OAS | CICAD

INTER-AMERICAN DRUG ABUSE CONTROL COMMISSION (CICAD)

PROTOCOL FOR HOUSEHOLD SURVEYS

INTER-AMERICAN UNIFORM DRUG USE DATA SYSTEM (SIDUC)

2021

General coordination of this document:
Marya Hynes

Primary authors and design:
Juan Carlos Araneda
Francisco Cumsille
Pernell Clarke
Paula Demassi
Marya Hynes
Daniela Ocaña

Suggested citation for this document: Inter-American Drug Abuse Control Commission, Organization of American States (CICAD/OAS), Inter-American Uniform Drug Use Data System (SIDUC): Protocol for Household Surveys 2021.

TABLE OF CONTENTS

1. **Formulae have not been translated into English.**
2. **Words or phrases where meaning is in doubt are highlighted in green.**

PROTOCOL HOUSEHOLD SURVEY 2021	Page
1.- Introduction	4
2.- Objectives	6
3.- Methodology	8
3.1.- Target population	8
3.2.-Coverage and representativeness	9
3.3.- Sampling issues	11
3.4.- Frequency	13
4.- Operational definitions	13
4.1.- Substances	14
4.2.- Variables, questions and indicators	15
5.- The questionnaire: organization of the questions into modules	19
6.- Statistical analysis	61
6.1.- Description of the sample	61
6.2.- Estimating indicators of substance use	63
6.3.- Comparison of indicators	65
Annex 1: Sampling	66
Annex 2: Statistical analysis	77
Annex 3: Instructions for the Fieldwork	132

PROTOCOL HOUSEHOLD SURVEY

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). It highlights two fundamental ideas: the role of evidence in public policies and the comparability of information among member states.

Public policies are decisions taken by a State to address an identified problem, regardless of whether the action occurs at the state or the community level. These decisions lead to interventions 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."***

The notion that evidence-based policy requires information gathered through valid scientific methods is reinforced by the Hemispheric Plan of Action on Drugs 2021-2025[3]. Member states agreed to "*Establish or strengthen national observatories on drugs, or similar technical offices, strengthen 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

[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

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

Another critical issue is the production of comparable information across countries. CICAD's Inter-American Observatory on Drugs (OID, by its Spanish acronym) is responsible for supporting member states' national drug observatories (NDOs) to produce timely scientific information and apply standardized methodologies that allow for cross-country trends comparison. The OID fulfills its mandate by producing regular reports on the state of the drug problem in the Hemisphere. At the same time, the NDOs are responsible for evaluating the state of the drug problem at a national level.

To support member states' responses to the Hemispheric Drug Strategy and in response to the OID's mandate, the OID developed the Inter-American Uniform Drug Use Data System (SIDUC, by its Spanish acronym). SIDUC comprises a series of standardized, epidemiology-based research protocols to build information on drug use comparable across countries. The OID provides technical assistance for national studies upon request, subject to the availability of resources.

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 risk factors that might foster substance use and protective factors that discourage drug use or delay use . This information is fundamental to the development, monitoring, and evaluation of drug demand reduction policies. Therefore, it is essential to stress the importance of conducting epidemiological 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 and changes in risk and protective factors.

This document lays out the research protocol to measure substance use and associated factors in the general population aged 12 to 65 residing in households.

Household surveys are fundamental to measuring drug use and changes over time because of their broad coverage. Indeed, persons residing in households may represent over half of the general population in many countries. Because household surveys sample a broad spectrum of ages, they can look into differences between age groups over time, drugs used across age groups, etc. The objective of the national household surveys is to contribute to a knowledge base so that each country can make informed decisions based on scientific evidence. They are also the cornerstone of regional and world reports on substance use. CICAD's Inter-American Observatory on Drugs (OID) and the United Nations Office on Drugs and Crime (UNODC) compile and prepare regular reports on the problem using this type of study as their basis.

Nevertheless, for these reports on survey data to be technically sound, the national studies must be conducted using methodologically equivalent bases. To that end, CICAD developed a series of standardized research protocols to measure drug use in various populations under the umbrella of SIDUC.

