OAS | CICAD

INTER-AMERICAN DRUG ABUSE CONTROL COMMISSION (CICAD)

PROTOCOL FOR A STUDY ON DRUG USE AMONG UNIVERSITY STUDENTS

INTER-AMERICAN UNIFORM DRUG USE DATA SYSTEM (SIDUC)

2021

TABLE OF CONTENTS

PROTOCOL UNIVERSITY STUDENTS

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 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

[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

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 a country's university population.

This ***study of university students*** is one of a country's most important studies, since it examines a population group (generally aged 18-24) in which the highest rates of substance use are found. The goal of studies that a country conducts of the university population is, like other research projects, to contribute to a knowledge base so that the country can take informed decisions. Studies like this are also an essential input to our understanding of the regional and world situation with regard to substance use. Thus, the regular reports that CICAD's Inter-American Observatory on Drugs (OID) compiles and prepares on the problem are based on these studies. As we said earlier, in order for these reports on survey data to be technically sound, the national studies must be conducted using methodologically equivalent bases. This, then, is the purpose of this Protocol, like the others that have been developed within the SIDUC context.

While few countries have conducted surveys of university students, the **Report on Drug Use in the Americas 2019**[4] prepared by the OID shows that in those countries that have such studies, the current use of alcohol (prevalence of past month use) is over 50% in a number of countries, with little difference between males and females. A high proportion of these students evidence problem alcohol use. The report shows that the prevalence rate for use of marijuana in the past year is over 20% in some countries, with an upward trend in those countries that have enough studies to enable them to assess a trend. In addition, the report shows that a very significant proportion of university students (over 60% in some countries) consider that it is easy to obtain marijuana. Another situation addressed in the report is the high level of use of LSD by university students, a trend that is rising over time.

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.

The following specific objectives are designed to achieve the overall objective.

SPECIFIC OBJECTIVES:

- To estimate[5] for the different substances:[6]
  - ***prevalence*** of "lifetime" use,
  - ***prevalence*** of "past year" use,
  - ***prevalence*** of "past month" use,
  - ***incidence*** of "past year" use,
  - ***incidence*** of "past month" use, and
  - ***Age of first use*** (onset).
- To estimate the prevalence of the use of **at least one licit or illicit substance** in the "past year" and in the "past month",
- To estimate the prevalence of the use of **at least one illicit substance** in the "past year" and in the "past month",

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

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

[6] As per the list in section 4.1.

- To examine the frequency and pattern of use of tobacco, alcohol, marijuana, cocaine, cocaine base paste, [7] crack[8] and ecstasy,
- To estimate the prevalence of substance use by socio-demographic characteristics: sex, age, occupation, education, socio-economic level,
- To estimate the proportion of people who present with alcohol use disorders,
- 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 the population who **perceive different levels of risk/harm** in the occasional or frequent use of each of the substances described,
- To analyze the association between substance use and the perception of their risk/harm,
- To analyze the relationship that exists between easy access to and availability of illicit substances and use of them,
- To estimate indicators of the use of e-cigarettes/vaporizers, and their relationship to current or previous use of tobacco,
- To investigate the **sources** of obtaining (controlled) prescription drugs.

Thus, in the context of SIDUC, there are six main indicators associated with substance use: **three** on prevalence of 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** that reports the age of first use.

As explained later in Section 5, the questionnaire consists of a set of questions that directly respond to the specific objectives listed above, and therefore, neither the wording of the questions nor of the answers should be changed. The questionnaire also contains other questions that, while optional, are recommended by the OID for inclusion, so that the respondents' opinions can be a source of new information that can inform discussions about optimum drug policies in the country.

3. METHODOLOGY

Central to any investigation are the mechanisms that are used to carry out the processes, that is, the methods that will enable the study to meet its objectives. These objectives should be borne in mind at all times. The following sections of this Protocol will set out and discuss what is needed for this purpose. The information is obtained by means of a self-administered confidential questionnaire using a centralized on-line platform, based on a random sample of the target population.

[7] Use the name most common in the country.
[8] The country should decide whether to include cocaine base paste, or crack, or both.

3.1 TARGET POPULATION

Post-secondary education takes a number of different forms, sometimes with different names in different countries. On the one hand, we have *universities* which grant academic degrees and professional qualifications, and on the other, there are *vocational or technical schools*. Most countries have a common understanding of what a university is, but there is greater diversity among vocational training institutes. There has been an explosion over the last few decades in enrollment in both types of post-secondary education.

Thus, in the context of the Inter-American Drug Use Data System (SIDUC), the **universe of study** are **university students in a country's public and private universities** who are studying for an academic degree or professional qualification. Graduate or post-graduate students are not included here.

3.2 COVERAGE AND REPRESENTATIVENESS

An important question to be decided when determining the overall objective of this study is the **level at which the investigation 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 determined from the very beginning, as it has different impacts on a number of different areas.

Given the particular population to be studied, the coverage is national, i.e., it covers all (undergraduate) university students in the country, and therefore the study results are representative nationally. The reports on the study are the result of a global/overall analysis, with the classic demographic segments, including how the universities are administered/funded (public or private.)

3.3 SAMPLING ISSUES

Studies that use SIDUC 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 of it, which is obtained by means of scientifically rigorous statistical procedures. These are what are called sampling techniques. The first issue to be decided is what is called the ***unit of analysis***, that is, the subject that will provide the information needed to fulfill the objectives of the study. In this case, the ***unit of analysis is a university student*** selected from among the population as previously defined.

This study can be conducted in two different ways. The first involves obtaining a complete list of the university students in the country, and selecting a random sample of students from that list. Although this option seems to be feasible, it has the disadvantage that the sample may be quite scattered among the universities--in other words, with a large number of universities with only a few students each, which will greatly complicate administration of the survey.

The second alternative is the one discussed in this Protocol and involves a **two-stage sampling process**. Universities, both public and private, will be selected in the first stage, and then in the second stage, a defined number of students will be selected from among those universities selected in the first stage.

For the first phase, a sampling with these characteristics requires a ***sampling frame of universities*** and their student enrollment, i.e., number of students. Based on this information, the universities that will form part of the study are selected by means of a random process. Each university is asked to provide a list of students that meet the pre-determined criteria, and students are then selected at random. We have two options here. The first is to generate a single long list of students based on the individual lists from the universities selected, and then select a random sample of students from that list. The second option is to obtain a random sample (size to be determined) in each of the universities selected in the first stage, and the grouping of these samples is the final study sample. Both options involve a two-stage sampling process.

For the purposes of SIDUC, the decision to use the second alternative is for two main reasons. *The first is that a random sample is obtained of each university in the sample, which in turn would enable each university to have its own information with which to develop policies on tobacco, alcohol, and other drugs*. This information will be made available only to the corresponding university. The second reason is related to the size of the sample: using this alternative, the sample size is somewhat larger than that obtained with the other method, so that sampling errors at the national level are fewer, which means that the estimates are more precise.

In short, the sampling methodology ***calls for a selection of universities in the first stage***, and then an ***independent random sampling of students in each of the selected universities***.

Any sampling technique that is different from a simple random sampling, such as the **two-stage** sampling described above, is generically what is called ***complex sampling***. This is the case with a survey of university students.

Another basic question concerning the sampling is related to the ***size of the sample*** for the study, that is, to determine the number of university students 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 the purposes of SIDUC and a comparison between countries, the size of the sample should be large enough to ensure good estimates **overall, and by type of university, by sex and age group**.

The level of disaggregation is not the only condition for determining the size of the sample. Another necessary condition has to do with 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 draw up the sample itself, the advice of a professional statistician with training and experience in sampling techniques will be needed. In addition, given that this is a complex sampling (not self-weighting), the advisor must, once the on-line questionnaire has been completed, decide on the **expansion factors** needed for a correct statistical analysis of the information obtained. The country's National Observatory on Drugs is responsible for performing these activities. However, the OID may provide technical assistance in drawing up the sample, deciding on the expansion factors, or on any other aspect of the study, if such assistance is requested and provided resources are always available for this.

Further details on this section are given in **Annex 1**. The organization of the work is described in **Annex 3.**

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

- Annex 1: *Sampling*. More information on sampling techniques for the study among university students (for household surveys—Spanish says "población general", including a determination of the sample sizes.
- Annex 2: *Statistical analysis*. Details about calculating the expansion factors and the use of them in the analyses.
- Annex 3: *Administration of the study.*

3.4 FREQUENCY

As stated earlier, a study of this kind enables us to gain an understanding of the status of substance use at a particular moment or point in time, and also has the goal 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, to ensure that any changes that may be necessary can be made in a timely fashion. The OID/CICAD, in the framework of SIDUC, recommends that the studies be conducted ***every two years***.

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

4. OPERATIONAL DEFINITIONS

This section will describe the substances that will be analyzed as a function of the objectives of the study, and will also discuss the variables and indicators associated with these substances, as well as others that address the context and potential variables that may explain substance use.

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 substances suggested in order to estimate the six indicators

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

→ **Tranquilizers (total)**	→ **Stimulants (total)**	→ **Analgesics (total)**
• **Alprazolam** (Alprazolam Intensol, Xanax and Xanax XR) • **Clonazepam** (Klonopin) • **Diazepam** (Diastat AcuDial, Diazepam Intensol, Diastat and Valium) • **Flunitrazepam** (Rohipnol) • **Chlordiazepoxide** (Klopoxid, Libritabs, Librium, Monthural, Multum, Novapam, Risolid, Silibrin, Sonimen, Tropium and Zetran)	• **Methylphenidate** (Ritalin, Concerta) • **Phenmetrazine** (Preludin) • **Amphetamine** (Adderall, Adderall XR, Mydayis, Evekeo, Zenzedi and Dexedrine) • **Dextroamphetamine** (Dexedrine, DextroStat) • **Pemoline** (Cylert)	• **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; however, it is suggested that the same six indicators be estimated.

List of substances suggested for estimating at least three indicators

➢ **<u>Methamphetamines</u>** (Meth, ice, crystal)	➢ **<u>Opium</u>**	➢ **<u>Anabolic steroids</u>**
➢ **<u>Amphetamine</u>** (fet, speed)	➢ **<u>Heroin</u>** (Paste, "H", white powder, skag and tar)	➢ **<u>Ketamine</u>** (Keta, vitamin K, super K, CK or Calvin Klein, horse, Mary Kay or MaryK)
➢ **<u>Synthetic Cannabinoids</u>** (synthetic marijuana, Spice, K2, Joker, Black Mamba, Kush or Kronic)	➢ **<u>Synthetic Cathinones</u>** (Bath salts)	➢ **<u>Aminoindanes</u>** **<u>(MDAI gold)</u>**
➢ **<u>Piperazines</u>** (BZP, mCPP, A2, Legal X and Pep X)	➢ **<u>Phencyclidine</u>** (PCP, angel dust, embalming fluid, hog, killer joints, love boat, ozone, peace pill, superweed, rocket fuel, estrafalaria)	➢ **<u>Hallucinogenic plants</u>** (Floripondio, angel's trumpets, campanita, borrachero or cacao sabanero; DMT, yagé or ayahuasca; mescaline or peyote; **Psilocybin,** hallucinogenic mushrooms or magic mushrooms; khat; salvia, salvia divinorum or María Pastora; scopolamine or burundanga)
➢ **<u>GHB</u>** (liquid x, liquid ecstasy, Georgia home boy, Oop, Gamma-oh, grievous bodily harm, Mils, "G", liquid G, Fantasía)	➢ **<u>Lean</u>** (Sprite mix, cough syrup and pastilles – also called purple drank or sizzurp)	➢ **<u>Caffeine products</u>** (caffeine pills, energy drinks, caffeine powder)
➢ **<u>Phenethylamines</u>** (Europa, 4-FMP, RDJ, 4-MMA, Methyl-MA, 2C-C-NBOMe, bomb, N-bomb, 251, Nexus, 2C-E and Blue mystic)		

Countries may reduce the lists or add new substances to either of the lists in light of their own situation and experience.

They should also adapt the names to those commonly used in the country.

4.2.- VARIABLES, QUESTIONS AND INDICATORS

It is important in an investigation of this type to remember three concepts, which are interrelated and which 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 to do this? Lastly, we need an ***indicator*** that will give us the result of the measurement of the variables. Note that these three concepts are interrelated and should fulfill the specific objectives described earlier.

The following groups of study variables should be considered:

- **General** variables: sex, age, marital status, socio-economic level (where pertinent), and others.
- Variables on **drug use (yes or no):** 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",
- Variables on **frequency and intensity of drug use:** in this case, the variables refer to:
    - "frequency of use of [Name of substance]"
    - "Days used [Name of substance] in the past 30 days"
    - "Amount (in ….) used [Name of substance] in the past 30 days"
- Variables to assess **harmful use of alcohol:** these are a set of variables that when taken together, can construct an indicator of alcohol use disorder.
- Variables related to the **perceived risk/harm** of the use of a particular substance; "occasional/experimental use" is considered separately from "frequent use."
- Variables related to the **ease of obtaining** the substances.
- Variables related to **offers of drugs received**.

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

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

Lastly, on a question that is 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*** (of a specific substance) is an indicator that measures the proportion of people that report having used the substance (most recently) at some particular point in time: lifetime ("ever used,") past year or past month. Prevalence indicators refer to the total sample (expanded) of individuals, and are generally expressed as a percentage.
- On the other hand, ***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 individuals 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 the other simple variables, some of which are demographic (% of males), others have to do with perceived risk (% that perceives that getting drunk is highly risky/of great harm to health), and others about offers of drugs (% who were offered ecstasy in the past year.) In addition, there are other percentages of interest that are the result of the analysis of a set of several variables; for example, we may want to estimate the ***percentage of people*** who drank alcohol in the past year whose **use of alcohol** was ***hazardous and harmful.*** This percentage is estimated by administering the ten-question AUDIT questionnaire (**A**lcohol **U**se **D**isorders **I**dentification **T**est), which was developed by the World Health Organization.[9] These 10 questions combined produce a score that can determine the percentage of people with hazardous or risky use of alcohol. Further details are given in Annex 2.
- Lastly, ***quantitative measurements*** are associated with quantitative variables such as "age of the respondent" or "age of first use of tobacco," where the **indicators** to evaluate the answers will be, for example, the **average or *mean*,** the **median,** and some **percentiles** (particularly the 25th and 75th percentiles.)

[9] https://www.who.int/publications/i/item/audit-the-alcohol-use-disorders-identification-test-guidelines-for-use-in-primary-health-care

Although the indicators of prevalence and incidence are measured using a percentage, the difference from the 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 a current view of the demand for drugs—whether it is rising or falling, in which specific groups these changes may be occurring, and also to have some objective criteria about some of the outcomes of drug policy in the country.

The **magnitud**e of drug use is measured by indicators called *prevalence* and *incidence*. Prevalence measures the total number of cases/subjects that have used a drug in a *period of time*, whether they were **new cases** that began during that period, or **old cases** with a history of substance use. On the other hand, incidence focuses only on **new cases** in a specified time period. The table below gives a set of questions that need to be answered in the survey, and shows the indicators that would be useful in eliciting an answer, taking one substance, alcohol, as the reference.

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

The definitions of these indicators are discussed in more detail later in this Protocol, but we must stress that they should be measured on a regular basis.

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

5. THE QUESTIONNAIRE: ORGANIZATION OF THE QUESTIONS INTO MODULES

This section provides the questionnaire, which will fulfill the specific objectives defined in SIDUC. This questionnaire will be available *on line, and will be administered by the country's National Observatory on Drugs and the Inter-American Observatory on Drugs (OID) of the Organization of American States.* The movement from one question to the next is automatic on line, but for the purposes of describing the questionnaire here, they are noted as "Skip to XX" where XX is the next question the student must answer. Further details on this are given in Annex 3.

As noted in the different modules of the questionnaire, the countries should adapt some of the questions to their own circumstances and make the necessary changes. The responses are absolutely confidential.

The **questionnaire** consists of the following modules:

MODULE 1. GENERAL INFORMATION

MODULE 2. SUBSTANCE USE:

- **TOBACCO**
- **ELECTRONIC CIGARETTES**
- **ALCOHOL**
- **PRESCRIPTION TRANQUILIZERS**
- **PRESCRIPTION STIMULANTS**
- **PRESCRIPTION ANALGESICS**
- **CANNABIS:**
  - **MARIJUANA**
  - **HASHISH**
- **COCAINE SUBSTANCES**
  - **COCAINE HYDROCHLORIDE**
  - **COCAINE BASE PASTE**
  - **CRACK**
- **ECSTASY**
- **LSD**
- **INHALANTS**
- **POPPERS**

- **METHAMPHETAMINE**
- **AMPHETAMINE**
- **HEROIN**
- **KETAMINE**
- **OTHER DRUGS**

MODULE 3: PERCEIVED RISK AND FACTORS ASSOCIATED WITH DRUG USE
MODULE 4. PREVENTION PROGRAMS AND INFORMATION ON TREATMENT

The table below shows for each topic the questions that are ***obligatory***: these are the questions that can respond to the objectives defined in Section 2 of this Protocol. As noted in that section, these questions and their possible answers must not be changed. The last column shows the questions that are ***optional*** for each area, and the country may therefore decide whether or not to include them. Some substances may or may not be included, depending on the circumstances in each country.

Modules[10]	**Obligatory questions**	**Optional questions**
General information	1, 2, 4, 5, 12, 14-19, 21-34	3, 6-11, 13, 20
Module 2: Tobacco	All	
Electronic cigarettes	All	
Alcohol	All	
Prescription tranquilizers	All	
Prescription stimulants	All	
Prescription analgesics	All	
Marijuana	MA1-MA11	MA12, MA13 and MA14-Ma17
Hashish	All, if the country decides to include this substance	
Cocaine	CO1 to CO8	CO9 and CO10
Cocaine base paste	PB1 to PB8	PB9 and PB10
Crack	CR1 to CR8, but will depend on the country	CR9 and CR10
Ecstasy	EX1 to EX8, EX11 to EX13	EX9 and EX10
LSD	All	
Inhalants	All	
Poppers	All, if the country decides to include this substance	
Methamphetamine	All	
Amphetamine	All	
Heroin	All, if the country decides to include this substance	
Ketamine	All	
Other drugs	All	
Module 3: Perceived risk and factors associated with drug use	All	
Module 4: Prevention and treatment programs		All

[10] Adjusting the names of drugs according to the language used in your country is recommended.

INFORMATION ON THE FIRST/START-UP SCREEN

Dear student,

Thank you for taking part in this ***Epidemiological survey of university students on public health and associated factors.*** It is part of a research project being conducted by the National Observatory on Drugs of COUNTRY X in cooperation with the Organization of American States (OAS), and is designed to gather reliable information on which to base future student welfare policies in the participating universities.

