OAS | CICAD

INTER-AMERICAN DRUG ABUSE CONTROL COMMISSION (CICAD)

MANUAL FOR SECONDARY SCHOOL STUDENT SURVEYS

INTER-AMERICAN UNIFORM DRUG USE DATA SYSTEM (SIDUC)

2021

TABLE OF CONTENTS

MANUAL FOR THE SECONDARY SCHOOL POPULATION 2021

1. INTRODUCTION:

"***Member states develop and implement evidence-based drug policies and strategies, and where possible, data that informs and evaluates strategies is collected in a format that permits comparison and analysis across countries."***

This statement is one of the core principles of the 2020 Hemispheric Drug Strategy[1] of the Inter-American Drug Abuse Control Commission (CICAD, by its Spanish acronym), which highlights two basic ideas: the role of **evidence** in public policies on drugs, and, **comparability** of information among member states.

Public policies are **decisions** taken by a State to address an identified problem, regardless of whether the action is taken by the state itself or by the community. These decisions lead to **interventions** that are designed to improve the situation that the problem has caused. To this end, **scientific evidence** plays a fundamental role in reducing the **uncertainty** that always accompanies any decision, thus ensuring that public policies achieve their objective.

Only through a clear **understanding** of an issue will it be possible to make proper decisions in response. This concept is clearly articulated in the resolution adopted by the OAS General Assembly at its 46th Special Session held in Guatemala in 2014, entitled "*Reflections and Guidelines to Formulate and Follow Up on Comprehensive Policies to Address the World Drug Problem in the Americas"1.*[2] In particular, the Assembly stressed the need for states to ***"Develop, according to the reality of each state and on the basis of an increased understanding of the causes of new challenges posed by the global drug problem, responses that prevent social costs or contribute to their reduction and, when appropriate, review traditional approaches and consider the development of new ones based on scientific evidence and knowledge."***

This requires gathering information through scientifically validated procedures, reinforced by the Hemispheric Plan of Action on Drugs 2021-2025[3], in which member states agree to "*Establish or strengthen national observatories on drugs, or similar technical offices, strengthening national drug information systems, and foster scientific research to generate, collect, organize, analyze, and disseminate information to inform the development and implementation of evidence-based drug policies and strategies.*" In short, by generating

[1] http://www.cicad.oas.org/main/aboutcicad/basicdocuments/OAS_Hemispheric_Drug_Strategy_2020_ENG.pdf
[2] https://www.oas.org/en/media_center/press_release.asp?sCodigo=E-387/14
[3] http://www.cicad.oas.org/main/aboutcicad/basicdocuments/Hemispheric_Plan_of_Action_on_Drugs_2021-2025_ENG.pdf

the knowledge necessary to understand the issue, authorities can design evidence-based policies, measure changes, and evaluate them.

Another important issue is the production of **information that is comparable** across countries. We therefore need to consider another important function of CICAD's Inter-American Observatory on Drugs (OID, by its Spanish acronym). The OID is the entity responsible for providing support to member states' national drug observatories (NDOs) as they produce timely scientific knowledge using standardized methodologies that allow trends to be analyzed within a country and allow for comparisons among countries. Therefore, the OID is able to fulfill its mandate to produce regular reports on the state of the drug problem in the Hemisphere, while the NDOs are responsible for evaluating the state of the drug problem at a national level. The OID/CICAD may provide technical assistance upon request, subject to the availability of resources.

With these issues in mind, the OID developed the **Inter-American Uniform Drug Use Data System (SIDUC, by its Spanish acronym),** which is designed to produce methodologically reliable information on drug demand reduction by providing epidemiological indicators and models that can help provide some responses to the issue of licit and illicit drug use. The OID therefore works with and assists NDOs, meeting with them on a regular basis to coordinate proper implementation of SIDUC methodologies.

These methodologies are designed to obtain reliable estimates of a variety of indicators related to drug use at a particular place and point in time, and to monitor trends over time. It is also important to determine and assess those factors that might foster substance use (risk factors) and those that, discourage drug use or delay first use (protective factors). All of this information is fundamental to the development, monitoring, and evaluation of drug demand reduction policies. It is therefore important to stress, from SIDUC's point of view, that it is critical that epidemiological population studies be conducted **on a regular basis** using standardized methodologies (see section 3.4 below), as the only way of detecting changes in the indicators on drug use as well as changes in risk and protective factors.

For these reasons, as part of SIDUC, the OID has regularly updated its protocols for epidemiological research involving different population groups as established in the Hemispheric Plan of Action referred to above.

One of these protocols, as described in the present document, addresses substance use and associated factors in the secondary school population, in the *8th, 10th,* and *12th grades*.

There are a number of reasons why this population group should be considered as a priority when producing information on the use of licit and illicit substances. As stated in the **Report**

on Drug Use in the Americas 2019[4] prepared by the OID (based on information produced by the NDOs), in most countries of the Americas, 25% or more of students reported that they drank an alcoholic beverage in the past month, with marginal differences between males and females. In a few countries, this figure rose to more than 50%. It should be noted that in a number of countries, more than 20% of 8th grade students reported that they drank alcohol in the past month. Overall, 1 of every 2 students who reported having drunk alcohol in the past month, said that they had had at least one episode of binge drinking.

The report also provides information on indicators relative to the use of other substances such as tobacco, marijuana, inhalants, prescription drugs, etc., and discusses other areas related to substance use, such as the perception of risk/harm of substance use, and ease of access to and supply of drugs, among other things. Lastly, the report identifies "***Four specific areas of relevance to hemispheric drug policy: early onset of drug use; drug use trends; changes in drug use by sex, and new challenges for drug policy.***"

The Report on Drug Use in the Americas 2019 as well as the studies published by member states demonstrate that a significant proportion of the population in each country has contact with and begins their use of licit and illicit substances during their school years. It is therefore critical to gather reliable information on this group as part of a broader assessment to inform the required targeted interventions.

2. OBJECTIVES

Overall objective

To estimate indicators related to the use of licit and illicit substances and their association with possible risk and protective factors.

Specific objectives:

- To estimate[5] the following areas for different substances:[6]
  - ***prevalence*** of "lifetime" use,
  - ***prevalence*** of "past year" use,
  - ***prevalence*** of "past month" use,
  - ***incidence*** of "past year" use,

[4]http://www.cicad.oas.org/main/pubs/Report%20on%20Drug%20Use%20in%20the%20Americas%202019.pdf

[5] For the purposes of this manual, the concept of "estimate" includes the precise estimate and the construction of Confidence Intervals in the context of a probability sample.

[6] As per the list in section 4.1.

- ***incidence*** of "past month" use, and
- ***age of first use*** (onset);

- To estimate the prevalence of use of **at least one licit or illicit substance** in the "past year" and in the "past month";
- To estimate the prevalence of use of **at least one illicit substance** in the "past year" and in the "past month";
- To estimate the prevalence of drug use by sex, grade and type of school;
- To estimate the proportion of students who present with harmful alcohol use, or frequent marijuana use;
- To describe the **profile** of substance users (licit and/or illicit substances);
- To describe the **profile** of users of any illicit substance;
- To estimate the percentage of students who **perceive different levels of risk/harm** while occasionally or frequently using a set of substances;
- To analyze the association between substance use and the perception of risk/harm of drugs;
- To analyze the relationship that exists between easy access to and availability of illicit substances and their use;
- To analyze patterns of use of alcohol and marijuana;
- To estimate indicators on the use of e-cigarettes/vaporizers, and their relationship to current or previous use of tobacco;
- To investigate the **sources** of (controlled) prescription drugs;
- To estimate the percentage of secondary school students exposed to substance use (i.e., questions such as the number of friends who use drugs and interaction with people who use) and relate it to indicators of use. Differentiate by grade;
- To estimate perception of close friends who use drugs. Differentiate by grade;
- To examine the relationship between academic performance and use of licit and illicit substances;
- To describe family relationships and examine their association with the use of licit and illicit substances.

Thus, in the context of SIDUC, there are six main indicators associated with substance use: **three** on prevalence use (lifetime, past year and past month), **two** indicators associated with new cases, i.e., incidence of use (in the past year and past month) and **one** on the age of first use.

With regard to the set of substances listed below in section 4.1, we suggest estimating ***all six indicators***; however, for the other group of substances, we suggest including ***at least three***: lifetime prevalence, past year, and past month. For this second group, the country

may, on the basis of its own circumstances and the knowledge it has gathered on substance use among the secondary school population, decide whether or not to increase the number of indicators.

3. METHODOLOGY

Central to any investigation are the mechanisms used to carry out the processes, that is, the methods to fulfill the objectives proposed for the study. This is fundamental: the most important point in any research is always to bear its objectives in mind. The sections below set out the elements needed in order to meet those objectives, bearing in mind that the information gathered on the students is obtained by means of ***a self-administered questionnaire (anonymous and confidential)*** and a study based on a random sample of students.

3.1. TARGET POPULATION

In the context of SIDUC, the target population for this study are students in urban areas with a population of at least 30,000[7] inhabitants, and who at the time of the study are in the ***8th, 10th and 12th grades*** (or the equivalent in each country), who are mainly between ages **13, 15, and 17**. In practice, depending on the situation in each country, definitions may be different from those described. For example, countries may conduct their study among a large number of grades and include 9th and 11th grade students, or even students in grades lower than the 8th grade. Some countries in the Americas end secondary education at the 11th grade. For the purposes of SIDUC and to ensure comparability of information among countries, only *8th, 10th, and 12th* grades will be considered in the first case*;* in the second case, the target populations will be students in the *8th, 10th, and 11th* grades which are those that will be included in the OID's comparative reports (with the pertinent notations).

3.2. COVERAGE AND REPRESENTATIVENESS

An important factor to be considered when determining the overall objective of this study is the **level at which the study is to be conducted,** and therefore the degree of disaggregation needed to fulfill the specific objectives described. The coverage that is determined for the study, i.e., the level of representativeness of the estimates must be decided at the very beginning since various aspects of the study will be impacted.

[7] While desirable, this requirement may be modified if the situation of a country's population makes it necessary.

So far countries of the Americas have conducted studies on secondary school students using several levels of representativeness, namely:

1. The national level only.
2. In two broad geographical groupings: the metropolitan area, and the rest of the country.
3. At the level of a state, province, department, or subregion (depending on the name in each country).
4. At the municipal or city level, or other name, depending on the country.

Most of the countries have chosen national level. However, examples exist of studies that looked at a state or province, and some examples of municipal or city level studies.[8] OAS member states have little experience with national studies at the community or city level (not including local initiatives aimed at examining a specific situation in a community).

Several elements should be considered when deciding which option to use: some relate to the size of the country, the available financial resources, the time available, while others relate to decisions on the actual use of the information, that is, how the information will be utilized. There may be countries whose geography and population size indicate significant diversity in substance-related issues which makes it advisable for policies to be developed that take that type of diversity into account. Other countries will present very differently.

The table below presents these four options and the potential advantages and disadvantages of each. We must, however, stress that in countries that ***have small homogeneous populations in terms of their socio-bio-demographic characteristics, the dilemma of choosing between these options simply does not exist***. In these cases, it is possible that drug policies are national in nature and are standard throughout the country, and that what is required are global indicators for the country as a whole. However, there will be other countries which have different internal realities and have to make these types of decisions. These countries can be helped in their decision-making by examining the table below.

Level of representativeness	Advantages	Disadvantages
National only	➢ Small sample size ➢ Less expensive	➢ Provides information only at the national level

[8] Countries that have conducted studies with regional or provincial representation have been able to show great differences in substance use and other important indicators. This demonstrates the importance of having this type of information.

	➢ Less time ➢ Simple to administer ➢ Can be replicated more often	➢ Does not take account of the situation on a smaller geographical level.
Metropolitan area, and the rest of country	➢ Provides estimates at the national level ➢ Provides information on different situations ➢ Sample size not too large ➢ Not costly ➢ Does not require much time ➢ Administration is moderately complex ➢ Can be replicated more frequently	➢ Does not take account of the situation in disaggregated geographical areas
States/provinces/ Subegions	➢ Provides estimates at the national level ➢ Provides estimates for states/provinces/subregions ➢ Provides information on different situations	➢ Large sample size ➢ High cost ➢ More time ➢ More complex administration ➢ Difficult to replicate more frequently
Cities/towns	➢ Provides estimates at the national level ➢ Provides estimates for states/provinces/regions ➢ Provides estimates for cities/towns ➢ Provides information on different situations	➢ Large sample size ➢ High cost ➢ More time ➢ More complex administration ➢ Difficult to replicate frequently

In short, based on the above, there will be countries in the Hemisphere that only require national-level information, while others will need to consider all options. Although it is important to have information on the country in general, it will probably be insufficient when using that information for decision-making on a smaller geographical area, such as the subregional or provincial level. We therefore highly recommend a specific level of representativeness that will provide information that is geographically disaggregated. Obviously, as stated earlier, such a decision has an additional cost in terms of requiring a sample size that is larger than that needed for a study that only envisages national representativeness. It is also more complex in terms of fieldwork and statistical analysis.

As to the requirements of the OID/CICAD for its hemispheric reports, countries are asked for information only at the national level, with a breakdown on grade, sex, age and/or type of school.

3.3. SAMPLING ISSUES

SIDUC studies are conducted by means of **sample surveys**. This means that they *do not collect* information from all individuals in the target population, but rather from a *fraction or subset* obtained by means of scientifically rigorous statistical procedures. These are called sampling techniques. The first issue to be determined is the *unit of analysis*, that is, the entity or unit that will provide the information needed to fulfill the objectives of the study. In this case, the unit of analysis is the student selected from the population as previously defined. It is necessary to define the *sampling unit*, or that entity used for the unit of analysis. In the case of this study, the sampling unit is the "class" to which the student belongs.

Generally speaking, we can identify two groups of sampling techniques: ***simple random sampling*** which is the easiest of all. It means using random procedures to select a subset (sample) of elements that make up the target population. Although it would be possible to obtain a list of all students in the grades under study from each country, in practice, the sample might include students who are widely scattered among cities and schools which would make the cost very high and the length of time for conducting the study too long. This sampling strategy is not advised for this type of study.

Any sampling technique that is different from the one described above is generically called ***complex sampling***. This is used in studies among secondary school students as discussed in this Manual. The primary information for a study of this type is the list of schools in the country (according to the decision taken about the size of the cities/towns), with basic data on enrolment rates (number of students) in each grade or form included in the population, and on the number of parallel sections in each grade. For example, in one school there may be 70 students in the 10th grade, divided into two sections. This is the information that is needed and constitutes what is called the **sampling frame** on the basis of which the final sample of students is obtained.

The sampling is done in two stages (two-phase sampling). *The schools are selected in the **first phase*** and in the ***second phase,*** *classes*[9] (grades and sections) are selected from among the schools selected in the first phase. It should be noted that all students present in the classroom at the time the study is conducted, form part of the sample, except for those who voluntarily decide not to participate. This means that there is no third phase of selecting students. In other words, it is not students that are selected but rather classes, and the students who are in the classes selected form the sample for the study.

[9] A particular grade in a school may have more than one section. For example, if there are 70 students in the eighth grade, there may be two sections, 8A and 8B. We are terming these two groups "classes."

Another basic question concerning sampling is related to the ***size of the sample*** for the study, that is, to determine the number of students (in the 8th, 10th and 12th grades) that are needed to meet the objectives of the investigation. The degree of disaggregation of the information for analysis must be borne in mind here; for SIDUC's purposes and to have a comparison between countries, the size of the sample should be large enough to ensure good **national-level** estimates **by sex, grade, age group, and type of school**. In countries where the study's representativeness is by subregions or provinces or at the national level, the size of the sample in each category should be such as to enable estimates to be made with acceptable sampling errors. The same is true if estimates are needed at the city or town level.

The level of disaggregation is not the only condition for determining the size of the sample. Another necessary condition relates to the **size of an important indicator in the study**, such as prevalence of past year use of any illicit substance.

It should be noted that in order to determine the size of the sample and to design the sample itself (i.e., to determine the schools and classes that make up the sample), the advice of a professional statistician with training and experience in sampling techniques will be needed. In addition, given that this is complex sampling (not self-weighting), the statistician must, once the fieldwork has been completed, decide on the **expansion factors** needed for a correct statistical analysis of the information obtained (see Annex 2 for more details about expansion factors).

Further details on this section are given in **Annex 1**. The organization of the fieldwork and the instructions for administering the survey are given in **Annex 3**.

Sampling issues will be dealt with again later in this Manual, as follows:

- Annex 1: *Sampling*. More information on sampling techniques for secondary school studies, including a determination of the sample sizes.
- Annex 2: *Statistical analysis*. Details about calculating the expansion factors and their use in the analyses.
- Annex 3: *Guidelines and instructions*. Issues related to the fieldwork.

3.4. FREQUENCY

As stated earlier, a study of this kind allows us to understand the status of substance use at a particular moment or point in time and has the objective of associating indicators on substance use with factors (determinants) that may be having a positive or negative impact on them. Therefore, the results of this study may also provide information on the development and scope of prevention programs conducted in the country. The studies

should therefore be conducted on a regular basis and any changes that may be necessary can be made in a timely fashion. The OID/CICAD, within the framework of SIDUC, recommends that these studies be conducted ***every two years***.

We need to emphasize that it is important to have information that is *timely,* that identifies changes that are occurring in substance use, and the appearance of new substances so that a rapid response can be provided. This is the reason for proposing that, ideally, these studies be conducted every two years; however, in difficult circumstances, the frequency should not be more than every four years.

4. OPERATIONAL DEFINITIONS

This section will discuss two central points about SIDUC. The substances to be included in any SIDUC study must be defined (along with those that are optional) in accordance with the degree of development of this type of study in the country. This points to two central issues. First, the substances included must be in accordance with the objectives of the study, and second, they must allow for estimates to be made with an acceptable sampling error.

4.1. SUBSTANCES

As noted in section 2 on the objectives of the study, all **six** indicators should be covered for a set of substances. The list of these substances appears below:

List of recommended substances to estimate the six indicators

➢ **Alcohol**	➢ **Tobacco**	➢ **Electronic cigarettes (total)** • **Containing nicotine products** • **Containing cannabis products** • **Containing flavoring products**
➢ **Cannabis (total)** • **Marijuana** • **Hashish**	➢ **Cocaine substances (total)** • **Cocaine hydrochloride** • **Cocaine base paste** • **Crack**	➢ **Inhalants (total)** • **Deodorants** • **Gasoline** • **Glue** • **Solvents** • **Aerosol paint**
➢ **Ecstasy**	➢ **LSD**	➢ **Poppers**

➢ **Controlled prescription drugs**		
→ **Tranquilizers (total)** • **Alprazolam** (Alprazolam Intensol, Xanax and Xanax XR) • **Clonazepam** (Klonopin) • **Diazepam** (Diastat AcuDial, Diazepam Intensol, Diastat and Valium) • **Flunitrazepam** (Rohypnol) • **Chlordiazepoxide** (Klopoxid, Libritabs, Librium, Mesural, Multum, Novapam, Risolid, Silibrin, Sonimen, Tropium and Zetran)	→ **Stimulants (total)** • **Methylfenidate** (Ritalin, Concerta) • **Phenmetrazine** (Preludin) • **Amphetamine** (Adderall, Adderall XR, Mydayis, Evekeo, Zenzedi and Dexedrine) • **Dextroamphetamine** (Dexedrine, DextroStat) • **Pemoline** (Cylert)	→ **Analgesics (total)** • **Fentanyl** (Duragesic, Ionsys, Subsys and Abstral) • **Tramadol** (ConZip and Ultram) • **Hydromorphone** (Dilaudid) • **Hydrocodone** (Lorcet, Vicodin, Hycet, Lortab) • **Oxycodone** (OxyContin, Xtampza ER, Oxaydo, Roxicodona, Primlev, Tylox, Endocet, Percocet and Percodan) • **Methadone** (Diskets, Metadona Intensol, Dolophine and Methadose) • **Codeine** (Codeisan, Codeisan jarabe, Fludan codeína, Histaverin, Notusin, Perduretas codeína and Toseina) • **Morphine** (MorphaBond ER, Arymo ER, Infumorph P/F, Astramorph-PF, Duramorph and MS Contin)

However, for the following substances, at least **three** indicators on prevalence should be estimated: lifetime use, past year use, and past month use.

List of recommended substances to estimate at least three indicators.

Some street names are included in parenthesis. Countries should add their own street names.

➢ **Methamphetamines** (Meth, ice, crystal)	➢ **Opium**	➢ **Anabolic steroids**
➢ **Amphetamine** (fet, speed)	➢ **Heroin** (Paste, "H", white powder, skag and tar)	➢ **Ketamine** (Keta, vitamin K, super K, CK or Calvin Klein, horse, Mary Kay or MaryK)
➢ **Synthetic Cannabinoids** (synthetic marijuana, Spice, K2, Joker, Black Mamba, Kush or Kronic)	➢ **Synthetic Cathinones** (Bath salts)	➢ **Aminoindanes** (MDAI gold)
➢ **Piperazines** (BZP, mCPP, A2, Legal X and Pep X)	➢ **Phencyclidine** (PCP, angel dust, embalming fluid, hog, killer joints, love boat, ozone, peace pill, superweed, rocket fuel)	➢ **Hallucinogenic plants** (Angel's trumpet, ayahuasca; mescaline or peyote; Psilocybin, hallucinogenic mushrooms or magic mushrooms, salvia divinorum)

➢ **GHB** (liquid x, liquid ecstasy, Georgia home boy, grievous bodily harm, "G", liquid G, Fantasía)	➢ **Lean** (Sprite mix, cough syrup and pastilles – also called *purple drank* or *sizzurp*)	➢ **Caffeine products** (caffeine pills, energy drinks, caffeine powder)
➢ **Phenethylamines** (Europa, 4-FMP, RDJ, 4-MMA, Methyl-MA, 2C-C-NBOMe, bomb, N-bomb, 251, Nexus, 2C-E and Blue mystic)		

Countries may add or remove substances to these lists in light of their own situation and experience. Names should be adapted to reflect those commonly used in the country.

4.2. VARIABLES, QUESTIONS, AND INDICATORS

It is important in a study of this type to remember three concepts, each of which are inter-related and have to do with the objectives of the study. The first has to do with the ***study variables***, that is to say, what is it that we want to measure? Second, what ***questions*** do we ask? Lastly, we need an ***indicator*** that will report the result of the measurement of the variables. Note that these three concepts are interrelated and should satisfy the specific objectives described earlier.

For example, the following groups of study variables should be considered:

- G**eneral variables** on the student: sex, age, grade, type of school, and region/district (in the event this coverage is contemplated in the study).
- Variables on **drug use:** in this case, the variables refer to:
  - "lifetime use of [name of substance]",
  - "past year use of [name of substance]",
  - "past month use of [name of substance]",
  - "use of [name of substance] for the first time in the past year",
  - "use of [name of substance] for the first time in the past month",
  - "frequent use of [name of substance]".
- Variables related to the **perceived risk/harm** of the use of a particular substance; "occasional/experimental" use is considered separately from "frequent use."[10]
- Variables related to the **family.**

[10] In relation to these questions, please note that there may be differences in the questions and answers between the English and Spanish versions of this Manual. This difference must be taken into account when making comparisons between countries. In Spanish, the question refers to "risk" in general, but the English version refers to "harm to health." The possible answers refer to "risk" in the first case and "harm to health" in the second.

- Variables related to **academic performance.**
- Variables related to the **ease of obtaining** substances.
- Variables related to **offers of drugs received**.

The questionnaire should have a question for each of the variables defined above so that the variable can be assessed.

For example, the variable related to "past year use of alcohol," the following question should be asked: **"Have you drank an alcoholic beverage in the past year?**" This is a binary question to which there are theoretically two possible answers: ***Yes*** or ***No***. For other types of questions, such as those that ask about access to drugs, the question: "**How easy or difficult would it be to obtain marijuana?**", there will be several alternative answers: ***easy***; ***difficult***; ***would not be able to obtain***; ***don't know***.