Because the Household Survey Protocol samples the population residing in households ages 12-64, and breaks the population into age groups, it allows researchers to focus on particular drugs in particular age groups. For example, the **Report on Drug Use in the Americas 2019**[4] prepared by the OID provides an overview of drug use prevalence from household surveys in the Western Hemisphere. According to the report, past month prevalence of alcohol use among people aged 18-34 is over 50% in many countries, followed by the 35-64 age group. A large majority of countries show past month prevalence of alcohol use over 40% among this latter age group. The largest proportion of tranquilizer use without a medical prescription is concentrated in the 18-34 and 35-64 age groups. Similarly, the highest past year prevalence of stimulant use without a prescription is concentrated in this age group. Finally, the report demonstrated that the highest past year of marijuana and cocaine use prevalence is found in people aged 18-34.

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).

[4] Report on Drug Use in the Americas, CICAD/OAS, Washington, D.C. 2019

[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 sample.

[6] As per the list in section 4.1.

- 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",
- 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 analyze patterns of use of alcohol and marijuana,
- 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,

SPECIFIC OBJECTIVES OF THE EMPLOYMENT-RELATED MODULE:

- To estimate substance use in the actively employed population.
- To relate substance use to the type of work performed.
- To estimate the percentage of people who had had a workplace accident in the past 12 months and to associate it with substance use.
- To estimate the average number of days absent from work, and to associate them with substance use.
- To estimate the percentage of people whose employer has rules governing substance use, and to associate it with substance use.
- To estimate the percentage of people who had participated in prevention programs, and to associate it with substance use.

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

- To estimate the percentage of people who work in a business that has programs to help with substance use, and to associate it with substance use.

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.

The questionnaire consists of a set of questions that directly respond to the specific objectives listed above. 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. Further details on the questionnaire are found in section 5.

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 ***face-to-face interview*** using a confidential structured questionnaire, with a study based on a random sample of the target population.

3.1 TARGET POPULATION

In the context of the Inter-American Uniform Drug Use Data System (SIDUC), the universe of study is the population aged ***12 to 65*** living in cities (***urban areas***) that have a population of at least 30,000 inhabitants. The size of the cities or towns may vary from one country to another, depending on the size of the country itself. Thus, for example, in a country with 40 or 50 million inhabitants, the size of the cities will be different from cities in a country with 400,000 inhabitants. Therefore, other population sizes may be decided on for the cities making up the universe of the study.

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.

So far, the countries of the region have conducted household surveys at several levels of representativeness, namely:

1. At the national level only.
2. In two broad geographical groupings: the metropolitan area, and the rest of the country.
3. At the level of a state, province, department or region (depending on the name in the particular country.)

Most of the countries have chosen the first option, that is, the national level. However, a few have opted for the second alternative, while others have moved to the third option.[9]

A number of elements should be considered when deciding which alternative to use: some have to do with the size of the country, the financial resources available, and the time available, and so forth, while others are related to decisions as to the use to which the information obtained will be put. There may be countries whose geographical size and size of the population suggest that there may be significant diversity in substance-related issues, which makes it advisable for policies to be developed that take that diversity into account. There will be other countries where that is not the case.

The table below shows these three options and the potential advantages and disadvantages of each. We must, however, stress that in countries ***having small populations that are homogeneous in terms of their socio-bio-demographic characteristics, a choice between these options simply does not exist***. In these cases, it is probable that drug policies are national in nature and are standard throughout the country, and that what is required are global indicators for the country as a whole. However, there will be other countries whose internal realities are different and that have to deal with this type of decision. It is these countries that may be helped in their decision by the table below.

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

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

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

Based on the above, some researchers may need only national-level information, and others may require more specific detail. Although it is important to have information on the country in general, it will probably not be sufficient when using that information for decision-making at a smaller geographical area, such as the regional or provincial level. Hence, we recommend choosing a representativeness level that permits geographically disaggregated information whenever possible. As noted earlier, such a decision has an additional cost in requiring a larger sample size than that needed for a study that only envisages national representativeness. It is also more complex in terms of fieldwork and statistical analysis.

When reporting drug use data to CICAD for hemispheric reports, the OID requests information only at the national level, by as sex and age.

3.3 SAMPLING ISSUES

Studies that use SIDUC are conducted through probabilistic sample surveys. Probabilistic sampling does not collect information from all individuals in the target population, as in a census, but rather from a fraction or subset obtained through scientifically rigorous statistical procedures called *sampling techniques*. The first issue to is the ***unit of analysis*** - the unit that will provide the information needed to fulfill the study's objectives. In this case, the *unit of analysis is a person aged 12-65* selected from among the population as previously defined. ***The sampling unit*** must be defined next. Since most countries do not have an exhaustive list of individuals with fixed addresses where they can be contacted, the most appropriate approach is by sampling **households;** therefore, the *sampling unit is the household*.