This research project is being conducted in different universities in the country. It is based on a random sample first of universities and then of students. Participants' data are protected according to bioethics and statistical secrecy standards. You are one of the students selected in this sampling process.

Your responses are voluntary, individual, and absolutely anonymous and confidential. We will be grateful for your complete cooperation in providing accurate answers. Each student's on-line responses will automatically populate a database which is stored on a server ENTER HERE WHAT WAS AGREED ON BY THE NATIONAL DRUG OBSERVATORY AND THE OID, which will make it doubly certain that the information provided is private. Your responses and those of the other students selected will be used for statistical purposes only. Individual participants cannot be identified.

Your opinions are very important to this research study, and we thank you again for volunteering to participate.

I agree voluntarily to participate in this study, and will answer the questionnaire truthfully. I understand that my responses will be completely confidential.

Click here to begin the questionnaire.

MODULE 1: GENERAL INFORMATION

1. Sex	**2. How old are you?**
1. Male 2. Female	o ……………years old
3. In what year did you begin your current course of study?	**4. How would you describe your financial situation?**
o In……………	1. Very good 2. Good 3. Not bad 4. Poor 5. Very bad/poor
5. Do you currently work as well as study?	**6. Approximately how many hours per week do you work?** Number should be between 1 and 60
1. Yes 2. No (Skip to Q9)	

7. Where are you living as a student?
Answer all that apply

1. I live in the university or on campus
2. I live outside the university/off campus, with friends
3. I live alone, outside the university/off campus
4. I live with my parents
5. I live in the house of a family member
6. I live with my partner
7. Other

8. How do you pay for your housing as a student?
Mark all that apply

1. My parents pay
2. The government pays
3. I work to support myself
4. I have a scholarship

9. How do you pay for your food, as a student?
Mark all that apply

1. My parents pay
2. The government pays
3. I work to support myself
4. I have a scholarship

10. How do you pay for your university fees, as a student?
Mark all that apply

1. My parents pay
2. The government pays
3. I work to support myself
4. I have a scholarship

11. How do you pay for books, photocopies, etc.?
Mark all that apply

1. My parents pay
2. The government pays
3. I work to support myself
4. I have a scholarship

12. Do you find it difficult to pay for your university studies?

1. Very difficult
2. Difficult
3. Not easy, not difficult
4. Easy
5. Very easy

13. How much is your family's monthly income, on average? (in round numbers.) If your family's income comes from more than one person, please enter the average income of the whole family.

	Income

14. How many years did you repeat during primary school?	**15. How many years did you repeat during <u>secondary school</u>?**
1. None 2. 1 3. 2 4. 3 5. 4 or more	1. None 2. 1 3. 2 4. 3 5. 4 or more

16. Did you have any academic difficulties in secondary school?

1. Never or rarely
2. Several times
3. Often

17. Did you have any disciplinary problems during secondary school?

1. Never or rarely
2. Several times
3. Often

18. How many times were you *expelled* from your secondary school?

1. None
2. Once
3. Twice
4. 3 times
5. 4 or more times

19. How would you describe your experience as a secondary school student?

1. Completely satisfactory
2. Generally satisfactory
3. Not bad
4. Generally unsatisfactory
5. Completely unsatisfactory

20. How old were you when you graduated from secondary school?

21. How many times have you changed university? Mark all that apply

1. None (skip to Q25)
2. Once
3. Twice
4. 3 times
5. 4 times
6. 5 or more times

22. What was the main reason or reasons you changed university?* Mark all that apply

1. I didn't like the university
2. I didn't like my course of study
3. I couldn't pay
4. I moved to another city
5. For academic reasons
6. It wasn't what I wanted to study
7. Other

23. How satisfied are you with what you are studying?

1. Very satisfied
2. Satisfied
3. Not bad
4. Not satisfied
5. Don't know

24. How many courses did you fail during university?

1. None
2. One
3. 2-3
4. 3-4
5. More than 5
6. Not applicable, am in my first term/semester or year

25. Have you ever thought about dropping out?

1. Never
2. Once or twice
3. Several times

26. Do you think you will easily complete your studies and graduate?

1. Yes, easily
2. Yes, but with some difficulty
3. Yes, but with a lot of difficulty
4. I don't think I'll be able to

27. How do you see your professional future?

1. Very optimistic
2. Optimistic
3. Pessimistic
4. Very pessimistic
5. It's not clear to me

28. What is your parents' marital status?

1. Married
2. Divorced
3. Separated
4. Widow/widower
5. Living together
6. Single
7. Not applicable

29. Who do you live with at the moment?

1. My parents
2. My mother
3. My father
4. My mother and my step-father
5. My father and my step-mother
6. Another family member
7. I live alone
8. Other--specify

30. What is your relationship like with your mother?

1. Better than before
2. Worse than before
3. Continues to be good
4. Continues to be bad
5. Not applicable

31. What is your relationship like with your father?

1. Better than before
2. Worse than before
3. Continues to be good
4. Continues to be bad
5. Not applicable

34. How much do your parents or family members monitor/check up on the following?

	A lot	Somewhat	Not much	Not at all
Your studies				
Your love life				
Your leisure activities				
Your friends				

MODULE 2: SUBSTANCE USE

TOBACCO (TA). **For each question, think about filter and nonfiltered cigarettes, cigars, pipes, hookahs, etc.**

TA1. Have you ever smoked tobacco?

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

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

	Years old

TA3 When was the first time you smoked tobacco?

	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

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

	0. Yes	
	1. No	(skip to TA8)
	9. Don't know/no opinion	(skip to TA8)

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

	1. Yes	
	2. No	(skip to TA7)
	9. Don't know/no opinion	(skip to TA7)

TA6. Do you currently smoke every day or less than every day?

	1. Every day	(skip to CE1)
	2. Only some days	(skip to TA7)
	9. Don't know	(skip to CE1)

TA7. In the past, did you smoke tobacco every day?

	1. Yes	Skip to CE1
	2. No	Skip to CE1
	9. Don't know	Skip to CE1

TA8. In the past, did you smoke every day or less than every day?

	1. Every day
	2. Only some days
	9. Don't know

ELECTRONIC CIGARETTES (CE)

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 ever 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 products?

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

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

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

CE11. Have you vaped e-cigarettes containing only flavoring products 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 flavoring products?

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

ALCOHOL *(alcoholic beverages such as beer, wine, champagne, brandy, whisky or other hard liquor (high concentration of alcohol) either alone, mixed with something else, or injected)*

AL1 Have you ever drunk an alcoholic beverage?	**AL2 How old were you when you first drank an alcoholic beverage?**
1. Yes 2. No (Skip to AL24)	Years old: ……………………
AL3 When was the first time you drank alcohol?	**AL4 Have you ever injected alcohol?**
1. During the past 30 days 2. More than 1 month ago but less than 1 year ago **3.** More than 1 year ago	1. Yes 2. No
AL5 Have you drunk alcohol in the past 12 months?	
1. Yes 2. No (Skip to AL24)	

The next questions have to do with your use of alcohol in the past year. Please mark the answers that are closest to your own. Note below the quantities of alcohol that represent the number of drinks:

1 drink	A bottle or individual can of beer (333 cc.). One glass of wine (140 cc.) One shot of hard liquor (40 cc.) (gin, rum, vodka, whisky) either alone or mixed with something else
1 drink and a half	Half a liter of beer
3 drinks	One liter of beer
6 drinks	One bottle of wine (750 cc.)
8 drinks	One liter of wine
18 drinks	One bottle of hard liquor (750cc.) (gin, rum, vodka, whisky)

AL6 How often do you drink an alcoholic beverage?	**AL7 How many drinks do you have on a typical day when you are drinking?**
1. Once a month or less 2. 2-4 times a month 3. 2-3 times a week 4. 4 or more times a week	1. 1 or 2 2. 3 or 4 3. 5 or 6 4. 7-9 5. 10 or more

Please answer the following questions.

	Never	*Less than once monthly*	*Monthly*	*Weekly*	*Every day or almost every day*
AL8 How often do you have 6 or more drinks in one day?	○	○	○	○	○
AL9 How often during the past 12 months have you found that you were not able to stop drinking once you had started?	○	○	○	○	○
AL10 How often during the past 12 months have you found that you failed to do what was normally expected from you because of drinking?	○	○	○	○	○
AL11 How often during the past 12 months have you needed a first drink in the morning to get yourself going after a heavy drinking session?	○	○	○	○	○
AL12 How often during the past 12 months have you had a feeling of guilt or remorse after drinking?	○	○	○	○	○
AL13 How often in the past 12 months have you been unable to remember what happened the night before because you had been drinking?	○	○	○	○	○

AL14 Have you or someone else been injured as a result of your drinking?	**AL15 Has a relative, friend, doctor or other health professional been concerned about your drinking or suggested that you cut down?**
1. No 2. Yes, but not during the past year 3. Yes, in the past year	1. No 2. Yes, but not in the past year 3. Yes, in the past year
AL16 During the past 12 months, have you driven a vehicle after having drunk alcohol? (Car, motorcycle, truck, etc.	**AL17 Thinking about the past 12 months, have you ever combined alcohol with energy drinks?*** such as Red Bull, Monster, Rockstar or similar
1. Never 2. From time to time 3. Often 4. Not applicable, I don't drive	1. Never 2. A few times 3. Regularly 4. Almost always 5. Always
AL18 Have you drunk alcohol in the past 30 days?	
1. Yes 2. No (Skip to AL24)	

In the past 30 days, what type of alcoholic beverage did you drink, and how often?

	Every day	On the weekends	Some days a week	Only on social occasions	I don't drink alcohol
AL19. Beer					
AL20. Wine					
AL21. Sparkling wine (champagne, or similar)					
AL22. Hard liquor, alone or mixed (rum and Coke, vodka and orange juice, etc.)					

AL23 Where do you most often drink alcohol?* Mark all that apply	**AL24 Do you have friends or family members who get drunk?**
1. At home 2. At a friend's house 3. At the university 4. In areas around the university 5. At work 6. In a shop 7. On a trip or outing 8. In a park or recreation spot 9. At a party or discotheque, or in a bar 10. At a concert, sporting event, etc.) 11. Other, Where?-----------------------	1. No 2. One 3. Two or more 4. Don't know
AL25 Have you seen students drinking alcohol in any of the classrooms or lecture halls of your university?*	**AL26 In the past 12 months, have you ridden in a vehicle with a driver who has drunk alcohol?** (Car, motorcycle, truck, etc.)
1. Never 2. From time to time 3. Often	1. Never 2. From time to time 3. Often 4. Not applicable
AL27 In the past 12 months, have you ridden in a vehicle driven by someone under the influence of alcohol? (Car, motorcycle, truck, etc.)	**Do you think that being drunk influences the following behaviors?**
1. Never 2. From time to time 3. Often	Yes No **AL28**. Having sex __ __ **AL29**. Having unprotected sex __ __ **AL30**. Using drugs __ __ **AL31**. Having an accident __ __

PRESCRIPTION TRANQUILIZERS

The next questions have to do with 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, Multum, Novapam, Risolid, Silibrin, Sonimen, Tropium and Zetran) or similar.

TR1. Have you ever taken a tranquilizer because a doctor prescribed it for you?

	1. Yes
	2. No

TR2. Have you ever taken a tranquilizer without a prescription?

	1. Yes	
	2. No	(skip to TR9)

TR3. How old were you the first time you 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 TR9)
	9. Don't know/no opinion	

TR6. 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** (Klopoxid, Libritabs, Librium, Monthural, Multum, 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. From a paramedic		4. A friend gave them to me
	2. In the street		5. A family member gave them to me
	3. At home		6. From the pharmacy

TR8. Have you taken a tranquilizer without a prescription in the past 30 days?

	1. Yes
	1. 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 AMEND THE LIST OF TRANQUILIZERS AS NECESSARY IN ORDER TO REFLECT THE NAMES USED LOCALLY IN THE COUNTRY.

PRESCRIPTION STIMULANTS

The following questions concern the use of prescription 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 a stimulant because a doctor prescribed it for you?

	1. Yes
	2. No

ES2. Have you ever taken a stimulant without a doctor's prescription?

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

Which stimulants have you taken 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. From a paramedic		4. A friend gave them to me
	2. In the street		5. A family member gave them to me
	3. At home		6. From the pharmacy

ES8. Have you taken a stimulant without a prescription in the past 30 days?

	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 ADAPT THE LIST OF STIMULANTS TO REFLECT THE NAMES USED IN THE COUNTRY.

PRESCRIPTION ANALGESICS

We're now going to ask you some questions about the use of some prescription analgesics or painkillers 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, 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) or similar.*

AN1. Have you ever taken or been given an analgesic (painkiller) because a doctor prescribed it for you?

	1. Yes	
	2. No	(skip to AN3)

And which analgesics have you taken or been given in the past 12 months?
Mark all that apply

AN2.1		**None in the past 12 months**
AN2.2		**Fentanyl** (Duragesic, Ionsys, Subsys and Abstral)
AN2.3		**Tramadol** (ConZip and Ultram)
AN2.4		**Hydromorphone** (Dilaudid)
AN2.5		**Hydrocodone** (Lorcet, Vicodin, Hycet, Lortab)
AN2.6		**Oxycodone** (OxyContin, Xtampza ER, Oxaydo, Roxicodona, Primlev, Tylox, Endocet, Percocet and Percodan)
AN2.7		**Methadone** (Diskets, Metadona Intensol, Dolophine and Methadose)
AN2.8		**Codeine** (Codeisan, Codeisan jarabe, Fludan codeína, Histaverin, Notusin, Perduretas codeína and Toseina)
AN2.9		**Morphine** (MorphaBond ER, Arymo ER, Infumorph P/F, Astramorph-PF, Duramorph, MS Contin)
AN2.10		Other:

AN3. Have you ever taken an analgesic without a doctor's prescription?

	1. Yes	
	2. No	(skip to AN9)

AN4. And in the past 12 months, which of the following analgesics have you taken?
Mark all that apply.

AN4.1		**None in the past 12 months**
AN4.2		**Fentanyl** (Duragesic, Ionsys, Subsys and Abstral)
AN4.3		**Tramadol** (ConZip and Ultram)
AN4.4		**Hydromorphone** (Dilaudid)
AN4.5		**Hydrocodone** (Lorcet, Vicodin, Hycet, Lortab)
AN4.6		**Oxycodone** (OxyContin, Xtampza ER, Oxaydo, Roxicodona, Primlev, Tylox, Endocet, Percocet and Percodan)
AN4.7		**Methadone** (Diskets, Metadona Intensol, Dolophine and Methadose)
AN4.8		**Codeine** (Codeisan, Codeisan jarabe, Fludan codeína, Histaverin, Notusin, Perduretas codeína and Toseina)
AN4.9		**Morphine** (MorphaBond ER, Arymo ER, Infumorph P/F, Astramorph-PF, Duramorph, MS Contin)
AN4.10		Other:

AN5. How old were you the first time you took an analgesic without a prescription?

	Years old

AN6. 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

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

	1. From a paramedic		4. A friend gave them to me
	2. In the street		5. A family member gave them to me
	3. At home		6. From the pharmacy

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 an analgesic 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 ADAPT THE LIST OF ANALGESICS TO REFLECT THE NAMES USED LOCALLY.

MARIJUANA (MA)

MA1. Have you ever used marijuana?

	1. Yes	
	2. No	(skip to MA13)

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. That first time you used marijuana, why did you do it?

	1. To try, out of curiosity
	2. Someone gave it to me to try
	3. Because my group of friends smoke marijuana
	4. Because I was feeling bad
	5. Other. What reason?________________
	6. I don't know why I tried it

MA5. Have you used marijuana in the past 12 months?

	1. Yes	(skip to MA14)
	2. No	

MA6. How often in the past 12 months have you used marijuana?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

MA7. How did you most often use marijuana in the past 12 months? Mark only one answer. THE COUNTRY SHOULD DECIDE.

	1. Smoked
	2. Vaped
	3. In food
	4. In oils or tinctures
	5. In pharmaceutical products or for medical use
	6. Other. Which?________________

MA8. Have you ever used marijuana in the past 30 days?

	1. Yes	
	2. No	(skip to MA14)

MA9. Thinking only about the past 30 days, how many days in the past 30 days did you use marijuana?

	Number of days (from 0 to 30)

MA10. How many marijuana cigarettes (joints or spliffs) do you smoke a month?

	Number of joints

MA11. How often 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

MA12. As far as you know, how much does a marijuana cigarette or joint (spliff) cost?

	In local currency **99=Don't know/no opinion**

MA13. How much did you spend on marijuana in the past 30 days?

	In local currency **99=Don't know/no opinion**

MA14. What is your opinion about the following measures?	Fully agree	Agree	Disagree	Completely disagree	Don't know/ no opinion
a) Allow marijuana to be used for therapeutic or medical purposes					
b) Allow marijuana to be used for religious purposes (e.g., Rastafarians)					
c) Allow marijuana to be grown in limited amounts in individual households					
d) Allow possession of marijuana in limited amounts for personal use					
e) Allow persons addicted to marijuana or other drugs who commit crimes such as theft to be put into a court-supervised drug treatment program instead of prison					
d) Allow marijuana to be cultivated for scientific research					

MA15. Some countries have permitted marijuana for recreational use. Have you heard about this?

Yes	
No	(skip to MA17)
Don't know/no opinion	(skip to MA17)

MA16. Do you agree or disagree with the laws in those countries that permit recreational use of marijuana?

Yes
No
Don't know/no opinion

MA17. Would you be in agreement with having a law in the country that would allow marijuana for recreational use?

1. Fully agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Very much disagree

Translator's note: (MA15-MA17) - The Spanish says "regulate the market", which is perhaps not very clear.

HASHISH (HA)

HA1. Have you ever used hashish?

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

HA2. How old were you when you used hashish for the first time?

	Years old

HA3. When was the first time you used hashish?

	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

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

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

HA5. How often have you used hashish?