Lastly, on a question more related to analysis, the answers must be quantified according to four statistical indicators: **prevalence, incidence, percentage, and quantitative measures**. The first two are related to substance use:-

- ***Prevalence*** (for a specific substance) is an indicator that measures the proportion of students that report having used the substance (most recently) at some particular point in time: lifetime, past year, or past month. Prevalence indicators refer to the total sample (expanded) of students, and are generally expressed as a percentage.
- ***Incidence*** is an indicator that focuses on the appearance of **new cases** at a specified moment in time, usually in the past year or past month. In this case, the indicator is determined with respect to those students who had not used drugs before the time period specified in the question; it is also expressed as a percentage.
- The third indicator, ***percentage***, refers to other variables some of which are demographic (% of males), others about perceived risk/harm (% that perceives that getting drunk is highly risky/or, of great harm to health), and others about offers of drugs (% who were offered ecstasy in the past year.)
- Lastly, ***quantitative measurements*** are associated with quantitative variables such as "age of student" or "age of first use of tobacco", where the **indicators** used to evaluate the answers will be, for example, the **mean,** the **median,** and some **percentiles** (particularly the 25^{th} and 75^{th} percentiles).

Although the indicators of prevalence and incidence are measured using a percentage, the difference with percentage indicators per se is that the former assess the **risk that an event will occur**, in this case, drug use. Further information on this can be found in section 6 and in **Annex 2** on Statistical Analysis.

Why are the indicators of prevalence and incidence of substance use so important?

The extent or magnitude of substance use and the changes in use in a particular population, as well as the use of new substances, show the impact that substance use has on that population. This is what has to be measured at a particular time and at several points in time in order to gain an up-to-date perspective of the demand for drugs—whether it is rising or falling, among which specific groups are these changes occurring, and to have some objective criteria about some of the outcomes of drug policy in the country.

The **magnitude or extent** of drug use is **measured** by the indicators that we have termed *prevalence* and *incidence*. Prevalence measures the total number of cases/subjects that have used a drug over a *period of time*, whether they were **new cases** that began during that period, or **old cases** with a history of substance use. Incidence focuses only on **new cases** in a specified time period. The table below provides questions in the survey, and shows the indicators that would be useful in eliciting an answer, using one substance, alcohol, as a reference point.

Question of interest	**Indicator**
What percentage of students has *used* alcohol ever in their lives?	*Prevalence* of lifetime use of alcohol
What percentage of students *has used* alcohol recently (past year)?	*Prevalence* of past year use of alcohol
What percentage of students *has used* alcohol currently (past month)?	*Prevalence* of past month use of alcohol
What percentage of students *used alcohol for the first time* in the year prior to the study?	*Incidence* of past year use of alcohol
What percentage of students *used alcohol for the first time* in the month prior to the study?	*Incidence* of past month use of alcohol

The definitions for these indicators are discussed further in this Manual, but we must stress that they should be measured on a regular basis.

In the past, far more attention was paid to prevalence than to incidence, but both are equally important. In the school population, it is particularly important to pay great attention to the dynamics of the first use of substances, since, to some extent, this behavior reflects the most immediate response to the interventions that a country or particular geographical area may be carrying out.

5. THE QUESTIONNAIRE: ORGANIZATION OF QUESTIONS INTO MODULES

The specific objectives of the SIDUC exercise will be clearly defined in the self-administered questionnaire. As noted in the different modules, countries should adapt some of the

questions to their own peculiar circumstances. The students' responses are anonymous and confidential.

The **questionnaire** consists of the following modules or sections:-

MODULE 1. INFORMATION ON THE SCHOOL AND THE CLASS BEING SAMPLED
MODULE 2. GENERAL INFORMATION ABOUT THE STUDENTS
MODULE 3. SUBSTANCE USE
MODULE 4. PERCEPTION OF HARM/RISK AND PROSPECTS OF SUBSTANCE USE
MODULE 5. RELATIONSHIP WITH PARENTS/GUARDIANS WITH WHOM THE STUDENT LIVES
MODULE 6. RELATIONSHIP WITH THE SCHOOL, ACADEMIC PERFORMANCE, AND FUTURE PROSPECTS

As we will see below, the questionnaire has many questions particularly in Module 3 on the use of psychoactive substances. However, hypothetically if the majority of students **do not use drugs**, the questions that they will have to answer are fewer in number. When a student has never used a particular drug in his/her lifetime, **the questionnaire will say "skip"** to the next substance.

QUESTIONNAIRE

Good day!

We are conducting a survey in different schools in the country on issues related to **public health.** Our objective is to gather information so that action can be taken to solve public health problems in the country. Your cooperation in this survey will be very useful. Your answers are **absolutely confidential and are completely anonymous**. No one will be able to identify you through your answers, so do not write any personal information on the questionnaire. We ask you to respond honestly and sincerely. The questions begin on the next page with Question 9.

MODULE 1. INFORMATION ON THE SCHOOL AND SAMPLE CLASS

1. REGION, (PROVINCE, COUNTY)	2. CITY/TOWN	3. NUMBER OF QUESTIONNAIRE

4. Type of school

	1. Public
	2. Private
	3. Other (specify)

6. Student's form or grade level

	1. 8th grade (2nd form)
	2. 10th grade (4th form)
	3. 12th grade (6th form)

5. Type of students in the school

	1. Males only
	2. Females only
	3. Mixed

7. Specify a letter to divide each grade or form into sections: for example, 8th grade A, or 4th form B, etc.

8. Number of students in the classroom

NOTE: THE COUNTRY SHOULD REVIEW THE NAMES OF THE GRADES IN QUESTION 6.

THE STUDENT STARTS FILLING OUT THE QUESTIONNAIRE HERE:-

PUT AN 'X' NEXT TO YOUR ANSWER, OR WRITE IN A NUMBER WHEN THE QUESTION ASKS YOU ABOUT AGE. IF YOU HAVE ANY QUESTIONS, PLEASE ASK THE FACILITATOR

MODULE 2. GENERAL INFORMATION ABOUT THE STUDENT

9. Gender

	1. Male
	2. Female

10. Age

Years old

11.What is your parents' marital status?

	1. Married or partners and living together
	2. Married, partners, NOT living together
	3. Separated, annulled, divorced, but living together
	4. Separated, annulled, divorced, but NOT living together
	5. Never lived together
	6. Widow(er)
	7. Other

12. With whom do you live at the moment?

	1. Mother and father
	2. Mother
	3. Father
	4. Mother and her partner
	5. Father and his partner
	6. Brother(s) and/or sister(s)
	7. Grandmother(s) and/or grandfather(s)
	8. Another adult

MODULE 3: SUBSTANCE USE. The questions below relate to smoking tobacco cigarettes.

TA1. Have you ever smoked cigarettes in your lifetime?

	1. Yes
	2. No **(skip to question CE1)**

TA2. How old were you when you smoked cigarettes for the first time?

	Years old

TA3. When was the *first time in your life* that you smoked cigarettes?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

TA4. Have you smoked cigarettes in the past 12 months?

	1. Yes	
	2. No	**(skip to question CE1)**
	9. Don't know/no opinion	

TA5. Have you smoked cigarettes in the past 30 days?

	1. Yes	
	2. No	**(skip to question CE1)**
	9. Don't know/no opinion	

TA6 How many days did you smoke cigarettes in the past 30 days?

	1. Never
	2. Only a few days
	3. Several days
	4. Almost every day
	5. Every day

TA7. About how many cigarettes have you smoked per day in the past 30 days?

	1. 1 to 5
	2. 6 to 10
	3. 11 to 20
	4. More than 20
	9. Don't know/no opinion

The following questions have to do with the use of electronic cigarettes (e-cigarettes).

CE1. Have you ever vaped e-cigarettes containing nicotine products?

	1. Yes	
	2. No	**(skip to question CE5)**

CE2. Have you vaped e-cigarettes containing nicotine products in the past year?

	1. Yes	
	2. No	**(skip to question CE5)**

CE3. Have you vaped e-cigarettes containing nicotine products in the past month?

	1. Yes	
	2. No	**(skip to question CE5)**

CE4. How many days in the past 30 days did you vape e-cigarettes containing nicotine products?

	1. Never
	2. Only a few days
	3. Several days
	4. Almost every day
	5. Every day

CE5. Have you every vaped e-cigarettes containing cannabis products?

	1. Yes	
	2. No	**(skip to question CE9)**

CE6. Have you vaped e-cigarettes containing cannabis products in the past year?

	1. Yes	
	2. No	**(skip to question CE9)**

CE7. Have you vaped e-cigarettes containing cannabis products in the past month?

	1. Yes	
	2. No	**(skip to question CE9)**

CE8. How many days in the past 30 days did you vape e-cigarettes containing cannabis products?

	1. Never
	2. Only a few days
	3. Several days
	4. Almost every day
	5. Every day

CE9. Have you ever vaped e-cigarettes containing only flavoring?

	1. Yes	
	2. No	**(skip to question AL1)**

CE10. Have you vaped e-cigarettes containing only flavorings in the past year?

	1. Yes	
	2. No	**(skip to question AL1)**

CE11. Have you vaped e-cigarettes containing only flavorings in the past month?

	1. Yes	
	2. No	**(skip to question AL1)**

CE12. How many days in the past 30 days did you vape e-cigarettes containing only flavorings?

	1. Never
	2. Only a few days
	3. Several days
	4. Almost every day
	5. Every day

The next set of questions relate to the use of alcoholic beverages.

AL1. Have you ever consumed alcoholic beverages?
(*Include wine, beer, hard liquor such as rum, whisky, vodka, etc. Don't answer "yes" if your parents once gave you a small sip to taste.*)

	1. Yes	
	2. No	**(skip to question AL10)**

AL2. How old were you when you consumed an alcoholic beverage for the first time?

	Years old

AL3. When was the first time you consumed an alcoholic beverage?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

AL4. Have you consumed alcoholic beverages in the past 12 months?

	1. Yes	
	2. No	**(skip to question AL10)**
	9. Don't know/no opinion	

AL5. Have you consumed alcoholic beverages in the past 30 days?

	1. Yes	
	2. No	**(skip to question AL10)**
	9. Don't know/no opinion	

AL6. In the past 30 days, how many days did you use/drink an alcoholic beverage?

	1. Never
	2. Only a few days
	3. Several days
	4. Almost every day
	5. Every day

AL7. In the past 2 weeks, how many times have you had 5 or more drinks at a single sitting?

	1. Never
	2. Only once
	3. 2 to 3 times
	4. 4 to 5 times
	5. More than 5 times
	9. Don't know/no opinion

In the past 30 days, what type(s) of alcoholic beverage(s) did you consume, and with what frequency?
(*Indicate your answers by marking an "X" for one option only for each* alcoholic beverage*)*

	Every day (1)	On the weekend (2)	Several days during the week (3)
AL8.1 Beer, Guinness, Shandy	1	2	3
AL8.2 Wine	1	2	3
AL8.3 Hard liquor (whisky, vodka, rum, ponche caribe, rum punch, liqueurs)	1	2	3

AL9. As you think about the past 30 days, indicate where you drank an alcoholic beverage. Mark all that apply.

	1. At home
	2. At a friend's house
	3. In a bar or pub
	4. In a restaurant
	5. At school
	6. In a park or other public place
	7. At school
	8. Somewhere else

AL10. How easy or difficult would it be for you to <u>buy</u> an alcoholic beverage?

	1. Easy
	2. Difficult
	3. Would not be able to buy
	4. Don't know

The next set of questions relate to the use of prescription tranquilizers such as *Alprazolam (Alprazolam Intensol, Xanax and Xanax XR), Clonazepam (Klonopin), Diazepam (Diastat AcuDial, Diazepam Intensol, Diastat and Valium), Flunitrazepam (Rohypnol), Chlordiazepoxide (Klopoxid, Libritabs, Librium, Mesural, Novapam, Risolid, Silibrin, Sonimen, Tropium and Zetran) or similar*.

TR1. Have you ever taken any of these tranquilizers *because a doctor prescribed it for you*?

	1. Yes
	2. No

TR2. Have you ever taken any of these tranquilizers *without a doctor's prescription*?

	1. Yes	
	2. No	**(skip to question TR9)**

TR3. How old were you when you first took a tranquilizer *without a prescription*?

	Years old

TR4. When was the first time you took a tranquilizer *without a prescription*?

	1. During the past 30 days
	2. More than 1 month ago, but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

TR5. Have you taken a tranquilizer *without a prescription* in the past 12 months?

	1. Yes	
	2. No	**(skip to question TR9)**
	9. Don't know/no opinion	

Which tranquilizers have you taken *without a prescription* in the past 12 months?
Mark all that apply.

TR6.1		**Alprazolam** (Alprazolam Intensol, Xanax and Xanax XR)
TR6.2		**Clonazepam** (Klonopin)
TR6.3		**Diazepam** (Diastat AcuDial, Diazepam Intensol, Diastat, Valium)
TR6.4		**Flunitrazepam** (Rohypnol)
TR6.5		**Chlordiazepoxide** (Librium, Novapam, Risolid, Silibrin, Sonimen, Tropium, Zetran)
TR6.6		Other

TR7. In the past 12 months, how did you obtain tranquilizers *without a prescription*?
MARK ALL THAT APPLY.

	1. At home without asking
	2. I took them from a friend without asking
	3. A family member gave them to me
	4. A friend gave them to me
	5. I bought them from a family member
	6. I bought them from a friend
	7. I had a prescription
	8. I bought them from someone I did not know who was selling them
	9. I bought them over the Internet

TR8. Have you taken a tranquilizer *without a prescription* in the <u>past 30 days</u>?

	1. Yes
	2. No
	9. Don't know/no opinion

TR9. If you wanted to, how easy or difficult do you think it would be to obtain a tranquilizer *without a prescription*?

	1. It would be easy
	2. It would be difficult
	3. Would not be able to obtain
	4. Don't know

NOTE: COUNTRIES SHOULD CHANGE THE LIST OF TRANQUILIZERS WHERE NECESSARY IN ORDER TO REFLECT THE NAMES USED LOCALLY IN THE COUNTRY.

The next set of questions relate to the use of prescribed stimulants such as *Methylphenidate (Ritalin, Concerta), Phenmetrazine (Preludin), Amphetamine (Adderall, Adderall XR, Mydayis, Evekeo, Zenzedi and Dexedrine), Dextroamphetamine (Dexedrine, DextroStat), Pemoline (Cylert) or similar.*

ES1. Have you ever taken one of these stimulants because a doctor prescribed it for you?

	1. Yes
	2. No

ES2. Have you ever taken one of these stimulants without a medical prescription?

	1. Yes	
	2. No	**(skip to question ES9)**

ES3. How old were you when you first took a stimulant without a prescription?

	Years old

ES4. When was the first time you took a stimulant without a prescription?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

ES5. Have you taken a stimulant without a prescription in the past 12 months?

	1. Yes	
	2. No	**(skip to question ES9)**
	9. Don't know/no opinion	

Which stimulants have you taken without a prescription in the past 12 months?
Mark all that apply.

ES6.1		**Methylphenidate** (Ritalin, Concerta)
ES6.2		**Phenmetrazine** (Preludin)
ES6.3		**Amphetamine** (Adderall, Adderall XR, Mydayis**,** Evekeo, Zenzedi, Dexedrine)
ES6.4		**Dextroamphetamine** (Dexedrine, DextroStat)
ES6.5		**Pemoline** (Cylert)
ES6.6		**Other**

ES7. In the past 12 months, how did you obtain stimulants without a prescription?
MARK ALL THAT APPLY

	1. At home without asking
	2. I took them from a friend without asking
	3. A family member gave them to me
	4. A friend gave them to me
	5. I bought them from a family member
	6. I bought them from a friend
	7. I had a prescription
	8. I bought them from someone I didn't know who was selling them
	9. I bought them over the Internet

ES8. Have you taken a stimulant without a prescription in the <u>past 30 days</u>?

	1. Yes
	2. No
	9. Don't know/no opinion

ES9. If you wanted to, how easy or difficult do you think it would be to obtain a stimulant *without a prescription*?

	1. It would be easy
	2. It would be difficult
	3. Would not be able to obtain
	4. Don't know

NOTE: COUNTRIES SHOULD CHANGE THE LIST OF STIMULANTS WHERE NECESSARY TO REFLECT THE NAMES USED LOCALLY IN THE COUNTRY.

The next set of questions relate to analgesic prescription drugs that are used as pain-killers, such as *Fentanyl (Duragesic, Ionsys, Subsys and Abstral), Tramadol (ConZip and Ultram), Hydromorphone (Dilaudid), Hydrocodone (Lorcet, Vicodin, Hycet, Lortab), Oxycodone (OxyContin, Xtampza ER, Oxaydo, Roxicodona, Primlev, Tylox, Endocet, Percocet and Percodan), Methadone (Diskets, Methadone Intensol, Dolophine and Methadose), Codeine (Codeisan, Codeisan jarabe, Fludan codeine, Histaverin, Notusin, Perduretas codeína and Toseina), Morphine (MorphaBond ER, Arymo ER, Infumorph P/F, Astramorph-PF, Duramorph and MS Contin) or similar.*

AN1. Have you ever taken any of these analgesics because a doctor prescribed them for you?

	1. Yes
	2. No

AN2. Have you ever taken any of these analgesics without a doctor's prescription?

	1. Yes	
	2. No	**(skip to question AN9)**

AN3. How old were you when you first took an analgesic *without prescription*?

	Years old

AN4. When was the first time you took an analgesic *without a prescription*?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

AN5. Have you taken an analgesic *without a prescription* in the past 12 months?

	1. Yes	
	2. No	**(skip to question AN9)**
	9. Don't know/no opinion	

Which analgesics have you taken *without a prescription* in the past 12 months?
Mark all that apply.

AN6.1		**Fentanyl** (Duragesic, Ionsys, Subsys and Abstral)
AN6.2		**Tramadol** (ConZip and Ultram)
AN6.3		**Hydromorphone** (Dilaudid)
AN6.4		**Hydrocodone** (Lorcet, Vicodin, Hycet, Lortab)
AN6.5		**Oxycodone** (OxyContin, Xtampza ER, Oxaydo, Roxicodona, Primlev, Tylox, Endocet, Percocet and Percodan)
AN6.6		**Methadone** (Diskets, Methadone Intensol, Dolophine and Methadose)
AN6.7		**Codeine** (Codeisan, Codeisan jarabe, Fludan codeína, Histaverin, Notusin, Perduretas codeína and Toseina)
AN6.8		**Morphine** (MorphaBond ER, Arymo ER, Infumorph P/F, Astramorph-PF, Duramorph, MS Contin)
AN6.9		Other

AN7. In the past 12 months, how did you obtain analgesics *without a prescription*?
MARK ALL THAT APPLY.

	1. At home without asking
	2. I took them from a friend without asking
	3. A family member gave them to me
	4. A friend gave them to me
	5. I bought them from a family member
	6. I bought them from a friend
	7. I had a prescription
	8. I bought them from someone I didn't know who was selling them
	9. I bought them over the Internet

AN8. Have you taken an analgesic *without a prescription* in the past 30 days?

	1. Yes
	2. No
	9. Don't know/no opinion

AN9. If you wanted to, how easy or difficult do you think it would be to obtain a stimulant *without a prescription*?

	1. It would be easy
	2. It would be difficult
	3. Would not be able to obtain
	4. Don't know

NOTE: COUNTRIES SHOULD CHANGE THE LIST OF ANALGESICS AS NEEDED TO REFLECT THE NAMES USED LOCALLY IN THE COUNTRY.

The next set of questions relate to the use of marijuana

MA1. Have you ever used marijuana?

	1. Yes	
	2. No	**(skip to question MA14)**

MA2. How old were you when you used marijuana for the first time?

	Years old

MA3. When was the first time you used marijuana?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

MA4. Have you used marijuana in the past 12 months?

	1. Yes	
	2. No	**(skip to question MA14)**
	9. Don't know/no opinion	

MA5. How often did you use marijuana in the past 12 months?

	1. 1-2 times
	2. 3-5 times
	3. 6-9 times
	4. 10-19 times
	5. 20-39 times
	6. 40 or more times

MA6. How did you use marijuana in the past 12 months?
Mark all that apply.

	1. Smoked
	2. Vaporized
	3. Eaten in food
	4. In oils or tinctures
	5. In pharmaceutical products or for medical use
	6. Other. Specify_____

MA7. Which was the most frequent way you used marijuana in the past 12 months? Choose one only.

	1. Smoked
	2. Vaporized
	3. Eaten in food
	4. In oils or tinctures
	5. In pharmaceutical products or for medical use
	6. Other. Specify_____

MA8. Have you used marijuana in the <u>past</u> 30 days?

	1. Yes	
	2. No	**(skip to question MA14)**
	9. Don't know/no opinion	

MA9. How many times did you use marijuana in the past 30 days?

	1. Never
	2. 1-2 times
	3. 3-5 times
	4. 6-9 times
	5. 10-19 times
	6. 20-39 times
	7. 40 or more times

MA10. How often in the <u>past</u> 30 days did you use marijuana?

	1. Never
	2. Only a few days
	3. Several days
	4. Almost every day
	5. Every day

MA11. During the past month, <u>how many marijuana cigarettes</u> did you smoke a day on average?

	1. None
	2. Less than 1 a day
	3. 1 a day
	4. 2 to 3 a day
	5. 4 to 6 a day
	6. 7 to 10 a day
	7. 11 or more a day

MA12. When you used marijuana in the past month, how long did you feel high?

	1. One to two hours
	2. Three to six hours
	3. Seven to 24 hours
	4. More than 24 hours
	5. I didn't really feel any effects

MA13. As you think about the past month, how did you obtain the marijuana that you used?
Mark all that apply.

	1. A friend or one of my brothers or sisters gave it to me, or someone I knew personally
	2. Someone I didn't know gave it to me
	3. It was shared with a group of friends
	4. I bought it from a friend or from someone I knew
	5. I bought it from someone I didn't know personally
	6. I obtained it in some other way

MA14. How easy or difficult would it be for you to obtain marijuana?

	1. It would be easy
	2. It would be difficult
	3. Would not be able to obtain
	4. Don't know

MA15. If you were offered marijuana, when was the last time it was offered to you, either to buy or to smoke? *Think only about the last time you were offered marijuana.* Choose one only.

	1. Never offered to me	**(skip to question HA1)**
	2. During the past 30 days	
	3. More than 1 month ago, but less than 1 year ago	
	4. More than 1 year ago	

MA16. If you have ever been offered marijuana, think about the last time it was offered to you. Where was it offered? ***Choose one only.***

	1. At home
	2. At school
	3. In the area around my school
	4. At a party, nightclub, concert
	5. In the area around my house
	6. By Internet, email or chat room
	7. At the house of friends or family

MA17. The last time you were offered marijuana, who offered it to you?

	1. A friend
	2. A family member
	3. Someone I know who is not a friend
	4. Someone I didn't know
	5. Someone else

The next set of questions relate to the use of hashish.

HA1. Have you ever used hashish?

	1. Yes	
	2. No	**(skip to question CO1)**

HA2. How old were you when you first used hashish?

	Years old

HA3. When was the first time you used hashish?

	1. In the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

HA4. Have you used hashish in the past 12 months?

	1. Yes	
	2. No	**(skip to question CO1)**
	9. Don't know/no opinion	

HA5. How often have you used hashish?

	1. Only once
	2. Several times during the past year
	3. Several times a month
	4. Several times a week
	5. Every day

HA6. Have you used hashish in the past 30 days?

	1. Yes
	2. No

The next set of questions relate to cocaine use.

CO1. Have you ever used cocaine?

	1. Yes	
	2. No	**(skip to question CO8)**

CO2. How old were you when you first used cocaine?

	Years old

CO3. When was the first time you used cocaine?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

CO4. Have you used cocaine in the past 12 months?

	1. Yes	
	2. No	**(skip to question CO8)**
	9. Don't know/no opinion	

CO5. What is the frequency of your use of cocaine?

	1. Only once
	2. Several times in the past year
	3. Several times a month
	4. Several times a week
	5. Every day

CO6. Have you used cocaine in the past 30 days?

	1. Yes	
	2. No	**(skip to question CO8)**
	9. Don't know/no opinion	

CO7. As you think about the past month, how did you obtain the cocaine that you used?
Mark all that apply

	1. A friend or one of my brothers or sisters gave it to me, or someone I knew personally
	2. Someone I didn't know gave it to me
	3. It was shared with a group of friends
	4. I bought it from a friend or from someone I knew
	5. I bought it from someone I didn't know personally
	6. I obtained it in some other way

CO8. How easy or difficult would it be for you to obtain cocaine?

	1. Easy
	2. Difficult
	3. Would not be able to obtain
	4. Don't know

CO9. If you were ever offered cocaine, when was the last time it was offered, whether to buy or to try? *Think only about the last time someone offered you* cocaine.
Choose one answer only.

