METHODS
Sample Design.
The GYTS Kenya 2001 was a school-based cross-sectional survey, which employed a two-stage cluster sample designed to produce a nationally representative sample of students, aged 13-15years. School going youth between standard 7 and form two were identified as the target group. All public primary and secondary schools, in the country, with enrolment of over 40 students, were incorporated in the sampling frame.

Stage 1: selection of schools.
CDC worked with the Research Co-ordinator to select the participating schools from an electronic file, consisting of a list of all schools in the sampling frame with their respective enrolment data, obtained from the Ministry of Education. Schools were selected with a probability proportional to enrolment size. This meant that large schools were more likely to be selected than small ones. Fifty-five schools with a target survey population of 4000 students were selected. 25 schools were selected in Nairobi an urban area, 6 in Mombasa an urban area, with strong influence from tourism and 24 from the rest of the country, which is predominantly rural setting. This facilitated a representation from both the urban and rural areas.

Stage 2: selection of classes. . In each selected school, the number of streams in each class, standards 7, 8 or form 1 and 2 were listed. From this list, classes were randomly selected based on the random start provided by CDC on the school level forms. In each school, depending on the number of classes listed one or two or three of those classes were selected. In each class selected, all students present were eligible for the survey.

The Questionnaire
This consisted of 57 'core' questions and 8 'optional' questions making a total of 65 questions. The core questions allow for comparison between countries and regions, and the optional questions allow for specific issues pertaining to Kenya. All questions had responses to choose from and apart from four questions that asked for background information such as age, gender, class and religion, the rest solicited information on the use of tobacco i.e. prevalence, access, brands of cigarettes and other tobacco products, knowledge and attitude towards smoking, environmental tobacco exposure, cessation, media and advertising and school curriculum. The questionnaire was pre-tested before it was administered in schools.

Data Collection
A training workshop for the Research Co-ordinators, representing ten African countries, of the WHO/AFRO region, was held in Harare, between June 5 – 7, 2001. The aim of the session was to familiarise the Research co-ordinators with the standard methodology to be used in all the countries, in implementing the GYTS. Tasks of the Research Co-ordinator were explained. These included overall management of the project, sample design and working in liaison with CDC to select the schools, development and pre testing of the country specific questionnaire, liasing with the various agencies to procure permission and ensure smooth implementation of the survey, making the initial contact with and securing participation of the selected schools, procurement of all supplies, recruiting of survey Administrators, training them and assigning them to selected schools, shipping of answer sheets to CDC for analysis and writing of the country report.

Because GYTS is a school-based survey, co-operation of the Ministry of Health and the Ministry of Education was necessary, especially the latter since government schools are under its immediate control. Although the Director of Education granted permission, to execute the survey, permission and co-operation had also to be obtained from the various Provincial Directors of Education and the respective School Heads.

Eleven Survey Administrators were identified and recruited by the Research Co-ordinator. These were mainly university students on vacation. The research co-ordinator organised a one-day training session for all the survey administrators, at Mathare hospital. The training covered the procedures to be followed prior to, during and after the survey. The survey administrators were issued with the necessary stationery and equipment for the survey and assigned to specific schools, five schools per person, they were also required to make contacts with the head teachers of the assigned schools in order to schedule the survey.

The survey was conducted between 11th and 24th of October 2001. During survey administration, one machine-readable answer sheet and a questionnaire were given to each student. Because the survey procedures were designed to ensure anonymity, students were not required to write their names on the Answer Sheet, or provide any other kind of identifying information. Two types of forms were filled by the survey administrators in respect of each school participating, the School-Level Form and the Classroom Level Form. The School-Level Form contained the name of the School, the sample size, and the School ID, the grades taught and the grades surveyed in the school, as well as the total number of eligible classes. The Classroom Level Form also showed the School name, the sample, the School ID and the Class ID. These two forms provided the necessary identification information and were the primary data management forms. The survey was conducted in class during normal lesson and duration of 45 minutes was allowed. On completion the survey Administrators delivered the completed answer sheets together with the documentation forms to the Research Co-ordinator who upon checking and confirming delivered them to CDC for scanning and analysis.

Analysis

During analysis, a weighting factor was applied to each student record to adjust for non-response and the varying probabilities of selection. The programs SUDAAN and EPI-info were used to compute rates and 95% confidence intervals for estimates. A weight was associated with each questionnaire to reflect the likelihood of sampling each student and to reduce bias by compensating for differing patterns of non-response. The weight used for estimation is given by:

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

Where,

W1=the inverse of the probability of selecting the school

W2=the inverse of the probability of selecting the classroom within the school

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

f2 =a class-level non response adjustment factor calculated for each school

f3 =a student-level non response adjustment factor calculated by class

f4 =a post stratification adjustment factor calculated by form