	1. Only once
	2. Several times in 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

COCAINE (CO)

CO1. Have you ever used cocaine? ☐ 1. Yes ☐ 2. No (skip to PB1**)**	**CO2. How old were you when you used cocaine for the first time?** ☐ Years old
CO3. When was the <u>first time</u> 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 <u>12 months</u>?** ☐ **1.** Yes ☐ **2.** No (skip to PB1**)**
CO5. How often did you use cocaine in the past 12 months? ☐ 1. Only once ☐ 2. Several times during the past 12 months ☐ 3. Several times a month ☐ 4. Several times a week ☐ 5. Every day	**CO6. Have you used cocaine in the <u>past 30 days</u>?** ☐ **1.** Yes ☐ **2.** No (skip to PB1**)**
CO7. Thinking only about the past 30 days, on how many days did you use cocaine? ☐ Number of days (from 0 to 30)	**CO8. How many grams/ounces of cocaine do you use in a month?** ☐ Number of grams/ounces
CO9. As far as you know, how much does a gram/ounce of cocaine cost? ☐ In local currency **99=Don't know/no opinion**	**CO10. Approximately how much did you spend on cocaine during the past 30 days?** ☐ In local currency **99=Don't know/no opinion**

COCAINE BASE PASTE (PB)

PB1. Have you ever tried cocaine base paste?

	1. Yes	
	2. No	(skip to CR1)

PB2. How old were you when you used cocaine base paste for the first time?

	Years old

PB3. When was the first time you tried cocaine base 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 ever used cocaine base paste in the <u>past 12 months</u>?

	1. Yes	
	2. No	(skip to CR1)

PB5. How often did you use cocaine base paste in the past 12 months?

	1. Only once
	2. Several times in the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

PB6. Have you used cocaine base paste in the <u>past 30 days</u>?

	1. Yes	
	2. No	(skip to CR1)

PB7. Thinking only about the past 30 days, how many days out of the past 30 days did you use cocaine base paste?

	Number of days (from 0 to 30)

PB8. How many units of cocaine base paste do you use a month?

	Number of units

PB9. As far as you know, how much does a unit of cocaine base paste cost?

	In local currency **99=Don't know/no opinion**

PB10. How much did you spend on cocaine base paste in the past 30 days?

	In local currency **99=Don't know/no opinion**

NOTE: EACH COUNTRY SHOULD USE THE NAME GIVEN TO THIS SUBSTANCE LOCALLY

CRACK (CR)

CR1. **Have you ever tried crack?**

	1. Yes	
	2. No	(skip to IN1)

Translator's note: This should probably be EX1, which is the next module in the questionnaire.

CR2. **How old were you when you used crack for the first time?**

	Years old

CR3. When was the first time you tried 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 past <u>12 months</u>?

	1. Yes	
	2. No	skip to IN1)

CR5. How often did you use crack in the past 12 months?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

CR6. Have you used crack in the <u>past 30 days</u>?

	1. Yes	
	2. No	(skip to IN1)

CR7. Thinking only about the past 30 days, how many days during the past 30 days did you use crack?

	Number of days (from 0 to 30)

CR8. How many rocks of crack do you use a month?

	Number of rocks

CR9. As far as you know, how much does a rock of crack cost?

	In local currency **99=Don't know/no opinion**

CR10. How much did you spend on crack in the past 30 days?

	In local currency **99=Don't know/no opinion**

Note: Each country should adapt the names to reflect those used locally.

MDMA – ECSTASY (EX)

EX1. Have you ever tried ecstasy?

	1. Yes	
	2. No	(skip to LS1)

EX2. How old were you when you tried ecstasy for the first time?

	Years old

EX3. When was the <u>first time</u> you tried 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 <u>past 12 months</u>?

	1. Yes	
	2. No	(skip to LS1)

EX5. How often have you used ecstasy?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

EX6. Have you used ecstasy in the <u>past 30 days</u>?

	1. Yes	
	2. No	(skip to LS1)

EX7. Thinking only about the past 30 days, how many days out of the past 30 days did you use ecstasy?

	Number of days (from 0 to 30)

EX8. How many ecstasy pills or tablets do you use in a month?

	Number of pills

EX9. As far as you know, how much does an ecstasy pill cost?

	In local currency

EX10. How much did you spend on ecstasy in the past 30 days?

	In local currency

EX11. Thinking about the times you have used ecstasy in the past 30 days, did you drink an alcoholic beverage at the same time as you were using ecstasy?

	1. Yes
	2. No

EX12. Thinking about the times you have used ecstasy in the past 30 days, did you use any other drug at the same time as you were using ecstasy?

	1. Yes
	2. No

EX13. Thinking about the times you wanted to use ecstasy in the past 30 days, did you end up using another similar drug because no ecstasy was available?

	1. Yes
	2. No

LSD (LS)

LS1. Have you ever used LSD (acid, trips, mellow yellow)?

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

LS2. How old were you when you used LSD for the first time?

Years old

LS3. When was the <u>first time you used</u> 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 <u>past 12 months</u>?

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

LS5. How often have you used LSD in the past 12 months?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

LS6. Have you used LSD in the <u>past 30 days</u>?

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

INHALANTS (IN). Substances such as glue, paint, varnish, deodorant, gasoline, petrol, benzene, toluene, paint thinner or something similar.

IN1. Have you ever Inhaled glue, paint, varnish, deodorant, gasoline, petrol, benzene, toluene, paint thinner or something similar?

	1. Yes	
	2. No	(skip to PO1)

IN2. How old were you when you used inhalants for the first time?

	Years old

IN3. When was the <u>first time</u> 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 <u>past 12 months</u>?

	1. Yes	
	2. No	(skip to PO1)

IN5. How often have you used inhalants?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

IN6. Have you used inhalants in the <u>past 30 days?</u>

	1. Yes	
	2. No	(skip to PO1)

IN7. And those times you used inhalants in the past 30 days in order to get high, what kind of substance did you use? *(Mark all that apply)*

	1. Glue
	2. Paint
	3. Deodorant
	4. Toluene
	5. Gasoline/petrol/benzene or paraffin
	6. Ether or acetone
	7. Lighter fluid
	8. Other Which?............................

IN8. And the times you used inhalants in the past 30 days, where did you get them from? *(Mark all that apply)*

	1. Supermarket or pharmacy
	2. Hardware store, DIY shop, gas station (petrol station)
	3. In my neighborhood shop
	4. At home
	5. At school
	6. At work
	7. From friends
	8. From someone who is not a friend
	9. Over the Internet

Note. The country should use the names most commonly used locally.

POPPERS (PO)

PO1. Have you ever used poppers?

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

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 one year ago
	3. More than 1 year ago
	9. Don't know/no opinion

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

	1. Yes	
	2. No	(skip to OD1) **ME1**

PO5. How often did you use poppers in the past 12 months?

	1. Only once
	2. Several times during the past 12 months
	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	(skip to OD1)

NOTES:

- ✓ **EACH COUNTRY SHOULD DECIDE WHETHER OR NOT TO INCLUDE POPPERS. IF THE DECISION IS TO DELETE THE SIX QUESTIONS ON POPPERS, THEN POPPERS SHOULD BE INCLUDED AS ONE OF THE SUBSTANCES IN QUESTIONS IN1 AND IN5 ON INHALANTS.**
- ✓ **COUNTRIES MAY ADD ADDITIONAL QUESTIONS FOR SPECIFIC SUBSTANCES.**
- ✓ **COUNTRIES MAY ADD A COMPLETE MODULE FOR A SPECIFIC SUBSTANCE THAT NEEDS SPECIAL ATTENTION.**

METHAMPHETAMINE (ME)

ME1. Have you ever used methamphetamine (crystal meth, ice)?

	1. Yes	
	2. No	(Skip to AF1**)**

ME2. How old were you when you first used methamphetamine?

	Years old

ME3. When was the first time you used methamphetamine?

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

ME4. Have you used methamphetamine in the past <u>12 months</u>?

	1. Yes	
	2. No	(Skip to AF1**)**

ME5. How often have you used methamphetamine?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

ME6. Have you used methamphetamine in the past <u>30 days</u>?

	1. Yes
	2. No

AMPHETAMINE (AF)

AF1. Have you ever used amphetamine (fet, speed)?

	1. Yes	
	2. No	(Skip to HE1**)**

AF2. How old were you when you first used amphetamine?

	Years old

AF3. When was the first time you used amphetamine?

	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

AF4. Have you used methamphetamine in the past <u>12 months</u>?

	1. Yes	
	2. No	(Skip to HE1**)**

AF5. How often did you use amphetamine?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

AF6. Have you used amphetamine in the past <u>30 days</u>?

	1. Yes
	2. No

HEROIN (HE)

HE1. Have you ever used heroin?

	1. Yes	
	2. No	(Skip to KE1**)**

HE2. How old were you when you first used heroin?

	Years old

HE3. When was the first time you used heroin?

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

HE4. Have you used heroin in the past <u>12 months</u>?

	1. Yes	
	2. No	(Skip to KE1**)**

HE5. How often did you use heroin?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

HE6. ¿Have you used heroin in the past <u>30 days</u>?

	1. Yes	
	2. No	(Skip to KE1**)**

HE7. How many days in the past 30 days did you use heroin?

	Number of days (from 0 to 30)

HE8. And on those days when you used heroin, how many doses did you use in a single day (maximum)?

	Number of doses

HE9. How (in what form) did you use heroin?

	1. Smoked
	2. Injected
	3. Smoked and injected
	4. Other
	9. Every day

KETAMINE (KE)

KE1. Have you ever used ketamine?

	1. Yes	
	2. No	(Skip to OD1**)**

KE2. How old were you when you first used ketamine?

	Years old

KE3. When was the first time you used ketamine?

	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

KE4. Have you used ketamine in the past <u>12 months</u>?

	1. Yes	
	2. No	(Skip to OD1**)**

KE5. How often did you use ketamine?

	1. Only once
	2. Several times during the past 12 months
	3. Several times a month
	4. Several times a week
	5. Every day

KE6. Have you used ketamine in the past <u>30 days</u>?

	1. Yes
	2. No

OTHER DRUGS

When was the last time you used any of these substances?	Never used	More than 1 month ago but less than 1 year ago	More than 1 year ago	In the past month
	(1)	(2)	(3)	(4)
OD1. Opium	○	○	○	○
OD2. GHB (Liquid X, liquid ecstasy, Georgia homeboy, Oop, Gamma-oh, grievous bodily harm, Mils, "G", Liquid G, Fantasia)	○	○	○	○
OD3. Synthetic cannabinoids (Synthetic marijuana, Spice, K2, Joker, Black Mamba, Kush or Kronic)	○	○	○	○
OD4. Synthetic Cathinones (bath salts)	○	○	○	○
OD5. Aminoindanes (MDAI gold)	○	○	○	○
OD6. Phencyclidine (PCP, angel dust, embalming fluid, hog, killer weed, love boat, ozone, peace pill, super weed, rocket fuel, estrafalaria)	○	○	○	○
OD7. Phenethylamines (Europa, 4-FMP, RDJ, 4-MMA, Methyl-MA, 2C-C-NBOMe, Bomb, N-Bomb N, 251, Nexus, 2C-E and Blue mystic)	○	○	○	○
OD8. Piperazines (BZP, mCPP, A2, Legal X and Pep X).	○	○	○	○
OD9. Hallucinogenic plants (Floripondio, angel's trumpet, 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).	○	○	○	○
OD10. Anabolic steroids	○	○	○	○
OD11. Caffeine products (caffeine pills, energy drinks, powdered caffeine**)**	○	○	○	○
OD12. Lean (Sprite mix, cough syrup and pastilles – also called purple drank o sizzurp)	○	○	○	○

MODULE 3: PERCEIVED RISK (PR) AND FACTORS ASSOCIATED (FA) WITH DRUG USE

PR1. Have you seen students using any of the following drugs on your campus?

	Never	From time to time	Often
1. Marijuana (weed, grass, pot)	○	○	○
2. Synthetic marijuana (Spice, K2, Bliss, Kronic, etc.)	○	○	○
3. Cocaine	○	○	○
4. Cocaine base paste	○	○	○
5. Heroin	○	○	○
6. Ecstasy	○	○	○
7. Amphetamines (speed)	○	○	○
8. Methamphetamines (meth, crystal meth, etc.)	○	○	○
9. Ketamine ("El key", vitamin K)	○	○	○
10. LSD (acid, blotter, mellow yellow)	○	○	○
11. Poppers	○	○	○
12. Inhalants (such as deodorants, glue, paint thinner, toluene, benzene, gasoline, petrol, etc.)	○	○	○
13. Hallucinogenic plants (peyote, ayahuasca or similar)	○	○	○

PR2. Do any of your family members use drugs such as marijuana, cocaine, etc.?

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

PR3. Do you have friends who use illicit drugs such as marijuana, cocaine, etc.?

1. None
2. A few
3. Some
4. Several
5. Don't know/no opinion

PR4. ¿Have you ever had the opportunity to try a drug such as marijuana, cocaine or other illicit substances?

1. Never
2. Once
3. Two or more times

PR5. Did you ever feel curious about trying an illicit drug?

1. Never
2. Yes, once
3. Yes, several times

PR6. If you had the opportunity, would you try a drug such as marijuana, cocaine or another illicit substance?

1. I've already tried
2. No, I wouldn't
3. I might
4. Yes, I would
5. Don't know/no opinion

PR7. Do you think that you will use an illicit drug in the near future?

1. I've already used
2. No
3. I might
4. Yes
5. Don't know/no opinion

PR8. In your opinion, what kind of risk a person runs when she or he does any of the following things?	No risk	Slight or moderate risk	High risk	Don't know
1. Smoking one or more packs of cigarettes a day	○	○	○	○
2. Drinking five or more drinks of alcohol a day	○	○	○	○
3. Taking medication without a doctor's prescription frequently	○	○	○	○
4. Smoking marijuana sometimes	○	○	○	○
5. Smoking marijuana often	○	○	○	○
6. Snorting cocaine sometimes	○	○	○	○
7. Snorting cocaine often	○	○	○	○
8. Using heroin sometimes	○	○	○	○
9. Using heroin often	○	○	○	○
10. Smoking cocaine base paste sometimes	○	○	○	○
11. Smoking cocaine base paste often	○	○	○	○
12. Using ecstasy sometimes	○	○	○	○
13. Using ecstasy often	○	○	○	○
14. Using methamphetamine sometimes	○	○	○	○
15. Using methamphetamine often	○	○	○	○
16. Using LSD sometimes	○	○	○	○
17. Using LSD often	○	○	○	○
18. Using Ketamine sometimes	○	○	○	○
19. Using Ketamine often	○	○	○	○

EN9. How easy or difficult would it be for you to obtain any of the following drugs?	Easy	Difficult	Would not be able to obtain	Don't know

1. Marijuana (weed, pot)	○	○	○	○
2. Synthetic marijuana (Spice, K2, Bliss, Kronic, etc.)	○	○	○	○
3. Cocaine	○	○	○	○
4. Cocaine base paste	○	○	○	○
5. Heroin	○	○	○	○
6. Ecstasy	○	○	○	○
7. Amphetamine (speed)	○	○	○	○
8. Methamphetamine (meth, crystal meth, etc.)	○	○	○	○
9. Ketamine ("El key", vitamin K)	○	○	○	○
10. LSD (acid)	○	○	○	○
11. Poppers	○	○	○	○
12. Inhalants (such as deodorants, glue, paint thinner, toluene, benzene, gasoline, petrol, etc.)	○	○	○	○
13. Hallucinogenic plants (peyote, ayahuasca or similar)	○	○	○	○

EN10 Have you ever been offered drugs such as marijuana, cocaine, cocaine base paste, ecstasy, amphetamine-type stimulants or other illicit substance, to try or to buy?

1. Yes
2. No (Skip to PP1)

EN11. When was the last time you were offered any of these drugs, to try or to buy?

	During the past 30 days	More than 1 month ago but less than 1 year ago	More than 1 year ago	Never offered to me
a) Marijuana (weed, pot)	○	○	○	○
b) Synthetic marijuana (Spice, K2, Bliss, Kronic, etc.)	○	○	○	○
c) Cocaine	○	○	○	○
d) Cocaine base paste	○	○	○	○
e) Heroin	○	○	○	○
f) Ecstasy	○	○	○	○
g) Amphetamine (speed)	○	○	○	○
h) Methamphetamine (meth, crystal meth, etc.)	○	○	○	○
i) Ketamine ("El key", vitamin K)	○	○	○	○
j) LSD (acid)	○	○	○	○
k) Poppers	○	○	○	○
l) Inhalants (such as deodorants, glue, paint thinner, toluene, benzene, gasoline, petrol, etc.)	○	○	○	○
m) Hallucinogenic plants (peyote, ayahuasca or similar)	○	○	○	○

MODULE 4: PREVENTION AND TREATMENT PROGRAMS (PP)

PP1. When you were in secondary school, were you given information, courses or workshops on prevention of drug use?

1. Yes
2. No (Skip to PP4)
3. Don't remember (Skip to PP4)

PP2. How often were you given information about drug use when you were in secondary school?

1. Only once
2. A few times
3. Several times
4. Many times
5. Don't remember

PP3. How would you describe the information you were given on drug use when you were in secondary school?

1. Very good
2. Good
3. Poor
4. Very poor
5. Don't remember

PP4. When you were in secondary school, did you receive any formal program of drug use prevention? ("formal" means as part of your regular schooling or classes)

1. Yes
2. No (Skip to PP6)
3. Don't remember (Skip to PP6)

PP5. How would you describe the drug prevention programs you received when you were in secondary school? Mark all that apply

1. Interesting
2. Alarmist, exaggerated
3. Boring
4. Incorrect
5. Incomplete
6. Don't remember

PP6. In your university, have you been given information, courses or workshops on drug use?

1. Yes
2. No (Skip to PP9)

PP7. In your university, how often have you been given information, courses or workshops on drug use?

1. Only once
2. A few times
3. Several times
4. Muchas veces

PP8. How would you describe the information on drug use that you received in your university?

1. Very good
2. Good
3. Not bad
4. Poor
5. Very poor

PP9. In your university, have you received any formal program of drug use prevention? ("Formal" means as part of your normal university course)

1. Yes
2. No (Skip to PP11)
3. Don't remember (Skip to PP11)

PP10. Which agencies or institutions conducted these drug prevention events?

1. Drug Commission
2. Health Ministry
3. Foundation or NGO
4. The university's student welfare department
5. Student union or association
6. Other

PP11. Do you think that the university should address the problem of drug use by systematically offering prevention events on an ongoing basis?

1. Yes
2. Perhaps
3. No

PP12. Do you think that the university should have a specific policy on preventing drug use by university students?

1. Yes
2. Perhaps
3. No (Skip to PP14)

PP13. Should this policy include alcohol?