To access to a sample of households, it is more efficient and less costly to select a *group of households.* There are several technical terms for a sampling group; for the purpose of this Protocol, we use the term "block."

A sample with these characteristics will be drawn in three phases or stages based on a ***sampling frame of households.*** This sampling frame is a list of households in the geographical areas where the study will be conducted as per the definition of the target population and must be as up-to-date as possible..

- Based on the sampling frame, the ***first phase*** means selecting the ***blocks*** in the cities as previously defined,
- The ***second phase*** consists of ***selecting the households*** in the blocks selected in the previous phase,
- Lastly, the ***third phase*** is where ***a person*** aged between 12 and 65 is selected from each of the households chosen in the second phase.

Generally speaking, we can identify two groups of sampling techniques: ***simple random sampling***, which is the easiest of all. It means using random procedures to select a subset (sample) of elements from a complete list of the elements that make up the target population. In *simple random sampling*, a list is obtained of all persons in the country aged 12-65, and their geographical location. Simple random sampling is not advised for this study because, while theoretically be possible, the sample would be widely scattered, making the fieldwork highly complex and time-consuming.

Any sampling technique that is different from a simple random sampling, such as the **three-phase** sampling described above, is called ***complex sampling***. Complex sampling is the recommended approach for this protocol.

Another basic question concerning the sampling is related to the sample siz***e*** for the study, i.e., determining the number of individuals aged 12-65 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. To ensure the study is nationally representative and meet the criteria for reporting to CICAD and international agencies, and the comparison between member states, the size of the sample should be large enough to ensure reasonable **national-level** estimates **by sex and age group**. In countries where the study representativeness is by regions or provinces in addition to the national level, the size of the sample in each subdivision should be such as to enable estimates to be made with acceptable sampling errors. The same is true if estimates are wanted at the city or town level.

The level of disaggregation is not the only condition for determining the size of the sample. Another necessary condition 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 fieldwork has been completed, decide on the **expansion factors** needed for a correct statistical analysis of the information obtained (see Annex 2 for more details about expansion factors.) The country's National Observatory on Drugs is responsible for performing these activities. However, the OID/CICAD 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 always provided resources are available for this.

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

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

- Annex 1: *Sampling*. More information on sampling techniques for household surveys, 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: *Manuals and instructions*. Issues related to the fieldwork.

3.4 FREQUENCY

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

List of substances suggested for estimating at least three indicators

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

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

They should also adapt the names 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? And lastly, we need an ***indicator*** that will report the result of the measurement of the variables. Note that these three concepts are interrelated and should fulfill the specific objectives described earlier.

The following groups of study variables should be considered:

- G**eneral variables**: sex, age, marital status, socio-economic level (where pertinent), region/district (in the event this coverage is contemplated in the study) 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",

- “frequent use of [Name of substance]”.

- ➢ 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 on 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 a 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 alternative 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*** (for 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, 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.[10] 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,** the **median,** and some **percentiles** (particularly the 25th and 75th percentiles.)

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 up-to-date 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.

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

The **magnitud**e **or extent** 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 (less than 24 years of age,) 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 details the questionnaire. The questionnaire consists of a set of questions that directly respond to the specific objectives defined in this protocol. While researchers may adapt some of the questions to the peculiar circumstances of their country, the importance of adhering to the original language as much as possible must be emphasized. *Neither the wording of the questions nor the answers should be changed.* Any departure from the questions or the methodology should be duly noted in the methods section of the research report.

The questionnaire also contains other questions that are recommended by the OID but are optional. Optional questions are noted in the protocol. The optional questions represent potential new information that can inform drug policies in the country, but may not be relevant in every national context.