	1. Never	**(skip to question PB1)**
	2. In the past 30 days	
	3. More than 1 month ago but less than 1 year ago	
	4. More than 1 year ago	

CO10. If you were ever offered cocaine, think about the last time it was offered to you. Where was it offered? *Choose one answer only.*

	1. At home
	2. At school
	3. In the area around my school
	4. At a party, nightclub, concert
	5. In the area around my house
	6. By Internet, email or chat room
	7. At the house of friends/family

CO11. The last time you were offered cocaine, who offered it to you?

	1. A friend
	2. A family member
	3. Someone I know who is not a friend
	4. Someone I didn't know
	5. Someone else

The next set of questions relate to the use of cocaine paste (pasta base/coca paste)

PB1. Have you ever used cocaine paste (pasta base/coca paste)?

	1. Yes	
	2. No	**(skip to question PB8)**

PB2. How old were you when you first used cocaine paste (pasta base/coca paste)?

	Years old

PB3. When was the first time you used cocaine paste (pasta base/coca paste)?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

PB4. Have you used cocaine paste (pasta base/coca paste) in the past 12 months?

	1. Yes	
	2. No	**(skip to question PB8)**
	9. Don't know/no opinion	

PB5. How often have you used cocaine paste (pasta base/coca paste)?

	1. Only once
	2. Several times in the past year
	3. Several times a month
	4. Several times a week
	5. Every day

PB6. Have you used cocaine paste (pasta base/coca paste) in the past 30 days?

	1. Yes	
	2. No	**(skip to question PB8)**
	9. Don't know/no opinion	

PB7. As you think about the past month, how did you obtain the cocaine paste (pasta base/coca paste) that you used? Mark all that apply.

	1. A friend or one of my brothers or sisters gave it to me, or someone I knew personally
	2. Someone I didn't know gave it to me
	3. It was shared with a group of friends
	4. I bought it from a friend or from someone I knew
	5. I bought it from someone I didn't know personally
	6. I obtained it in some other way

PB8. How easy or difficult would it be for you to obtain cocaine paste (pasta base/coca paste)?

	1. Easy
	2. Difficult
	3. Would not be able to obtain
	4. Don't know

PB9. If you were offered cocaine paste (pasta base/coca paste), when was the last time it was offered to you, either to buy or to try? Think only about the last time you were offered cocaine paste (pasta base/coca paste). Choose one answer only.

	1. Never	**(skip to question CR1)**
	2. During the past 30 days	
	3. More than 1 month ago, but less than 1 year ago	
	4. More than 1 year ago	

PB10. If you were ever offered cocaine paste (pasta base/coca paste), think about the last time it was offered to you. Where was it offered? Choose one answer only.

	1. At home
	2. At school
	3. In the area around my school
	4. At a party, nightclub, concert
	5. In the area around my house
	6. By internet, email or chat room
	7. At the house of friends/family

PB11. The last time you were offered cocaine paste (pasta base/coca paste), who offered it to you?

	1. A friend
	2. A family member
	3. Someone I know but is not a friend
	4. Someone I didn't know
	5. Someone else

NOTES:

- ✓ **EACH COUNTRY SHOULD SELECT THE NAME(S) USED IN THE QUESTIONS ABOVE, i.e., CHOOSING BETWEEN COCAINE PASTE, PASTA BASE, AND COCA PASTE.**
- ✓ **IF IN ADDITION TO THIS SUBSTANCE, THE COUNTRY DECIDES TO INCLUDE CRACK, THEN IN QUESTION PB9, THE STATEMENT "SKIP TO QUESTION EX1" SHOULD BE CHANGED TO: "SKIP TO QUESTION CR1."**

The next set of questions relate to the use of crack.

CR1. Have you ever used crack?

	1. Yes	
	2. No	**(skip to question CR8)**

CR2. How old were you when you <u>first</u> used crack?

	Years old

CR3. When was the <u>first time</u> you used crack?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

CR4. Have you used crack in the <u>past 12 months</u>?

	1. Yes	
	2. No	**(skip to question CR8)**
	9. Don't know/no opinion	

CR5. How often have you used crack?

	1. Only once
	2. Several times in the past year
	3. Several times a month
	4. Several times a week
	5. Every day

CR6. Have you used crack in the past 30 days?

	1. Yes	
	2. No	**(skip to question CR8)**
	9. Don't know/no opinion	

CR7. As you think about the past month, how did you obtain the crack you used?
Choose all that apply

	1. A friend or one of my brothers or sisters gave it to me, or someone I knew personally
	2. Someone I didn't know gave it to me
	3. It was shared with a group of friends
	4. I bought it from a friend or from someone I knew
	5. I bought it from someone I didn't know personally
	6. I obtained it in some other way

CR8. How easy or difficult would it be for you to obtain crack?

	1. Easy
	2. Difficult
	3. Would not be able to obtain
	4. Don't know

CR9. When was the last time you were offered crack, whether to buy or to try? Think only about the last time you were offered crack.
Choose one answer only.

	1.	Never	**(skip to question EX1)**
	2.	During the past 30 days	
	3.	More than 1 month ago but less than 1 year ago	
	4.	More than 1 year ago	

CR10. If you were ever offered crack, think about the last time it was offered to you. Where was it offered? *Choose one answer only.*

	1. At home
	2. At school
	3. In the area around my school
	4. At a party, nightclub, concert
	5. In the area around my house
	6. By Internet, email or chat room
	7. At the house of friends/family

CR11. The last time you were offered crack, who offered it to you?

	1. A friend
	2. A family member
	3. Someone I know who is not a friend
	4. Someone I didn't know
	5. Someone else

NOTE:
IN THOSE COUNTRIES WHERE BOTH CRACK AND COCAINE PASTE ARE USED, BOTH SETS OF QUESTIONS SHOULD BE INCLUDED IN THE QUESTIONNAIRE. OTHERWISE, SELECT THE SECTION ON THE RELEVANT SUBSTANCE.

We will now ask some questions about the use of ecstasy.

EX1. Have you ever used ecstasy?

	1. Yes
	2. No **(skip to question EX8)**

EX2. How old were you when you first used ecstasy?

	Years old

EX3. When was the first time you used ecstasy?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

EX4. Have you used ecstasy in the past 12 months?

	1. Yes	
	2. No	**(skip to question EX8)**
	9. Don't know/no opinion	

EX5. How often have you used ecstasy?

	1. Only once
	2. Several times during the past year
	3. Several times a month
	4. Several times a week
	5. Every day

EX6. Have you used ecstasy in the past 30 days?

	1. Yes	
	2. No	**(skip to question EX8)**
	9. Don't know/no opinion	

EX7. As you think about the past month, how did you obtain the ecstasy that you used? Choose all that apply.

	1. A friend or one of my brothers or sisters gave it to me, or someone I knew personally
	2. Someone I didn't know gave it to me
	3. It was shared with a group of friends
	4. I bought it from a friend or from someone I knew
	5. I bought it from someone I didn't know personally
	6. I obtained it in some other way

EX8. How easy or difficult would it be for you to obtain ecstasy?

	1. Easy
	2. Difficult
	3. Would not be able to obtain
	4. Don't know

EX9. If you were ever offered ecstasy, when was the last time it was offered to you, whether to buy or to try? Think only about the last time you were offered ecstasy.
Choose one answer only.

	1. Never	**(skip to question LSD1)**
	2. During the past 30 days	
	3. More than 1 month ago but less than 1 year ago	
	4. More than 1 year ago	

EX10. If you were ever offered ecstasy, think about the last time it was offered to you. Where was it offered? Choose one answer only.

	1. At home
	2. At school
	3. In the area around my school
	4. At a party, nightclub, concert
	5. In the area around my house
	6. By Internet, email or chat room
	7. At the house of friends/family

EX11. The last time you were offered ecstasy, who offered it to you?

	1. A friend
	2. A family member
	3. Someone I know who is not a friend
	4. Someone I didn't know
	5. Someone else

This next set of questions relate to the use of LSD

LS1. Have you ever used LSD?

	1. Yes	
	2. No	**(skip to question IN1)**

LS2. How old were you when you first used LSD?

	Years old

LS3. When was the first time you used LSD?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

LS4. Have you used LSD in the past 12 months?

	1. Yes	
	2. No	**(skip to question IN1)**
	9. Don't know/no opinion	

LS5. How often have you used LSD?

	1. Only once
	2. Several times during the past year
	3. Several times a month
	4. Several times a week
	5. Every day

LS6. Have you used LSD in the past 30 days?

	1. Yes
	2. No

We will now ask some questions about the use of inhalants.

IN1. Have you ever used inhalants, such as glue, paint, varnish, solvents, deodorants, gasoline, perfume spray, propellants or something similar?

	1. Yes	
	2. No	**(skip to question PO1)**

IN2. How old were you when you first used inhalants?

	Years old

IN3. When was the first time you used inhalants?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

IN4. Have you used inhalants in the past 12 months?

	1. Yes	
	2. No	**(skip to question PO1)**
	9. Don't know/no opinion	

Which inhalants have you used in the past 12 months?

IN5.1		**Glue**
IN5.2		**Paint**
IN5.3		**Varnish**
IN5.4		**Solvents**
IN5.5		**Deodorants**
IN5.6		**Gasoline**
IN5.7		**Perfume spray**
IN5.8		**Propellants**
IN5.9		**Other**

IN5. How often have you used inhalants?

	1. Only once
	2. Several times during the past year
	3. Several times a month
	4. Several times a week
	5. Every day

IN6. Have you used inhalants in the past 30 days?

	1. Yes
	2. No

Next, we will ask some questions about the use of poppers.

PO1. Have you ever used poppers?

	1. Yes	
	2. No	**(skip to question OD1)**

PO2. How old were you when you first used poppers?

	Years old

PO3. When was the first time you used poppers?

	1. During the past 30 days
	2. More than 1 month ago but less than 1 year ago
	3. More than 1 year ago
	9. Don't know/no opinion

PO4. Have you used poppers in the past 12 months?

	1. Yes	
	2. No	**(skip to question OD1)**
	9. Don't know/no opinion	

PO5. How often have you used poppers?

	1. Only once
	2. Several times during the past year
	3. Several times a month
	4. Several times a week
	5. Every day

PO6. Have you used poppers in the past 30 days?

	1. Yes
	2. No

NOTES:

- ✓ **EACH COUNTRY SHOULD CONSIDER WHETHER OR NOT TO INCLUDE POPPERS. IF THE COUNTRY DECIDES TO DELETE THIS BLOCK OF SIX QUESTIONS, THEN POPPERS SHOULD BE INCLUDED AS ONE OF THE SUBSTANCES AMONG THE QUESTIONS FROM IN1 TO IN5 ON INHALANTS.**
- ✓ **COUNTRIES MAY ADD ADDITIONAL QUESTIONS FOR SPECIFIC SUBSTANCES.**
- ✓ **COUNTRIES MAY ADD A COMPLETE MODULE ON A SPECIFIC SUBSTANCE THAT REQUIRES SPECIAL ATTENTION.**

Here are questions about the use of other substances.

When was the last time you used any of these substances?	Never used (1)	More than 1 month ago but less than 1 year ago (2)	More than 1 year ago (3)	In the past month (4)
OD1. Methamphetamine (Meth, ice, crystal)	1	2	3	4
OD2. Amphetamine (fet, speed)				
OD3. Heroin (Paste, "H", white powder, skag and tar).	1	2	3	4
OD4. Ketamine (Keta, vitamin K, super K, CK or Calvin Klein, Mary Kay or María-K).	1	2	3	4
OD5. Opium	1	2	3	4
OD6. GHB (Liquid X, liquid ecstasy, Georgia homeboy, Oop, Gamma-oh, grievous bodily harm, Mils, "G", Liquid G, Fantasy).	1	2	3	4
OD7. Synthetic cannabinoids (synthetic marijuana, Spice, K2, Joker, Black Mamba, Kush or Kronic).	1	2	3	4
OD8. Synthetic cathinones (bath salts).	1	2	3	4
OD9. Aminoindanes (MDAI gold).	1	2	3	4
OD10. Phencyclidine (PCP, angel dust, embalming fluid, hog, killer weed, love boat, ozone, peace pill, super weed, rocker fuel, quirky).	1	2	3	4
OD11. Phenethylamines (Europa, 4-FMP, RDJ, 4-MMA, Methyl-MA, 2C-C-NBOMe, Bomb, N-Bomb N, 251, Nexus, 2C-E and Blue mystic).	1	2	3	4
OD12. Piperazines (BZP, mCPP, A2, Legal X y Pep X).	1	2	3	4
OD13. Hallucinogenic plants (Floripondio, angel trumpets, bell, borrachero o cacao sabanero; DMT, yagé or ayahuasca; mescaline or peyote; **Psilocybin,** hallucinogenic mushrooms or magic mushrooms; Khat; Salvia, salvia divinorum María Pastora; scopolamine or burundanga).	1	2	3	4
OD14. Anabolic steroids	1	2	3	4
OD15. Caffeine products (caffeine pills, energy drinks, powdered caffeine)	1	2	3	4
OD16. Lean (Sprite mixed with cough syrup and candy – also called *purple drank* o**r** *sizzurp*)	1	2	3	4

NOTE: COUNTRIES MAY DELETE SOME OF THESE OR INCLUDE OTHERS

MODULE 4. PERCEPTION OF HARM/RISK, AND POSSIBILE SUBSTANCE USE. The following questions relate to your opinion on the effects of using certain drugs.

What risk or harm does a person experience if he or she engages in the following behaviors?					
MARK YOUR ANSWER WITH AN X ON EACH LINE.	Not harmful	Slightly harmful	Moderately harmful	Very harmful	Don't know
12.1 Smoking cigarettes frequently	1	2	3	4	5
12.2 Smoking one or more packs of cigarettes a day	1	2	3	4	5
12.3 Vaping nicotine sometimes	1	2	3	4	5
12.4 Vaping nicotine frequently	1	2	3	4	5
12.5 Drinking alcoholic beverages from time to time	1	2	3	4	5
12.6 Drinking five or more drinks one or more times every weekend	1	2	3	4	5
12.7 Getting drunk	1	2	3	4	5
12.8 Often taking medication without a doctor's prescription	1	2	3	4	5
12.9 Inhaling glue, paint, varnish, deodorant, gasoline or similar once in your life	1	2	3	4	5
12.10 Frequently inhaling glue, paint, varnish, deodorant, gasoline or similar	1	2	3	4	5
12.11 Smoking marijuana sometimes	1	2	3	4	5
12.12 Smoking marijuana frequently	1	2	3	4	5
12.13 Vaping cannabis sometimes	1	2	3	4	5
12.14 Vaping cannabis frequently	1	2	3	4	5
12.15 Smoking cocaine paste (pasta base/coca paste) or crack once in your life	1	2	3	4	5
12.16 Smoking cocaine paste (pasta base/coca paste) or crack frequently	1	2	3	4	5
12.17 Snorting cocaine sometimes	1	2	3	4	5
12.18 Snorting cocaine frequently	1	2	3	4	5
12.19 Using ecstasy sometimes	1	2	3	4	5
12.20 Using ecstasy frequently	1	2	3	4	5
NOTE: COUNTRIES MAY ADD OTHER SUBSTANCES TO QUESTION 12.					

13. Have you ever felt curious about trying a drug? (marijuana, cocaine, cocaine paste, ecstasy or similar)		14. If you had the chance, would you try drugs? (marijuana, cocaine, cocaine paste, ecstasy or similar)	
	1. Yes		1. Yes
	2. No		2. No
	3. Perhaps		3. Perhaps
	9. Don't know/no opinion		9. Don't know/no opinion

MODULE 5. The following questions relate to your relationship with your parents or with the people you live with.

15. How often does your mother, father or the adult that you live with not know where you are after school or on weekends. Let's say for an hour or more.

	1. They never or almost never know where I am
	2. Sometimes they don't know
	3. They always or nearly always know where I am

16. In general, does your father, mother or the adult you live with pay attention to the programs you watch on television, the Internet sites you visit, the social networks you use, the online games you play or the streaming programs or Internet sites you look at?

	1. Yes
	2. No

17. How closely does your father, mother or the adult you live with pay attention to what you are doing at school?

	1. Very closely
	2. Closely
	3. Somewhat
	4. Not at all

18. How many days per week do you sit to eat together with your mother, father or other adult you live with, at the same table for breakfast, lunch, dinner/supper or tea (Mark only one option)

	0. Never		4. Four days
	1. One day only		5. Five days
	2. Two days		6. Six days
	3. Three days		7. Every day

19. On weekends, does your father, mother or other adult you live with check on the time you get home at night?

	1. Yes
	2. No

20. When you go out at night or on weekends, does your mother, father or adult living with you ask you where you are going, or expect that you will tell them?

	1. Yes
	2. No

21. Generally, how well do you think your father, mother or adult you live with know your best friends?

	1. Very well
	2. More or less
	3. Slightly
	4. Not at all

How do you think your father and mother would react if they caught you coming home tipsy or drunk?

22.Your father	Extremely angry/upset	Very angry	Somewhat angry	Not very angry	I don't know how he would react	Not applicable
	1	2	3	4	5	6
23.Your mother	Extremely angry	Very angry	Somewhat angry	Not very angry	I don't know how she would react	Not applicable
	1	2	3	4	5	6

How do you think your father and mother would react if they found out you smoke marijuana?

24.Your father	Extremely angry	Very angry	Somewhat angry	Not very angry	I don't know how he would react	Not applicable
	1	2	3	4	5	6
25.Your mother	Extremely angry	Very angry	Somewhat angry	Not very angry	I don't know how she would react	Not applicable
	1	2	3	4	5	6

YOUR RELATIONSHIP WITH YOUR PARENTS	**Excellent**	**Very good**	**Good**	**Not very good**	**Poor/bad**	**Not applicable**
26. How would you describe your current relationship with your father?	1	2	3	4	5	6
27.How would you describe your current relationship with your mother?	1	2	3	4	5	6
28. How would you describe the relationship between your parents? Describe it	1	2	3	4	5	6

here, even if they don't live together.						

29. Have you talked seriously with either of your parents (or with the adult living with you) about the dangers of drug use?

	1. Yes
	2. No

30. Do you think that your parents used drugs when they were young? (do not include cigarettes or alcohol)

	1. Yes
	2. No
	9. Don't know

31. Does either of your parents regularly smoke at least one cigarette a day?

	1. Yes, my father
	2. Yes, my mother
	3. Yes, both of them
	4. No, neither of them
	9. Don't know/don't live with them

32. As far as you know, do your brothers or sisters or anyone else living in your house currently use drugs? (do not include cigarettes or alcohol)

	1. Yes
	2. No
	9. Don't know

33. How would you describe your father's *drinking habits*? (wine, beer, hard liquor)

34. How would you describe your mother's *drinking habits*? (wine, beer, hard liquor)

	Father	Mother
Never drinks alcohol	1	1
Only on special occasions	2	2
Only on weekends, but never during the week	3	3
Drinks alcohol every day, one or two drinks	4	4
Drinks alcohol every day, more than two drinks	5	5
Not applicable	6	6

MODULE 6. THESE QUESTIONS RELATE TO YOUR RELATIONSHIP WITH SCHOOL, ACADEMIC PERFORMANCE, AND FUTURE PROSPECTS

35.How happy do you feel when you go to school?

	1. Very happy
	2. Fairly happy
	3. Neither happy nor unhappy
	4. Unhappy
	5. Very unhappy

36.Generally speaking, would you say that you feel a sense of belonging at school?

	1. Yes
	2. No

37. This year, did you play truant or not go to school for most or all of the day?

	1. Never
	2. A few times
	3. Several times
	4. Often

38.What is your average grade/mark at the end of the school year? Choose one of the answers below. EACH COUNTRY SHOULD INCLUDE THE RANKINGS OR GRADES GIVEN IN THE EXAMINATION SYSTEM.

	1. Less than XX
	2. Between XX and XX
	3. Between XX and XX
	4. Etc.
	5. Etc.

39.During the year, how many days have you failed to go to school for whatever reason? Choose one of the answers below.

	1. Less than 5 days
	2. 5 to 10 days
	3. 11 to 20 days
	4. 21 to 30 days
	5. More than 30 days

40. How would you generally describe your relationships with your teachers?

	1. Very good
	2. Good
	3. Average
	4. Bad
	5. Very bad

41. Do you work as well as attend school?

	1. Yes
	2. No **(Skip to Question No. 43)**

42. How many hours per week do you work?

	Hours

43. How likely is it that you will finish your last year of high school/secondary school?

	1. Very likely
	2. Likely
	3. More or less likely
	4. Not very likely
	5. Very unlikely or impossible
	9. Don't know/no opinion

44. How likely is it that you will attend university or college or other institute of higher education?

	1. Very likely
	2. Likely
	3. More or less likely
	4. Not very likely
	5. Very unlikely or impossible
	9. Don't know/no opinion

45. How many grades/forms have you had to repeat throughout your school years?

	1. None
	2. One
	3. Two or more
	9. Don't know/no opinion

46. Have you had any behavioral or disciplinary problems at school?

	1. Never
	2. A few times
	3. Often
	9. Don't know/no opinion

47. If your close friends knew that you smoked marijuana, do you think that…

	1. They would criticize you, or they would tell you to stop?
	2. Some would criticize you but others wouldn't?
	3. They wouldn't criticize you at all or wouldn't say anything?
	9. Don't know/no opinion

48.Generally speaking, do you think that drugs are available in your school? Do some students bring drugs to school or try drugs or deal drugs inside the school?

	1. Yes
	2. No
	9. Don't know

49. Do you think that there are drugs in the area around your school? In other words, some students try drugs, or buy or deal them in the neighborhood around your school?

	1. Yes
	2. No
	9. Don't know

50. Have you personally seen a student selling drugs or dealing drugs inside the school or in the area around your school?

	1. Yes
	2. No
	9. Don't know

51. Have you personally seen a student using drugs at school or in the area around your school?

	1. Yes
	2. No
	9. Don't know

THINKING ABOUT YOUR FRIENDS	None	Less than half	Half	More than half	All or almost all	Don't know
52. How many of your friends regularly drink alcohol? Let's say, every weekend or more often	1	2	3	4	5	9
53. How many of your friends regularly smoke marijuana? Let's say every weekend or more often	1	2	3	4	5	9

54. If you had ever tried marijuana, would you have said so when you answered this questionnaire?

	1. Yes, I just said so
	2. Definitely yes
	3. Probably yes
	4. Probably no
	5. Definitely would not say so

55. If you had ever tried cocaine or coca paste, would you have said so when you answered this questionnaire?

	1. Yes, I just said so
	2. Definitively yes
	3. Probably yes
	4. Probably not
	5. Definitely would not say so

THANK YOU VERY MUCH FOR PARTICIPATING IN THIS SURVEY. PLEASE PUT YOUR QUESTIONNAIRE IN THE BOX PROVIDED.

6. STATISTICAL ANALYSIS

Two basic points should be borne in mind when performing a statistical analysis of the data collected in the survey:- (1) the objectives of the survey, and (2) the results that will be communicated by various means. A plan of analysis must therefore be prepared.

In general, the plan of analysis will cover three broad areas which will in turn become the basis for the report(s) that will be published. In summary, these areas are: (1) Description of the sample, (2) Estimation of indicators on substance use, and (3) Comparison of Indicators.

6.1 DESCRIPTION OF THE SAMPLE

The sample (including the expanded target population) is described by means of general variables such as the *student's grade or form*, *sex*, *age*, *type of school, and geographical distribution,* if the study has coverage at that level. A table such as the one below should be produced:-

Table 1: Distribution of the sample by demographic variables and population represented

Variables	Sample size	Population represented	%
Gender			
Male			
Female			
Grade*			
8th			
10th			
12th			
Age**			
14 and younger			
15-16			
17 and older			
Type of school*			
Public			
Private			
TOTAL	**n**	**N**	**100**

* Use the classification as defined for the study in the country

** Another classification could be used.

The "*sample size*" column describes the number of cases in the *effective sample* (having eliminated the cases that were present but that were considered invalid for inclusion in the

analysis, as explained below.) Thus, **n** represents the total number of cases in the sample that are considered **valid** for the study.

The "*Population represented*" column corresponds to the number of students in the population classified according to the corresponding variables, where **N** is the total number of students in the country represented in the study obtained by means of the **expansion factor** applied to each case in the sample. This concept is discussed further below.

Lastly, the "%" column represents the percentages of each category in relation to the total population represented (N).