1. Yes
2. Perhaps
3. No

PP14. Have you ever received any kind of treatment for drug or alcohol abuse? Do not include treatment to stop smoking cigarettes

1. Yes, for alcohol
2. Yes, for drugs
3. Yes, for alcohol and drugs
4. No, I've never been in treatment (Skip to PP16)
5. I don't use drugs (Skip to PP18)

PP15. In the past 12 months, have you received any kind of treatment for alcohol or drug abuse? Do not include treatment to stop smoking cigarettes

1. Yes, for alcohol
2. Yes, for other drugs
3. Yes, for alcohol and other drugs
4. No

PP16. In the past <u>12 months</u>, have you felt the need to receive any kind of treatment to cut down on or stop using alcohol or drugs?

1. Yes, for alcohol
2. Yes, for drugs
3. Yes, for alcohol and drugs
4. I'm afraid to ask for help
5. I don't need any kind of help

PP17. If you use alcohol or other drugs, have you seriously considered making an effort voluntarily to moderate your use or stop entirely?	**PP18. If you had ever tried marijuana, would you have said so on this questionnaire?**
1. Yes 2. Perhaps 3. No	1. Yes, I've just said so 2. Probably yes 3. Probably not 4. Definitely would not say so
PP19. If you had ever tried cocaine, would you have said so on this questionnaire?	
1. Yes, I've just said so 2. Probably yes 3. Probably not 4. Definitely would not say so	

PP20. Would you like to add anything? Maximum of 500 characters	
PP21. I hereby declare that I am providing this information on a voluntary basis	
1. Yes	2. No

THANK YOU VERY MUCH FOR YOUR TIME
The information that you have provided is absolutely confidential and will be used for statistical purposes only.

SEND

6. STATISTICAL ANALYSIS

Two basic points should be borne in mind when performing a statistical analysis of the data collected in the study, namely, the objectives of the study, and the results that will be communicated by a variety of means. An ***analysis plan*** must therefore be prepared.

In general, the analysis plan 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:

6.1 DESCRIPTION OF THE SAMPLE

The sample (including the expansion to the target population) is described by means of general variables such as *sex*, *age, and type of university* or other important variables included in the study. Thus, for example, if we consider only sex and age, we have a table such as the one below:

Table 1: Distribution of the sample by sex and age, and population represented

Variables	Sample size	Population represented	%
Sex			
Male			
Female			
Age			
18 or younger			
19-20			
21-22			
23-24			
25 or older			
TOTAL	**n**	**N**	**100**

The column "*Sample size*" describes the number of cases in the *effective sample* (having eliminated the cases that were present but that were considered not valid for inclusion in the analysis, as explained below.) Thus, **n** represents the total number of cases in the sample that are considered to be **valid** for the study.

The column "*Population represented*" corresponds to the number of people in the population classified according to the corresponding variables, where **N** is the total number of university 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 these concepts 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 planning 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 that are 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 2,000 cases in each university selected for the study (corrected for possible rejections or other contingences.) The students are selected on the basis of this sample size. It may be that the student cannot be contacted, or if contacted, he or she decides not to participate. Therefore, even though provision may have been made for certain contingencies, the number of successful interviews may be fewer than planned—1,100, for example. We now have a “second” sample size.

b) **Exclusion criteria:** there are reasons why some questionnaires should be excluded from the analysis: for example, if the respondent answered only the demographic questions, and then refused to continue with the questionnaire and did not answer the minimum number of questions needed. It is important that those directing the study, that is, the professional staff of the National Observatory on Drugs and the Inter-American Observatory on Drugs, decide in advance on the **exclusion criteria** that will be used to remove cases from the analysis. Having performed this “cleaning,” we obtain the ***valid cases*** that make up the **final sample** or the **effective sample** for statistical analysis. Again, let us suppose that in the previous example, we have 1,000 valid cases.

c) **Expansion factor**: as noted in Section 3.3 and as explained more fully in Annex 2, there are different sampling designs. A survey of university students household survey (Spanish says “estudio en población general) in particular is what is called “complex sampling.” This involves different stages in the process of selecting the students who will constitute the sample, and at each stage, the **probability of selection** is defined for each individual in the population. According to some probability theories, the final probability that a student will become part of the sample consists of multiplying probabilities at each stage of the sampling process. Mathematically speaking, ***the expansion factor associated with an individual in the sample corresponds to the reciprocal of the probability of selection of that***

individual. This means that the result ***represents a particular number of students in the population***. Each individual in the effective sample will be associated with a particular expansion factor. The **sum** of the expansion factors for the individuals in the effective sample will give the total number of the university student population in the country (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. First, it should be noted that this type of study, which is based on samples, only allows for **estimates** of what occurs in the population under study; it does not make 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—and this must be stated explicitly when describing the estimates. Fortunately, the degree of error can also be estimated on the basis of the sample data, which should be presented together with the estimates of the indicators. The way of presenting these errors is by means of what are called ***standard errors*** of the estimate, which are used to build the ***Confidence Intervals*** for the estimate. The population indicators (unknown) are termed **parameters**, which is what we are trying to estimate by means of a sampling process. We therefore have three concepts in the estimation process:

6.2 ESTIMATING INDICATORS ON SUBSTANCE USE

1. The **parameter** to be estimated (for example, the prevalence of past month use of alcohol in the country's university population),
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 the standpoint of **SIDUC, 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 suppose we have a study in a country where the population represented is 140,000 university students; five universities are selected, and the effective sample is 5,339 students. By correctly using the expansion factors and the sample design (further details in Annex 2,) the results of the study show that 57.9% of the students reported that they had used an alcoholic beverage in the past month, with a standard error of 1.7%. Thus, the prevalence of alcohol use in the past month is 57.9%, with a 95% Confidence Interval of between 53.0% and 62.7%. The 95% Confidence Interval to estimate a proportion (expressed as a percentage) of a population is expressed as:

p-1,96*ee(p) ; p+1,96*ee(p) (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/accuracy** of the estimate.

Thus, a Confidence Interval may be expressed as:

p-d ; p+d (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	57.9	53.0 – 62.7

As well as reporting the overall prevalence (55.3% in this case), it will always be necessary to disaggregate that indicator by at least sex and age group. This information is given in the table below for the example:

Table 3: Past month prevalence of alcohol use and 95% Confidence Intervals, by sex and age

Variables		Prevalence (%)	95% CI
Sex	Male	59.8	56.3 – 63.2
	Female	56.2	49.3 – 63.1
Age	18 and younger	45.6	39.2 – 52.1
	19-20	59.0	53.3 – 64.8
	21-22	65.5	59.6 - 71.3
	23-24	63.7	58.3 – 69.0
	25 and older	59.8	56.1 – 63.5
Total		**57.9**	**53.0 – 62.7**

As we advance further with the analysis of the data, it may often become necessary to develop indicators broken down by other categories. For example, it might be of interest to estimate and compare indicators on drug use by type of university, or according to the perceived risk or harm of substance use, or according to the perception of ease of access to substances, and so on. It is essential 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 questions are discussed again in Annexes 1 and 2.

6.3 COMPARISON OF INDICATORS

A third area of interest in the analysis process is the comparison of indicators. Thus, for example, depending on the objectives of the study, we want to find out whether there are differences in the prevalence of past month alcohol use between males and females or between people of different ages. Similarly, as stated earlier, the analysis might focus on studying the relationship between substance use and other characteristics such as perceived risk, 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 respond to 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. It is ***undergraduate students[11] who are enrolled in the university in the year the study is conducted***.

Once the target population has been defined, the next question is how to access that population.

Sampling frame: As we said in Section 3.3, the best way of finding the students in the target population is by means of a **sampling of universities** as the first stage. This means having a list of the country's universities and their enrollment This ***up-to-date sampling frame of universities is the basis of a research study*** such as that described in this Protocol.

Having defined the target population and once the sampling frame of universities has been drawn up, a sample of the universities must be obtained.

The sampling technique used in this study is a **two-stage** cluster sampling. We use the following to describe these stages:

Let:

- **M** be the number of universities in the country.
- **m** be the number of universities in the sample, which are selected in the first stage.
- $\mathbf{N_i}$ be the number of students enrolled or registered in university i, where i=1….M
- **N** total number of students enrolled in the M universities, i.e., the size of the population under study.
- $\mathbf{n_i}$ the size of the sample in university i
- **n** = total size of the sample

First stage

Let us use the previous example to describe the selection of universities in the first stage of sampling:

A country has 10 universities, each having the number of undergraduate students shown in column 2 of Table A.1.1. Total enrollment in the country is 140,000 students. Column 3 represents the percentage of students in each university vis-à-vis the total enrollment in the country.

[11] There may be other ways of referring to students whose courses will lead to an academic degree or professional qualification.

Table A1.1: Distribution of students in 10 universities

University	Number of students	%
1	2	3
1	5,000	3.57
2	25,000	17.86
3	10,000	7.14
4	8,000	5.71
5	30,000	21.43
6	10,000	7.14
7	15,000	10.71
8	25,000	17.86
9	4,000	2.86
10	8,000	5.71
Total	**140,000**	**100**

Let us now suppose that it is decided to select a sample of 5 universities for the study. There are generally two options for selecting these 5 universities: the first is a selection with an **equal probability of selection**; this means that university 1 with 5,000 students has the same probability of being selected as university 5, which has 30,000 students. For this option, it is only necessary to make a random selection of 5 numbers from numbers 1 to 10.

The other option, which is the one used in this Protocol, is to make a selection with a **probability proportional to size (PPS) of the university**. Here, for example, university 5 with 30,000 students will have a probability of being in the sample six times greater than university 1 with 5,000 students.

More information therefore needs to be added to the preceding table. This is shown in the following table:

Table A1.2: Distribution of students in 10 universities, cumulative frequency and range

University	Number of students	%	Cumulative number of students	Range
1	2	3	4	5
1	5,000	3.57	5,000	1 – 5,000
2	25,000	17.86	30,000	5,001 – 30,000
3	10,000	7.14	40,000	30,001 – 40,000
4	8,000	5.71	48,000	40,001 – 48,000
5	30,000	21.43	78,000	48,001 – 78,000
6	10,000	7.14	88,000	78,001 – 88,000
7	15,000	10.71	103,000	88,001 – 103,000
8	25,000	17.86	128,000	103,001 – 128,000
9	4,000	2.86	132,000	128,001 – 132,000
10	8,000	5.71	140,000	132,001 – 140,000
Total	140,000	100		

Column 4 shows the cumulative number of students. The first value in this column represents the enrollment in university 1 (5,000); the second value is the sum of the enrollments in universities 1 and 2, that is, 5,000+25,000=30,000. The third value, 40,000, is the sum of the enrollments in universities is the sum of the 1, 2 and 3, and so on, up to the last value, 140,000, which is the total enrollment in the country.

Let us now imagine a long list of the country's students, numbered from 1 (first student in university No. 1) to 140,000 (last student in university No. 10.) In university 1, we have students from 1 to 5,000; in university 2, we have students from 5,001 to 30,000, and so on up to university 10, where we have students from 132,001 to 140,000. This is what is shown in column 5, called "Range."

Columns 1 and 2 are generated from the information gathered by the National Observatory on Drugs. Columns 3, 4 and 5 are constructed on the basis of the information in column 2.

In order to select the 5 universities with probability proportional to size (PPS), we need a list of random numbers in the range of students in the population, that is, numbers from 1 to 140,000. Tables of this kind are easy to generate, as shown below. If the first number selected is in the range of 1 to 5,000, then the first university selected will be Number 1. On the other hand, if the first number is in the range of 5,001 to 30,000, then it will be university 2 that is selected.

Returning to our example, a list of random numbers can easily be generated in an Excel spreadsheet, using the function ***RANDBETWEEN().*** The initial and final numbers are put inside the parentheses, in this case, 1 and 140,000. Placing the cursor in one of the cells of the spreadsheet, we obtain the first value by means of the function fx=RANDBETWEEN(1,140000); the cursor is then moved down to obtain as many random numbers as are wanted. For this example, column 6 gives the first 10 random numbers generated between 1 and 140,000.

Table A1.3: **Distribution of students in 10 universities, cumulative frequency, range and random numbers to select universities**

University	Number of students	%	Cumulative number of students	Range	Random numbers
1	2	3	4	5	6
1	5,000	3.57	5,000	1 to 5,000	26,999
2	25,000	17.86	30,000	5,001-30,000	81,118
3	10,000	7.14	40,000	30,001-40,000	6,640
4	8,000	5.71	48,000	40,001-48,000	98,374
5	30,000	21.43	78,000	48,001-78,000	91,968
6	10,000	7.14	88,000	78,001-88,000	71,346
7	15,000	10.71	103,000	88,001-103,000	126,429
8	25,000	17.86	128,000	103,001-128,000	52,344
9	4,000	2.86	132,000	128,001-132,000	15,312
10	8,000	5.71	140,000	132,001-140,000	60,253
Total	140,000	100			

Five universities are selected on the basis of the numbers in column 6. The first number selected, 26,999, is in the range (column 5) of between 5,001 and 30,000, and therefore university 2 is the first to be selected. The second random number selected is 81,118 which is in the range of 78,001 to 88,000, which corresponds to university 6; the third number (6,640) is in the range corresponding to the second university, but since it was already selected, we move to the next number, 98,374, which is in the range of university 7. We continue in this way until we select the 5 universities in the sample, which in this example are universities 2, 5, 6, 7 and 8. Note that these universities have the largest enrollment, and therefore have the highest probability of being selected. This does not mean, however, that a university with a small number of students cannot be selected, but it will have a lower probability of being selected than a university with a higher enrollment. Note that of the 10 random numbers given in Column 6, only 7 were used.

This means of selecting the universities with PPS assumes that there is a single list of universities. However, if the study objectives include gathering differentiated information about universities that are public and universities that are private, the methodology will require two independent lists, one of public universities and one of private universities. The procedure for each will be the same as above.

Having selected the universities, the students are then selected.

Second stage

The **second stage** is the selection of students in the **m** universities selected to make up the sample.

Let $\mathbf{n_i}$ be the size of the sample in i university selected. At this second stage, there are two options for deciding on the size of $\mathbf{n_i}$ and hence determining the probability of selection of the students in the universities selected. The first option is ***to determine a sample size that is independent in each university,*** as described below. The second option is to make a selection ***proportional to the size of the university***, i.e., to determine an overall sample size for the study and then assign the size of the sample in each university according to their weight in the total number of enrollments in the universities in the sample.

This Protocol uses the ***first alternative***, since we want an appropriate sample size in each university in the sample so that they can have the information on their own university and themselves analyze internally the information provided by their students. It is important to note that this approach does not alter the fact that the study is anonymous, and therefore the universities cannot match a response to a specific student.

Note that the overall size of the study in all of the universities that is obtained by means of the first alternative will be larger than the size that would have been obtained using the second option.

According to this criterion, i.e., selection of universities with PPS, we see that the probability of selecting student j in university i of the sample is:

$$\boldsymbol{P_{ij} = m * \frac{n_i}{N} \quad i = 1, 2 \ldots m \qquad j = 1, 2 \ldots . n_i} \tag{3}$$

where $\mathbf{n_i}$ is the size of the sample in i university, **m** is the number of universities in the sample, and **N** is the size of the population, that is, the enrollment of university students in the country. The expression above means that all of the students in the same university have the same probability of being in the sample; the probability changes from university to university.

Returning to the previous example, universities 2, 5, 6, 7 and 8 were selected for the sample. The next step is to obtain a complete list of students in each of these universities and then obtain a random sample in each of them.

Sample size

In order to select the students in each university selected, it is first necessary *to determine the size of the sample* in each of the universities selected. It will be recalled that the main purpose of the study is ***to estimate the prevalence*** of the use of a substance or group of substances. For example, we might say that this indicator is the *prevalence of the use of an illicit substance in the past 12 months*. The value of that indicator in the target population is called a **parameter**; the value obtained as a result of a random sample is called the **estimator**. Of course, the parameter is an unknown value, but is what we are trying to estimate by means of a sample, i.e., by means of a subset of elements of which the population is composed.

The elements that need to be considered in determining the sample sizes are described below (these are estimates at the level of the universities):

1. The first element has to do with the **sampling design or method** that will be used. As stated earlier, within each university selected in the first stage, a ***simple random sampling*** of students will be used. This means having a sampling frame of the enrollment, i.e., a list of students in the university, and then making a random selection of a particular number of them. The main feature of a simple random sampling is that ***each student in the population has a known probability of being in the sample and it (the probability) is the same for all of the students***. Thus, for example, in a university that was selected in the first stage, the enrollment is 10,000 students, and it has been decided that the sample in that university is 2,000, then all of the students have a probability of (2,000/10,000)=0.2.
   Generally speaking, if in a university of $\mathbf{N_i}$ students (population), a simple random sample of $\mathbf{n_i}$ students is selected, then using this sampling design, the probability that a student will be in the sample is $\mathbf{n_i/N_i}$.
   How to select the students? Let us assume that the students in the university are listed in some kind of order (enrollment number, or alphabetically by last name, for example) and that they are numbered from 1 to 10,000 (as in the previous example), then a computer program can be used to generate random numbers to select 2,000 from the population of 10,000. As described above, Excel offers this option using ***RANDBETWEEN(1,10000).*** The first 2,000 cases are selected from the list generated in this way.

2. The second factor looks at the **variability** in the population with respect to an indicator considered to be the most important. We might, for example, assume that the indicator of the greatest interest is the **prevalence of use of an illicit substance in the past 12 months**. Different magnitudes of the estimator will produce different sample sizes; in fact, the prevalence (in %) in the university population of one country may be 1% for that indicator, while in another country it may be 10% or perhaps 50%. If we call that prevalence **P**, then the variability will be the product of **P** and its complement, that is, **Q=100-P**, or **P*Q**. A P value of 50% means that the population is divided into

two equal parts, which means maximum variability. This variability decreases to the extent that P moves further away from 50%. The size of the sample is directly proportional to the variability, which means that the ***greater the variability, the larger the required sample will be***. In order to determine the size of the sample, it is necessary to have an approximation to the value of **P** in the population. This may be done by drawing on estimates obtained in previous studies or on results from other countries in a similar situation. In the event such alternatives are not available, it will be necessary to resort to the scenario that will produce the largest sample size, that is, to assume that P=50%.

