: [Name of the coordinating agency]

School: [Name of the school]

School ID: [Unique school identifier]

GYTS School ID: [Unique identifier for sampling at CDC]

Class ID: _____

        (Use the random class number shown on the School-Level Sample Information Form.)

How many students are ENROLLED in this class? _____

What is the GRADE in this class?   [list eligible grades]   Other
                                                                (Circle one)

How many students in this class participated in this survey? _____

If the CLASS did not participate in this survey, please explain:

Note: Send this form in with your Header Sheet and Answer Sheets/Booklets for this class. After completing this form, please send it to Global Tobacco Control Branch in the Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

Appendix F: School-Level and Class-Level Sample Information Forms - Template for Double Draw

Global Youth Tobacco Survey (GYTS) / Survey X
[Country] – [Region], [Year]
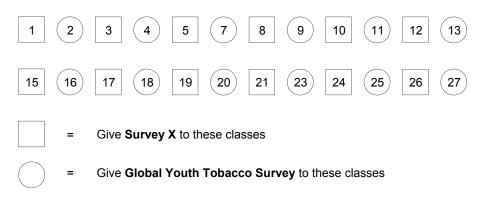**School-Level Sample Information Form**

PAGE 1 OF 2

GYTS/Survey X Coordinating Agency: [Name of the coordinating agency]

School: [Name of the school]

School ID: [Unique school identifier]

GYTS School ID: [Unique identifier for sampling at CDC]

What grades are taught in this school? _____
                                              (Examples: K-12, 9-12)

School participation status: PARTICIPATING   REFUSING   INELIGIBLE
                                                      (Circle one)

If the school is ineligible, please explain: _____

If this school refused to participate or is ineligible, you do not need to complete the rest of this form.

Obtain a list of classes in this school to use for sampling. Number the classes on the list beginning with 1.
The list of classes must be such that each student in the surveyed grades must be represented on the list ONE AND ONLY ONE time. Refer to the GYTS Implementation Instructions for more information.

Enter the TOTAL NUMBER of eligible classes on the list: _____

For each number below, select the corresponding class on your list. Keep selecting classes until you reach the END of your class list. If the first number listed below is greater than the number of classes on your list OR if you run out of numbers before you reach the end of your list of classes, contact [Name of the research coordinator and contact information with telephone and email address].

1	②	3	④	5	⑦	8	⑨	10	⑪	12	13

15	⑯	17	⑱	19	⑳	21	㉓	24	㉕	26	㉗

☐   =   Give **Survey X** to these classes

◯   =   Give **Global Youth Tobacco Survey** to these classes

Please keep your class list until you receive your report from CDC in case there are any questions concerning these classes. If you prefer not to keep this list, you may send it to [Name of the research coordinator] with these forms. On the next page is a class tracking form. For each class selected, enter the class ID (use the number above that was used to select the class) and the teacher's name. As you receive completed questionnaires and class-level sample information forms, put a check mark in the space provided.

School policy questionnaire completed   Yes   ☐   No, please explain: _____

Global Youth Tobacco Survey (GYTS) / Survey X

[Country] – [Region], [Year]
School-Level Sample Information Form

GYTS/Survey X Coordinating Agency: [Name of the coordinating agency]

School: [Name of the school]

School ID: [Unique school identifier]

GYTS School ID:[Unique identifier for sampling at CDC]

CLASS TRACKING FORM

Class ID	Teachers Name	Completed answer sheets received	Completed Class-Level Sample Information Form Received

If you need more lines, make a copy of this page.

Note: Send this form in with your Header Sheet and Answer Sheets/Booklets for this class. After completing this form, please send it to Global Tobacco Control Branch in the Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

-------------------------------- FOR OFFICE USE --------------------------------
School sampling interval: 13.6282993416307
Within-School sampling interval: 1.52395048073902
Random number: 0.4076209068

Survey X
[Country] – [Region], [Year]

Class-Level Sample Information Form for Classes Completing the Survey X

*These classes were identified with the following symbol: ☐

Note: Make enough copies of this form so there is one for EACH selected class in each school.