Some of the concepts described above are discussed below:-

a) **Effective sample**: the concept of "sample" is commonly used in sample surveys at different points in the study, but it involves different questions. The first time the word "sample" is used is during the organization of the study, to answer the question *"How many cases do I need to study?"* in order to achieve the objectives of the study. There are formulae for this associated with certain pre-determined conditions. Let us assume that in accordance with those conditions, and using the appropriate formulae, we decide on a sample size of 5,000 (to be amended for possible rejections or other contingences). Based on this sample size, the schools and classes (classrooms) are selected for study according to grade. Let us assume that 150 classes are selected from different grades. In the fieldwork with the 150 classes, we may find that the total number of students differs from the original 5,000 (the actual number may be larger or smaller). Some simple explanations for this difference may be that on the day that the study is administered, some students are absent due to illness, or the number of students present in a classroom is different from the number in the sample frame at the moment the sample was drawn. There may be other circumstances that create differences between the number of cases determined during the planning and organization and the actual number of students present in the classroom when the questionnaire is administered. Continuing with the same example, let us assume that of the 5,000 in the sample, 4,800 students responded to the survey questionnaire. We now have a "second" sample size which will determine the number of questionnaires available, in this case, 4,800.

   Lastly, there are reasons why some questionnaires should be excluded from the analysis. For example, if a student answered the demographic questions only, or answered "yes" to past month use of all of the substances specified in the questionnaire. It is important that those directing the study, that is, the professional staff of the NDO, determine the

exclusion criteria in advance to remove cases from the analysis. Having performed this "elimination," we will obtain *valid cases* that comprise the **final sample** or the **effective sample** for statistical analysis. Again, let us assume that in the previous example, we have 4,700 valid cases.

b) **Expansion factor**: as noted above and as explained more fully in Annex 2, there are different sampling designs. The study among secondary school students, in particular, corresponds to what is called "complex sampling." This involves different stages in the process of selecting students who will constitute the sample, and at each stage, the probability of selection is defined for each student in the population. According to some probability theories, the final probability that a student becomes part of the sample consists of multiplying probabilities at each stage of the sampling process. Mathematically speaking, ***the expansion factor associated with a student in the sample corresponds to the reciprocal probability of selection of that student***. This means that the result ***represents a particular number of students in the population***. Each student in the designated sample will be associated with a particular expansion factor, which will be the same for each student in the same class. In other words, the students from the same class will be associated with the same expansion factor, but it will be different from the expansion factor for the other classes. The **sum** of the expansion factors for the students in the effective sample (4,700 in the example here) will produce the total number of students in the target population, shown as **N** in the table above.

6.2 ESTIMATING INDICATORS ON SUBSTANCE USE

A second area of analysis concerns the estimates of the prevalence and incidence of substance use. Firstly, it should be noted that this type of study, which is based on samples, allows for **estimates** of what occurs in the population under study, rather than an exact, error-free determination of the situation of drug use in the population. This means that there will be some uncertainty or error associated with these estimates—something that must be stated explicitly when describing estimates. Fortunately, the degree of error can also be estimated on the basis of the sample data which should be presented with the estimates of the indicators. These errors can be presented by means of ***standard errors*** of the estimate which are used to build the ***Confidence Intervals***. The population indicators (unknown) are termed **parameters**, which is what we are trying to estimate by means of a sampling process. We therefore have to consider three concepts in the estimation process:

1. The **parameter** to be estimated (for example, the prevalence of past month use of alcohol among 8th, 10th and 12th grade students combined),
2. The **estimate** of the parameter on the basis of the sample, and
3. The **standard error** of that estimate.

The Confidence Interval **(CI)** is constructed on the basis of the estimate and the standard error. From SIDUC's standpoint**, we strongly recommend this as the strategy for analysis and presentation of** the results, rather than simply giving the estimate of the indicator (that is, without the standard errors and Confidence Intervals.)

For example, let us assume that we have a study in a country where the population represented is 1,000,000 students in the 8th, 10th and 12th grades, and that the designated sample is 2,000 students. By correctly applying the expansion factors and the sample design (further details provided in Annex 2), the results of the study will show that 20% of the students report that they had used an alcoholic beverage in the past month, with a standard error of 1%. Thus, the prevalence of alcohol use in the past month is 20%, with a 95% Confidence Interval of between 18.04% and 21.96%. The 95% Confidence Interval used to estimate a proportion (expressed as a percentage) of a population is expressed as:

$$p-1.96*ee(p)\ ;\ p+1.96*ee(p) \quad (1)$$

where:

- 1.96 is the value of the normal distribution of an estimate with a confidence level of 95%, and
- ee(p) is the **standard error** of the estimate,
- **p-1.96∗ee(p)** as the ***lower limit*** and **p+1.96∗ee(p)** is the ***upper limit*** of the 95% Confidence Interval, and
- **d=1.96∗ee(p)** is called the **precision** of the estimate.

Thus, a Confidence Interval may be expressed as:

$$p-d\ ;\ p+d \quad (2)$$

Going back to the previous example, the classical way of presenting this information in the report on the study is given in the following table:

Table 2: Prevalence of past month alcohol use and 95% Confidence Intervals

Variable	Prevalence (%)	95% CI
Past month alcohol use	20	18.04 – 21.96

In addition to reporting on the overall prevalence (20% in this case), it will always be necessary to disaggregate that indicator by sex, grade, and type of school. This information is given in the table below using fictitious data:

Table Nº 3: Past month prevalence of alcohol use and 95% Confidence Intervals by sex, grade and type of school

Variables		Prevalence (%)	95% CI
Sex	Male	25	22.5 – 27.5
	Female	14	11.4 – 16.6
Grade	8th	8	5.2 – 10.8
	10th	19	15.9 – 22.1
	12th	30	27.5 – 32.5
Type of school	Public	18	16.2 – 19.8
	Private	21	18.4 – 22.6
Total		**20**	**18.04 – 21.96**

As we advance further with the data analysis, it may be necessary to generate indicators broken down by other categories. For example, it might be of interest to estimate and compare indicators on drug use among students who were exposed (or not) to prevention programs in their schools, or according to the students' perception of risk or harm of substance use, and so on. It is important to bear in mind that the errors of the estimate increase as the size of the sample becomes smaller, and this has a direct impact on the width of the Confidence Interval in question. A Confidence Interval that is too wide will not be very informative and should therefore be avoided. These issues are discussed in Annex 1 and 2.

6.3 COMPARISON OF INDICATORS

A third area of interest in the analysis process is the comparison of indicators. For example, depending on the objectives of the study, we need to know whether there are differences in the prevalence of past month alcohol use between male and female students. Similarly, we might be interested in comparing the use of alcohol or other substances by grade, type of school, etc. The analysis could also focus on examining the relationship between substance use and other characteristics such as academic performance, access to substances, etc.

These topics will be discussed again in Annex 2. However, it must be stressed that the ***statistical analysis*** of a survey ***must focus on the objectives of the study and must capture each of them***.

ANNEX 1: SAMPLING

In this Annex, we return to some of the concepts described in section 3.3 above. First, the **target population** for the study must be determined, which means:-

- Determining geographical areas (according to the size of cities/towns, urban/rural areas), and,
- Determining the grades of the students who are subjects of the study (in accordance with SIDUC, for example, those in the 8th, 10th and 12th grades. As discussed earlier, the grades might be different from country to country.)

Once the target population has been defined, it is necessary to gather information about that population which is called the sampling frame.

Sampling frame: Let us assume that the first item (geographical area) is defined as *schools in cities of 30,000 inhabitants or more*, and the second item is defined as the *8th, 10th and 12th grades*. This clearly determines the target population, and it is therefore possible to determine the **sampling frame**, that is, the ***list of schools*** that have these characteristics. This list should be available from the Ministry of Education in the country. The information needed in order to select the sample is given in the table below, using fictitious data:-

Table A1.1: Outline of a sampling frame

Information on the school					Information on classes and students					
					Grade 8		Grade 10		Grade 12	
Code	Name, address, telephone, name of contact person	State, department or subregion	Town	Type (public/private)	**Number of students**	**Number of classes**	**Number of students**	**Number of classes**	**Number of students**	**Number of classes**
1				Public	80	2	60	2	50	2
2				Public	90	3	70	2	60	2
3				Private	40	2	40	2	35	2
4				Public	120	4	100	3	90	3
5				Private	65	2	60	2	60	2
6				Private	80	3	75	3	70	3
..				..	..	..	..	..	..	..
..				..	..	..	..	..	..	..

Having defined the target population and developed the corresponding sampling frame, the means of obtaining the samples will depend on the representativeness that is desired for

the study. If the study has only ***national representation***, then there will be only **one sample** (which includes the three grades) which will be selected directly from a **sampling frame** like the one shown in the table above.

However, it is possible that depending on the objectives of the study, specific information is required on two levels (in addition to national-level representation), for example, from the country's metropolitan area and from the remaining geographical areas. This means dividing the earlier sampling frame into **two sampling frames**, one with all the schools in the cities/towns in the metropolitan area, and the other with schools in other areas of the country. Two random samples are thus selected independently using the corresponding sampling frames. This strategy will yield three levels of indicators: ***national level***, ***metropolitan areas*** and the ***remaining areas grouped together***. Since we want estimates for two geographical divisions, it must be stressed that the size of the sample in each segment must be large enough to produce estimates with acceptable errors.

The same procedure is used when estimates for a smaller geographical area are planned. If a country has 20 regions or departments, and it has been agreed to estimate indicators at that level, then the earlier sampling frame must be divided into 20 sampling frames, each containing information on the schools in their respective region or department. A random sample of classes in the three grades under study will be obtained from each of these.

Generating *a priori* the sampling frames according to a particular criterion (generally a geographical area) is known as pre-stratification. However, when analyzing the survey data, it becomes necessary to generate sub-groups in order to examine specific characteristics, for example, males and females, or type of school, etc. In these cases, we talk about post-stratification. The main difference between pre-stratification and post-stratification is that in the former, the size of the sample of each sub-group (stratum) is determined from the study's design. The sizes of the samples obtained from post-stratification are not known in the design and may be insufficient to obtain robust estimates at the desired levels. For example, if the specific objectives required estimates by type of school, let us assume that 20% of students in a country go to private schools and 80% to public or state schools. If this situation is not taken into account in the design, we would expect similar proportions in the sample, and it may therefore be a case of the sample size being insufficient to obtain estimates in private schools with acceptable sampling errors.

It should be clear that discussion on representativeness, sample sizes and precision of the estimates should form part of the ***planning of the study***. The decisions taken on these areas including levels of representativeness will have a direct impact on the size of the sample, on the complexity of administration of the study, and undoubtedly, on the overall cost.

Having determined the levels of representativeness and the divisions required for the corresponding estimates, the next question relates to the size or sizes of the sample(s). The size of the overall sample (total number of cases/subjects to be studied) depends on a number of factors:-

- The first factor to consider is the level of **representativeness** of the sample, that is, whether estimates are needed only at the national level or in a particular area of the country determined a priori during the planning process.
- The second relates to the **sampling or design method** that will be used.
- The third examines the **variability** of the population with regard to the indicator considered to be the most important. For example, we might assume that the indicator of the greatest interest is the **prevalence of past year use of an illicit substance**, or the **past month prevalence of alcohol use**.
- The fourth relates to the **precision** of the estimates which accounts for the width of the Confidence Interval or the **coefficient of variation** associated with the estimate of interest.
- Finally, the fifth factor relates to the desired **level of confidence** in the estimate.

In order to determine the size of the sample, it is necessary to determine these conditions a priori. It should first be borne in mind that the principle interest of the study is to ***estimate*** the prevalence of a substance or group of substances. We could therefore say that this indicator is the *prevalence of the use of an illicit substance in the past twelve months*.

The value of that indicator in the target population is called a **parameter**, and the value obtained by means of a random sample is called the **estimator**. Of course, the parameter is an unknown value; this is what we are trying to estimate by means of a sample, i.e., a subset of elements that make up the population.

On this basis, it is possible to analyze each of the factors that determine the size of the sample. Perhaps the simplest way is to determine the **confidence level** which defines the probability that the confidence interval constructed for the parameter of interest includes that indicator. Usually, that probability is 0.95 (or 95% as it is often written) and based on a standard normal distribution of the estimator for the parameter of interest; the percentile that represents that probability is 1.96, defined as **z**, or z=1.96. *The greater the confidence in the estimate, the larger the sample size will be.*

Another item mentioned above is the **variability** found in the population with respect to the variable under study. For example, if the variable is "*use of an illicit drug in the past 12 months"*, it may be found in different forms within a population: in one population, the prevalence (in %) for that variable may be 1%, and in another population, it may be 10% or

50%. If we call that prevalence P, then the variability will be given by the product of P and its complement, that is Q=100-P, or P*Q. A value of P=50% means that the population is divided into two equal parts, which means maximum variation. This variation decreases to the extent that P moves away from 50%. The size of the sample is directly proportionate to variability, which means that *the more variation, the larger the required sample size.*

A factor that has a significant impact on the size of the sample is **precision**. This concept is related to the width of the confidence interval to be constructed. For example, a confidence interval between 10% and 50% is less informative than a confidence interval between 25% and 35%, or between 29% and 31%, and will require a smaller sample size than the latter two. The *greater the desired precision* of the estimate (the narrower the Confidence Interval), *the larger the sample size should be*.

In summary, there are many important points to consider when determining the size of the sample during the planning phase of the study, as we have described. One, in particular, is the sampling design: as stated earlier, we may have ***simple random sampling*** and ***complex sampling***, such as in the school study.

When estimating a proportion (P) in simple random sampling, the variance of that estimate[11] is given by:

$$V_{mas}(p) = \frac{p(1-p)}{n} \quad (3)$$

where **n is the size of the sample**.

Unlike a simple random sampling, the study among secondary school students is a complex sample basically because of its design. Schools are selected first, and then classes within the selected schools are chosen. The total number of students in that class form part of the final sample.

The first point to note in the case of simple random sampling, is that all of the units in the sample are independent of each other. However, that is not the case with the design of the student survey. When a class is selected (it may be termed a "cluster"), the variable under study in the case of substance use may be impacted by the overall circumstances of the school, which means that the responses from the students in that class are not independent of each other. For example, if the school from which that class is a part has an ongoing drug use prevention program, that program might be having the same impact on all of the

[11] Throughout the annexes a sufficiently large population size is assumed, therefore the correction factor for a finite population is not applied.

students participating in it; therefore their behavior with regard to substance use would be affected similarly. This is not to say that all students would have the same response to drug use (they either use or they do not use), but rather, that program adds value to the student's personal decision about drug use. This situation (derived directly from the sampling design used) means that there is a degree of correlation between the answers of the students in the same class, which is known as an ***intra-class or intra-cluster correlation coefficient*** (ICC), which is called **ρ**. In the case of a simple random sample, this coefficient[12] would equal zero.

Due to the ICC, the variance for estimating a parameter in a complex sample is greater than the variance based on a simple random sample; this will have a direct impact on the size of the sample in complex sampling. This higher value (inflation) of the variance is expressed as the ***Variance Inflation Factor*** (FIV is the Spanish acronym)-or ***Design Effect*** (ED: Spanish acronym) which is defined as follows:-

$$\boldsymbol{FIV = ED = 1 + (m - 1) * \rho} \tag{4}$$

where m=number of clusters and ρ=ICC.

In the case of a simple random sample, where the units are independent of each other, ρ=0 and thus ED=1.

In the case of a complex sample, the variance of a proportion corresponds to the following expression:-

$$\boldsymbol{V_{mc}(p) = \frac{p(1-p)}{n} * ED} \tag{5}$$

This means that the variance in a complex sample corresponds to the variance of a simple random sample multiplied by the design effect, that is:-

$$\boldsymbol{V_{mc}(p) = V_{mas}(p) * ED} \tag{6}$$

In other words, according to the preceding formula, the design effect is the quotient of the variance of the complex sample and the variance of the simple random sample, that is:-

[12] In statistics, a correlation coefficient has values between -1 and 1, where the value 0 indicates no correlation.

$$ED = \frac{V_{mc}(p)}{V_{mas}(p)} \quad (7)$$

Both the design effect and other indicators are unknown a priori and therefore must somehow be determined in order to calculate the size of the sample needed for the study.

For this purpose, let us again consider the definition of precision that we saw above: d=1.96∗ee(p)

According to expressions (1) and (2), the precision is:

$$d = z * \sqrt{\frac{\mathbf{p}(\mathbf{1}-\mathbf{p})}{\mathbf{n}} * ED} \quad (8)$$

such that the size of the sample is defined as:-

$$n = \frac{z^2 * p * (1-p)}{d^2} ED \quad (9)$$

To solve this equation, it is necessary to:-

i. Determine the **confidence level** of the estimate. Identified as the value of ***z***. Based on the approximation of the normal distribution associated with the sample distribution of a proportion, and given that 95% confidence intervals are those most often used, the value for z is 1.96. Other values used are 1.64 for 90% confidence intervals, and 2.58 for 99% confidence intervals. Other probability distributions are used instead of the approximation to the normal distribution, particularly when working with statistical packages. One of these distributions is a binomial distribution.

ii. Determine a value for ***p***, the most important indicator to be estimated in the study, such as, prevalence of the use of an illicit drug in the past year. There are some options available for approximating this value: if the country has already conducted an equivalent study, then the prevalence from that study may be used. If the country does not have any previous equivalent studies, there may be local studies that could be helpful, or indicators that could be used from countries with similar characteristics, or regional or sub-regional averages given in the Reports on Drug Use prepared by the OID/CICAD/OAS. A less than optimal option would be to resort to the criterion of "maximum variation" which occurs when **p** (prevalence) is 50%. This is the worst option since it is highly unlikely (highly undesirable) that the prevalence of the use of an illicit

substance in the secondary school population would reach such a figure. In order to properly identify this value, what is needed is information relevant to the situation in the country as well as and considerable judgment on the part of the researchers in the NDO. It is critical to determine a value.

iii. Assign the value of the third component of the formula (9) ***d***, or precision**.** This value determines the width of the confidence interval and is closely related to the prevalence discussed in the point above—and in fact, it can be determined on the basis of the p-value. For example, to declare that d will not exceed 10% of p. That is, if we assume that p=10%, then d=1% such that the Confidence Interval is between 9% and 11%. However, if p is estimated at 20% then d=2 and the CI would be 18% to 22%. If it is determined that d shall not exceed 20% of the p-value, then d=2% in the first case and d=4% in the second. Various alternatives are given in the table below which reflect how they impact the size of the sample.

iv. Determine the last item to consider in calculating the sample size, the design effect. Again, it is necessary here to resort to other sources of information, whether through earlier studies conducted in the country or similar studies in other countries. As discussed above, the value of ED is more than 1, and may have different values. Again, determining the possible value of ED requires judgment and expertise.

To explain the issues discussed above, Table A1.2 below shows the results of sample sizes under different conditions. The first column gives 3 values for the design effect (1.5, 3, and 5); the second column simulates 4 different values (5%, 10%, 20%, and 30%) for the indicator of greatest interest (e.g., prevalence of the use of an illicit substance in the past year), and the last four columns show the sample sizes as a function of the desired precision (1%, 2%, 5%, and 10%).

It is important to note that **the question of interest here is the size of the sample that the study should have,** and in order to do so, for example, we will focus on the line highlighted in yellow, which allows us to analyze the contents of the table. This line assumes that the prevalence should be around 10% and that the design effect is 3. Therefore, if:-

- ✓ The desired precision is 1%, which means that the confidence interval will be between 9% and 11%, the size required under these conditions should be 10,372 students.
- ✓ If the desired precision is 2%, the sample size should be 2,593 students. In this case, the confidence interval will be 8% to 12%.
- ✓ So, if the precision is 5%, that is, an interval of between 5% and 15%, the study should be conducted on a total of 415 students which would result in the loss of precision.

✓ Lastly, with a precision of 10%, the number of cases required is 104, but the interval will be too wide (0% to 20%) and therefore, not very informative.

Table A1.2: sample sizes for combinations of prevalence (p), precision (d) and design effect (DE)

ED	p (%)	Precision d (%)			
		1	2	5	10
1.5	5	2,737	684	109	27
1.5	10	5,186	1,297	207	52
1.5	20	9,220	2,305	369	92
1.5	30	12,101	3,025	484	121
3	5	5.474	1,369	219	55
3	10	10,372	2,593	415	104
3	20	18,440	4,610	738	184
3	30	24,202	6,051	968	242
5	5	9,124	2,281	365	91
5	10	17,287	4,322	691	173
5	20	30,733	7,683	1,229	307
5	30	40,337	10,084	1,613	403

This table shows a number of elements about the size of the sample:

✓ For the same prevalence (p) and precision (d), as the design effect increases, the size of the sample increases.

✓ For the same design effect and the same prevalence, the greater the precision (lower d-value), the larger the sample size.

✓ For the same design effect and the same precision, the greater the value of prevalence p (i.e., greater heterogeneity in the characteristic under study), the larger the size of the sample.

It should be stressed that the sample size determined is the value that satisfies the requirements for estimates at the national level. This means that for information disaggregated by region, sex, or grade, for example, the precision of those estimates on the basis that sample size will be smaller than those obtained at the national level. All of these considerations should be taken into account when making the decision about the size of the sample.

Another important point for decision-making, concerns the errors of the estimate mentioned earlier. Generally speaking, there are two ways of looking at these errors: First, we have **absolute error (AE)** of the estimate, which is the same value we defined earlier as

precision, that is, **1.96*ee**[13] for estimates at 95% confidence (using the approximation to a normal distribution). Second, we have **relative error (RE)**, which is the *quotient of the absolute error and prevalence* and is usually expressed as a percentage. Thus:-

$$EA = 1.96 * ee \qquad y \qquad ER = \frac{EA}{p} * 100 \tag{10}$$

Looking again at the yellow line in Table A1.2, where p=10%, the table below shows the relative errors associated with the absolute errors:

Table A1.3: Relative errors (RE) for a prevalence of p=10% with different absolute errors (AE). All values are expressed in percentages

AE	1	2	5	10
RE = $\frac{AE}{10} * 100$	10	20	50	100

We see in Table A1.3 that for a prevalence of 10%, if the absolute error (expressed by the precision) is 5% (i.e., a confidence interval of 10% ± 5%, or between 5% and 15%), the relative error will be 50%, which is excessively high. Remember that the size of the sample in this case was 415, as seen in Table A1.2. There are no rules about acceptable values of relative error, but it is suggested that it be no higher than 30%. This places a new condition on determining the size of the sample for the study and is something to be considered in the process.

As we must appreciate, determining the sample size in a population study is no small feat. We must consider all of the elements discussed earlier. Therefore, starting with planning the study, the human resources responsible for determining all issues related to the sample must have training in statistics and particularly in sampling, must be on board. Their work is essential to the success of the study. It should focus on the following (this is not an exhaustive list):-

- ✓ Determine **the size of the sample** on the basis of the objectives of the study, the conditions described and the information available, and on that basis, determine the number of classes in each grade that will be included in the sample,
- ✓ Select **the sample**, i.e., select the schools and classes that include the students that comprise the units of analysis,
- ✓ **Calculate the expansion factors**, values needed for the subsequent statistical analyses, an issue that will be discussed in more detail in the next annex.

[13] ee refers to standard error.

ANNEX 2: STATISTICAL ANALYSIS

This Annex will go into some detail about the ***statistical analysis*** of the results of the school survey conducted by means of a self-administered questionnaire. As stated earlier, the analyses ***must respond to the objectives of the study*** and are both central to preparation of the relevant **reports**.

Analysis of the results of the study requires the following:-

1. Human and technical resources

The statistical analysis will require **human resources** that have the necessary training and expertise, and preferably those who have experience in the entire process starting from the planning of the study. **Statistical software** will be needed to perform the ***analysis of complex samples***. Such software includes SPSS,[14] STATA,[15] SAS,[16] and R.[17] The first three require a license, but R is free. A statistical analysis cannot be carried out without the appropriate computer software that can cater to the demands of the methodology.

2. Databases

What is a database? It is a ***matrix*** of the responses to the questions on the questionnaire included as a database in the software that is used.

The ***observations*** which represent the subjects of the study are found in the lines of the matrix, and the ***variables*** in the columns of the matrix correspond to the questions on the questionnaire. The matrix thus reflects each subject's ***responses*** to each variable. The following may serve as a general outline:-

	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	…
Observation 1						
Observation 2						
…..						

Generally, there will be *two databases*: the first will show the raw data of the answers to the questionnaire, without adjustment or manipulation.