3. A third component refers to the **precision/accuracy** of the estimates, which determines the width of the Confidence Interval. The greater the precision required for the estimates, the larger the size of the samples needed. For example, a Confidence Interval of between 10% and 50% is less informative that a Confidence Interval of 25%-35%, or of 29%-31%, and therefore will require a smaller size than the latter. The greater the precision desired for the estimate (i.e., a narrower Confidence Interval), ***the larger the sample size must be***. As defined earlier, precision d is **d=z*ee(p)**, where z represents the value of the normal distribution for a certain level of confidence, and ee(p) is the standard error of the estimate (which is the square root of the variance.)
4. The fourth factor to be borne in mind is related to the desired **level of confidence** in the estimate, which defines the probability that the Confidence Interval constructed for the parameter of interest will include that indicator. Based on the approximation to the normal distribution, in association with the sampling distribution of a proportion, and given that it is the 95% Confidence Intervals that are used most often, 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 also used instead of the approximation to the normal distribution, particularly when working with statistical packages. One of these distributions is the binomial distribution. ***The greater the confidence in the estimate, the larger will be the size of the sample.***
5. Lastly, we must consider the size of the population, in this case, the student enrollment in the university.

Let us suppose that the main concern of the study is to determine the prevalence of the use of an illicit substance in the country's university population. Let us call that prevalence ***P*** for a specific university. For the purposes of explaining the determination of the sample size in a **simple random sample,** it is necessary to define the variance of the estimate of ***P*** in each university, expressed as follows:

$$V_{mas}(p) = \frac{P(1-P)}{n_i} \left(\frac{N_i - n_i}{N_i - 1}\right) \quad (3)$$

where **N_i is the size of the population of i university in the sample, and n_i is the size of the sample** for that university.

Let us again consider the definition of precision that we saw above: **d=z*ee(p).** For practical purposes, we shall replace the parameter **P** with the estimator **p**. Then, for the particular university **i** we have:

$$d = z * \sqrt{\frac{p(1-p)}{n_i} * \left(\frac{N_i - n_i}{N_i - 1}\right)} \quad (4)$$

If the size of the population N_i is large, then an initial approximation to the size of the sample in that university is:

$$n_o = \frac{z^2 * p * (1-p)}{d^2} \quad (5)$$

Correcting for the size of the population, the final sample size in each university is given by:

$$n_i = \frac{n_0}{1 + ({n_0}/{N_i})} \qquad i = 1, 2 \ldots . m \quad (6)$$

To illustrate how the size of the sample is determined, let us consider the previous example, where the first university selected was Number 2, with an enrollment of 25,000 students, or N_2 = 25,000. We want to estimate the prevalence of the use of any illicit drug, and we have information from previous studies in which the estimated prevalence was 25%, or p=25%. It is decided that the precision/accuracy will not be greater than 10% of the estimator, i.e., 10% of 25%, or d=2.5%, and that the estimate will use a 95% Confidence Level, that is, z=1.96.

With this in mind, the initial sample size given by equation (5) is expressed as:

$$n_o = \frac{1,96^2 * 25 * (1-25)}{2,5^2} = 1.152$$

and in accordance with (6), the final sample size is:

$$n = \frac{1.152}{1 + ({1.152}/{25.000})} = 1.102$$

Note that in expression (3), the same value of **P** has been used for all of the universities in the sample, meaning a national global indicator for this population that is common to all of the country's universities. Bearing this in mind, the table below shows the initial and final sample sizes (adjusted by size of the university) for the 5 universities selected in the first stage of the example.

Table A1.4: Sample sizes in 5 universities in the sample

University	Number of students	Initial sample size	Final sample size
2	25,000	1,152	1,102
5	30,000	1,152	1,110
6	10,000	1,152	1,033
7	15,000	1,152	1,070
8	25,000	1,152	1,102
Total	105,000	5,762	5,417

The final sample size for each university is the number needed to carry out the analysis under the conditions defined. However, OID's experience with studies of university students in a number of countries has shown that a not insignificant proportion of students do not respond to the invitation to participate in the study; therefore, in order to obtain the desired size according to expressions (5) and (6), an oversampling of 100% should be planned.

Thus, for example, in university 2 on Table A1.4, in order to obtain a response from the 1,102 students in the sample, a sample of 2,204 students should be selected. Thus, the number of students selected in each university, which we will call **n´** is:

$$n_i{}' = 2 * \frac{n_0}{1 + \left(n_0/N_i\right)} \qquad (7)$$

Lastly, the total sample size for the study in the country will be the sum of the sizes determined for each university.

ANNEX 2: STATISTICAL ANALYSIS

This Annex will go into some detail about the ***statistical analysis*** of the results of the survey of university students, conducted by means of a two-stage sampling. As stated earlier, the analyses ***must respond to the objectives of the study***, and together are central to preparation of the relevant **reports**.

Analysis of the results of the study requires the following:

1. Human and technology resources

The statistical analysis requires **human resources** who have the necessary training and know-how and who for preference will have participated in the entire process, starting with the planning of the study. **Statistical software** will be needed to perform the ***analysis of complex samples***. Such software includes SPSS,[12] STATA,[13] SAS,[14] and R.[15] The first three require a license but R is available without cost. A statistical analysis cannot be carried out without appropriate computer software that can meet the demands of the methodology.

2. Databases

As stated earlier, the students selected for the study will respond to the questionnaire available on the server agreed by the National Drug Observatory and the OID. Once the response process has been completed, all of the students' responses become part of a database in SPSS, and the data analysis process can begin.

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 on the answers to the questionnaire, without adjustment.

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

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 three is explained below:

Cleaning the database

This means essentially deciding on certain criteria for inclusion, that is, deciding on the reasons why certain questionnaires should be excluded from the database, and therefore which should be retained. The people directing the study must decide on these criteria in advance. Given that the information-gathering process is via an on-line questionnaire, the questions must be answered in the sequence in which they appear in the questionnaire.

The criterion to be used in this Protocol is that all those cases that have complete answers for the Alcohol module should be included; in other words, that the following modules are complete: General Information, Tobacco, Electronic Cigarettes and Alcohol.

As a result, *observations or lines will be deleted* from the database.

a) Generating new variables in the database

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

For the purposes of the questionnaire, 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 are variables of "first use" of any substance.

Regarding "first use", we are interested in finding out about the point in time when people 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 30 days
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.

AUDIT: It may also be the case that a variable will need to be represented by more than one question on the questionnaire. Psychometric scales and other ad hoc scales are typical examples of this. For example, in 1992, the World Health Organization (WHO) developed the **A**lcohol **U**se **D**isorders **I**dentification **T**est (**AUDIT).**[16] Questions **AU1** through **AU10** on the questionnaire in Section 5 of this Protocol are taken from the AUDIT. These questions are asked of those who responded affirmatively to the question on use of alcohol in the past year. The first eight questions have five possible answers, each with a score of 0, 1, 2, 3 or 4. The other two questions offer only three possible answers, each with scores of 0, 2 or 4.

The **sum** of the results of the ten questions gives a score of between **0 and 40**. The authors recommend that scores of 8 or more are "***indicators of hazardous and harmful alcohol use."*** They add that "since the effects of alcohol vary with average body weight and differences in metabolism, establishing the cut-off point for all women and men over age 65 one point lower at a score of 7 will increase sensitivity for these population groups."

On the basis of the ten questions, we develop a new variable, which we will call the ***AUDIT score*** (or sum) ranging from 0 to 40. We can then develop a further variable that defines ***hazardous and harmful alcohol use*** (among those who said they had used alcohol in the past year) with the following values:

- 1 if the AUDIT score for **males** is 8 or higher, and 7 or higher for **females**
- 0 if that is not the case.

For those persons who did not drink alcohol in the past year, the value of this variable in the database is considered to be a missing value.

Therefore, for alcohol, we add **two new variables** to the original database: *AUDIT* score with values of 0-40, and *hazardous and harmful alcohol use* with values of 0 and 1 (and the missing value if the person did not drink alcohol in the past year.)

The WHO document also enables us to create other classifications of alcohol users: for example, a score of 1 or more to questions AU4-AU6 implies ***signs of dependence***. This leads us to consider a new variable, "signs of dependence" with a value of 1 if the score for

[16] https://auditscreen.org/cmsb/uploads/1992-audit-the-alcohol-use-disorders-identification-test-guidelines-for-use-in-primary-health-care-geneva-world-health-organization-1992.pdf

any of the three questions is 1 or more, and 0 if that is not the case. Again, the variable is considered a missing value if the person did not drink alcohol in the past year.

Another case we find frequently concerns questions designed to find out about the risk that people perceive of using certain substances. For example, in the case of the question “What kind of risk do you think someone runs if he or she does any of the following things?” the possible answers are:

1.- No risk
2.- Slight risk
3.- Moderate risk
4.- High risk
9.- Don’t know

We might be interested in examining only “high risk”. We could therefore create a new variable for this situation; let us call this variable **r_mar**, which has a value of 1 if the person gave response No. 4 to the original question, and otherwise has the value 0. There are many similar cases that can lead to development of new variables based on those in the original questionnaire.

Thus, the *second database* contains all the variables associated with the questions on the questionnaire, plus all those variables constructed on the basis of the original questions that may be necessary for subsequent analyses. It should also contain a variable that is called the ***expansion factor***, described below.

b) Expansion factor

Because of the importance of the issue, determination of the expansion factors is discussed here in a special section.

The **expansion factor** is a variable that is the ***inverse of the probability of selection of each person,*** and it is theoretically different in each case. 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 same expert who worked on the design of the sample, and should be entered into the database.

Expression (3) in Annex 1 gave the probability of selection of each student in the sample in a two-stage sampling with a probability proportionate to the size of the universities in the first stage, and a simple random sample of students in the second stage. From there, we determine the expansion factor:

$$f_{ij} = \frac{N}{m} * \frac{1}{n'_{ki}} \quad i = 1, 2 \ldots m \qquad j = 1, 2 \ldots . n_i \tag{8}$$

where:

N = size of the population

m = number of universities in the sample

n´ki = number of students in the sample of university i that meet the criteria for inclusion.

Note that within a university in the sample, the expansion factor is the same for all of the students in the sample. The factor N/m is constant for all of the students selected, independently of the university.

Let us illustrate this with the example given in Table A1.4. The constant factor N/m is the quotient of the size of the population and the number of universities in the sample, i.e., 140,000/5=28,000.

The table below gives the sample sizes (including the oversampling) of the 5 universities selected, and also a hypothetical list with the effective sample, that is, those who responded and who also meet the criteria for inclusion, and the expansion factors.

Table A2.1: Sample sizes, expansion factors in 5 universities in the sample

University	Number of students	Size of sample with oversampling	Size of effective sample **n´k**	Expansion factor **(N/m)*(1/n´k)**	Sum of expansion factors
2	25,000	2,204	980	28,57	28,000
5	30,000	2,220	1,021	27,42	28,000
6	10,000	2,066	998	28,06	28,000
7	15,000	2,140	1,200	23,33	28,000
8	25,000	2,204	1,140	24,56	28,000
Total	105,000	10,834	5,339		140,000

Note that in each university, the sum of the expansion factors is equal to the constant factor defined earlier (N/m). This is due to the fact of having selected the universities in the first stage with probabilities proportionate to their size. In addition, the sum of the expansion factors for all of the universities gives the size of the population, that is, the total enrollment in the country's 10 universities.

As we said earlier, the expansion **factor** variable is an essential part of what we have called the second database, in that it ensures that ***unbiased estimates of the parameters*** defined in the study objectives can be obtained.

3. Statistical analysis

With these conditions in place, we can move to the analysis itself. This will require an analysis plan that basically covers three areas that complement each other but that 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, age, and type of university. But there will also be a description of other variables of interest, such as perception of risk or harm of substance use, substances offered, user profiles, etc.
- ✓ Construction of **confidence intervals** for the estimates of the most important indicators such as prevalence or incidence of drug use, both nationally and also by sex and age.
- ✓ 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, age, or type of university.

To better explain the analyses that will be described below, the next table shows the variables in a database containing responses from 5,339 students. We then give a list with a small part of the database (only 23 cases and 13 variables), which we will use for subsequent examples.

Table A2.2 Description of variables in a database of 5,339 cases

Name of variable	Description
UNIVERSIDAD	Identification of the university
TIPO_UNIV	Type of university (public/private)
SEXO	Sex (1=male, 2=female)
EDAD	Age
PTAB_MES	Past month prevalence of tobacco
PBEB_MES	Past month prevalence of alcohol
AUDIT	AUDIT score
PCUA_VID	Lifetime prevalence of any drug
PCUA_ANO	Past year prevalence of any drug
EXP	Expansion factor
RIESGO_MAR	Perceived risk of frequent use of marijuana, 1=no risk 2=slight risk 3=moderate risk 4=high risk
FACIL_MAR	Factor conseguir de expansión
P50	First use of alcohol

	B	C	D	E	F	G	H	I	J	K	L	M
1	TIPO_UNIV	SEXO	EDAD	PTAB_MES	PBEB_MES	AUDIT	PCUA_VID	PMAR_ANO	EXP	RIESGO_MAR	FACIL_MAR	P50
2	0	2	22	0	0	0	0	0	28,57	2		3
3	0	1	19	1	1	1	1	1	28,57	3	1	3
4	0	1	23	1	1	0	1	1	28,57	3	1	3
5	0	2	17	1	0	0	0	0	28,57	1	0	2
6	0	2	20	0	1	0	1	0	28,57	3	1	1
7	0	2	18	0	0	0	1	0	28,57	3	0	2
8	0	2	18	0	1	0	0	0	28,57	2	1	3
9	0	2	21	0	1	0	1	1	28,57	2	1	3
10	0	2	24	0	0	0	0	0	28,57	2	1	3
11	0	1	18	0	1	1	1	1	28,57	3	1	3
12	0	2	22	0	1	0	0	0	28,57	2	1	3
13	0	2	26	0	0		0	0	28,57	3	1	3
14	0	1	22	0	1	0	1	1	28,57	3	1	2
15	1	2	17	0	0	0	0	0	23,33	3	0	1
16	1	2	19	0	0	0	0	0	23,33	3	0	1
17	1	2	22	0	0	0	0	0	23,33	2		3
18	0	1	24	0	0		0	0	23,33	3	1	3
19	1	2	19	0	0		0	0	23,33	2	0	3
20	1	2	19	1	1	1	1	1	23,33	3	0	3
21	0	2	23	0	0	0	1	0	23,33	2	1	1
22	1	2	20	0	0	0	0	0	23,33	3	0	1
23	0	2	22	0	1	0	0	0	23,33	1	0	3
24	1	2	19	1	1	0	1	1	23,33	2	1	1

What we see in these 23 cases is that 13 of them come from university 2 and the other 10 are from university 7. The expansion factor is 28.57 for all the students in the university, but the factor is 23.33 for the other students.

When we refer to the **description** of some variables, we think, for example, of ***sex and age***. On the other hand, for variables such as ***past month use of alcohol*** *(pbeb_mes) or* ***marijuana use in the past year*** *(pmar_ano),* the **Confidence Interval** should be included in the analysis. Lastly, when we want to study the **association** between the ***perceived risk of the occasional use of marijuana*** *(riesgo_mar)* and ***use of marijuana in the past year*** *(pmar_ano)* we must construct some statistical models. The following sections will examine the three strategies for analysis.

The analyses must always bear in mind the objectives of the study
All of the analyses must by weighted by/for the expansion factor
For the purposes of a national report, we suggest focusing on sections 3.1 and 3.2, 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 look at the 23 cases in the previous list. The sum of the values in the last column and the expansion factors gives a total of 604.71, which corresponds to the population represented by this sub-sample of 23 students. We see in the database that there are 5 males and 18 females. If we do **not weight** for expansion factors, we would conclude that the distribution is 21.7% males and 78.3% 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 the one hand and of females on the other would give us 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 28.57; 28.57; 28.57; 28.57 and 23.33. The sum of these expansion factors is 137.61.

In the case of females, the expansion factors are 28.57 for 9 females and 23.33 for the remainder. The sum of these values is 467.1. The percentage of males and females are then calculated, resulting in 22.8% and 77.2% respectively, which are quite different from the values that would have been obtained if not weighted for the pertinent factors.

Formally speaking, let us recall that f_{ij} is the expansion factor for student j selected in university i.

For the purpose of describing the sample and for the next stages of the analysis, we shall assume that the principal variables of the study are of three possible types:

- ✓ **Binary variables**, namely, those that have only **two possible answers,** with values of 1 and 0. For example, *use or non-use* of substance, where the variable has the value 1 for substance use, and 0 for non-use (see variable *pmar_ano* in the previous example.)
- ✓ **Qualitative variables with more than two possible answers**, for example, perceived risk of drug use, which has five possible answers (No risk, slight risk, moderate risk, high risk, and don't know,) or other such as ease of access with three possible answers (easy, difficult and don't know.)
- ✓ **Quantitative variables**, such as age of the person, age of first use of alcohol, or number of days alcohol used in the past month.

How are the 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. Thus, for example, for the variable sex, 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 **mean**, **median**, **standard deviation** or different **percentiles**.

Let us look first at the case of a binary variable, such as *alcohol use in the past month* (pbeb_mes with values 1=yes, and 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*[17] as:

$$\boldsymbol{p} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{x_{ij}} * \boldsymbol{f_{ij}}}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{f_{ij}}} \qquad (9)$$

which is equal to:

$$\boldsymbol{p} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{x_{ij}} * \boldsymbol{f_{ij}}}{\boldsymbol{N}} \qquad (10)$$

where **X** represents the variable *use of alcohol in the past month*, that is, **X**=pbeb_mes where:

$$\boldsymbol{X_{ij}} = \begin{cases} 1 & if\ the\ student\ selected\ reports\ having\ used\ alcohol\ in\ the\ past\ month, \\ 0 & if\ the\ contrary\ is\ true. \end{cases}$$

$\boldsymbol{f}_{ij}$ = *expansion factor of student j selected from university i in the sample*

Given that in certain cases, the variable X has the value 0 (when the person has not used alcohol in the past month), the numerator is only the *sum of the expansion factors of the respondents for whom X=1, that is, who said they had drunk alcohol* in the past month. But the denominator is *the sum of the expansion factors of all of the cases in the sample*, i.e., corresponds to the population represented.

The second case is analysis of a **quantitative variable**, such as age.