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**
- **OTHER SUBSTANCES**

MODULE 3: PERCEIVED RISK AND FACTORS ASSOCIATED WITH DRUG USE

MODULE 4. INFORMATION ON TREATMENT

MODULE 5. INFORMATION ON DRUG TRAFFICKING/DEALING

MODULE 6. DRUG PREVENTION AND CONTROL MEASURES

MODULE 7: INFORMATION ON EMPLOYMENT

The table below shows for each topic the questions that are ***obligatory***: these are the questions that can respond to the objectives defined in Chapter 2 of this Protocol. As noted in that chapter, 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	QUESTIONS Obligatory	Optional questions
General information	All	
Tobacco	All	
Electronic cigarettes	All	
Alcohol	All	
Prescription tranquilizers	All	
Prescription stimulants	All	
Presciption analgesics	All	
Marijuana	MA1 to MA10	MA11 and MA12 and MA13 to Ma16
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	EX9 and EX10
LSD	All	
Inhalants	All	
Poppers	All, if the country decides to include this substance	
Other drugs	All	
Module 3: Perceived risk and factors associated with drug use	PR1 to PR17 FA21 to FA23 FA24 to FA27	FA1 to FA20
Information on treatment		All
Information on drug trafficking/dealing	TD8 to TD11	TD1 to TD7
Drug prevention and control measures		All
Information on employment		All

1.1.1.1 GEOGRAPHICAL IDENTIFICATION

Region/ Department	Town/ District	Area code	Code of household	Number of questionnaire	Letter for selection

1.1.1.2

Address (Street and number, district/town, post code, etc.)

1.1.1.3

1.1.1.4 KISH GRID FOR SELECTION

a) Write in the names of all members of the household who are between the ages of **12 and 65**, beginning with the **oldest**.
b) Assign a number to each person in the column entitled "Number." (numbers should be consecutive).

Name	Age	Number	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
			1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
			1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3
			1	1	1	2	2	1	2	2	2	3	3	3	4	3	3	4	4	4
			1	1	1	2	2	3	2	2	3	3	4	4	3	4	4	5	5	5
			1	2	1	2	2	2	3	3	3	4	4	4	5	5	5	6	5	6
			1	1	2	1	2	3	4	3	4	4	5	4	5	6	7	6	7	7
			1	1	2	2	3	4	4	3	4	5	6	5	5	6	7	7	8	8
			1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9
			1	1	2	2	3	3	4	5	4	6	7	6	8	8	9	9	10	10
			1	2	3	4	2	4	5	5	6	6	7	7	8	10	8	9	10	11
			1	2	3	4	2	4	5	6	5	7	8	7	9	10	9	11	10	12

PERSON SELECTED

Sex of respondent		**Age of respondent**	
	1. Male		**Years old**
	2. Female		
Contact information			

1.1.1.5 FIELD LOG

Date and time of visit	Interviewer code	Outcome	Interviewer signature
SUPERVISION	**Supervisor code**	**Outcome**	**Supervisor signature**
Field Assistant			
Supervisor			

MODULE 1: GENERAL INFORMATION

Good day!
We are carrying out a survey of thousands of households in the country on issues related to public health. Our goal is to obtain information that will help solve public health problems in this country. Your cooperation in this survey will be very helpful to this effort. Your answers will be absolutely confidential and completely anonymous, and will be used for statistical purposes only.

1. What is your current employment situation?

	1.	Work full time
	2.	Work part time
	3.	Occasional work
	4.	* Unemployed, <u>looking</u> for work
	5.	* Student
	6.	* Unemployed, <u>not</u> <u>looking</u> for work
	7.	* Homemaker
	8.	* Retired
	9.	*Self-employed
	0.	*Don't know/no opinion

**** SKIP TO QUESTION No. 3***

2. Which of the following best describes your current job or employment situation?

SHOW CARD NO.1. EACH COUNTRY SHOULD EDIT ACCORDING TO ITS OWN CIRCUMSTANCES.

	1. Senior government official, civil servant, senior management of a company. a senior officer in the armed forces, owner of large business, owner of large farm
	2. Professional, scientist, university professor, armed forces officer, medium-sized business, medium-sized farm
	3. Technician, teacher, small business owner, middle management, administrative or service personnel, small farmer
	4. Office staff in business, industry, services, civil service, or similar.
	5. Small business owner/small farmer.
	6. Skilled worker (plumber, hairdresser, air-conditioning technician, etc.)
	7. Laborer, farm worker
	8. Apprentice.
	9. Armed forces enlisted personnel, private security guards.
	10. Domestic service/street vendor.
	11. Pensioner (disability, old age, or similar.)
	99. Don't know/no opinion

3. Which of the following apply in your case (READ AND MARK ONE WITH AN X)

	1. You are the only one bringing in income
	2. You bring in the most income but there is also other income
	3. You bring in half of the household income
	4. You bring in some income, but it's not the main household income
	5. You don't bring in any income

4. Are you the head of household?

	1.	Yes
	2.	No

5. **In a normal month, was is the approximate monthly income of your household. If you are a minor, please try to get this information from an adult.**
SHOW CARD NO.2. USE LOCAL CURRENCY AMOUNTS.