The second database contains the information needed for the analysis as adjusted according to certain criteria: **cleaning**, **generating new variables**, and **expansion factors**. Each of these is explained below:-

[14] https://www.ibm.com/analytics/spss-statistics-software
[15] https://www.stata.com/
[16] https://www.sas.com/
[17] https://www.r-project.org/

a) **Cleaning the database**

This essentially means deciding on the criteria for exclusion, that is to say, to determine the reason why certain questionnaires should be excluded from the database. The national drug observatory (NDO) must determine these criteria in advance. Some typical criteria for exclusion should be:-

- ✓ Questionnaires that contain only the demographic variables, or where those variables, principally sex and age, are missing,
- ✓ Questionnaires that are incomplete,
- ✓ Questionnaires with affirmative answers on drug use (principally past year or past month) for all **or most** substances.

As a result of this, *observations will be deleted* from the database, i.e., lines will be deleted.

b) **Generating new variables in the database**

What do we mean by a variable? Basically, a ***variable*** is a *characteristic* that we want to study among the subjects of the study, in this case, the country's secondary school population. The variables to be included in the study derive directly from the study's objectives.

The variables of interest become ***questions*** on the questionnaire.

Generally speaking, a variable is represented by only one question in the questionnaire. However, it is possible that two or more variables may be represented by a single question. A typical case of the latter is variables of "first use" of any substance.

Regarding "first use", we are interested in finding out about the point in time when the students had their first experience, and hence there are two variables of interest: "*first use in the past month*" and *first use in the past year*", with the time the survey is conducted as the reference point. However, these two variables may be derived from *a single question* on the questionnaire:

When was the first time you used [NAME OF THE SUBSTANCE]?

	1. Never used
	2. During the past month
	3. More than 1 month ago but less than 1 year ago
	4. More than 1 year ago

The second answer enables us to find out who used the substance for the first time in the "past month," while affirmative responses 2 or 3 provide information on first use in the

"past year." This particular situation is examined in more detail in the section on statistical analysis.

A variable of interest may also need to be represented in the questionnaire by more than one question. A typical example of this is the psychometric scales or other ad hoc scales. For example, the "level of parental involvement" is expressed in seven questions on the questionnaire, and the combination of these questions produces a variable termed *score on the scale of parental involvement*.

The *second database* contains all the variables associated with the questions on the questionnaire, plus all those variables that are constructed on the basis of the original questions and are needed for the subsequent analyses. It should also contain a variable called an ***expansion factor,*** which is described below:-

c) **Expansion factor**

Due to the importance of the issue, determination of the expansion factors is discussed below in a special section.

The **expansion factor** is a variable that is the *inverse of the probability of selection of each student*, and it will have the same value for each student in the same class. The sum of the expansion factors of the total number of cases in the sample should be the same as the size of the target population. The expansion factors should be calculated by the expert who worked on the design of the sample and should be entered into the database.

Let us examine the following simple example (fictitious data, not actual data):- the total number of students in the country in the grades of interest to the "study" is 600,000, of whom 220,000 are in 8^{th} grade, 200,000 in 10^{th} grade, and 180,000 in 12^{th} grade. Let us assume that the size of the sample determined according to the considerations discussed in Annex 1 is 3,000 students, and that it is decided to produce a sample of 1,000 students for each grade. We also know that classes have, on average, 30 students. Therefore, we must select 33 classes per grade (giving a result of 990 students instead of 1,000).

Table A2.1 shows the distribution of students and number of classes per grade according to the information provided by the school authorities (assuming an average of 30 students per class):-

Table A2.1: Distribution of students and number of classes, by grade (fictitious exercise)

Grade	Number of students	Number of classes
8	220,000	7,330
10	200,000	6,670

12	180,000	6,000
Total country	600,000	20,000

For example, if we had available the complete list of the 7,330 classes in the eighth grade, the probability that each would be selected is:

$$Probability\ of\ selection\ of\ a\ class = \frac{number\ of\ classes\ in\ the\ sample}{number\ of classes\ in\ the\ population}$$

$$Probability\ of\ selection\ of\ an\ 8th\ grade\ class = \frac{33}{7330} = 0.00450205$$

Therefore, the expansion factor for each 8th grade student in the sample would be the inverse of that number:

$$Expansion\ factor\ 8th\ \ grade\ students \frac{7330}{33} = 222.12$$

Given that there are 33 8th grade classes in the sample, and on average, each class has 30 students, there will be 990 students from that grade, each with an expansion factor of 222.12. This expansion factor multiplied by the 990 students in the sample gives a total of 219,900. The same method should be applied to the other two grades.

Table A2.2 shows the situation for all three grades:

Table A2.2: Sample of classes and students, probabilities of selection and expansion factors, by grade

Grade	Sample Number of classes	Number of students in the sample	Probability of selection of class	Expansion factor class	Sum of expansion factors*
8	33	990	0.00450205	222.121212	219,900
10	33	990	0.00494753	202.121212	200,100
12	33	990	0.00550000	181.818182	180,000
Total country	99	2,970	---	---	600,000

*It is assumed that the number of students who answer the questionnaire is 30, which was the national average. In practice, the number will be the variable in each class selected in the sample.

This is a very simple example to illustrate the calculation of the probabilities of selection of the classes, and the expansion factors for each student. There are many variants that must be taken into account by those responsible for the selection of the sample, since any of them would vary the probabilities of selection and therefore the corresponding expansion factors. Some of the variants (which are not mutually exclusive) are:-

- ✓ Generate strata based on a pre-determined geographical distribution. The simplest is to generate two strata, the first with the schools in the country's metropolitan area, and the second using schools in the rest of the country, selecting the same number of classes in each stratum. In this scenario, the expansion factors for a specific grade will be different for students in that grade whose school is in the metropolitan area or the rest of the country.
- ✓ Select schools in the first phase and then classes within those schools. This scenario may reduce the number of schools in the sample, and this will no doubt have administrative benefits. As to the probabilities of selection, there will be two stages: schools and classes; hence the probability of each student will be the product of both probabilities, and the expansion factor will be reciprocal of this latter probability.
- ✓ Select independent samples of each grade combining geographical areas and types of school (public and private.) Again, the probabilities of selection will be dependent on the structure of the strata.

We see from the explanation above that the ***expansion factors depend entirely on the sampling strategy*** *determined* to select the sample of students. The sampling may be done in one stage (like the simple example given above), or in two or more phases. What is important is that it should be a random sample that ensures ***unbiased estimates of the parameters*** of interest defined in the study's objectives.

As stated above, the variable **expansion factor** is an essential part of the second database.

3. Statistical analysis

Having met these conditions, we can move to the analysis itself. This will require an analysis plan that basically covers three areas that complement each other but will be described separately in this Protocol. These areas are:-

- ✓ **Description** of certain characteristics of interest. For example, a description of the sample by sex, grade, age, type of school, or geographical areas, if pertinent. There will also be a description of other variables of interest, such as perception of risk or harm of substance use, family relationships, user profiles, etc.
- ✓ Construction of **confidence intervals** for the estimates of the most important indicators such as prevalence of incidence of drug use, both nationally and also by sex, grade, and type of school.
- ✓ Construction of **statistical models** to analyze the association between substance use and factors of interest, and to compare indicators on substance use, such as a comparison by sex, grade, and type of school.

To better explain the analyses that will be described, the table below shows a segment of a database. In this example, let us assume that it is the complete database, where:-

ID=Identification of the case (subject) (consecutive number)
X_1=Age (in years)
X_2=Sex (1=male; 2=female)
X_3=Grade (8=8th; 10=10th; 12=12th)
X_4=Past month use of alcohol (1=yes; 0=no)
X_5=Perception of risk/harm of drinking alcohol frequently (1=no risk/harm; 2=slight risk/harm; 3=moderate risk/harm; 4=high risk/harm; 5=don't know)
X_6=Easy to buy alcohol (1=very easy/easy; 2=difficult/very difficult; 3=don't know)
X_7=Expansion factor.

Table A2.3: List of 10 fictitious cases

ID	Age (X_1)	Gender (X_2)	Grade (X_3)	Past month alcohol use (X_4)	Perception of risk/harm (X_5)	Easy to buy alcohol (X_6)	Expansion factor (X_7)
1	13	2	8	0	4	3	124.2
2	18	1	12	0	2	2	130.2
3	14	2	8	1	3	3	51.4
4	15	1	10	1	4	1	87.4
5	17	1	12	1	1	1	23.9
6	16	2	10	0	4	3	87.4
7	13	2	8	0	4	2	34.5
8	15	2	10	0	2	2	85.9
9	17	1	12	1	4	2	154.1
10	16	1	10	0	2	1	64.0

When we refer to the description of certain variables, we think for example of *age* (x_1), *sex* (x_2) or *grade* (x_3). For variables such as *past month alcohol use* (x_4), the confidence interval must be included in the analysis. Lastly, when we want to examine the association between x_4 and x_2, or between x_4 and x_3, or else between x_4 and x_5 and x_6, we must construct statistical models. The following sections will examine the three analysis strategies:-

The analyses must always bear in mind the objectives of the study
All of the analyses must by weighted using the expansion factor
For the purposes of a national report, we suggest focusing on sections 3.1 and 3.2, that is to say, the description including the Confidence Intervals for the principal indicators. Section 3.3 can be used for specific reports.

3.1 Description

To explain the description of some of the variables, let us examine a database of 10 cases. The sum of the values in the last column and the expansion factors give a total of 843, which corresponds to the population represented by the sample of 10 students. We see in the database that there are 5 males and 5 females. If we do **not weight** by expansion factors, we would conclude that the distribution is 50% males and 50% females.

What does weighting mean? As stated earlier, the expansion factor corresponds to the number of individuals in the population that are represented by each case in the sample. Therefore, the sum of the expansion factors of males on one hand, and females on the other, would provide the male and female populations represented in the sample, and on that basis, the corresponding percentages for each group are determined. Thus, the expansion factors for males are 130.2; 87.4; 23.9; 154.1 and 64.0, the sum of which is 459.6. In the case of females, the sum of the expansion factors is 383.4, meaning that the total population represented by these 10 students is 843. The percentages of males and female respectively are calculated on this basis, and similarly with the variable ***Grade***. The results are shown in Table A2.4:-

Table A2.4: Description by sex and grade of 10 simulated cases

Variables	Sample size	Population Represented*	%
Sex			
Male	5	460	54.5
Female	5	383	45.5
Grade			
8^{th}	3	210	24.9
10^{th}	4	325	38.5
12^{th}	3	308	36.6
TOTAL	10	843	100

*values approximate to the whole

The percentage of students who perceive frequent use of alcohol as being of ***high risk/harm*** (variable x_5) can also be estimated. Of the 10 cases in the sample, five answered 4 “high risk”, which, if **not weighted** by expansion factors, one might conclude that 50% chose that particular answer. However, weighting the sum of the factors of those five cases, we have a total of 487.6, which gives 57.8% of the population represented by contrast to the 50% obtained without weighting.

In formal terms, we need to use specific nomenclature:

- ✓ It is assumed that H schools were selected
- ✓ J_h classes were selected within school h, where h=1,2,….H

- ✓ The students in each class are identified by the sub-index i=1, 2, 3…n_{jh} where n_{jh} represents the number of students in class j of school h.
- ✓ Thus, each student in the sample is identified by sub-indices ijh, or student i in class j in school h.

For the purposes of describing the sample, and also for the next stages of the analysis, let us assume that the study's principle variables can be presented in three possible ways:-

- ✓ **Binary** variables, namely, those that have only **two possible responses,** with values 1 and 0. For example, the use or non-use of a particular substance, where the variable has the value of 1 in the case of drug use, and the value 0 if there is none (see variable x_4 in the previous example).
- ✓ **Quantitative variables**, such as the age of the student, age of first use of alcohol, or number of days the student drank alcohol in the past month.
- ✓ **Qualitative variables with more than two responses possible:** for example, the student's grade (three options), perception of risk/harm of drug use, which has five possible responses (variable x_5 in the example), or ease of access with three options (x_6).

How are indicators estimated?

Let us first define what we mean by an **indicator**: ***an indicator is a measurement that summarizes*** the responses obtained for a variable. For example, where the variable "sex" is concerned, (which is binary), the traditional measurement would be the **percentage** of males or females. If the variable is "*use of marijuana in the past year*", then the appropriate indicator is called **prevalence** (of marijuana use in the past year), which is also expressed as a percentage. Another case might be the variable *age*, and indicators that summarize the responses may be the **average**, **median**, **standard deviation** or different **percentiles**.

Let us first examine the case of a binary variable, such as x_4=alcohol use in the past month (1=yes, 0=no).

Let us assume that we are interested in estimating **P**=*prevalence of the use of alcohol in the past month at the population level*, and to do so we define **p**=*prevalence of the use of alcohol in the past month at the sample level*[18] as:-

$$p = \frac{\sum_{h=1}^{H}\sum_{j=1}^{J_h}\sum_{i=1}^{n_j} \boldsymbol{x_{4ijh}} * \boldsymbol{f_{ijh}}}{\sum_{h=1}^{H}\sum_{j=1}^{J_h}\sum_{i=1}^{n_j} \boldsymbol{f_{ijh}}} \qquad (11)$$

[18] P (upper case) is often used to denote the parameter, that is, the value at the population level p (lower case) the estimator of P, i.e., the value obtained in the sample.

where:

$$\begin{cases} X_{4ijh} = 1 \; if student \; i \; of \; class \; j \; of \; school \\ h \; reports \; having \; used \; alcohol \; in \; the \; past \; month, and \\ X_{4ijh} = 0 \;\; if \; the \; contrary \; is \; true \end{cases}$$

F_{4ijh} represents *the expansion factor of student i in class j of school h stating that he/she had used alcohol in the past month.*

It should be noted in this example that the numerator corresponds to the *sum of the expansion factors of the students who reported that they had drank alcohol* in the past month, i.e., when x_4=1. The denominator corresponds to *the sum of the expansion factors of all cases in the sample*, i.e., it corresponds to the population represented. Thus, we have:

$$p = \frac{51.4+87.4+23.9+154.1}{124.2+130.2+\cdots+154.1+64} = \frac{316.8}{843} = 0.3758 \; (37.58\%)$$

We see from the above that the estimate of the prevalence of alcohol use in the past year is 37.58%. Note that if it had not been weighted by the expansion factor, the result would have been 40% (four affirmatives out of a total of 10 cases). This same result is shown below using the SPSS statistical software:

Table A2.5: Prevalence of past month alcohol use (X_4)

X_4		Frequency	Percentage	Valid Percentage	Cumulative Percentage
Valid	No	526	62.42	62.4	62.4
	Yes	317	37.58	37.6	100.0
	Total	843	100.0	100.0	

In the case of variable x_2=Sex, which has been coded with values 1 and 2, for the purpose of using the previous formula, the numerator should include the expansion factors for males, for example, to obtain the weighted percentages of males, and therefore the percentage of females. The results are given below in SPSS:-

Table A2.6: Distribution of the population represented by sex (X_2)

X_2		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	Male	460	54.52	54.5	54.5
	Female	383	45.48	45.5	100.0
	Total	843	100.0	100.0	

The second case is an analysis of a **quantitative variable**, such as x_1=Age.

Formula (11) operates in the same way. The result is as follows where $\bar{x}_1$ represents the **average** age in the weighted sample:

$$\bar{x}_1 = \frac{13 * 124.2 + 18 * 130.2 + \cdots \ldots + 16 * 64}{124.2 + 130.2 + \cdots . . + 64} = \frac{13{,}174.2}{843} = 15.63 \text{ years}$$

Again using SPSS, we obtain the following results:-

Table A2.7: Description of average age (X_1)

	N	Mean
X_1	843	15.63
N valid (list)	843	

However, the average is not the only indicator that can be used for quantitative variables; in fact, it is desirable to add four indicators: standard deviation, and the 25th, 50th and 75th percentiles.

The standard deviation shows the degree of variability of the variable under study. *The 25th percentile shows the value below which 25% of the observations are found, the 50th percentile, or mean indicates the value below which 50% of the cases are found, and the 75th percentile is the value below which 75% of the observations are found.*

The **standard deviation** is of particular importance since it is on the basis of this value that the Confidence Interval is determined. This indicator should be calculated in accordance with **the complex sampling design** used to conduct the study.

As an illustration of this and other topics to be examined later in this annex, we will use a portion of the sample of a study in the school population of a country. This sub-sample consists of *6,088 cases* (representing a population of *232,836*); First, we will determine the average age of the students in three different situations using SPSS:-

a.- **Without weighting for the expansion factors, i.e., assuming a simple random sample of 6,088 students**:

Table A2.8: Description of average age (X_1) using simple random sampling

	N	Mean	
	Statistical	Statistical	Standard error
Age	6088	15.1041	.02031
N valid (list)	6088		

b.- **Weighting for expansion factors, but without taking into account the sample design:**

Table A2.9: Description of average age (X_1)

with expansion factors

	N	Mean	
	Statistical	Statistical	Standard error
Age	232836	14.6934	.00325
N valid (list)	232836		

c.- **Weighting for expansion factors and taking the sample design into account:**

Table A2.10: Description of average age (X_1)
using a complex sampling design (with expansion factors)

				95% Confidence Interval	
		Estimate	Standard error	Lower	Upper
Mean	age	14.6934	.08812	14.5175	14.8694

It is important to note that in the first case (Table A2.8), the average age is 15.10. However, in the other two cases (Tables A2.9 and A2.10), it is 14.69. The first estimate is incorrect **because it does not consider the weights**. The second and third estimates of the average **are correct**. The difference between these last two cases relates to the manner in which the standard errors are calculated. In the second case, the standard error is very low (ee=0.00325) since it considers the weights as an expansion of the cases according to the expansion factor and therefore assumes the population as the sample itself, but not considering the sample design. By contrast, the result in A2.11 takes the sample design into consideration, and the weights are the inverse of the probabilities of selection. As a result, the standard error is ee=0.08812 which is much higher than the previous one. **This is the correct standard error** which leads to a 95% confidence interval for the average age with limits of 14.52 and 14.87 years.

The *design effect (DEFF)* can also be obtained together with the previous estimate:

Table A2.11: Estimate of the design effect

			95% Confidence Interval		
	Estimate	Standard error	Lower	Upper	Design effect
Average age	14.6934	.08812	14.5175	14.8694	19.741

As seen in the previous example, the average of the age variable can be correctly obtained by also applying expression (11) but with the quantitative variable X. However, to derive a correct calculation of the standard error and a correct determination of the respective Confidence Interval, a *statistical package for analysis of a complex sampling, such as SPSS or similar is required*. We shall return to this in point 2 below when we discuss the question of confidence intervals in more detail.

So far, we have reviewed the procedures for estimating a *binary* variable (past month use of alcohol and sex in the test example), as well as a quantitative variable (age in the sub-sample of 6,088 cases).

Lastly, the third situation in the analysis refers to a **qualitative variable with 3 or more possible responses**, such as the student's grade. As with a binary variable, the expansion factors of the students in each grade should be added, and then divided by the population represented. This would be fairly time-consuming if we used expression (11). But the table below shows the results using SPSS:-

Table A2.12: Percentage distribution of the expanded sample, by grade

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	8th	109694	47.1	47.1	47.1
	10th	88061	37.8	37.8	84.9
	12th	35080	15.1	15.1	100.0
	Total	232836	100.0	100.0	

According to the above, it can be deduced that using the expression (11), it is possible to obtain estimates for a proportion (for example, use of alcohol in the last month) when the corresponding variable has been encoded with values 0 and 1, and for an average (for example, age), through calculations that can be done by using a spreadsheet in Excel. In the same way, but with additional work, it is possible to obtain a description of the frequency of a qualitative variable with two or more levels of response (for example, risk perception or ease of access to some substance). However, in order to advance the analysis, it is necessary to have a statistical computer package.

Special mention should be made of the indicators on **incidence** of drug use. As stated earlier, incidence measures the proportion of **new cases.** For example, the "*incidence of marijuana use in the past year"* corresponds to the percentage of people who used *marijuana for the **first time** during the past year.* An important difference between prevalence and incidence indicators is that the definition of the former uses **the entire population** as the denominator, while incidence refers to the **population exposed**, that is, those who had not previously used the substance during the period defined by the study.

Let us review these concepts using the previous example:- we are interested in analyzing the **prevalence and the incidence of alcohol use in the past year**. The relevant questions would be:-

✓ **Past year prevalence**:

Have you used an alcoholic beverage in the past 12 months?

	1. Yes
	2. No

✓ **Past year and past month incidence**:

When was the first time you used an alcoholic beverage?

	5. Never used
	6. During the past month
	7. More than 1 month ago but less than 1 year ago
	8. More than 1 year ago

Let us assume that the answers to each question are as follows:-

➢ **To estimate past year <u>prevalence</u>**:

Table A2.13: Percentage distribution of the question: Have you used an alcoholic beverage in the past 12 months?

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	No	97989	42.1	42.1	42.1
	Yes	134847	57.9	57.9	100.0
	Total	232836	100.0	100.0	

➢ **To estimate past year and past month <u>Incidence</u>** (also weighted cases):

Table A2.14: Percentage distribution of the question: When was the first time you used an alcoholic beverage?

	Frequency	Percentage	Valid percentage	Cumulative percentage
Never used	91,761	39.4	39.4	39.4
During the past 30 days	41,391	17.8	17.8	57.2
More than 1 month ago but less than 1 year ago	34,667	14.9	14.9	72.1
More than 1 year ago	65,016	27.9	27.9	100.0
Total	232,836	100.0	100.0	

In the case of prevalence (Table A2.13), the result is direct: 134,847/232,836=0.579 (57.9%); that is, almost 58% of the students reported having used an alcoholic beverage in the past 12 months.

As for incidence, Table A2.14 demonstrates that the calculation is somewhat more complicated. First, for incidence in the past year, we must distinguish between those students **who are exposed to** alcohol use during the time period indicated, that is, those who have not yet used alcohol in the past year, and those students who **had used** alcohol in the past, i.e., prior to the year of the study. Therefore, those exposed correspond to the sum of answers 1, 2, and 3, that is, 91,761+41,391+34,667 = 167,819 students. The 65,016 students (weighted figures) who reported having used alcohol for the first time "more than 1 year ago" (since the study was conducted) are excluded, since the event of interest had already occurred within this sub-group. Therefore, the **incidence of alcohol use during the past year** corresponds to those who used for the first time "during the past 30 days," plus those who drank alcohol for the first time "more than 1 month ago but less than 1 year ago," divided by the total number of students exposed, i.e., the total population represented minus those who reported that they had used alcohol for the first time "more than one year ago." In other words:-

$$Incidence\ of\ past\ \ year\ alcohol = \frac{41391 + 34667}{232836 - 65016} = \frac{76{,}058}{167{,}820} = 0.453\ (45.3\%)$$

If, we want to estimate the incidence of alcohol use in the past month (instead of in the past year), the method should be the same: the numerator includes those students who said that they had used alcohol for the first time during the past 30 days, i.e., 41,391 students. In the denominator, we include f the total population minus those who had used alcohol prior to the specified time period, i.e., prior to the 30 days before the study. These cases relate to those who reported having used alcohol for the first time "more than 1 year ago" plus those who said they had used for the first time "more than 1 month ago but less than 1 year ago." The incidence of the use of alcohol in the past month is therefore expressed as:

$$Incidence\ of\ alcohol\ past\ month = \frac{41{,}391}{232{,}836 - 65{,}016 - 34{,}667} = \frac{41{,}391}{133{,}153}$$
$$= 0.311\ (31.1\%)$$

Expression (11) above allows any prevalence indicator to be estimated, where in all cases the denominator will be expressed as the total population represented.

We now provide an expression that accounts for the **incidence** of use of a substance in a **defined period of time**. There are essentially two time periods of interest: ***use for the first time (first use) in the past month and use for the first time in the past year***. In a cross-sectional study such as the one discussed in this Manual, the question in the questionnaire for estimating incidence during either of those two time periods is as stated earlier:-

When was the **first time** you used NAME OF SUBSTANCE?

	1. Never used
	2. During the past month
	3. More than 1 month ago but less than 1 year ago
	4. More than 1 year ago

Let N be the population represented (i.e., the sum of the expansion factors that is the denominator of expression (11) above), and let A, B, C, and D represent the total number of cases weighted for responses 1, 2, 3, and 4 respectively, as shown in the table below:

Table A2.15: Outline for the percentage distribution of the question: When was the first time you used/tried NAME OF THE SUBSTANCE?