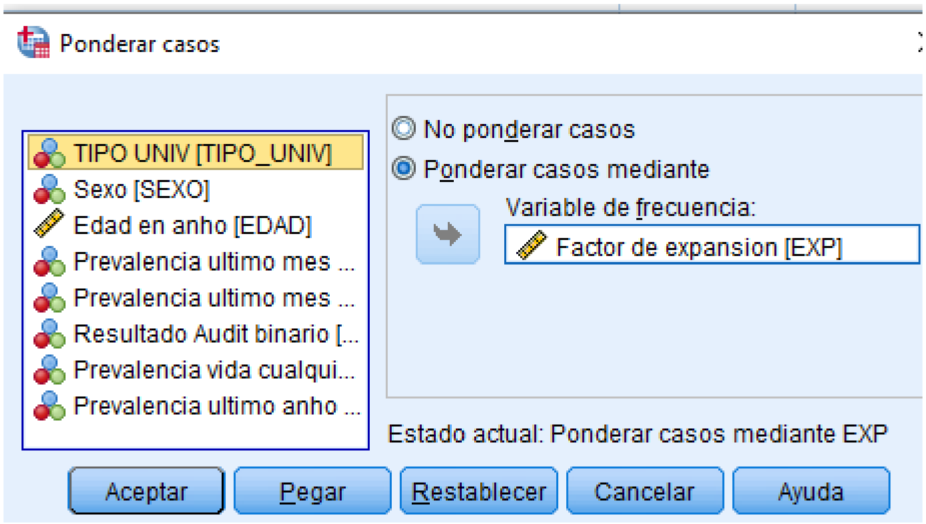
[17] P (upper case) is often used to denote the parameter, that is, the value at the population level, p (lower case) the estimator of P, i.e., the value obtained in the sample.

Formula (10) operates in the same way. The result is as follows, where $\bar{x}$ represents the **mean** age in the weighted sample (for the 23 cases shown in the previous table):

$$\bar{X} = \frac{22 * 28{,}57 + 19 * 28{,}57 + \cdots + 19 * 23{,}33}{28{,}57 + \cdots + 23{,}33} = \frac{12.473{,}22}{604{,}71} = 20{,}6 \text{ años}$$

Henceforth, we shall use the SPSS statistical software to illustrate the calculations of interest, and will go back to the complete database of 5,339 cases.

In order to **describe** some variables, in SPSS it is necessary only to **weight** the database for the corresponding **expansion factor**, and then calculate percentages or averages. Under *Data*, go down to *weight cases,* and a window like the one below will pop up. Select the variable that represents the expansion facto, ***EXP*** in this case, and press *Done*.



Let us suppose that we want to determine the distribution by sex of the population represented. In the main SPSS menu, select *Analyze*, then *Descriptive Statistics* and lastly *Frequencies*. Select the variable of interest (sex, in this case), and then *Done*. The result for the variable sex is given in Table A2.2, which shows us that the population represented is 140,000, with 46.6% male and 53.4% female.

Table A2.2. Distribution by sex of the population represented

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	Male	65206	46.6	46.6	46.6
	Female	74794	53.4	53.4	100.0
	Total	140000	100.0	100.0	

In the same way, we can estimate the prevalence of past month use of alcohol, but by doing this, *we only obtain the estimate*, but not the corresponding Confidence Interval, a subject that will be addressed below in Section 2.2 of this Annex. According to Table A2.3, 57.9% of the university population represented said they had used alcohol in the past month (the value "yes" in the variable pbeb_mes indicates a positive case.) This value is the isolated estimate of the prevalence in the population, but it is all that we are able to obtain by this route.

Table A2.2: Prevalence of past month use of alcohol (pbeb_mes)

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	No	59002	42.1	42.1	42.1
	Yes	80997	57.9	57.9	100.0
	Total	140000	100.0	100.0	

Using SPSS and the complete database, let us now analyze a **quantitative variable** such as age. Again, from the main SPSS menu, choose *Analyze*, then *Descriptive Statistics* and then *Descriptives.* Select the variable of interest, in this case age, and *Done*. The result is shown in Table A2.4.

Table A2.4: Description of age

	N	Minimum	Maximum	Mean
Age	140000	16	30	20.74
N valid (by list)	140000			

This table shows that the ages of the people in the sample range from 16 to 30, with a mean age of 20.7.

However, the mean is not the only indicator that can be used for quantitative variables; in fact, it is preferable to add four additional indicators: the 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 median, 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.*

Looking at the preceding data, we can obtain some percentiles in SPSS (using the option *Explore* instead of *Descriptives*) with the following results:

Table A2.4_1: Age percentiles

	Percentiles						
	5	10	25	50	75	90	95
Age	17	17	19	20	22	25	27

We can therefore say that 50% of the population represented is aged at most 20 (50th percentile), 25% are aged 19 or less (25th percentile), and 25% are aged 22 or more (75th percentile). We also know that 5% are aged 17 or less (5th percentile), and that 10% of the university population is aged 25 or more (90th percentile.)

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.

To illustrate this and also to discuss other issues that will be examined later in this Annex, we shall continue to use the previous database of 5,339 cases, which represent a university population of 140,000. First, using SPSS, we shall determine the mean age of students in three different situations.

a.- ***Without weighting for the expansion factors, i.e., assuming a simple random sampling of 5,339 students.***

Table A2.5: Description of mean age using simple random sampling

	N	mean	
	Statistical	Statistical	Standard error
Age	5339	20.72	,042
N valid (list)	5339		

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

Table A2.6: Description of mean age with expansion factors

	N	Median	
	Statistical	Statistical	Standard error
Age	5339	20.74	,008
N valid (list)	5339		

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

Table A2.7: Description of mean age using a complex sampling design (with expansion factors)

	Estimate	Standard error	95% Confidence Interval	
			Lower	Upper
Mean age	20.74	,177	20.25	21.23

We see that in the first case (Table A2.5), the mean age is 20.72. However, in the other two cases, (Tables A2.6 and A2.7) it is 20.74. The first estimate is **incorrect since it does not take into account the weighting factors**. The second and third estimates of the mean age are **correct**. The difference between these two latter cases lies in the way in which the standard errors are calculated. In the second, the standard error is very small (ee=0.008) since it takes into account the weightings as an enlargement of the cases in accordance with the expansion factor and therefore assuming the population as the sample itself, but does not take into account the sampling design. On the other hand, the result in Table A2.6 takes the sample design into consideration, and the weighting factors are considered the inverse of the probabilities of selection, and as a result, we have ee=0.177, which is much higher than the previous one. **This is the correct standard error** with a 95% Confidence Interval for the mean age, with limits of 20.25 and 21.23 years of age.

As we saw in the previous example, the mean of the age variable can be obtained by again using expression (10) but with a quantitative variable X. However, in order to properly calculate the standard error and therefore the correct Confidence Interval, a *statistical package such as SPSS or similar is needed to analyze complex samplings*. We shall return to this in point 2 below when we discuss the subject of Confidence Intervals in more detail.

Thus far, we have reviewed the procedures for estimating a *binary* variable (past month alcohol use and sex), as well as a quantitative variable (age).

Lastly, the third analysis refers to a **qualitative variable with 3 or more possible answers**, such as the variable that has to do with the first time the respondent drank an alcoholic beverage, variable P99 in the list above. As with a binary variable, the expansion factors of the individuals in each category should be added up, and then divided by the population represented, which would be fairly time-consuming if we use expression (10).

The following table shows the results from using SPSS:

Table A2.8: Percentage distribution of the expanded sample, by first use of alcohol

First use of alcohol	Frequency	Percentage	Valid percentage
Never used	11286	8.1	8.1
In the past 30 days	22101	15.8	15.8
More than 1 month ago but less than 1 year ago	27790	19.9	19.9
More than 1 year ago	78823	56.3	56.3
Total	140000	100.0	100.0

From the above, we can see that using expression (10), we can obtain estimates for a proportion (for example, *use of alcohol in the past month*) when the variable has values 0 and 1, and for an average (such as *age*)—these calculations can be done using an Excel spreadsheet. Similarly, but with somewhat more work, we can obtain a description of the frequency of a qualitative variable with two or more levels of response (for example, *first use of alcohol*). However, to pursue this analysis, a statistical package is necessary.

Special mention should be made of the indicators on **incidence** of drug use. As stated earlier, incidence measures the proportion of **new cases.** Thus, for example, the *incidence of alcohol use in the past year* corresponds to the percentage of people who used *alcohol for the **first time** during the past year, of those persons who had never used alcohol*. An important difference between the prevalence and incidence indicators is that the definition of the former uses **the entire population** as the denominator, while incidence considers only the **population exposed**, that is, those who had not previously used the substance under study (in the period defined).

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

- ✓ For **past year prevalence**, the question is:

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

  1. Yes
  2. No

- ✓ For **incidence in the past year and incidence in the past month**, the question is:

  When was the first time you drank an alcoholic beverage?

  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 us suppose that the answers to each question were as follows:

- ➢ **To estimate past year <u>prevalence</u>** (weighted cases):

Table A2.9: Percentage distribution of the question: Have you used an alcoholic beverage in the past month?

	Frequency	Percentage	Valid percentage	Cumulative percentage
No	59002	42.1	42.1	42.1
Yes	80997	57.9	57.9	100.0
Total	140000	100.0	100.0	

In the case of prevalence (Table A2.9), the result is direct: 80997/140000=0.579 (57.9%); that is, almost 58% of the university population said they had drunk an alcoholic beverage during the past month.

To estimate past year and past month <u>incidence</u> (also weighted cases), we shall use the results in Table A2.8 shown earlier. As to incidence, the calculation is somewhat more complicated. First, for ***incidence in the past year***, we must distinguish between those persons **who are exposed to** occurrence of the event during the time period indicated, that is, those who have not yet used alcohol in the past year, and those persons 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, 11,286 + 22,101 + 27,790 = 61,177 students. Excluded are the 78,823 students (weighted figures) who said they had drunk alcohol for the first time “more than 1 year ago,” that is, more than one year prior to the study, which is the subgroup in which the event of interest had already occurred. Thus, the **incidence of alcohol use during the past year** is those who used for the first time "during the past 30 days," plus those who said they had used for the first time "more than 1 month ago but less than 1 year ago," divided by the total number of persons exposed, that is to say, all the population represented less those who said that they had used alcohol for the first time "more than 1 year ago." That is:

$$Incidence\ alcohol\ past\ year = \frac{22{,}101 + 27{,}790}{140{,}000 - 78{,}823} = \frac{49{,}891}{61{,}177} = 0.816\ (81.6\%)$$

We proceed in the same way if we want to look at the ***incidence* of *alcohol use in the past month*** (instead of in the past year): the numerator includes those persons who said they used alcohol for the first time during the past 30 days, that is, 22,101 students. The denominator includes the entire population less those who had drunk alcohol prior to the period defined, i.e., prior to the 30 days prior to the study. These cases are: those who said they had used 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 alcohol use in the past month is expressed as:

$$Incidence\ alcohol\ past\ month = \frac{22{,}101}{140{,}000 - 78{,}823 - 27{,}790} \frac{22{,}101}{33{,}387} = 0.662\ (66.2\%)$$

In formal terms, expression (10) given earlier allows us to estimate any prevalence indicator, where in all cases, the denominator *will be expressed by the total of the population represented*.

Let us now look at an expression that gives us the **incidence** of the use of a substance in a **particular period of time**. The time periods of interest are essentially two: ***used for the first time in the past month,*** and ***used for the first time in the past year***. In a cross-sectional study such as the one discussed in the present Protocol, the question asked in the questionnaire to estimate the incidence at either period of time is the following:

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

1. Never used
2. During the past 30 days
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 (10) 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 A2.10: Outline for the percentage distribution of the question "When was the first time you used [NAME OF 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 shown below in the following expressions:

$$Past\ year\ incidence = \frac{B + C}{N - D} \tag{11}$$

and

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

When using computer programs such as those mentioned earlier, the recommendation is to generate missing values for those cases "not exposed," and then compute a proportion in each case, determining the corresponding Confidence Intervals using the study's sample design.

Thus, for example, for past year incidence, a new variable is developed, and missing values are defined in all of those cases in which the response was 4, i.e., in all those 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 those 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 A2.11: Outline for estimating *past year incidence*

Used NAME OF SUBSTANCE 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 way 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.

The operation is similar in the case of past month incidence. A new variable is defined with missing values if the answers were 3 or 4, and there will thus be a new number for exposed cases (that is, those who did not use the substance prior to the month prior to the study.) 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 A2.12: Outline for estimating *past month incidence*

First used [NAME OF SUBSTANCE] 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, and what was said above also applies in this case.

In short, based on the question "When was the **first time** you used NAME OF SUBSTANCE?" two new variables should be created to reflect past year and past month incidence, using expressions (11) and (12).

To illustrate this, let us again use the results of Table A2.8, where:

Table A2.13: Percentage distribution of the expanded sample, by first use of alcohol

First use of alcohol	Frequency	Percentage	Valid percentage
Never used	11286	8.1	8.1
During the past 30 days	22101	15.8	15.8
More than 1 month ago but less than 1 year ago	27790	19.9	19.9
More than 1 year ago	78823	56.3	56.3
Total	140000	100.0	100.0

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

First, we declare as missing values those persons who first used alcohol "More than 1 year ago", from which we obtain the following table A2.14:

Table A2.14: Distribution (frequency and percentage) to estimate incidence *in the past year* of use of an alcoholic beverage

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	No	11286	8.1	18.4	18.4
	Yes	49891	35.6	81.6	100.0
	Total	61177	43.7	100.0	
Missing values		78823	56.3		
Total		140000	100.0		

The frequency shown as 1 in the table above (*Used for the first time during the past month*), i.e., 49,891 cases, is the sum of the cases in categories "*in the past month*" and "*more than 1 month ago but less than 1 year ago*" from Table A2.13, that is, 22.101+27,790.

The column "Valid percentage" shows incidence in the past month, 81.6%, which was what was obtained earlier.

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

In this case, we say that those respondents 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:

Table A2.15: Distribution (frequency and percentage) to estimate incidence *past month* of use of an alcoholic beverage

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	No	11286	8.1	33.8	33.8
	Yes	22101	15.8	66.2	100.0
	Total	33387	23.8	100.0	
Missing values		106613	76.2		
Total		140000	100.0		

In this case, the frequency for category 1 in the previous table (*Used for the first time during the past month*), that is, 22,101 cases, is the same value as the category "*in the past month*" in Table A2.13.

Again, in the column "valid percentage" we obtain the incidence in the past month, 66.2%.

Using this approach, namely, constructing two new variables for past year incidence and past month incidence, it should be relatively easy and direct to construct the respective

Confidence Intervals using appropriate statistical software—a subject we shall address in the next Section.

Returning to the definition of prevalence, but now in general terms, let us suppose that we want to estimate the prevalence of use of a substance at some point in the person's lifetime, **P**. If x represents the variable "substance use," then:

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

Using formula (10), we see that the estimator of the prevalence of substance use for the period is:

$$\boldsymbol{p} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} * f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}} \qquad (13)$$

where the subindices i,j identify student j from university i selected in the first stage of the sampling; the denominator is the sum of the expansion factors, i.e., the population represented in the study.

3.2 Confidence Intervals

Confidence Intervals have generally been used for two purposes: on the one hand, to supplement the description of an indicator, and second, as a criterion for comparing two or more indicators.

Regarding the first case, *under SIDUC, it is strongly recommended that Confidence Intervals be used when estimating the parameters of the principal variables for analysis*, according to the study objectives. Thus, for example, the mandatory specific objectives of a study of this type 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, among others: alcohol, tobacco, tranquilizers and stimulants without medical prescription, inhalants, marijuana, cocaine base paste, cocaine and ecstasy.)*

Earlier, we gave the expressions for these indicators, namely, for prevalence and incidence. So why would we use Confidence Intervals? Let us suppose 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 (13) (i.e., with expansion factors) is 50%. If we also know that the error of the estimate (what we have called "precision/accuracy") for a 95% Confidence level is

2%, then the Confidence Interval is between 48% and 52%. On the other hand, if the error of the estimate was 10%, the Confidence Interval is from 40% to 60%.

Thus, 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 2% or an error of 10%--in other words, *we will not know how precise the estimate is*. An estimate of 50% ± 2% does not provide the same information as an estimate of 50% ± 10%. There is no doubt that an interval of between 48% and 52% is much more precise and informative than an interval of between 40% and 60%.

Estimates for specific subgroups in 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 essential that it be included in the reports, it should be accompanied by a note pointing out the low level of precision of that particular estimate**. It should be recalled that the size of the sample was determined under certain conditions, and therefore, when disaggregating the sample into subgroups, the sample size of those subgroups will not allow for estimates at the same level of precision used at the time the size of the sample was determined overall.

Section 6 above 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 necessary to be clear about the sampling design, remembering that this study is a complex sampling. We shall give below some examples using SPSS software. We shall again use as an example the sample of 5,339 students. For the purposes of the examples that are given below, we shall use the same database with the variables described in Table A2.2.

We first need to give a brief description of the sample for the two demographic variables that are in the database, i.e., sex and age. The idea is to develop estimates of the principal indicators for the subgroups of these variables, and compare the results, if possible. A typical case is a comparison of the prevalence rates of the use of a particular substance by males and females, or by age groups, for example. ***For this, we shall construct a new variable for age, using the following age categories: up to the age of 18, 19-20, 21-22 and 23 and older.*** The following table gives a description of the sample for these two variables:

Table A2.16: Description of the sample by sex and age, and percentages of the population represented

Variables	Sample size	Population represented	%
Sex			
Male	2,493	65,206	46.6
Female	2,846	74,794	53.4
Age			
18 and younger	1,342	34,966	25.0
19-20	1,542	40,278	28.8
21-22	1,318	34,552	24.7
23 and older	1,137	30,204	21.6
TOTAL	5,339	140,000	100

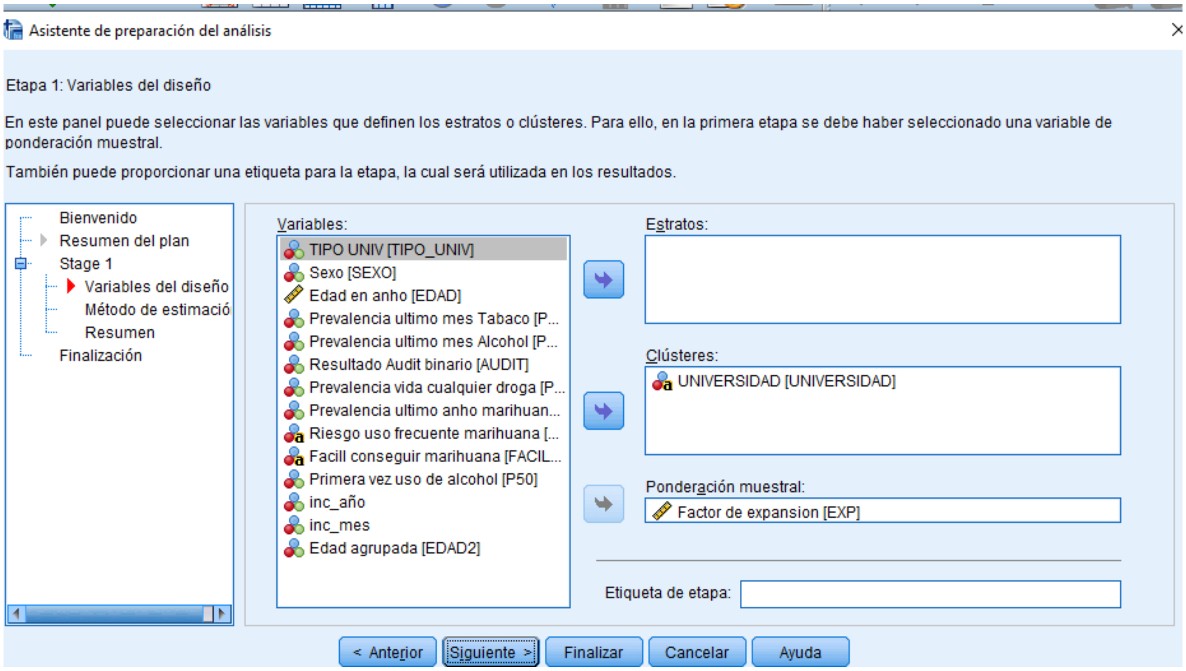
As shown in the table, in the expanded sample there are just over 53% females and around 25% are students aged 18 or younger.