	1. Less than $200		8. $2000-$2500
	2. $201-$400		9. $2501-$3000
	3. $401 y 600		10. $3001-$3500
	4. $601 y 800		11. $3501-$4000
	5. $801 y 1000		12. $4001-$5000
	6. $1001 y 1500		13. More than $5000
	7. $1501 y 2000		99. Don't know/no opinion

6. **What is the highest educational level that you have achieved?**
SHOW CARD No. 3, ADAPTED FOR EACH COUNTRY

	1. Never attended school
	2. Did not complete primary school
	3. Completed primary school
	4. Did not complete secondary school
	5. Completed secondary school
	6. Technical/vocational school
	7. Higher education/did not complete university
	8. Higher education/completed university
	9. Graduate/post-graduate

7. **What is the highest educational level achieved by the head of household?**
SHOW CARD No. 3, ADAPTED FOR EACH COUNTRY

	1. Never attended school
	2. Did not complete primary school
	3. Completed primary school
	4. Did not complete secondary school
	5. Completed secondary school
	6. Technical/vocational school
	7. Higher education/did not complete university
	8. Higher education/completed university
	9. Graduate/post-graduate

8. **What type of health insurance do you have?**

	1. No insurance
	2. State/government insurance
	3. Private insurance
	4. Other type of insurance
	9. Don't know/no opinion

9. **Your marital status**

	1. Single, lives alone
	2. Single, lives with partner
	3. Married, lives with spouse
	4. Separated, divorced, lives alone
	5. Separated, divorced, lives with partner
	6. Widow/widower, lives alone
	7. Widow/widower, lives with partner
	9. Don't know/no opinion

NOTES:
1.- Countries should adapt questions 2, 5, 6, and 7 to their circumstances.
2.- Countries may add questions about nationality, ethnicity, and/or religious affiliation to this module.

MODULE 2: SUBSTANCE USE

TOBACCO (TA). Now I am going to ask you some questions about tobacco use. 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?

	1. Yes	
	2. 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 used e-cigarettes containing nicotine, marijuana, or flavoring products?

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

CE2. Have you ever vaped e-cigarettes containing nicotine products?

	1. Yes	
	2. No	**(Skip to question CE6)**

CE3. Have you vaped e-cigarettes containing <u>nicotine products</u> in the past year?

	1. Yes	
	2. No	**(Skip to question CE6)**

CE4. Have you vaped e-cigarettes containing <u>nicotine products</u> in the past month?

	1. Yes	
	2. No	**(Skip to question CE6)**

CE5. How many days in the past 30 days did you vape e-cigarettes containing <u>nicotine products</u>?

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

CE6. Have you ever vaped e-cigarettes containing <u>cannabis products</u>?

	1. Yes	
	2. No	**(Skip to question CE10)**

CE7. Have you vaped e-cigarettes containing cannabis <u>products</u> in the past year?

	1. Yes	
	2. No	**(Skip to question CE10)**

CE8. Have you vaped e-cigarettes containing <u>cannabis products</u> in the past month?

	1. Yes	
	2. No	**(Skip to question CE10)**

CE9. How many days in the past 30 days did you vape e-cigarettes containing <u>cannabis products?</u>

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

CE10. Have you ever vaped e-cigarettes containing only flavoring <u>products</u>?

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

CE11. Have you vaped e-cigarettes containing <u>only flavoring products</u> in the past year?

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

CE12. Have you vaped e-cigarettes containing <u>only flavoring products</u> in the past month?

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

CE13. 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 (AL): I am going to ask you some questions about the use of alcoholic beverages. Unless I say otherwise, for all of the questions, think about wine, beer, and hard liquor such as whisky, gin, rum, vodka, tequila, brandy, etc.

THE COUNTRY SHOULD ADAPT THE LIST OF ALCOHOLIC BEVERAGES TO ITS SITUATION

AL1. Have you ever drunk an alcoholic beverage?

	1. Yes	
	2. No	(skip to TR1)
	9. Don't know/no opinion	

AL2. How old were you when you drank for the first time?
(Leave blank if the respondent only had a single sip to taste or try.)

	Years old

AL3. When was the first time you drank alcoholic beverages?