First use of NAME OF SUBSTANCE	Weighted sample/Population represented
1. Never used	A
2. During the past 30 days	B
3. More than 1 month ago but less than 1 year ago	C
4. More than 1 year ago	D
Total	N

Thus, the **past year incidence**, and the **past month incidence** are provided below in the following expressions:

$$Past\ year\ incidence = \frac{B + C}{N - D} \tag{12}$$

and

$$Past\ month\ incidence = \frac{B}{N - C - D} \tag{13}$$

When using computer programs such as those mentioned earlier, the recommendation is to generate missing values for those cases (i.e. students) “not exposed,” and then compute

a proportion in each case, determining the corresponding Confidence Intervals with appropriate use of the study's sample design.

For example, where past year incidence is concerned, a new variable is developed, and missing values are defined in all cases in which the response was 4, i.e., in all cases that reported having used alcohol for the first time "more than 1 year ago." In this case, the number of weighted cases is reduced for this new variable; this number is designated N_1 (note that N_1=N-D in the previous table). Cases with answers 2 and 3 are grouped together, i.e., all cases that responded affirmatively to the use of the substance for the first time in the past year. The frequency table for the new variable is shown in the following table:

Table A2.16: Outline for estimating
past year incidence

NAME OF SUBSTANCE used for the first time in ***the past year***	Weighted sample/Population represented
1. Yes	B+C
2. No	A
Total	N_1=B+C+A

The calculation of incidence in this manner is reduced simply to a calculation of a proportion, and again, by using a statistical program, we can determine the corresponding Confidence Interval in accordance with the sample design.

This calculation is similar to the case of past month incidence. If the answers were 3 or 4, a new variable with missing values will be defined, and there will be a new number of exposed cases (that is, those who did not use the substance prior to the month when the study was conducted. That number is termed N_2, where N_2=N-C-D=A+B. The new variable and the corresponding result are shown in the following table:

Table A2.17: Outline for estimating
past month incidence

NAME OF THE SUBSTANCE first used in the ***past month***	Weighted sample/population represented
1. Yes	B
2. No	A
Total	N_2=B+A

As with past year incidence, we have a proportion. What was said above also applies in this case. To illustrate this, let us use an earlier example. Using SPSS we obtain the following results:-

- **Original table**:

Table A2.18: Distribution (frequency and percentage) for the question: When was the first time you tried an alcoholic beverage?

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	Never	91761	39.4	39.4	39.4
	During the past 30 days	41391	17.8	17.8	57.2
	More than 1 month ago but less than 1 year ago	34667	14.9	14.9	72.1
	More than 1 year ago	65016	27.9	27.9	100.0
	Total	232836	100.0	100.0	

- **Calculation of past year incidence (ibeb_ano)**:

We must declare as missing values those students who first used alcohol "more than 1 year ago," from which we obtain the following table A2.19:

Table A2.19: Distribution (frequency and percentage) to estimate *past year* incidence of the use of an alcoholic beverage

ibeb_ano		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	No	91761	39.4	54.7	54.7
	Yes	76059	32.7	45.3	100.0
	Total	167820	72.1	100.0	
Missing values		65016	27.9		
Total		232836	100.0		

The column "Valid percentage" shows incidence in the past year, 45.3%.

- **Calculation of incidence in the past month (ibeb_month)**:

In this case, we say that those students who used alcohol for the first time "more than 1 year ago" or "more than 1 month ago but less than 1 year ago" are missing values. This gives us the following:

Table A2.20: Distribution (frequency and percentage) to estimate *past month incidence* of use of an alcoholic beverage

ibeb_month		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	No	91761	39.4	68.9	68.9
	Yes	41391	17.8	31.1	100.0
	Total	133152	57.2	100.0	

Missing values	99683	42.8		
Total	232836	100.0		

Again, in the column "Valid percentage" we obtain the past month incidence, 31.1%.

Using this strategy, that is, by constructing two new variables for past year and past month incidence, it should be relatively easy to construct the respective confidence intervals using appropriate statistical software, a subject to which we shall return in the following section.

We need to underscore the definition of prevalence but in more general terms. Let us assume that we want to estimate the prevalence of lifetime use of a particular substance, **P**. If x represents the variable "substance use," then:-

$$\begin{cases} x_{ijh} = 1 \; if \; the \; person \; (ijh) \; used \; the \; substance \; during \; the \; period \; under \; study \\ x_{ijh} = 0 \; if \; the \; person \; (ijh) \; did \; not \; use \end{cases}$$

The following formula (11) therefore provides the estimator of the prevalence of substance use during the period as:

$$p = \frac{\sum_{i=1}^{n_j} \sum_{j=1}^{J_h} \sum_{h=1}^{H} \boldsymbol{x_{ijh}} * \boldsymbol{f_{ijh}}}{\sum_{i=1}^{n_j} \sum_{j=1}^{J_h} \sum_{h=1}^{H} \boldsymbol{f_{ijh}}} \qquad (14)$$

where the sub-indices i,j,h identify student i, in class j of school h; the denominator is the sum of the expansion factors, that is, the population represented in the study.

Age of onset

To conclude this section on the description of variables, let us examine those variables that relate to the **age of onset** or **age of first use** of a substance. This forms part of one of the specific objectives of epidemiological studies on drug use. It is clearly a very sensitive and frequently used indicator that can cause a fair amount of confusion and controversy. For example, let us assume that the average age of onset of a particular substance (licit or illicit) is 13. Generalizations are often made that **"students begin to use drugs at the age of 13."** The indicator "average age of onset" must not be analyzed individually; rather, we recommend that it be used in combination with prevalence indicators.

In other words, if the average age of first use of a substance in a country is 13, and lifetime prevalence is 40%, this is probably conceptually different and poses different challenges from another country where the average age of first use is the same but the lifetime prevalence is 5%. In other words, if we only examine the age of onset, one might conclude that the situation is similar in both countries (or in the same countries at different points in

time), but if that indicator is examined with the indicator on prevalence, it is likely that the conclusion would be somewhat different.

The statistical "average" may be influenced by extreme values that might confuse the result. We therefore recommend using the ***"median" as the indicator of the variables associated with the age of first use, which, with prevalence indicators*** will produce a more balanced picture of "early onset" of the use of substances. It is also recommended that we include the percentage of 8th grade students who say they have used the substance being analyzed.

In summary, to have a more complete analysis of the onset of substance use, three elements should be considered: 1) the **median** age of onset, 2) the **prevalence** of use with the frequency (i.e., how many students have used substances) and 3) of note, the indicators of **prevalence** and **incidence** in **8th grade** students.

3.2 Confidence Intervals

Confidence intervals have generally been used for two reasons:- (1) to supplement the description of an indicator, and (2) as a criterion for comparing two or more indicators.

Regarding the first, *under SIDUC, it is highly recommended that confidence intervals be used when estimating the parameters of the principle variables for analysis* relative to the study's objectives. For example, the mandatory specific objectives of a study of this nature include ***estimating the prevalence of "lifetime use", "past year use", and "past month use", in addition to the incidence of use "in the past year" and "in the past month"*** *(of the following substances, and others: alcohol, tobacco, tranquilizers, stimulants without a medical prescription, inhalants, marijuana, cocaine base paste, cocaine, and ecstasy.)*

Earlier, we introduced the expressions for these indicators, namely, prevalence and incidence. So why use confidence intervals? Let us assume that we want to estimate the prevalence of past month use of alcohol, and that the value obtained from the sample by means of expression (14) (i.e., with expansion factors) is 10%. If we also know that the error of the estimate (what is known as "precision") for a 95% Confidence level is 1%, then the confidence interval is between 9% and 11%. Similarly, if the error of the estimate was 5%, the confidence interval is between 5% to 15%.

Therefore, if an indicator as important as prevalence of past month use of alcohol, or any other indicator is described only by means of a precise estimate (in statistical terms), we will not know whether that estimate has an error of 1% or an error of 5%, in other words, we will not know how precise the estimate is. An estimate of 10% ± 1% does not provide the same information as an estimate of 10% ± 5%. There is no doubt that an interval

between 9% and 11% is much more precise and informative, than an interval between 5% and 15%.

Estimates for specific sub-groups within the sample are made very frequently, and it is here that we may find confidence intervals that are *very wide, i.e., with very little precision, or with a very high relative error*. **This type of estimate should be avoided, but if it is necessary to be included in reports, it should be accompanied by a note highlighting the low level of precision of that particular estimate**. We must remember that the size of the sample was determined under certain conditions; therefore, when disaggregating the sample into sub-groups, the sample size of those sub-groups will not provide estimates with the same level of precision used at the time the size of the sample was first determined .

In Section 6 above, we discussed what a confidence interval is, and expressions (1) and (2) were given to define it. The central point in that definition lies in the calculation of the **standard error** which requires a statistical package. It is also important to be clear about the sampling design knowing that this study is complex sampling.

Some examples using SPSS software will be provided below. We will again use as an example the sub-sample of 6,088 students. The sample is taken from 12 geographical areas (known as strata) from which are chosen 83 public and private schools. It was determined that in each school, Grades 8, 10, and 12 would form part of the study. If there is more than one class in a specific grade, only the first class would be considered. For example, if there are two classes in the 8th grade, 8A and 8B, the first one (8A) will be considered. In this particular example, there is no selection of classes within a selected school, rather, a prior decision made that a particular class would form part of the sample.

The geographical areas are therefore considered as strata, and the selected schools as clusters. The indicators of interest will be the past month prevalence of alcohol, past year prevalence of marijuana, and past year incidence of alcohol. Analysis will be done on a general level, and also by sex, grade, and type of sch]ool.

The variables in the database for the purpose of this example are, as follows:

- ✓ School: number of schools (80 schools)
- ✓ Area: geographical area (12 geographical areas)
- ✓ Sex: 1=male, 2=female
- ✓ Grade: grades 8, 10, 12
- ✓ Age in years
- ✓ Type_of school: 1=public, 2=private
- ✓ Expansion: expansion factor
- ✓ pbeb_month: use of alcohol in the past month, 1=Yes, 0=No

- ✓ ibeb_year: first use of alcohol in the past year, 1=Yes, 0=No
- ✓ pmar_year: past year marijuana use, 1=Yes, 0=No
- ✓ Ac_alc: ease of buying alcohol: 1=Easy, 2= Difficult, 4=Don't know
- ✓ R_beb: perception of great risk/harm of getting drunk: 1=High risk, 0=Not high risk.

A description of the sample is shown in the following table:-

Table A2.21: Description of the sample by sex, grade and type of school, and percentages of the population represented

Variables	Sample size	Population Represented	%
Sex			
Male	2,826	109,578	47.06
Female	3,262	123,258	52.94
Grade			
8th	2,220	109,694	47.11
10th	2,081	88,062	37.82
12th	1,787	35,080	15.07
Type of school			
Public	4,545	164,257	70.55
Private	1,543	68,579	29.45
TOTAL	6,088	232,836	100

As shown in the table, in the expanded sample, approximately 53% are females, 47% of the students are in 8th grade, and roughly 70% come from public schools. In terms of the estimators of the parameters of interest, we show the statements below in SPSS:-

- ✓ The sample design must be determined first; basically in this specific example, we define the strata (geographical areas) and the clusters (schools), as well as the variable having to do with the expansion factor:-

Once the file is open in SPSS, under "***Analyze",*** select "**Complex Samples"** and then "***Prepare for Analysis".*** Then select "***Create a plan file"*** and in "***Browse"*** choose a name to generate and save the definition of the sampling model. "***Save"*** and then select "***Continue"***. In this new window, select the variables corresponding to the "***Strata"***, the variable that identifies the "***Clusters"***, and lastly, the variable corresponding to ***Sample Weight*** (expansion factors). Then click "***Done"*** to save the sampling plan that can be used each time a new analysis is required.

It is very important to highlight that this format is based on the design of this particular study. Therefore, ***given that each study may have a specific sampling strategy, the instructions above should accurately reflect the sample design***.

- ✓ The statements are then used to obtain the desired estimates.

Again in SPSS, under "***Analyze",*** select "***Complex Samples"***, where you will find different options, including:

→ Frequencies…

→ Descriptives…

→ Logistic regression….

In order to estimate the prevalence of past month use of alcohol, choose "**Frequencies"** and select the analysis plan described above, the corresponding database, and the variable associated with the indicator for which the respective estimate is required, and also include the confidence interval option. The results are as follows:-

Table A2.22: Estimate of prevalence of past month alcohol use

pbeb_mes		Estimate	Standard error	95% Confidence Interval	
				Lower	Upper
% of total	No	64.1%	1.2%	61.6%	66.4%
	Yes	35.94%	1.20%	33.58%	38.37%
	Total	100.0%	0.0%	100.0%	100.0%

The prevalence of the use of alcohol in the past month is 35.94%. The error of the estimate is 1.2%, which gives a 95% Confidence Interval with values of between 33.58% and 38.37%. The error of the estimate and the Confidence Interval are correct and are properly computed on the basis of the sample design used.

The design effect (DEFF) can also be determined selecting this option from the SPSS menu.

Table A2.23: Design effect on indicator of prevalence of past month use of alcohol

pbeb_mes		Standard error	Estimate	Design effect
% of total	No	64.1%	1.2%	3.910
	Yes	35.94%	1.20%	3.910
	Total	100.0%	0.0%	.

As shown above, the DEFF=3.91, i.e., the variance of the model used is less than 4 times the variance obtained if a simple random sampling had been used. ***This is important to bear in mind when determining the sample size for future similar studies***.

Using expression (10) in Annex 1, the absolute and relative errors can be estimated by calculating the standard error. However, since different statistical programs use different probability distributions to define the 97.5% percentile of that distribution, rather than using the formula (10), it is more effective to use the confidence limits of the interval obtained.

If Ls represents the upper limit of the confidence interval, and Li represents the lower limit, then the absolute error and the relative error can be obtained as follows:-

$$EA = \frac{Ls - Li}{2} \qquad \text{and} \qquad ER = \frac{EA}{p} * 100 \tag{15}$$

In the example:

$$\boldsymbol{EA = \frac{38{,}37 - 33{,}58}{2} = 2.39\%} \text{ and } \boldsymbol{ER = \frac{2{,}39}{35{,}94} * 100 = 6.7\%}$$

Both values are fairly low, which means that the estimate is fairly precise.

Using the same criteria and the same software, estimates can be obtained for past month use of alcohol and 95% confidence intervals for the categories sex, grade, and type of school. The results are given in the following table:

Table A2.24: prevalence of past month alcohol use and 95% Confidence Intervals (CI) by sex, grade, and type of school

Variables	Prevalence (%)	Standard error (%)	CI (%)
Sex			
Male	34.28	2.00	30.41 – 38.37
Female	37.41	1.52	34.44 – 40.48
Grade			
8th	26.21	1.68	23.00 – 29.92
10th	42.99	1.80	39.45 – 46.60
12th	48.64	2.38	43.91 – 53.40
Type of school			
Public	36.63	1.46	33.78 – 39.59
Private	34.26	1.84	30.69 – 38.03
TOTAL	35.93	1.20	33.58 – 38.37

The first point to notice in the above table is that the standard error for the total sample is the lowest or the smallest of all, and the reason is, the estimate considers the total sample: the larger the size of the sample, the smaller the error of the estimate (for similar prevalence.) When estimates are made at the level of sections of the sample, the precision of those estimates will be less; for example, among 12th grade students, the width of the confidence interval (upper limit minus lower limit) of almost 10 percentage points will double the result at the national level

Looking at the precise estimates of prevalence, we see that the prevalence of past month alcohol use is higher among females than among males, 37.4% and 34.3% respectively. We also see that the prevalence increases as the students move into higher grades (which is associated with age), from 26.2% in the 8th grade to almost 49% in the 12th grade. Lastly, prevalence is higher among students in public schools when compared to private schools, 36.6% and 34.3% respectively.

Given that these are estimates based on random samples of the population, the question we need to ask is whether there are differences between past month prevalence of alcohol use at the level of the population on the basis of the results obtained from the samples. For example, are there differences between males and females? Between grades? Between public and private schools?

In other words, ***are the differences observed attributed to the sampling process or do they truly reflect what is happening at the population level?***

The correct way to answer these questions is by performing ***statistical tests to test the hypotheses***; for example, a possible hypothesis of interest is that the prevalences of past month alcohol use at the level of the population are the same for males and for females.

However, this example and similar hypotheses are often tested by comparing the respective confidence intervals and examining whether the intervals are overlapping. Using this criterion, if, for example, the CI for a particular sub-group is between 10% to 15%, and for the other sub-group, it is between 18% to 24%, then, given that there is no overlap between the confidence intervals, the conclusion would be that there is a statistically significant difference between the prevalences in the two sub-groups. If the CI for the second sub-group was between 13% to 18%, then there is an overlap of the two intervals, and the conclusion would therefore be that there are no statistically significant differences between the two sub-groups. A third situation might occur if there is a slight or very small overlap of the intervals, for example, 10% to 15% and 14.9% to 16%, or 10% to 15% and 15.1% to 16%. In the first case, there is a slight overlapping, and in the second, little or no overlap.

Although this criterion is used frequently, it is always recommended to ***use statistical procedures to provide the best possible response to the hypothesis proposed*** about the comparison of indicators between two or more sub-groups. Returning to the previous examples, in those situations in which there is clear overlapping of the intervals, or a clear distance between them, the qualitative conclusion is probably the same as that obtained by methods that test the particular hypothesis. However, in the latter situation, the conclusions may be different. In any event, when Confidence Intervals are used as a criterion, it will only be possible to detect differences, but the size of that difference cannot be assessed in probability terms.

The section below discusses methods of testing hypotheses that compare indicators for two or more sub-groups, which is ***the procedure that is strongly recommended***.

3.3 Comparison between sub-groups

As with the previous analyses, to compare indicators of interest among population sub-groups, statistical software is needed as well as the skillset to use and interpret it correctly.

As an example, let us assume that:-

P_1*=P (past month alcohol use/male)=Prevalence in the population of past month alcohol use among male students,* and

P_2=P (past month alcohol use/female)=Prevalence in the population of past month alcohol use among female students.

Our aim is to examine the hypothesis that both prevalences are the same, i.e.:-

$H_0 : P_1=P_2$ called the null hypothesis,

which is contrasted with another hypothesis that suggests that the prevalences in the population among males and females are not equal, i.e.,

$H_1 : P_1 \neq P_2$ which is called the alternative hypothesis.

Other alternative hypotheses could also be proposed when there is evidence of that, for example,

$H_1 : P_1 > P_2$ higher prevalence among males than among females, or

$H_1 : P_1 < P_2$ lower prevalence among males than among females.

Based on the random sample of students and the classification of cases of male students and female students, it must be determined whether the null hypothesis is to be rejected or not. (This would mean that there is insufficient evidence to reject it). The former means that the test for such a determination assumes that the null hypothesis is true. This is what causes the large difference in the use of confidence Intervals discussed earlier.

Returning to the previous example, and once again using SPSS, the test of the hypothesis can be resolved in *two ways*:-

a) The **first** solution looks only at the estimated prevalences (based on the sample) for each sub-group, i.e., males and females:

Let p_1=prevalence of past month alcohol use in the sub-sample of males, and

Let p_2=prevalence of past month alcohol use in the sub-sample of females.

The previous results show that: p_1=34.28% and p_2=37.41%.

The hypothesis can be solved by means of Pearson's chi-squared test (test of independence), the result of which in SPSS as shown in Table A2.25:-

Table A2.25: Test of independence for comparison of past month alcohol use by sex

pbeb_month * sex	Chi-squared	F corrected	gl1	gl2	Sig.
Pearson	6,466	1,469	1	66	,230

Odd ratio	6,473	1,470	1	66	,230

The table above gives two results for Pearson's chi-squared statistic: the first column of Table A2.25 (chi-squared=6,466) does not consider the sample design, whereas the second column (***F corrected=1.469***) is accurate since it considers the sample design for the study. In this case, the value of chi-squared (corrected), which we shall call Q, is equal to 1.469 (Q=1.469) with an associated probability (p-value) of 0.23. If the significance level is defined as 5%, that is, α=0.05, given that the p-value>0.05, the null hypothesis is accepted since there is no evidence for stating that at the level of the population, there are differences between males and females in the prevalence of past month use of alcohol.

In other words, even though there is a higher prevalence among women at the level of the sample, there is no evidence for stating that the difference observed in the samples is also found at the level of the population, and it is therefore assumed to be "by chance" and not explained by sex. It could also be said that the difference is "within the margin of sampling error."

Another example is to compare prevalences by grade where we have three groups rather than two. The hypothesis in this case is:

H_0: $P_8=P_{10}=P_{12}$, i.e., the prevalence of past month alcohol use is the same in the three grades, and

H_1: at least two indicators are different.

Again using SPSS, we have the following results:-

Table A2.26: Test of independence for comparison of past month alcohol use by grade

pbeb_month *grade	Chi-square	F corrected	gl1	gl2	Sig.
Pearson	231,949	42,708	1,783	117,691	,000
Odds ratio	234,217	43,126	1,783	117,691	,000

The results in Table A2.24 show that the sample prevalences of past month alcohol use by grade are (in %): p_8=26.21%, p_{10} =42.99% and p_{12}=48.64%.

In Table A2.26, we see that Pearson's chi-square using the conditions of the sample design is **Q=42.71** with p<0.0001. This means rejecting the null hypothesis and concluding that there are great differences (statistically significant) in the prevalences of past month alcohol use among the three grades. We can deduce from this that there are differences, but we cannot identify where those differences are between pairs of grades. The values suggest that there would be statistically significant differences between the 8th grade, the

10th, and 12th grades, but it is not possible to make a definitive conclusion, nor determine a difference between the last two grades. A more general option including a solution to the previous situation will be presented next.

b) The **second** solution involves a **logistic regression model** which provides an analysis of the previous hypothesis and also helps us to construct explanatory models about a particular variable of interest, such as the use of an illicit drug in the past year. We will examine below different options associated with this type of model.

A regression model is an equation that *relates a response* or dependent variable *with a set of factors* or independent variables. Depending on the type of response (i.e., of the dependent variable), different regression models are defined, some of which are described below:-

1. If the response is a **continuous variable**, then the model is called a **linear regression *model.***

   Variable: *age of first use of any substance.*
   Possible responses: 12, 13, 14, etc.

2. If the response is **binary (two possible alternatives),** the model is called a **binary logistic regression model**.

   Variable: *past month alcohol use.*
   Possible responses: Yes or No.

3. If the response is a **qualitative ordinal (more than two possible alternatives that do not have a natural order),** then the model is called an **ordinal logistic regression model**.

   Variable: *frequency of use of illicit drugs in the past year*.
   Possible responses: Never, only a few times in the past year, several times in the past year, often in the past year.

4. If the response is a **qualitative nominal (more than two possible alternatives that do not have a natural order)**, then the model is called a **nominal logistic regression model**.

   Variable: *type of drugs used in the past year.*
   Possible responses: None, only alcohol, only marijuana, only tobacco, only alcohol and tobacco, etc.

In all previous cases, when there is only one independent variable, we speak of **a simple or univariate regression model**. If there are two or more independent variables, we then have a **multiple or multivariate regression model.**

The dependent variable is **Y**, and the independent variables are $\mathbf{X_1, X_2, X_3}$, etc.

In the field of drug use, and depending on the objectives of the study, the response of interest **Y** generally represents the use of a substance, such as the *use of any illicit*

substance in the past year. The variable Y will have the value 1 if the person has used, and 0 if he/she has not used. **Y is therefore a binary variable**, and the most appropriate regression model in these cases is the **binary logistic regression model** (from here on, we shall speak only of logistic regression). The idea is to model response Y: some people use and others do not, and the pertinent question is precisely to identify those characteristics that make for Y=1, that is, to find factors where the probability that Y=1 is greater than for other factors. In other words, to find those factors associated with the use of an illicit substance.

Let us first consider a single factor or dependent variable, which we shall call X, where if X=1 that factor is assumed to be present in the individuals, and if X=0, that factor is absent.

If we are interested in analyzing the use of any illicit substance in the past year among the country's secondary school population, we have:-

- ✓ **P_1=P(Y=1/X=1)** represents the probability (likelihood) that a student *had used an illicit substance in the past year* given that the factor *is present*, and
- ✓ **Q_1=P(Y=0/X=1)=1-P_1** is the likelihood that the student *did not use an illicit substance in the past year*, if the factor is *present*.

Similarly:

- ✓ **P_0=(Y=1/X=0)** represents the probability that a student *had used an illicit substance in the past year*, given that the factor is *absent*,
- ✓ **Q_0=P(Y=0/X=0)=1-P0** is the probability that the student *did not use an illicit substance in the past year*, if the factor is *absent*.

The hypothesis of interest concerning the association between the factor X and the variable of interest (response of interest) Y may be expressed as:-

$$H_0: P_1=P_0$$

Clearly, if the factor X **is not associated** with response Y, the probability that a student used an illicit substance **does not depend on** (is independent) whether the factor is present (X=1) or absent (X=0), and then $P_1=P_0$.