The estimators of the parameters of interest are shown below using SPSS syntax:

- ✓ The sample design must be determined first; basically, in this specific example, we indicate the **clusters** (i.e., universities), as well as the variable having to do with the **expansion factor.**

- ✓ Once the data set (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 variable that identifies the ***Clusters***, and lastly, the variable corresponding to ***Sample Weight*** (expansion factors.) Then ***Done*** to save the sampling plan that can be used each time a new analysis is required.

✓



It is very important to point out 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 that sample design***. Thus, for example, if the study calls for sampling by type of university (public or private), then that must be included in the area labelled **Strata**.

✓ Then we use the SPSS syntax to obtain the desired estimates.

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

- ➢ Frequencies…
- ➢ Descriptives…
- ➢ Logistic regression….

Thus, for example, in order to estimate the prevalence of use of alcohol in the past month, select *Descriptives* and then select the analysis plan file already saved, the corresponding database, and then the variable associated with the indicator for which we want the estimate, and also include the Confidence Interval option. The option Descriptives gives the result in terms of an average of the variable pbeb_mes, which has coded values of 0 and 1. Therefore, the results in the table below are expressed as a proportion rather than a percentage:

Table A2.17: Estimated prevalence of past month alcohol use

	Estimate		95% Confidence Interval

			Standard error	Lower	Upper
Mean	pbeb_mes	.5785	.0174	.5303	.6268

The prevalence of the use of alcohol in the past month is 57.9%, which is the same as was obtained in Table A2.9. The error of the estimate is 0.94%, which gives a 95% Confidence Interval with values of between 53.0% and 62.7%. The error of the estimate and the Confidence Interval are correct and are properly computed on the basis of the sample design used.

Based on the limits of the Confidence Interval, two magnitudes of error should be calculated to show the quality of the estimate. On the one hand we have the ***absolute error*** of the estimate (**AE**), and on the other, the ***relative error*** of estimate (**RE**).

- If Ls represents the upper limit of the Confidence Interval,
- Li represents the lower limit, and
- p is the estimate of the indicator of interest,

then the absolute error and the relative error can be obtained as follows:

$$\boldsymbol{EA} = \frac{\boldsymbol{Ls - Li}}{\boldsymbol{2}} \quad \text{and} \quad \boldsymbol{ER} = \frac{\boldsymbol{EA}}{\boldsymbol{p}} * \boldsymbol{100} \qquad (14)$$

In the example:

$$\boldsymbol{EA} = \frac{62{,}68 - 53{,}03}{2} = 4{,}83\% \text{ and } \boldsymbol{ER} = \frac{4{,}83}{57{,}82} * 100 = 8{,}4\%$$

Both values are fairly low, which means that in this particular case, the estimate is sufficiently precise.

Using the same criteria and the same software, estimates can be made of the past month use of alcohol and 90% Confidence Intervals for the categories sex and age group. The results are given in the table below: Translator's note: The Spanish says "90% CI" but in light of table below, should perhaps be "95% CI".

Table A2.18: Prevalence of past month alcohol use and 95% Confidence Intervals (CI), by sex and age groups

Variable	Prevalence (%)	Standard error (%)	CI (%)
Sex			
Male	59.8	1.24	56.3 – 63.2
Female	56.2	2.49	49.3 – 63.1
18 or less	45.6	2.32	39.2 – 52.1
19-20	59.0	2.07	53.3 – 64.8
21-22	65.5	2.10	59.7 – 71.3
23 or more	61.7	1.33	58.0 – 65.4
TOTAL	57.9	1.74	53.0 – 62.7

Looking at the precise estimates of prevalence, we see that the prevalence of past month alcohol use is higher among males than among females, 59.8% and 56.2% respectively. We also see that the highest prevalence rate is among students in the 21-22 year age group, at 65.5%, a little more than 15 percentage points than among students aged 18 or less.

Given that these are estimates based on random samples of the population, the question now is whether there are differences by sex or age group between the past month prevalences 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 age groups?

In other words, ***are the differences observed explained by 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 whether the prevalences of past month alcohol use at the level of the population are the same for males and for females.

However, this and other 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 subgroup is 30% to 40%, and for the other subgroup it is 45% to 60%, then, given that there is no overlap between the Confidence Intervals, the conclusion would then be that there is a statistically significant difference between the prevalences in the two subgroups. If the CI for the second subgroup had been 35% to 45%, then there is an overlap of the two intervals, and the conclusion would therefore be that there are no statistically significant differences between the groups. A third situation might occur if there is a slight or very small overlap of the intervals, for example 30% to 40% and 39.5% to 44%, or 30% to 40% and 40.1% to 44%. In the first case, there is a slight overlapping, and in the second, a slight lack of overlap.

Although this criterion is used frequently, ***the use of statistical procedures is always recommended to provide the best possible response to the hypothesis proposed*** about the comparison of indicators between two or more subgroups. Returning to the previous examples, in those situations in which there is a 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 the criterion, it will be possible only to detect differences, but the size of that difference cannot be assessed, in terms of probability.

The section below will discuss methods of testing hypotheses that compare indicators of two or more subgroups, which is ***the procedure that is strongly recommended***.

3.3 Comparison between subgroups

As with the previous analyses, in order to compare indicators of interest among population subgroups, statistical software is needed, along with the skill to use it and interpret it correctly.

As an example, let us suppose that:

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

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

Our interest is in examining 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 for 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 and the classification of the cases (subjects) into males and females, it must be decided whether the null hypothesis is rejected or not rejected (this latter case means that there is not sufficient evidence to reject it); the former means that the test for such a decision is based on the assumption that the null hypothesis is true, and that 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 subgroup, i.e., for males and for females:

Let p_1=prevalence of past month alcohol use in the subsample of males,
and
Let p_2=prevalence of past month alcohol use in the subsample of females.

The previous results in Table A2.18 show that: p_1=59.8% and p_2=56.2%.

The hypothesis can be solved by means of Pearson's chi-squared test (test of independence), the result of which in SPSS is shown in Table A2.19:

Table A2.19: Test of independence for comparison of past month use of alcohol by sex

	Chi-squared	F corrected	gl1	gl2	Sig.
Pearson	7,081	3,755	1	4	,125
Odds ratio	7,087	3,757	1	4	,125

The above table gives two results for the Pearson's chi-squared statistic: the first column of results of Table A2.19 (chi-squared=7,081) does not consider the sample design, whereas the second column (***F corrected=3,755***) is correct since it does take into account the sample design. Therefore, the value of chi-squared corrected (known as Rao-Scott chi-squared correction), which we shall call Q, is equal to 3,755 (Q=3,755) with an associated probability (p-value) of less than 0.125. If the significance level is defined as 5%, that is, α=0.05, given that the p-value=0.125, the null hypothesis is not rejected, that is to say that 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 prevalence of the use of alcohol. In other words, *there are no statistically significant differences in the prevalence rates of past* *month alcohol use between males and females,* and we can therefore say that the difference seen in the samples is explained by sampling errors associated with this type of study.

Another example is to compare prevalences by age groups, where we now have four groups to compare rather than two. The hypothesis in this case is:

> **H_0 : P_1=P_2=P_3=P_4** that is, prevalence of past month alcohol use is the same for the four age groups, and
>
> **H_1**: at least two indicators are different.

Again using SPSS, we obtain the following results:

Table A2.20: Test of independence for comparison of past month alcohol use by age category

	Chi-squared	F corrected	Sig.
Pearson	121,232	50,978	,0002
Odds ratio	120,680	50,746	,0002

According to the results in Table A2.20, the prevalences in the sample of past month alcohol use by age group are (in %): p_1=45.6%, p_2=59.0%, p_3=65.5% and p_4=61.7%.

In Table A2.20 we see that Pearson's chi-square value according to the conditions of the sampling design is **Q=50,978** with $p<0.005$. This means that we reject the null hypothesis and conclude that there are large differences (statistically significant) in the prevalences of past month alcohol use between the four age groups. We are able to deduce from this only that there are differences, but we cannot identify where those differences are between pairs of age groups. The prevalences and Confidence Intervals given in Table A2.18 suggest that there may be statistically significant differences between the 18 and under age group and the other three groups. We next present a much more general option which provides a way of analyzing what the data suggest in this particular case. We next present a much more general option which provides a way of analyzing what the data suggest in this particular case.

b) The **second** solution involves a **logistic regression model**, which allows for analysis of the previous hypothesis and can also enable us to construct explanatory models about a particular variable of interest, such as the use of an illicit drug in the past year. Below, we will look at the different uses 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:

- ✓ If the response is a **continuous variable**, then the model is called a ***linear regression model*.**
  - Variable: *age of first use of a substance:*
    - 12, 13, 14, etc.....
- ✓ If the response is **binary (two possible alternatives),** the model is called a **binary logistic regression model**.
  - Variable: *past month use of alcohol:*
    - Yes or No.
- ✓ If the response is a **qualitative ordinal (more than two possible alternatives that do have a natural order)**, the model is called the **ordinal logistic regression model**.
  - Variable: *frequency of use of illicit drugs in the past year*:
    - Never, only a few times in the past year, several times in the past year, often in the past year.
- ✓ 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*:
    - None, only alcohol, only marijuana, only tobacco, only alcohol and tobacco, etc.

In all of the above 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 a person has used, and 0 if the person has not used. Therefore, **Y is a binary variable** and the most appropriate regression model in these cases is the **binary logistic regression model** (from here on, we shall simply speak of logistic regression.) The idea is to model response Y: thus, 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 that are 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.

Thus, if we are interested in analyzing the use of any illicit substance in the past year among the country's university population, we have:

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

On the other hand:

- ✓ **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 (not present)*,
- ✓ **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 (not present*.)

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 factor X **is not associated** with response Y, the probability that a person used an illicit substance **does not depend on** (is independent of) 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[18] 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 in accordance with the objectives):

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

Odds ratio~~[19]~~: epidemiological studies on prevalence of drug use—the category to which a study of the university population belongs—are cross-sectional studies conducted at a specific moment in time, in which past and present events are investigated. Let us suppose that X represents the variable "drug use by people who live in the same household," where X has the value 1 for those respondents who answer "Yes", and has the value 0 when no-one in the household uses any drug.

Translator's note: Footnote 19 not needed in English and should be deleted.

The hypothesis behind a question of this nature is that for those university students who state that there are people in their household who use drugs, i.e., X=1, the probability that they themselves will also use a substance is greater than in those cases in which no individuals in the household 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:

$$RD = \frac{P_1/Q_1}{P_0/Q_0} = \frac{P_1 * Q_0}{P_0 * Q_1} \qquad (15)$$

If the hypothesis is true, then $P_1>P_0$ and therefore $Q_0>Q_1$ and the numerator in formula (20) will be greater than the denominator. In the event $P_1=P_0$ then RD=1, (OR=1) called **null value**. Given that all the expressions that make up the OR 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

[19] ~~La medida en inglés es odds ratio, y existen varias otras traducciones al español: *razón de ventajas, razón de oportunidades, razón de posibilidades, razón de momios, razón de productos cruzados, razón de desigualdades. Aquí usaremos razón de disparidad, aun cuando SPSS usa razón de ventaja.*~~

model, assuming that there is only one independent variable or exposure (X), where P is the probability that the event of interest will occur for a particular value of X:

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

Following the previous example where X has two values, 1 if the factor is present, 0 if not, then we will determine the relevant probabilities in accordance with expression (21).

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:

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

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:

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

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

Based on the information collected from the individuals 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 again return to the previous example, where Y=past month alcohol use, and X=sex. The hypothesis is that prevalence among males (P_1) is equal to the prevalence among females (P_2),[20] that is, **H_0: P_1=P_2.** Using SPSS, again, select from the menu *Analyze* and then *Complex Samples*. And then ***Logistic Regression***, and it is here that 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. We give below the result of the estimate of the coefficient of the model, or $\hat{\beta}_1$:

Table A2.21: Estimate of parameters of logistic regression model for past month alcohol use and sex

Parameter	B	Standard error	95% Confidence Interval		Contrast hypothesis	
			Lower	Upper	t	Sig.
(Intersection)	-,396	-,540	-,253	-7,662	,002	-,396
[sex=,00]	**,148**	-,064	,360	1,937	,125	,148
[sex=1,00]	,000[a]	.	.	.	.	,000[a]

The most important result in Table A2.21 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, $\hat{\beta}_1$=0.148 with p<0.148 and a 95% Confidence Interval of between 0.360 and 1.937. Given that the p-value is over 5%, the conclusion is that there are no statistically significant differences between the prevalence rates of past month alcohol use among males and females, which is fairly consistent with the results shown in Table A2.19.

The estimator of the odds ratio is:

$$R\widehat{D} = e^{\hat{\beta}1} = \exp(\hat{\beta}_1) = \exp(0{,}148) = 1{,}16$$

with a 95% Confidence Interval of 0.938 to 1.434, and given that the null value 1 is within the limits of the CI, we reach the same conclusion as above. The same results in SPSS are

[20] It does not matter which subindex is used. Usually, the subindex 0 is used to indicate the absence of a factor. In this case, the variable sex does not represent a risk or protective factor, and therefore, we will use subindices 1 and 2 rather than 1 and 0.

shown in Table A2.22, which gives the estimate of the odds ratio, and the corresponding Confidence Interval:

Table A2.22: Estimate of odds ratio for past month alcohol use and sex based on logistic regression model

Odd ratio	95% Confidence Interval	
	Lower	Upper
1.16	0.938	1.434

The values for the variable sex are *1 for males and 0 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 males is compared to the prevalence among females. Given that the prevalence among males is higher than among females, it is therefore to be 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 under study 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 or by age group, or by region or area, these are conditions that cannot be changed, and therefore, if statistically significant differences are found between males and females, this information will be useful in targeting the interventions, but not to modify the variable as such.

On the other hand, if X represents the *perceived risk of substance use*, on the understanding that the greater the perceived risk the less the probability of substance use, the policies can be geared to increasing the number of individuals who perceive that substance use is highly risky, and thus reduce the probability of use. This condition could in fact be modified, and it is in these cases where we speak of risk and protective factors.

Let us now suppose that the response of interest is *past year use of marijuana*. Shown below are the prevalences by sex and age group.

Table A2.23: Prevalence of past year marijuana use and 95% Confidence Intervals (CI), by sex and age group

Variables	Prevalence (%)	Standard error (%)	CI (%)
Sex			
Male	28.6	2.3	22.2 – 34.9
Female	17.5	0.8	15.3 – 19.6
Age2*			
18 and less	16.6	1.3	13.0 – 20.1
19-20	22.0	1.4	18.1 – 25.9
21-22	26.1	2.4	19.4 – 32.8
23 and more	26.5	2.0	21.0 – 32.0
TOTAL	22.6	1.3	19.0 – 26.3

*Age2 is a new variable that has four possible values depending on age.

Analyzed by sex, i.e., comparing the prevalence among males with the prevalence among females, we see that the Confidence Intervals do not overlap, which would lead us to think that there is a 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 A2.24 and A2.25):

Table A2.24: Estimate of <u>parameters</u> in logistic regression model for past year use of marijuana and sex

Parameter	B	Standard error	95% Confidence Interval		Contrast hypothesis	
			Lower	Upper	t	Sig.
(Intersection)	0,917	0,112	0.606	1.227	8,200	0,001
Male vs female	0,635	0,073	0.432	0.839	8,669	0,001

Table A2.25: Estimated odds ratio for past year alcohol use and sex based on a logistic regression model

Comparison	Odds ratio	95% Confidence Interval	
		Lower	Upper
Male vs female	1,887	1.635	2.179

The RD̂=1,887 with p=0.001 (see previous table), less than α=0.05 and with a Confidence Interval of between 1.635 and 2.179 (does not include null value 1), which leads to the conclusion that there is a statistically significant difference in prevalences of past year marijuana use among male and female students.

In the two preceding examples, the independent variable (sex) could be answered in two ways. Let us now look at the case where this variable has more than two levels or categories, such as, for example, the variable **riesgo_mar** (perceived risk of frequent use of marijuana, see Table A2.2), which we shall analyze on the basis of four possible responses: no risk, slight risk, moderate risk, and high risk.

First, let us look at the prevalences of past year marijuana use as a function of perceived risk, which is shown in the table below:

Table A2.26: Prevalence of past year marijuana use and 95% Confidence Intervals, by perceived risk of frequent use of marijuana

Perceived risk of occasional use of marijuana	Prevalence (%)	Standard error (%)	CI (%)
No risk	30.4	1.9	25.2 – 35.6
Slight risk	24.4	2.3	18.0 – 30.8
Moderate risk	21.6	1.5	17.4 – 25.7
High risk	18.8	2.4	12.1 – 25.5
TOTAL	22.9	1.4	19.1 – 26.8

As shown in the above table, the prevalences of marijuana use tend to decline to the extent that the perceived risk of frequent use of the substance increases, going from 30.4% of those who perceive no risk to a 18.8% of those who see high risk. The Confidence Intervals only suggest a difference in prevalences in the groups at each end. As we said earlier, these findings and possibly others should be confirmed by means of appropriate statistical methods. To do this, we shall use a logistic regression model with the category “high risk” as the reference, since it is the category with the lowest prevalence rate, which means that the model will be comparing each of the other three categories with this one.

Adjusting the model, we conclude that there are statistically significant differences in the prevalence of past year marijuana use between the categories of perceived risk of marijuana use, as shown in the table below, where F=52.249 with p=0.019.