A Class-Level Sample Information Form MUST be completed for each of the classes selected for your survey, whether or not the class participated.

Survey X Coordinating Agency: [Name of the coordinating agency]

School: [Name of the school]

School ID: [Unique school identifier]

GYTS School ID: [Unique identifier for sampling at CDC]

Class ID: _____

      (Use the random class number shown on the School-Level Sample Information Form.)

How many students are ENROLLED of this class? _____

What is the GRADE in this class?   [list eligible grades]   Other
                                                    (Circle one)

How many students in this class participated in this survey? _____

If the CLASS did not participate in this survey, please explain:

Note: Send this form in with your Header Sheet and Answer Sheets/Booklets for this class. After completing this form, please send it to Global Tobacco Control Branch in the Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, GA.

Global Youth Tobacco Survey (GYTS)
[Country] – [Region], [Year]

Class-Level Sample Information Form for Classes Completing the GYTS

*These classes were identified with the following symbol: ◯

Note: Make enough copies of this form so there is one for EACH selected class in each school.

A Class-Level Sample Information Form MUST be completed for each of the classes selected for your survey, whether or not the class participated.

GYTS Coordinating Agency: [Name of the coordinating agency]

School: [Name of the school]

School ID: [Unique school identifier]

GYTS School ID:[Unique identifier for sampling at CDC]

Class ID: _____

         (Use the random class number shown on the School-Level Sample Information Form.)

How many are ENROLLED in this class?   _____

What is the GRADE of this class?   [list eligible grades]   Other
                                                        (Circle one)

How many students in this class participated in this survey?   _____

If the CLASS did not participate in this survey, please explain:

Note: Send this form in with your Header Sheet and Answer Sheets/Booklets for this class. After completing this form, please send it to Global Tobacco Control Branch in the Office on Smoking and Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

Appendix G: Sampling Description Template

Global Youth Tobacco Survey (GYTS)
[Country name], [Year]

Sample Description and Weighting Procedures

SAMPLE DESCRIPTION:
All schools containing grades [list eligible grades] that contained [min. no of students] or more students were included in the sampling frame. A two-stage cluster sample design was used to produce a representative sample of students in grades [list eligible grades].

School Level - The first-stage of sampling consisted of selection of schools containing grades [list of eligible grades]. Schools were selected with probability proportional to school enrollment size.

Class Level - The second stage of sampling consisted of systematic equal probability sampling (with a random start) of classes from each selected school that participated in the survey. All classes in the participating schools were included in the sampling frame. All students in the selected classes were eligible to participate in the survey.

OVERALL RESPONSE RATES:

Schools -   [XX.X%]   [XX] of the [XX] sampled schools participated

Classes -   [XX.X%]   [XX] of the [XX] sampled classes participated

Students -   [XX.X%]   [XXXX] of the [XXXX] sampled students participated

Overall response rate -   [School response rate] * [class response rate] * [students response rate] = [XX.X%]

WEIGHTING:
A weight has been associated with each participating student record to reflect the likelihood of sampling each student and to reduce bias by compensating for differing patterns of nonresponse. The final weight used for estimation is given by:

$$W = W1 * W2 * f1 * f2 * f3 * f4$$

W1 =   the inverse of the probability of selecting the school

W2 =   the inverse of the probability of selecting the class within the selected school

f1 =   a school-level nonresponse adjustment factor calculated by school size category (small, medium, large)

f2=   a class-level adjustment factor calculated by school

f3 =   a student-level nonresponse adjustment factor calculated by class

f4 =   a post stratification adjustment factor calculated by gender and grade

USE OF THE WEIGHTED RESULTS:

The weighted results can be used to make important inferences concerning tobacco use risk behaviors of students in grades [list eligible grades] in [**Country**].