Based on the above, a measure of association[19] can be defined between a factor (what in epidemiology is called "exposure") and a response, that is, use or non-use of an illicit substance (or any other response of interest given the objectives):-

Disparity Ratio (Odds Ratio, OR): epidemiological studies on prevalence of drug use—a category to which a school population study belongs—are cross-sectional studies

[19] There are other measures of association that depend on the study's design, particularly longitudinal studies.

conducted at a specific moment in time, in which past and present events are investigated. Let us assume that X represents the variable "drug use by siblings or people who live in the same house," where X has the value 1 for those students who respond "Yes", and has the value 0 when no one in the home uses any drug.

The hypothesis behind a question of this nature is that for those students who state that there are people at home who use drugs, i.e., X=1, the probability (likelihood) that the student will also use a substance is higher than in those cases in which no individuals at home use drugs, i.e., when X=0.

In a cross-sectional study such as this, the best *measure of association* for analyzing this hypothesis is the ***odds ratio***, which is defined as:-

$$OR = \frac{P_1/Q_1}{P_0/Q_0} = \frac{P_1 * Q_0}{P_0 * Q_1} \quad (15)$$

If the hypothesis is true, then $P_1>P_0$ and therefore $Q_0>Q_1$ and the numerator in formula (15) will be greater than the denominator. In the event $P_1=P_0$ then OR=1, called **null value**. Given that all the expressions that make up the Odds Ratio are probabilities, the value of the ***OR*** will always ***be greater than or equal to zero***. In summary:-

- ✓ If OR=1 it means that there is no exposure effect on the response of interest.
- ✓ If OR>1 it means that there is a positive exposure effect on the response of interest, i.e., if there is exposure, then the probability of an affirmative response increases. In these cases, we say that X is a *risk factor*,
- ✓ If OR<1 it means that there is a negative effect, i.e., if there is exposure, then the probability of a positive response is reduced, and we say that X is a *protective factor*.

Having decided on an appropriate measure of association, the next step is to determine the logistic regression model and then examine how that measure relates to the model. The expression below is one of the possible ways of representing the logistic regression model, assuming that there is only one independent or exposure variable (X), where P is the probability that the event of interest will occur for a particular value of X:

$$P = P(Y = 1/X = x) = \frac{e^{\beta_0+\beta_1*X}}{1 + e^{\beta_0+\beta_1*X}} \quad (16)$$

Following the previous example where X has two values, 1 if the factor is present, 0 if there is none, then we will determine the relevant probabilities in accordance with expression (16).

If X=1 then

$$P_1 = P(Y = 1 \,/\, X = 1) = \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}} \tag{17}$$

and

$$Q_1 = 1 - P_1 = P(Y = 0 \,/\, X = 1) = 1 - \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}} = \frac{1}{1 + e^{\beta_0+\beta_1}} \tag{18}$$

From (17) and (18) we have:

$$P_1/Q_1 = e^{\beta_0+\beta_1} \tag{19}$$

However, if X=0 then:

$$P_0 = P(Y = 1 \,/\, X = 0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \tag{20}$$

and

$$Q_0 = 1 - P_0 = P(Y = 0 \,/\, X = 0) = 1 - \frac{e^{\beta_0}}{1 + e^{\beta_0}} = \frac{1}{1 + e^{\beta_0}} \tag{21}$$

From (20) and (21) we have:

$$P_0/Q_0 = e^{\beta_0} \tag{22}$$

Lastly, going back to expression (15) for the odds ratio (OR), we have:

$$OR = \frac{P_1/Q_1}{P_0/Q_0} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1} \tag{23}$$

This expression therefore defines the link between the logistic regression model and the odds ratio.

Based on the information collected from the students in the sample, and in particular the information on variables X and Y, a logic regression model can be used (with the appropriate software) and on that basis, estimate the β_1 coefficient of the model, and thus derive the estimate of the OR.

We return once again to the previous example where Y=past month alcohol use, and X=sex. The hypothesis is that prevalence among males (P1) is equal to the prevalence among females (P_2),[20] that is, **H_0: $P_1=P_2$.** Using SPSS, select from the menu *"Analyze"* and then *"Complex Samples"*. There is the ***"Logistic Regression"*** *model*, and this is where the variables of the model should be defined; in this case, the dependent variable is past month use of alcohol, and the factor of analysis is the variable "sex". Below we provide the result of the estimate of the coefficient of the model, or $\hat{\beta}_1$:

Table A2.27: Estimate of parameters of logistic regression model for past month alcohol use and by sex

Parameter	B	95% Confidence Interval		Contrast hypothesis		
		Lower	Upper	t	gl	Sig.
(Intersection)	,515	.385	.644	7,951	66,000	,000
[sex1=,00]	,136	-.088	.361	1,212	66,000	,230
[sex1=1,00]	,000[a]	.	.	.	.	.

Table A2.28 shows the estimated odds ratio and the corresponding Confidence Interval:

Table A2.28: Estimate of odds ratio for past month alcohol use and by sex based on logistic regression model

Odds ratio	95% Confidence Interval	
	Lower	Upper
1.146	.915	1.435

The most important result in Table A2.27 is the coefficient associated with the variable sex, which we will call $\hat{\beta}_1$ (the symbol "^" is used to identify the estimator of a parameter).

That is, $\hat{\beta}_1$=0.136 with p=0.23 and a 95% confidence interval of between -0.088 and 0.361. Given that the p-value is greater than 0.05, the conclusion is that there are no statistically significant differences between the past month prevalence of alcohol use between males and females.

The estimator of the odds ratio is:

[20] It does not matter which sub-index is used. Usually, the sub-index 0 is used to indicate the absence of a factor. In this case, the variable sex does not represent a risk or protective factor. Therefore, we will use sub-indices 1 and 2 rather than 1 and 0.

$$\widehat{RD} = e^{\hat{\beta}1} = \exp(\hat{\beta}_1) = \exp(0{,}136) = 1{,}146$$

With a 95% confidence interval of 0.915 to 1.435, and given that the null value 1 is within the limits of the CI, we arrive at the same conclusion as above. The same results in SPSS are given in Table A2.28.

The values for the variable sex are 1 for males and 2 for females, and the comparison is therefore made between the higher value of X and the lower value of X, i.e., the prevalence among females is compared to the prevalence among males. Given that the prevalence among females is higher than among males, it is therefore expected that the estimated value of the odds ratio will be greater than 1.

As stated earlier, the value of the odds ratio may be less than, greater than or equal to 1. With this in mind, and assuming that the independent variable represents a factor that may intervene in the response, then if the confidence interval for the odds ratio:

- ✓ Has both limits **greater than 1**, and therefore the CI does not include the null value of 1, then the presence of the factor (X=1) significantly **increases** the probability that the event of interest will occur.
- ✓ Has both limits **lower than 1**, and therefore the CI does not include the null value of 1, then the presence of the factor (X=1) significantly **decreases** the probability that the event of interest will occur.
- ✓ The lower limit is less than 1 and the upper limit is greater than 1, and therefore the CI includes the null value of 1, then the presence of the factor **does not affect** the probability that the event of interest will occur.

It is important to differentiate between the characteristic associated with the independent variable or Factor x. In some situations, the variable cannot be modified, but in others, it is feasible to do so. For example, when substance use is analyzed by sex, by grade, or by type of school (public or private), these are conditions that cannot be changed, and therefore, if statistically significant differences are found between males and females, or between grades or types of school, this information will be useful in targeting the interventions, but not to modify the variable as such.

If X represents the *perception of risk/harm of substance use*, on the understanding that the greater the perception of risk, the less the probability of substance use, policies can be geared to increasing the number of students who perceive that substance use is highly risky, and thus reduce the probability of use. This situation could in fact be modified; in such cases we speak of risk and protective factors.

Let us now assume that the response of interest is *past year use of marijuana*. Shown below are the prevalences by sex, grade, and type of school:

Table A2.29: prevalence of past year marijuana use and 95% confidence intervals (CI) by sex, grade and type of school

Variables	Prevalence (%)	Standard error (%)	CI (%)
Sex			
Male	7.36	0.76	5.98 – 9.03
Female	4.91	0.53	3.95 – 6.09
Grade			
8th	4.91	0.63	3.79 – 6.34
10th	6.92	0.71	5.63 – 8.48
12th	7.51	0.99	5.76 – 9.74
Type of school			
Public	6.14	0.54	5.15 – 7.30
Private	5.88	073	4.58 – 7.51
TOTAL	6.06	0.43	5.25 – 6.99

Analyzed by sex, i.e., comparing the prevalence among males with the prevalence among females, we see that there is a slight overlapping of the confidence intervals, which would lead us to think that there is no difference between them.

Let us now examine the hypothesis that both prevalences are equal when we adjust a logistic regression model. Again using SPSS, we obtain the following results (Tables A2.30 and A2.31):

Table A2.30: Estimate of parameters in logistic regression model for past year use of marijuana and by sex

		95% Confidence Interval		Contrast hypothesis		
Parameter	B	Lower	Upper	T	gl	Sig.
(Intersection)	-2,964	-3.192	-2.736	-25,952	66,000	,000
[sex1=,00]	,431	.100	.762	2,603	66,000	,011
[sex1=1,00]	,000[a]	.	.	.	.	.

Table A2.31: Estimated odds ratio for past month alcohol use and by sex based on a logistic regression model

		95% Confidence Interval	
pmar_year	Odds ratio	Lower	Upper
Yes	1.539	1.106	2.142

We see from this table that the OR=1.539 with p=0.011, less than α=0.05 and with a confidence interval of between 1.11 and 2.14 (not including the null value 1), which leads to the conclusion that there is a statistically significant difference between the prevalences of past year marijuana use among males and females. Note that this conclusion is different from that obtained via the qualitative comparison between the respective Confidence Intervals. ***This is the reason for making comparisons using statistical procedures that test the corresponding hypothesis***.

Comparing by grade, we obtain the following results using the 8th grade as the reference group; therefore the odds ratios give the comparisons between the 10th and the 8th grade, and between the 12th and the 8th grade. First of all, it is assumed that the prevalences in the three grades are the same, and then the 10th and 12th grades are compared with the 8th grade by estimating the coefficients and the odds ratios.

Table A2.32: Test of effects of the logistic regression model for past year marijuana use by grade

Origin	gl1	Wald F	Sig.
(Corrected model)	2,000	6,409	,003
(Intersection)	1,000	1088,148	,000
grade	2,000	6,409	,003

Secondly, according to the results in Table A2.32, the hypothesis that the prevalences in the three grades are the same is rejected using the F (Wald) statistic=6.409 with p=0.003, which is less than the level of α=0.05.

We will now examine which grades show these differences, and to do so, we will adjust a logistic regression model where the 8th grade is the reference group.

Table A2.33: Estimate of parameters of the logistic regression model for past year marijuana use by grade. The reference group is the 8th grade

pmar_year			Contrast hypothesis		
	Parameter	B	t	gl	Sig.
	(Intersection)	2,964	21,856	66,000	,000
	[Grade=10th	-,364	-1,936	66,000	,057
	[Grade=12th	-,453	-3,439	66,000	,001
	[Grade= 8th	,000[a]	.	.	.

Table A2.34: Odds ratio of the logistic regression model for past year marijuana use by grade. Reference group is the 8th grade

pmar_year	Odds ratio	95% Confidence Interval	
		Lower	Upper
10 versus 8	1.439	.989	2.096
12 versus 8	1.573	1.209	2.046

We see in Tables A2.33 and A2.34 that there are no statistically significant differences between grades eight and ten (OR=1.439 when p=0.057 and a CI between 0.989 and 2.096), but there is a difference between the 8th and 12th grades (OR=1.573 when p<0.001 and a CI of between 1.209 and 2.046). The result of the comparison between grades 8 and 10 is important since the p-value and the confidence interval show that this is a borderline situation.

It is impossible to determine from the above analyses whether or not there are differences between grades 10 and 12; therefore it is necessary to execute the statement in SPSS but now with the 12th grade as the reference group, and comparing the prevalences for the 8th and 12th grades (although it was solved before) and the prevalences for the 10th and 12th grades. The results are given below:-

Table A2.35: Odds ratio for the logistic regression model for past year marijuana use by grade. Reference group is the 12th grade

	Odds ratios	95% Confidence Interval	
		Lower	Upper
8 versus 12	.636	.489	.827
10 versus 12	.915	.614	1.365

In Table A2.35, when we examine the line comparing prevalences in the *10th and 12th* grades (highlighted in yellow), we find that the RD̂=0.915 with a confidence interval that includes the null value, we can conclude that there is no statistically significant difference between the prevalence rates in those two grades. It should be noted that in the comparisons given here, the reference group was grade 12, and given that a higher prevalence rate is found in that grade, it is reasonable for the odds ratios to be less than the value of 1. By way of example, if we invert the value of the odds ratio for the comparison between the 8th and 12th grades, that is, 1/0.636, we obtain 1.57 which is the odds ratio when the 8th grade was the reference group, and the 12th grade was thus compared with that reference.

As mentioned earlier, the regression model (any of those defined) takes us a little further into the analysis. We can therefore conclude from the above analyses that:

→ the prevalence rate of past year use was significantly higher among males than among females.

→ In terms of grades, there was a statistically significant difference between grades 8 and 12, but not between those grades and the 10th .

It is reasonable to ask whether the distribution between males and females is homogeneous in the three grades to avoid a situation in which there is a variable (in this case, sex) that may be interacting in the relationship of the variables "*use of marijuana"* and "*grade"*. If that were to occur, from the point of view of the epidemiological analysis, a ***confounding variable*** (sex) would be present. Let us therefore examine the corresponding distributions which are shown in the table below:-

Table A2.36: Distribution of the sample expanded by sex, by grade

Grade	sex		Total
	Males	Females	
8th	49.30%	50.70%	100.0%
10th	45.49%	54.51%	100.0%
12th	44.01%	55.99%	100.0%
Total	47.06%	52.94%	100.0%

We know that there is less marijuana use among females. The table above shows that the 10th grade has almost 4% more females that the 8th grade, and this may be the reason why no differences were found between the two grades.

We can now analyze the differences by grade ***adjusting*** for the variable "sex", i.e., assuming that this variable is not interfering with the ratio under study. To do this, we need to modify expression (16) and incorporate the two independent variables into the model (grade and sex), assuming that Y=past year marijuana use. The variable grade should be represented by two variables to account for the contrasts being made given that the 8th grade is the

reference category, the variable *grade₁* assesses the difference between the 10th grade and the 8th, and the variable *grade₂* is the comparison of the 12th grade with the reference grade. The model is thus expressed by the following equation:

$$P = P(Y/\ Sex, grade) = \frac{e^{\beta_0+\beta_1*Grado_1+\beta_2*Grado_2+\beta_{3*}Sexo}}{1+e^{\beta_0+\beta_1*Grado_1+\beta_2*Grado_2+\beta_{3*}Sexo}} \quad (24)$$

In SPSS, the variable grade is expressed as a Factor, and Sex as a covariate. The results of the adjustment to this new model are as follows:-

Table A2.37: Estimate of parameters of the logistic regression model for past year use of marijuana by grade, adjusted for sex. Reference is the 8th grade

Parameter	B	95% Confidence Interval		Contrast hypothesis		
		Lower	Upper	t	gl	Sig.
[Grade=10th]	.382	.002	.763	2.005	66,000	.049
[Grade=12th]	.478	.218	.738	3.671	66,000	.000
[Grade=8th]	.000	.	.	.	.	.
sex	-.450	-.781	-.119	-2.714	66,000	.008

The first line of Table A2.37 shows the estimate of the coefficient of the comparison between the 10th and the 8th grades (the 8th grade is the reference grade) adjusted for sex. We see that $\hat{\beta}_1$=0.382 where p=0.049 which has a 5% significance level but is on the borderline. In the second line, we have the comparison between the 12th grade and the 8th grade adjusted for sex, with a $\hat{\beta}_2$=0.478 and p<0.001. The last line shows the coefficient for the covariable/covariate sex with $\hat{\beta}_3$=-0.45 and p=0.008.

We are interested in the comparison between grades 10 and 8, where, adjusted for sex, there is a slight difference in the corresponding prevalences. The earlier analysis (not adjusted for sex) for this comparison shows that the OR=1.439 with a confidence interval of between 0.989 and 2.098 (includes the null value). We shall call this the crude OR or unadjusted OR, OR_{crude}. According to the following table, adjusted for the sex variable, the $\hat{RD}$=1.466 with an interval that contains the value 1, which leads us to a decision different from the preceding one. This odds ratio is called an adjusted OR, $OR_{adjusted}$=1.466.

Table A2.38: Odds ratios of the logistic regression model for past year use of marijuana by grade, adjusted for sex. Reference group is the 8th grade

pmar_year	Odds ratio	95% Confidence Interval	
		Lower	Upper
10 versus 8	1.466	1.002	2.145
12 versus 8	1,613	1.244	2.092

Apart from the results, the idea here is to show the potential for using regression models for analysis, not only because they provide us with more precise decisions when comparing indicators on substance use (rather than using confidence intervals), but we can also make adjustments and control potential confounding variables.

In addition, we might think of an explanatory model of substance use by ***simultaneously*** including different levels of possible variables that affect the likelihood of substance use. Some variables relate to the family environment, the school setting, close friends, the perception of risk/harm of substance use, and so forth, may be utilized in an effort to find an overall explanation of substance use, and to estimate the individual contribution of those factors that were included in the model.

Special attention should be given to the *factors associated* with substance use as mentioned in the introduction of this Manual. This leads us directly to discuss some of the specific objectives described above: "Describe the type of family relationship and examine its association with the use of licit and illicit substances" and "Analyze the association between substance use and perception of risk." There are other objectives to be achieved from this association, for example, analyze the relationship between school performance and use of licit and illicit substances.

When we refer to *factors associated* with substance use, risk ***and protective factors are important considerations*** where the former relate to those factors that, if present, we assume a greater probability of substance use, and by contrast, the latter are factors that indicate no substance use , i.e., if they are present, the probability of substance use is lower.

If Y represents a variable that identifies substance use, and X is a factor associated with Y, we will say that X is a ***risk factor*** for Y if:

$$P(Y = 1 \,/\, X = 1) \; > P(Y = 1/X = 0) \tag{25}$$

In other words, if the factor is present (X=1), the likelihood of substance use (Y=1) is ***greater*** than when the factor is not present (X=0).

If Z is another factor associated with substance use, we will say that Z is a ***protective factor*** for Y if:

$$P(Y = 1 \,/\, Z = 1) \; < P(Y = 1/Z = 0) \tag{26}$$

In other words, if the factor is present (Z=1), the probability of substance use (Y=1) is ***less*** than when the factor is not present (Z=0).

It is important in epidemiological studies such as outlined in this Manual, to include in the study's objectives an analysis of risk and protective factors which are theoretically important, not only to study or confirm their association with substance use, but equally important, to examine the *trends in those factors*. When risk and protective factors are

confirmed by scientific research, then it is expected that interventions will focus on producing changes in those conditions as well as changes in substance use.

Let us examine questions related to parental involvement (module 5 of the questionnaire) and other questions concerning the perception of risk associated with substance use (module 4 of the questionnaire). We will examine four questions in particular:-

- How often does your mother, father or adult you live with, not know where you are after school or on weekends? Let's say for an hour or more." The answer "They always, or nearly always, know where I am" is the answer associated with the desired situation (less substance use.)
- In general, how well do you think your father, mother or adult you live with know your close friends? The answer "A fair amount" is the answer associated with the desired situation (less substance use.)
- Perception of harm of using marijuana sometimes. The answer "Very harmful" is the answer associated with the desired situation (less marijuana use).
- Perception of frequent use of marijuana. The answer "Very harmful" is the answer associated with the desired situation (less marijuana use).

For the four questions, a "Yes" is the answer that in theory leads to less substance use, while "No" means the opposite. The table below shows the distribution of these responses to the four questions above.

Table A2.39: Percentage distribution of questions

Question	Answer		
	Yes	No	Total
How often does your mother, father or adult you live with not know where you are after school hours or on weekends? Let's say for an hour or more.	74.0	26.0	100
In general, how well do you think your father, mother or adult you live with know your close friends?	40.9	59.1	100
Perception that using marijuana sometimes is very harmful	27.2	72.8	100
Perception that frequent use of marijuana is very harmful	58.1	41.9	100

We see from the table above that 74% of the students stated that their mother or father or adult they live with always or almost always know where they are; almost 41% stated that their father, mother or adult they live with know their close friends well, and the remaining 59%, slightly or not at all.

Further, 27.2% perceive that using marijuana sometimes is very harmful; this figure rises to 58.1% when the question refers to frequent use of the substance.

Let us look now at the prevalence of past year marijuana use for each possible answer to these four questions, recalling that the overall prevalence is 6.06% according to Table A2.29.

Table A2.40: Prevalence of past year marijuana use based on response alternatives for each question.

Question	Answer	
	Yes	No
How often does your mother, father or adult you live with not know where you are after school hours or on weekends? Let's say for an hour or more.	4.5	10.7
In general, how well do you think your father, mother or adult you live with know your close friends?	4.3	7.4
Perception of marijuana use sometimes	3.1	7.2
Perception of frequent use of marijuana	3.7	9.3

As shown in the table above, of those who stated that the responsible adult *always or almost always* knows where he/she is when they leave school, 4.5% report that they had used marijuana in the past year; this figure rises to 10.7% among those who say that the responsible adult *never, almost never, or only sometimes* knows where he/she is.

Something similar occurs regarding the responsible adult knowing the student's closest friends. Of the students who stated that parents know their friends *well*, 4.3% say they had used marijuana in the past year, a prevalence rate that rises to 7.4% when the responsible adult knows his/her close friends *more or less, slightly or not at all*.

Similarly for the other two questions: of those students who perceive *great harm* (high risk) from marijuana use (sometimes or frequently), the past year prevalence is lower than among those who *do not perceive great harm*. Considering the four variables analyzed, the differences in past year prevalence rates for marijuana are statistically significant which confirms the hypotheses associated with these variables.

One can conclude that if there are positive changes in the variables analyzed, in other words, if the distribution described in Table A2.39 improves, there ought to be an impact on the indicators of the prevalence of marijuana use in the past year (i.e. a reduction) if the other conditions that also influence those indicators remain constant. Hence the importance of including variables such as these in order to ***analyze trends*** beyond their association with substance use.

Considering all that has been discussed in this Annex, from a practical point of view, we suggest first describing the sample and performing individual analyses of the indicators

of substance use overall and disaggregated according to socio-demographic variables. Models for analysis should be adjusted for each of those indicators as dependent variables and each of the factors that are conceptually linked as independent variables.

From the findings found in the adjustments of these models, the next step would be to adjust a multivariate model with all relevant variables or predictors, either by the analyses themselves, or because the theory or literature demonstrates their importance.

The latter (modelling) is likely to go beyond what is required in a general report on the results of the study but should be carried out to achieve a better explanation of the phenomena associated with the use of the substance to better target prevention strategies in this specific population.

ANNEX 3: GUIDELINES AND INSTRUCTIONS

A requirement for success of any study among the secondary school population relates to the conduct of fieldwork, that is, *all activities needed to ensure completion of the questionnaire by students*. The fieldwork should be planned, organized and closely supervised by the national drug observatory (NDO).

The quality of the information obtained from the study will largely depend on how the fieldwork was conducted. It is therefore very important to avoid possible biases in the information obtained. Biases may be related to different situations during the process, such as:-

- Poor communication with the country's education authorities or with the schools selected for the sample; this may lead to the authorities refusing to have the study conducted in those schools.
- Poor communication by the facilitator (the person responsible for conducting the process in the classroom) with the students present, or poor answers given to questions posed by the students could have an impact on the responses to the questionnaire.
- Poor management in terms of cleaning answers on the questionnaire could also affect the final results of the study.

In addition to this, there must be full transparency throughout the process so that the results are in no way suspect. The data produced by a study of this type are quite sensitive, therefore there should not be a scintilla of doubt about their reliability or transparency.

These outcomes can be avoided if the fieldwork is performed by an ***institution*** *(university or company)* that is ***independent*** of the NDO and has experience with this type of work. The NDO is responsible for choosing or selecting the institution, and ensuring, for example, that facilitators do not have a connection to schools selected in the sample.

The fact is, given specific circumstances in some countries of the Americas, there will be special situations in which the NDO will have a more active role to play in the process, but the greatest possible transparency should be guaranteed. In situations such as this, we recommend close coordination with the OID/CICAD.