Table A2.27: Estimate of parameters of the logistic regression model for past year marijuana use and perceived risk

Origin	gl1	F Wald	Sig.
(Corrected model)	3,000	52,249	0,019

(Intersection)	1,000	247,074	0,000
RIESGO_MAR	3,000	52,249	0,019

Table A2.28 gives the odds ratios for comparison of each category with perceived high risk.

Table A2.28: Estimated odds ratio for past year marijuana use and perceived risk based on logistic regression model with "high risk" as reference

Comparison	Odds ratio	95% Confidence Interval	
		Lower	Upper
No risk vs. high risk	1,886	1.332	2.670
Slight risk vs. high risk	1,395	0.695	2.800
Moderate risk vs. high risk	1,187	0.901	1.564

As shown in the table, the odds ratio for the comparison of "No risk" to "High risk" is associated with a Confidence Interval that does not contain the null value, which means that there is a statistically significant difference between the two groups. However, comparing the other two groups with the "high risk" reference category, there are no significant differences.

In Tables A2.26 to A2.28, we have developed two examples of independent qualitative variables: the first was based on a binary variable (sex) and the second on an ordinal variable with four possible categories.

The next step is to look at a continuous variable, such as the age of the respondent, and we will examine the possible ways of studying the association of a binary dependent variable (in this case, past year use of marijuana) with a continuous independent variable, by using a logistic regression model. Let us recall here expression (16) on a logistic regression model:

$$P = P(Y = 1/X = x\,) = \frac{e^{\beta_0+\beta_1 * X}}{1 + e^{\beta_0+\beta_1 * X}} \qquad (24)$$

If P is the probability that the event under study *will occur*, then Q=1-P represents the probability that the event *will not occur*, and, in accordance with the previous expression:

$$Q = 1 - P = \frac{1}{1 + e^{\beta_0+\beta_1 * X}} \qquad (25)$$

Formula (24) is not the only representation of the model. Indeed, if we combine the two previous expressions, we find that a logistic regression model can also be expressed by means of the following, where instead of modeling P, we model the *natural logarithm of* *the quotient of P and Q*, (which is called the logit of P) and which gives the following ***linear relationship***:

$$\ln(P/Q) = \beta_0 + \beta_1 * x \qquad (26)$$

This is important because it assumes a ***linear relationship*** between a function of P and the continuous variable X; if age increases by one year, there is a change of some sort in the **ln(P/Q)**, and this change is independent of whether the increase in age is between people aged 18 and 19, or between people aged between 23 and 24. This may therefore often be a fairly robust assumption.

Let us illustrate this by means of an example in which the dependent variable is again the *use of marijuana in the past year*, and the independent variable is the *age* of the respondents. The results of the adjustment to the model are shown in the following two tables:

Table A2.29: Estimated parameters in logistic regression model for past year marijuana use and age

Parameter	B	Standard error	95% Confidence Interval		Contrast hypothesis	
			Lower	Upper	t	Sig.
(Intersection)	-2,184	0,205	-2.754	-1.614	-10,638	0,000
age	0,046	0,011	0.016	0.076	4,246	0,013

In Table A2.29, we see that the estimate of the age coefficient is **0.046** with a p value of 0.013, which shows that there is a statistically significant increase in the *ln (P/Q)* as the age increases. We conclude from this that the probability of marijuana use in the past year, i.e., P, also increases as the age increases, which is consistent with what we saw in Table A2.23.

As we have seen, the use of analytic methods to compare groups is no small matter: a number of issues need to be taken into account, some of them theoretical and some of them practical in nature. In the first case, the linear relationship defined in equation (26) must be borne in mind; it involves an assumption which, if not true, might lead to some imprecise conclusions. From a practical point of view, the most important suggestion is always to perform a good descriptive analysis and bear it in mind when moving to more complex statistical solutions.

As we said earlier, a regression model (any of those defined above in Section 3.3 of this Annex) enables us to move forward somewhat with our analyses. It was concluded from the above analyses (Tables A2.24 and A2.25) that the prevalence rate of past year use of marijuana was significantly higher among males than among females. It was also concluded that age is associated with marijuana use.

A reasonable question to look at is whether the difference observed between males and females might be influenced by possible differences in the age distribution in the two groups. For example, if the mean age of male students were higher than females, and given that a higher rate of marijuana use was found among males, then the difference by sex could be explained in part by age. Questions such as these can be resolved by the use of this type of model. The solution is relatively simple: simply to adjust a new model with the variables sex and age (in their original state, i.e., continuous), and examine whether there is any change in the odds ratio associated with sex.

The new model will be as follows:

$$P = P(Y/\,Sex, Age) = \frac{e^{\beta_0+\beta_1*Sex+\beta_2*Age}}{1+e^{\beta_0+\beta_1*Sex+\beta_2 Age}} \qquad (27)$$

The important thing in this analysis is to examine the change (from the model adjusted only by the variable sex) in the odds ratio for the variable sex, **adjusted for age**; this means, to study the difference in male and female prevalence rates for people of the same age, i.e., where age is not confounding the relationship between marijuana use and sex.

The results are given below:

Table A2.30: Estimated <u>odds ratios</u> in logistic regression model for past year use of marijuana and sex, adjusted for age

Comparison	Odds ratio	95% Confidence Interval	
		Lower	Upper
Male vs. female	1.853	1.521	2.259

The odds ratio shown in Table A2.30, that is, $\hat{RD}$=1.853 is called an **age-adjusted** odds ratio, which is different from that given in Table A2.25 where $\hat{RD}$=1.887, which is the **crude** or **unadjusted** odds ratio. The important point is that the adjusted model examines the comparison between prevalence rates in males and females without being influenced by age.

Beyond the results themselves for a particular example such as the ones given above, the idea is to show the potential for using regression models for the analysis, not only because they enable us to take more precise decisions when comparing indicators on substance use (rather than using Confidence Intervals), but also because we can make adjustments and control for 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 probability of substance use. Some demographic variables, or variables related to the family environment, or close friends who are substance users, and so forth, may be modeled in an effort to find an overall explanation of substance use, and also to estimate the individual contribution of those factors that were included in the model.

The same could apply to other variables of interest, such as modeling the variable "*substance use disorder*" as a dependent variable, and by means of this type of model find factors that could differentiate between the sub-categories "disorder" and "not disorder" among those persons who said they had used in the past year.

Special attention should be paid to the *factors associated* with substance use, as mentioned in the introduction to this Protocol. This leads us directly to discuss some of the specific objectives described above, such as: "Analyze the association between substance use and perception of risk." Hay otros objetivos en esta misma dirección, como por ejemplo la. Translator's note: Something missing in the Spanish.

When we refer to *factors associated* with substance use, we are thinking about what are called ***risk and protective factors***, where the former are those factors that, if present, we assume a greater probability of substance use, and by contrast, the latter are factors that operate in the other direction, 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:

$$P(Y = 1 \,/\, X = 1) \; > P(Y = 1/X = 0) \qquad (28)$$

In other words, if the factor is present (X=1), the probability 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:

$$P(Y = 1 \,/\, Z = 1) \; < P(Y = 1/Z = 0) \qquad (29)$$

In other words, if the factor is present (Z=1), the probability of substance use (Y=1) is ***lower*** than when the factor is not present (Z=0).

It is important in epidemiological studies such as the one discussed in this Protocol to include in the study objectives an analysis of the risk and protective factors that are felt, in the theory, to be important. This is not only in order to study or confirm the association with substance use, but, equally important, to examine the *trends in those factors*. When these risk and protective factors are confirmed through scientific research, then it is to be hoped that any interventions will focus on producing changes in these conditions, which hopefully would also produce changes in substance use.

Let us look, as an example, at the perceived risk of occasional use of marijuana. Table A2.27 showed the prevalence of past year marijuana use for each category of perceived risk, values that range from 18.8% of those who perceive occasional use of marijuana to be "high risk," to a prevalence of use of 30.4% among those who perceived "no risk." Looking at a variable where there are two options ("high risk" and "not high risk") the table below shows the corresponding prevalences:

Table A2.31: Prevalence of past year marijuana use by perception of high risk of occasional use of marijuana

Perceived risk of occasional marijuana use	Prevalence (%)
High risk	18.8
Not high risk	23.7
TOTAL	22.9

We conclude from the table that among those who perceived that the occasional use of marijuana is "high risk," the prevalence of marijuana use in the past year was 18.8%, almost five percentage points less than among those who perceived a lower risk, 23.7%.

We need now to know what proportion of cases in the expanded sample falls into each of these two categories. Having a prevalence rate of 18.8% associated with 60% of the population represented is very different from having it correspond to 5% of that population. Table A2.32 describes this distribution:

Table A2.32: Distribution of perception of risk of occasional use of marijuana

Perception of risk of occasional marijuana use	N	(%)
High risk	22,933	16.4
Not high risk	117,067	83.6
TOTAL	140,000	100.0

Some 83.6% of the university population represented *does not perceive* occasional use of marijuana to be "high risk."

It may be concluded that if there are positive changes in the variable analyzed, in other words, if the distribution described in Table A2.32 improves and the percentage of cases in the "high risk" category increases, then this should have an impact on the indicator of the prevalence of marijuana use in the past year (reduction in size) provided that the remaining conditions that also influence this indicator remain constant. Hence the importance of including variables such as these in order to ***analyze trends***, beyond their association with substance use.

In light of everything 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. Then models for analysis should be adjusted for each of those indicators as dependent variables* and each of the factors that are conceptually linked to it, as independent variables.

Depending on what is found in the adjustment of these models, the next step will be to adjust a *multivariate model* with all of the variables or predictors that have been shown, either as a result of the analysis or because of the literature or theory, to be important. This latter (adjustment of the models) probably goes further than what is needed in a general report on the results of the study, but should be carried out in order to find a better explanation of the phenomena associated with substance use, in order to target the prevention strategies appropriately for this specific population.

ANNEX 3: INSTRUCTIONS FOR ADMINISTRATION OF THE QUESTIONNAIRE

The main feature of the study of the university population is that the selected students will answer the questionnaire on line. The questionnaire will be stored on a server agreed to by the National Drug Observatory and the OID.

The study is coordinated and managed by the country's National Observatory on Drugs.

The different activities needed to conduct a study of this nature are laid out below, in chronological order, along with the responsibilities of each institution.

1. The country's National Observatory on Drugs decides, as part of its information-gathering strategy, to conduct a study of the university population using the present Protocol, and will discuss it with the OID.
2. The National Observatory and the OID agree on the study objectives (on the basis of the objectives described in this Protocol), and make adjustments as necessary to the questionnaire provided in Section 5. A decision will also need to be made as to whether the study should be stratified according to type of university (public or private.)
3. The questionnaire that will be uploaded onto the platform decided on by the National Drug Observatory and the OID is finalized.
4. The National Observatory must obtain information about all of the country's universities, and the number of undergraduate students enrolled in the year in which the study is to be conducted. The list should include the following, at a minimum:
   a. Name of the university
   b. Type of university (public or private) if necessary, as per point 2 above.
   c. Number of students enrolled in the particular year.
   d. A sampling frame of all of the universities is drawn based on the above information, or else two sampling frames by type of university, depending on the decision taken in point 2.
5. If there is only one sampling frame, then the National Observatory should select the 12 universities for the study. The first 10 selected will form the sample, and the other two are replacements in case any of the first ten does not agree to participate. In the event it is decided to have two sampling frames, one with public universities and the other with private universities, then six of each should be selected, based on the same argument. In either case, the proposed approach is to select the universities with probability proportional to size (PPS) as described in Annex 1.

6. Once the universities have been selected, the university authorities should be contacted and asked for their approval of the study. See below for a model letter to the university authorities:

Date……………………..

Dr.______________________________
Rector/Chancellor
Name of university
City, Country

Dear Dr.___________:

I am pleased to write to advise you that the NAME OF THE COUNTRY'S DRUG COMMISSION and the Inter-American Observatory on Drugs (OID) of the Organization of American States (OAS) will conduct an ***Epidemiological Study on Drug Use*** among the country's university students.

We have made a random selection of universities in NAME OF COUNTRY, and have selected your university to take part in this study. This research project is designed to gather current information that may provide guidance on the development of future student welfare policies. The study will be conducted in ten universities, chosen by means of a random sample, first of universities and then of students in the universities selected. It will gather information on the use of alcohol and other drugs, as well as related behaviors, perceptions and risk factors. We anticipate that this study will provide reliable information that can be used to develop or enhance policies on drug use prevention and treatment among university students.

The study will be conducted via an on-line platform that students can use to access a standardized questionnaire, which is stored on a server decided on by the National Drug Observatory and the OID. This will ensure that the participants' answers will be protected by statistical secrecy and will therefore be completely confidential.

Each participating university will be offered a copy of the database containing the responses of its own students (assuring their anonymity) for its own analysis, and thus contribute to the university's own policies to mitigate the effects of drug-related problems, if needed.

This research study will be possible only with the firm commitment of the selected universities, including your university, and we therefore ask you to advise us whether you wish your university to participate in this research project. If you do wish to participate, please ***designate a study coordinator***, who will be the liaison with NAME OF COORDINATOR IN NATIONAL OBSERVATORY ON DRUGS (telephone XXXXX, e-mail:

XXXX). The study will involve only a few small expenditures for your university, such as paper and photocopies, and coordination of the study. Most of the cost will be borne by the DRUG COMMISSION NAME.

We will shortly contact the person you have appointed as study coordinator to answer any questions he or she may have, and to coordinate the activities, one of which will be an on-line or in person workshop scheduled in principle for DATE XX de XX of this year in NAME OF CITY. We will advise you shortly of the time and place for the workshop. The expenses of this workshop will be defrayed by the NATIONAL DRUG COMMISSION NAME.

We look forward to having your university participate in this important study,

Yours sincerely,

SIGNED……………

7. In the event that it is not possible to obtain 10 affirmative responses from the 12 universities selected, we recommend a second selection process.

8. The universities will need to designate a coordinator. A training workshop for the coordinators in each university, conducted by the National Observatory, should be offered. Experience has shown that good coordination and strong commitment to the study will increase the number of responses from the students.

9. The National Observatory should request the list of students in each university. The list should contain at a minimum contact information and an e-mail address. The identification may be the matriculation number, and full name. Importantly, this identification will be used only for the selection process, and will not be stored in the database.

10. The National Observatory on Drugs will determine the size of the sample for each university using expressions (5), (6) and (7).

11. In accordance with the sample size determined in point 10, a list will be drawn up for each university with the ***user name and password*** that the students will use to access the questionnaire. The user name should have six fields: the first two identify the university, and the next four identify the student. The National Observatory on Drugs will develop a list of the universities in the sample, each of which will be assigned a number from 01 to 10. These are the first two digits of the user name. The next fields will be from 0001 up to the maximum number of the sample size in the university. Thus, if the size of the sample in the **first** university is **2,100**, then the students will have an identification number from **010001** to **012100**.

12. In accordance with the sample size determined in point 10, the National Observatory will select the sample using a simple random process.
13. With the sample selected in point 13 and the user names and passwords generated in point 11, a new list will be drawn up that will be used to contact the students selected. This list contains the original information about the students, plus the ***user name and password.***
14. These processes will be done university by university as the responses are received from the universities. It is not necessary to have the responses from all of the universities.
15. The coordinator in the National Drug Observatory who is leading the study will contact the university coordinator and send him or her the list with the sample described in point 14.
16. The study coordinator in the university should contact the students in the sample, by sending them an e-mail or by other means. The important point in this stage is to reach the maximum number of students in the sample, and the university coordinator will decide on the best approach for his or her university. The university study coordinator should report back to the coordinator from the National Drug Observatory. Below is a sample letter to the students:

Dear student:

I am pleased to inform you that our university has been selected to participate in a. Epidemiological Study on Drug Use among University Students. This research project has been agreed on by our university and the NATIONAL DRUG COMMISSION, in cooperation with the Organization of American States (OAS). The goal is to obtain reliable information that will help guide future student welfare policies in the participating universities.

The research project is being conducted in ten of the country's universities. A **random sample** has been taken first of universities and then of students. In the case of our university, XXX number of students have been selected from various faculties, and you are one of them. We strongly urge you to participate.

The participating students' data are protected by bioethical and statistical secrecy standards. This means that your **responses are absolutely confidential.** We ask for your fully cooperation in answering the questionnaire accurately.

The questionnaire, which is the same for all of the participating universities, should be answered on line. The answers will be stored on an OAS server, which will further protect the secrecy of the information you provide. Your responses, and those of the other students selected, will be used for statistical purposes only. **<u>Individual respondents cannot be identified</u>**.

To respond to the questionnaire, please enter **www.XXXXXX** and access the questionnaire using the following:

USER NAME: XXXXXX
PASSWORD: XXXXX

The survey will be available for six weeks, starting on **XXXX and ending on XXXX.**

Your views are very important for this research project, and so thank you again for participating.

If you have any further questions, please contact NAME, TELEPHONE AND E-MAIL OF THE STUDY COORDINATOR IN YOUR UNIVERSITY.

We again encourage you to participate in this research project.

Yours sincerely,

SIGNED UNIVERSITY AUTHORITY

17. This email is the first interaction with the students and follow-up on the responses. Contacts will be made by the coordinator from the National Observatory. Remember that the student can respond to the questionnaire in several on-line sessions. The National Observatory on Drugs will monitor and follow up on the process by reviewing the database on a regular basis to examine the status of each student, and classify them as follows:
    a. ***Completed***, which means that the student answered all of the questions and sent in the questionnaire.
    b. ***Entered***/Accessed but not finished; the student may have responded to part of the questionnaire but has not yet pressed the SEND button.
    c. ***Not accessed***.
18. The National Drug Observatory will create individual lists for each university containing the above information but excluding those students who have already completed the questionnaire. These lists will be distributed to the respective universities so that the local coordinator can send another letter to those students who have not accessed the questionnaire or who have not completed it. Again, the university coordinator may use a different approach.
19. This process should be repeated up to the final closing date specified in the letter to the students, at which point the process ends and students can no longer access the questionnaire.

20. The database in the system becomes a database in SPSS. The following needs to be done:

    a. The National Observatory on Drugs should perform a cleaning of the database as described in paragraph 2 of Annex 2.

    b. The National Observatory should calculate the expansion factors for each university using expression (8) considering the effective sample.

    c. The National Observatory should develop standardized indicators on prevalence and incidence of the use of the various substances, and incorporate them into the database.

    d. The National Observatory should create individual databases for each university. The variable **student identification** must be deleted from all databases.

21. The National Observatory will forward the databases to each university.

22. The National Observatory will prepare a report on the study.