In order to plan and conduct fieldwork, the responsible institution will have the sample***, i.e., the list of schools and classes selected***, which cannot be modified. In some countries, the NDO will ask the same institution (university or company) to select the sample and to perform the fieldwork, but in other countries, these two activities may be carried out by

different institutions. Either option is acceptable provided they have sufficient knowledge and experience for the task.

The institution responsible for fieldwork ***will receive the questionnaire*** to be administered. It cannot be altered unless with the expressed authorization of the NDO.

As an example, Table A3.1 gives a hypothetical list of schools and classes in a sample. This list forms part of the material that the NDO will provide to the institution and will be the basis of the fieldwork.

Table A3.1: Hypothetical list of schools and classes in the sample

Information about the school					Classes (grades/sections) selected by school					
Code of the school	Name, address, phone, Name of contact	Department or region	Town	Type (public/private)	Grade/ section	Number of students	Grade/ section	Number of students	Grade/ section	Number of students
2	ABC				8A	35	12A	28	-	-
4	DEF				10B	29	-	-	-	-
7	GHY				8B	30	10B	29	12A	25
..	...				..	..	..	..	..	..
125	XYZ				12C	35	-	-	-	-

1. ORGANIZATION OF THE FIELD WORK

The institution in charge of the study will appoint a ***National Coordinator*** responsible for coordinating the study, including the fieldwork. He or she will remain in constant contact with the Director of the National Drug Commission; execution of the study is the responsibility of the team selected by the Coordinator, that is, the ***supervisors*** and ***facilitators***. The ***degree of complexity of the organization of the workload and the number of people performing each function will depend on the size of the study***. A national study in a country that has few administrative regions and is based on a small sample (small, but sufficient to meet the specific conditions in that particular country) will require a smaller administrative structure than that of a country where representativeness is regional as well as national, and where the sample is much larger.

1.1 The Survey Coordinator

The Survey Coordinator is the person responsible for coordination of all activities in the study. The Survey Coordinator will supervise and monitor all procedures related to the study and will keep abreast of all work carried out by supervisors and facilitators during the collection of information. As stated above, he or she should be in constant contact with the team from the NDO.

1.2 The Survey Supervisor

The Survey Supervisor is the person responsible for fieldwork in a specific area. There should be at least one Supervisor for each geographical area in which the study is conducted . The Supervisors answer directly to the Study Coordinator. Details of tasks should be outlined in the relevant handbook.

1.3 The Facilitator

The Facilitator is the person responsible for conducting the survey in each class assigned to him/her, ensures that the information is gathered by means of a self-administered questionnaire, and supervises the students' completion of the questionnaire in the classroom. The number of facilitators will depend on the number of classes in the sample.

2. INSTRUCTIONS FOR THE COORDINATOR

2.1 General tasks of the Survey Coordinator

The Coordinator is responsible for:

- ✓ Organizing and leading the survey in the country,
- ✓ Hiring supervisors and facilitators,
- ✓ Checking the supervisors' work,
- ✓ Ensuring compliance with all technical and methodological procedures and the instructions given by the NDO throughout the process,
- ✓ Checking envelopes and their contents before distributing them to the supervisors so that they in turn can distribute them to the facilitators,
- ✓ Ensuring the safety of the material,
- ✓ The quality of the data collected,
- ✓ The discipline of personnel under his or her supervision,
- ✓ Carrying out instructions from the NDO.

2.2 Tasks to be completed before the fieldwork (i.e., before collecting information from the students to be surveyed).

The Coordinator will complete the following tasks in preparation for fieldwork:

- ✓ Write to Principals of the schools selected for the sample requesting permission to conduct the study, explaining in detail the study's objectives, and providing information about the agency or institution conducting the study. The letter should stress the importance of gathering this type of information for the country. Depending on the country, the letter may also be signed by an official from the Ministry of Education.
- ✓ In some countries, parents must be informed before any study is conducted. If this is the case, the Study Coordinator should write to parents explaining the objectives of the study, stressing that the information is confidential, and that students' answers will be anonymous.
- ✓ Organize all details of the survey.
- ✓ Maintain the timetable that has been set.
- ✓ Organize a training course for the study's facilitators and supervisors to inform them of all details of the organization of the study, and their own responsibilities. The Coordinator must also review each question on the questionnaire, be ready to answer students' questions, and answer any concerns about the conduct of the fieldwork. The training course should be at least four hours with no more than 100 people in each course.
- ✓ The Coordinator should satisfy himself/herself that the facilitators are capable of answering students' questions (see Facilitator's introduction below), and that they have all material needed for the survey.
- ✓ The Coordinator should also ensure that Question 3 on the questionnaire, "Number of questionnaire" is filled out in the Coordinator's office after the data collection process. The left-hand box identifies the "school", the middle box identifies the "class", and the right-hand box contains a number from 1 up to the number of students in that class:

3. NUMBER OF QUESTIONNAIRE

School	Class	Sequential number

For example, if the identification number of a school is 111, the grade of the class in the sample is the 8th grade, and there are 30 students in that class, the numbers of the questionnaires for that specific class will be 111-08-01 to 111-08-30.

2.3 Coordinator's tasks

The Coordinator will organize the fieldwork on the basis of a random sample of schools and grades/sections received from the NDO, i.e., the **List of schools and classes** given in Table A3.1**;** to do this, the Coordinator should prepare a **List of classes in the sample**. Let us

assume that the total number of classes in the survey is 100. Referring to the sample in Table A3.1, the list of classes is shown in the table below:

Table A3.2: List of 100 classes in the sample, by school

Class	Information about the school					**Selected classes by school selected for sample**	
	Code of school	Name, address. Phone, name of contact	Department or region	County/parish or town	Type (public/private)	Grade/ Section	Number of students
1	2	ABC	f			8A	35
2	2	ABC				12A	28
3	4	DEF				10B	29
4	7	GHI				8B	30
5	7	GHI				10B	29
6	7	GHI				12A	25
..	..	…				..	..
..	..	…				..	..
100	125	XYZ				12C	35

Based on the information obtained earlier, the Coordinator will prepare the material that each Facilitator requires for the fieldwork.

a) An ***envelope for each class*** that will participate in the survey.

b) Place a ***label*** on each envelope (see example in Table A3.3), containing information similar to that given in section 5 for each questionnaire in the same class.

c) Place enough questionnaires in each *envelope* for all students in the class, using the number of students noted in the **List of Classes** (Table A3.2) plus an additional 10%.

d) Pencils to be given to the students; one per student.

e) Record on the "**Material Control Form"** the number of questionnaires and pencils to be given to each class.

f) Complete the information identifying the class on the label placed on the envelope.

Three pieces of information should be placed on the label:-

a. The information needed by the Facilitator to identify the class in which the survey will be conducted,

b. The information that will enable the Supervisor to check that all students have been surveyed, and

c. The details needed to identify the information that will be added to the questionnaires when they are digitized.

Label on Envelopes

Below is an example of a label based on the first line of the list of classes in Table A3.2. The Coordinator and his/her team should fill in the information required in capital (block) letters (except for the date of the survey and the number of students surveyed).

Table A3.3: Example of a label to be placed on an envelope for each class in the sample

1. Name of school					
2. Code of school					
3. Address of the school					
4. State, department or region					
5. County/parish/town					
6. Type of school	1 Public	2 Private	3 Other		
7. Type of school by sex	1 Male	2 Female	3 Mixed		
8. Date of survey	dd	mm	year		
9. Grade/section					
10. Number of students	Original sample	Number surveyed			
	35	Males	Females	Total	

Type of school, and **Type of school by sex** are data that the Coordinator or Supervisor would have obtained during their visit to the schools or through their contacts.

The **Date of the survey** will be written in by the Supervisor upon delivery of the envelope to the Facilitator. The **Number of students surveyed** will be written in or added at the time the survey is conducted.

Once the fieldwork has been completed, the Coordinator will send the following to the NDO:

a) The complete **Database** in the format defined (Excel, SPSS or other), provided this was assigned to the Coordinator.

b) Completed questionnaires and answer sheets

c) Supervisors' reports

d) Facilitators' reports

e) Final report, including detailed information on the work performed.

IMPORTANT: THE INFORMATION HIGHLIGHTED IN YELLOW ON THE LABEL (TABLE A3.3) IN ADDITION TO THE ANSWERS OF EACH STUDENT IN THAT CLASS SHOULD BE ENTERED BY AN INDIVIDUAL INDEPENDENT OF DATABASE ENTRY. THE SAME INFORMATION IS REPEATED FOR EACH STUDENT WHO ANSWERED THE QUESTIONNAIRE IN THE RESPECTIVE CLASS.

3. INSTRUCTIONS FOR SUPERVISORS

3.1. GENERAL

The Study's Supervisors must fully understand all documents, instruments, and instructions about the fieldwork, and must understand and apply concepts, definitions and data collection procedures. As stated above, the number of Supervisors will depend on the size of the study. There should be at least one Supervisor for each region in which the study is conducted, and each Supervisor will be responsible for a specific number of facilitators.

The present manual contains specific instructions for the tasks that Supervisors must deliver in a country when the study is conducted.

Supervisors are in charge of all stages of fieldwork. They are also responsible for developing uniform criteria for data collection to ensure that the work is carried out with as few errors as possible. This manual is a practical guide for the supervision of the data collection process.

3.2. IMPORTANCE OF THE SUPERVISOR

Supervisors play a very important role in the data collection process. They are responsible for executing all plans and procedures for the preliminary phases of the survey, controlling the sample, supervising the field work and the verification process, ensuring the safety of the documents, and sending them to the Coordinator.

3.3. MATERIALS

The Supervisor will be given the following:-

- Identification card
- Facilitator's instructions
- Supervisor's instructions
- Standardized questionnaire and answer sheets.
- Material delivery and receipt form
- Updated data form
- Pencils, erasers, and pencil sharpeners.
- List of the sample of schools, grade (class), groups or sections, address, etc. in the region where the Supervisor will work.
- Envelopes or a box in each classroom in which the students will deposit their completed questionnaires.

3.4. GENERAL AND SPECIFIC TASKS

Supervisors must take note of the material they give to facilitators for each day's workload; this will be returned to them once the fieldwork is completed.

3.4.1.- Tasks prior to the field work (collection of information from the students surveyed).

Supervisors must review all survey materials, and must:-

- Be completely familiar with the Supervisor's and the Facilitator's manuals.
- Distribute material to the facilitators.
- Inform facilitators of possible ways of travelling to the school assigned to them.

3.4.2.- Supervisors' tasks during the collection of information.

The work of Supervisors during the collection of information is, as follows:-

- Place the facilitators in their respective areas of work.
- Check that the facilitators are working in the school and class assigned to them.
- Pay random visits to the schools during the information gathering process for monitoring purposes.
- Intervene if there are any refusals and make every effort to reverse the situation.
- Record any errors made by the facilitators that may be detected. The Study Coordinator must be informed of any such errors. If a facilitator makes repeated errors, he or she should be replaced by someone else who can perform the work properly.
- Following the data collection process, visit schools as deemed appropriate (or at random) to check on the facilitators' work.
- If they find any omissions in schools, they should advise the facilitator and ask him or her to return to the school to review the work. If necessary, the Supervisor should accompany the facilitator.

3.4.3.- Tasks subsequent to collection of information

Supervisors must receive completed work in accordance with the timetable. They will use the material delivery and receipt form for each facilitator.

- Replace any facilitator who has not appeared for work (for whatever reason.)
- Review the work performed on the basic questionnaire.
- Review the number of questionnaires administered, noting the total number of questionnaires or answer sheets completed.
- Detect errors or omissions that may have occurred in the course of collecting the information.
- Carefully review the questionnaires to examine the consistency of information and ensure that they have been completely filled out.
- Prepare a detailed report on his or her work for the Survey Coordinator.

- Finally, deliver the material collected (completed forms, extra forms, and facilitators' reports) to the Survey Coordinator.

The Supervisor will also perform some review and coding work:-

Although the Supervisor has some review and codification to do, the most important part of this process is performed by the Coordinator. However, the Supervisor must perform the following specific tasks of review and coding:-

- ✓ The first task of the review is to go over the completed questionnaire in a general way and then carefully, question by question, correct any errors that may have occurred.
- ✓ In the event an Answer Sheet is used, the questionnaire should be coded prior to conducting the training courses and the fieldwork. This codification process will be the responsibility of the Study Coordinator, but it is the Supervisor who is responsible for verifying the codes during fieldwork.
- ✓ The Supervisor should conduct two reviews of the questionnaires. The first is a general review to check that the ***answers are consistent***. For example, a student may answer "No" to question TA1 ("Have you ever smoked cigarettes?") but then write down "aged 12" in question TA2 about the student's age when he or she first smoked a cigarette. That would be inconsistent—and there are many other examples. The reviewer/coder must find and correct such errors. In cases such as this, the questionnaire should be thrown out.

The Supervisor will then go through the questionnaire, question by question. The following errors are often found in this process:

- ✓ Some students (the younger they are, the more cases/errors are found) do not obey the instruction to "skip" a question. For example, if the student's answer to question EX1 is "No", meaning that he/she has never used ecstasy, they ought to have gone directly to question EX8, but instead answered some questions between EX2 and EX7. In such a case, if the error (or errors) committed by the student(s) DOES NOT AFFECT the content nor consistency of the questionnaire, the supervisor should correct the error with a red pen and mark it with a tick to show that the questionnaire is valid. Otherwise, it should be discarded.
- ✓ A student may also commit other errors such as writing errors. For example, in spite of the instruction to use a pencil, the student uses a pen. If that is the only error and the questionnaire has no other problems or errors, it will be considered valid.
- ✓ There may also be a case where the student does not follow the instructions about marking. That is to say, instead of a cross, the student marks a tick, or enters a line or colors in the space, or uses a pencil that does not write properly. If these are the only

writing errors, the questionnaire will be considered valid. Otherwise, it should be discarded.

- ✓ Questionnaires on which students have written jokes or insults should be thrown out, as should questionnaires that contain impolite or scurrilous remarks.

4. INSTRUCTIONS FOR THE FACILITATOR

The facilitator is the person responsible for obtaining the information (by means of the self-administered questionnaire) and for supervising the process of completing the questionnaires.

4.1 Importance

The success of the study largely depends on the *facilitator's* sense of responsibility and interest in the assignment.

The work of *facilitator* requires special personal attributes such as empathy, strong interpersonal skills, flexibility to adapt to different situations, as well as a sense of responsibility, discretion and a serious approach given that he/she will have access to confidential information.

4.2 Line of authority

The *facilitator* answers directly to the supervisor to whom he or she will report on the fieldwork and from whom he/she should ask for additional instructions.

4.3 Obligations

The *facilitator* has the obligation to:

- ➢ Assume the job responsibly and with dedication.
- ➢ Locate the schools and classes that were selected earlier.
- ➢ Fully understand this manual and any additional instructions.
- ➢ Request permission from the school Principal.
- ➢ Agree with the Principals on the dates and times for the survey to be administered in the selected classes.
- ➢ Respect the source of information in any investigation.
- ➢ Check that there are no students from other grades or classes in the selected class (coordinate with the school Principal and/or teacher).

- Clearly explain to students the objectives of the survey and how it will be conducted. Also inform them about the institution responsible for the survey (see Facilitator's Introduction below).
- Instruct the students how to complete the questionnaire (according to this manual.) No teacher from the school should be present during this process.
- Check that the students are completing the questionnaire with a pencil and not a pen.
- Complete the label on the envelope with the information on the class, and check that all information on the label is correct.
- Complete the form on the container (box, urn or envelope, depending on the country).
- Prepare a detailed technical report for the National Coordinator (once the **information** gathering process is completed.)

4.4 Facilitator materials

- Identification card.
- Facilitator's instructions.
- List of the addresses of selected schools, general information on said schools, and the classes selected in each school.
- Forms on the distribution of the number of students per school and grade level.
- Forms for delivery and receipt of the material.
- Envelopes with the standardized questionnaire.
- Pencils with erasers so that the students can complete the questionnaire.

4.5 Concepts and definitions

School

Primary and/or secondary schools, colleges or academies. Schools may be all boys, all girls or mixed. They are classified as follows:-

- **Public (state) schools:**

Primary or secondary schools financed by the state, province or the city/county. Sometimes these are called city schools.

- **Private schools**:

Primary or secondary schools privately funded or endowed. They may also be called parochial schools, fee-paying schools, independent schools, or preparatory schools.

- **Mixed schools**:

Schools that are both public and private, typically financed by religious organizations, trusts or non-governmental organizations that often receive a subsidy from the government.

Education System

Education systems are organized in different ways. Most countries use the following system:-

- **Primary education**:

First level of education, grades 1 to 5 or 6, depending on the country.

- **Secondary education**:

Secondary schools typically go from grades 8 through 12 (1st to 6th forms)

Based on the above, in most cases, education systems comprise 12 years of instruction; however, in some countries it is only 11 years.

4.6 Filling out the standardized questionnaire

The questionnaires should be completed using a pencil. The Facilitator will distribute the questionnaire with a pencil and an eraser to each student in the selected class. Each questionnaire has two parts:- the first, Questions 1-8, contain general information about the class selected. These should be completed by the supervisor and the facilitator. These initial questions are not directed to the individual student but relate to the class in general. The questions that the students must answer start with Question 9. These questions are specific to each student, but as we have said, **are anonymous and confidential**.

NOTE. The facilitator will inform the students that they should not answer Questions 1-8. Those questions should be filled out by the facilitator. The students will begin their answers with Question 9.

4.6.1 For the facilitator:

In Question 1, write down the name of the region (1) and the city/town/parish (2) where the school is located. Question 3 is a blank box, "Number of questionnaire." Students should not write anything in this box. It will be completed by the Study Coordinator.

In Question 4 (Type of school), put an "X" in the box next to the type of school as specified in these instructions to define a public, private, or other type of school. The type of school must be specified.

In Question 5, the facilitator will specify whether the school is all boys, all girls, or mixed.

In Question 6, the facilitator should mark with an "X" the grade or form in which the students selected for the survey is currently located. Then in Question 7, the facilitator will identify the class, that is, the grade/form and section of that grade. In Question 8, the facilitator will write in the number of students that are present in the class and are responding to the questionnaire.

Once the facilitator enters the classroom and is introduced by the teacher, the facilitator should introduce himself/herself to the students and give some instructions. The teacher does not participate in the completion of the questionnaire.

A **Model introduction by the Facilitator** is given below.

General instructions

The facilitator will introduce himself/herself by name and inform the students about the agency that he/she represents, and will then discuss completion of the questionnaire. As far as possible, the facilitator should not make any direct mention of the fact that he/she works for a national drug commission.

The following is a simple guideline:-

Facilitator's introduction

Good day, **students:**

My name is (full name). We are conducting a study in which we are surveying secondary school students in a number of schools in **NAME OF COUNTRY**, as well as in other countries in Latin America and the Caribbean.

We will distribute a questionnaire to each of you and ask you please to complete it. But first, we want you to assure you that:

1. This is not an exam; it is not a test. So, there are no right answers or wrong answers. Just answer the questions honestly.
2. Do not consult with your classmates. We need to know your own, honest opinion.
3. You will notice that there is no place on the questionnaire for you to write your name. That means that **no one will be able to find out who completed which questionnaire**. It is like a **secret ballot**. **No one can be identified**. Since you answer the questions by

marking crosses, you cannot even be identified by your handwriting. The results will be reported on an overall basis, not individually.

4. The questionnaire is organized into "modules", which are groups of questions on the same topic.

5. There are some questions where if you answer "No", you will skip or jump directly to the next module. For example, in question TA1, if you answer "No", then you should not complete anything in the next set of questions and should go directly to question CE1.

Another question where you can skip answers is question CE1, where if you answer "No", you need to skip to question CE5 in the same module.

But these are not the only questions where you can skip answers. Look carefully at others like this and be careful when you have to skip answers. **This is very important**.

6. Any questions?.... Good! Let us continue…….

The facilitator should add the following:-

- ✓ If the students answer the questions directly **on the questionnaire**:

When you have finished, ***put the questionnaire*** in this envelope (or box or urn) and leave the material I have just given you (pencil, eraser and pencil sharpener) on your desk.

- ✓ If they answer on the **answer sheet**:
- ✓ When you have finished, ***put the answer sheet*** in this envelope (or box or urn) and leave the material I have just given you (pencil, eraser and pencil sharpener) on your desk.

4.6.2 For the students:

Before beginning to complete the questionnaire, the students need to be informed about the characteristics of the questionnaire, and how to skip questions when necessary.

The facilitator will explain to the students that question 9, which refers to the student's sex, is the first one they will answer. They should put an "X" in the correct box.

4.6.3 Brief description of the questionnaire

Questions 9-12 correspond to **Module 2,** general information about the student. Question 10 asks about the student's age on the day before the survey.

In Question 11, they should put an "X" next to the box describing their parents' marital status, and specifying whether they live together or not, since there may be parents who are married but do not live together, or others who are separated but do live together.

Question 12 asks whether the student currently lives with one or both parents, or not.

In **Module 3**, there is a series of questions about the use of each substance. This is the central portion of the questionnaire, which will respond to the objectives of the study.

The questions ask about different substances. The number of questions about each substance varies. There are questions for estimating prevalence of substance use at three points in time:

- ✓ Lifetime prevalence, asking the question:- Have you ever used NAME OF THE SUBSTANCE? (Example: question TA1)
- ✓ Past year prevalence, asking the question:- Have you used NAME OF THE SUBSTANCE in the past 12 months? (Example: question TA4)
- ✓ Past month prevalence, asking the question:- Have you used NAME OF THE SUBSTANCE in the past 30 days? (Example: question TA5)

There is also a question designed to estimate incidence at two points in time:-

- ✓ Incidence in the past year,
- ✓ Incidence in the past month.

The same question is used for both cases: When was the first time you used NAME OF THE SUBSTANCE? (Example: Question TA3).

There are also questions for assessing the frequency and intensity of use of some substances (Examples: questions TA6 and TA7).

For some illicit substances, the students are also asked about obtaining the substances and substances that may have been offered to them.

Particularly in the case of marijuana, there are other questions about the types of substance and how they are used.

In **Module 4,** Questions 12-14 ask specifically about the students' perception of the degree of risk or harm in using a particular drug. The perception of risk/harm has shown in a number of studies to have a strong association with drug use. Those students that do not perceive a high risk or great harm have higher rates of drug use and by contrast, those who see greater risk or harm are those who are less involved in drug use.

In **Module 5**, Questions 15-21 relate to a set of questions about parental involvement, including different elements of parental supervision and monitoring based on the students' own statements about their parents' normal behavior. Many of these questions are currently used in research on the association between the family and drug use, particularly those that refer to parental supervision of what the student does outside the home, which is the variable *parental monitoring* (do parents always know where their children are? do they ask or expect to be told where the students are going when they go out? do they pay attention to what their children are doing at school? and, do they know enough about their children's friends?) In addition, the National Center on Addiction and Substance Abuse at Columbia University has shown the importance of certain behaviors by parents inside the home, particularly the custom of eating together regularly, checking the time their children come home at night, and monitoring what their children watch on television or when surfing the Internet (CASA, 2005).

Questions 22-25 refer to the students' perceptions of how angry or upset their parents would be if they found them using alcohol or marijuana.

Questions 26-28 ask about the type of relationship the student has with his/her parents, and the type of relationship between his/her mother and father.

Questions 29-34 refer to the student's preventive environment, that is to say, whether they talk to their parents about the dangers of using drugs, whether their parents had used any drug in the past, or whether their parents currently use alcohol or other drugs at home.

Questions 35-55 are in **Module 6**. Questions 35-40 refer to the school itself, and how the student feels about attending school:- is he or she well adapted? the relationship with teachers and in general?, whether they feel part of the school? The student is also asked about his/her average grades during a normal year, and the level of truancy.

Questions 41-47 relate to other responsibilities the students may have in addition to their studies, such as working. They also ask about expectations for the future, and whether the students think they will complete secondary education or go to university. These questions also ask about academic problems such as repeating grades or having disciplinary problems at school.

Questions 48-51 ask whether there are drugs in the school itself, whether there is drug use, whether drugs are offered, or all of the above.

Questions 52 and 53 relate to the influence of peer groups on drug use, and ask whether the student's closest friends are preventing them from drug use or not, and whether the student's friends use alcohol and/or marijuana.

Questions 54 and 55 ask whether the student is curious about the use of illicit drugs, and whether they would use them if they had the opportunity.