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

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

	1. Yes	
	2. No	**(Skip to** TR1)
	9. Don't know/no opinion	

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

	1. Yes	
	2. No	**(Skip to** AU1)
	9. Don't know/no opinion	

AL6. Thinking only about the past 30 days, how many days did you drink alcohol?

	Number of days

AL7. In the past 30 days, what type(s) of alcoholic beverage(s) did you drink, and how often?
(*Indicate the answer by marking with an "X" only one option for each alcoholic beverage)*

	Every day 1	Some weekdays 2	Weekends 3	Never 4
a) Beer	1	2	3	4
b) Wine	1	2	3	4
c) Hard liquor (whisky, vodka, rum, gin, tequila, etc.)	1	2	3	4

NOTE: EACH COUNTRY SHOULD LIST THE TYPES OF HARD LIQUOR (HIGHER ALCOHOL LEVEL) DRUNK MOST OFTEN IN THE COUNTRY

AL8. Over the past 2 weeks, how many times have you had 5 drinks or more on a single occasion?

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

AL9. Thinking about the past 12 months, did you drive a car or other vehicle after drinking an alcoholic beverage?

	1. I don't have a driver's license
	2. No
	3. Yes, once
	4. Yes, several times
	9. Don't know/no opinion

AUDIT (AU) QUESTIONNAIRE ON ALCOHOL USE. SHOW CARD NO. 4 IN ORDER TO ESTIMATE THE NUMBER OF DRINKS

1 drink	1 drink and a half	6 drinks	18 drinks
-One bottle or 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	Half a liter of beer	One bottle of wine (750 cc.)	One bottle of hard liquor (750 cc.)
	3 drinks	**8 drinks**	
	One liter of beer	One case of wine (1 liter)	

	Never	Monthly or less	2 to 4 times a month	2 to 3 times a week	4 or more times a week
AU1. How often do you have a drink containing alcohol? (If **"**never" skip to TR1)	0	1	2	3	4

	1-2 drinks	3-4 drinks	5-6 drinks	7-8 drinks	9 or more drinks
AU2. How many drinks containing alcohol do you have on a typical day when you are drinking? SHOW CARD NO. 4	0	1	2	3	4

	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
AU3. How often do you have 6 or more drinks on one occasion? SHOW CARD NO. 4	0	1	2	3	4
AU4. How often during the past 12 months have you found that you could not stop drinking once you had started?	0	1	2	3	4
AU5. During the past 12 months, have you ever failed to do what was normally expected of you because of drinking?	0	1	2	3	4
AU6. 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?	0	1	2	3	4
AU7. How often during the past 12 months have you felt guilt or remorse after drinking?	0	1	2	3	4
AU8. In the past 12 months, have you ever been unable to remember what happened the night before because you had been drinking?	0	1	2	3	4

	No		Yes, but not in the past year		Yes, in the past year
AU9. Have you or someone else been injured as a result of your drinking?	0		2		4
AU10. Has a relative, friend, doctor, or other health professional been concerned about your drinking or suggested that you cut down?	0		2		4

TRANQUILIZERS

The following questions have to do with tranquilizers like those shown on the card. SHOW CARD NO. 5:

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 <u>without a prescription</u>? ☐ 1. Yes ☐ 2. No (skip to TR9)	**TR3. How old were you <u>the first time</u> you took a tranquilizer <u>without a prescription</u>?** ☐ **Years old**

TR4. When was the <u>first time</u> you took a tranquilizer <u>without a prescription</u>?

	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 <u>12 months</u>?

	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?
<u>SHOW CARD NO.5.</u> 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** (Rohipnol)
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 <u>without a prescription</u>?
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 <u>past 30 days</u>?

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

TR9. How easy or difficult do you think it would be to obtain a tranquilizer without a prescription if you wanted to?

	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 TO REFLECT THE NAMES USED LOCALLY IN THE COUNTRY.

STIMULANTS

The following questions have to do with prescription stimulants like the ones shown on this card. SHOW CARD NO.6

ES1. Have you ever taken a stimulant <u>because a doctor prescribed it for you</u>?

	1. Yes
	2. No

ES2. Have you ever taken a stimulant without a doctor's prescription?	ES3. How old were you when you first took a stimulant without a prescription?

	1. Yes	
	2. No	(skip to ES9)

	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	

ES6. Which stimulants have you taken in the past 12 months? SHOW CARD NO. 6
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 using some prescription analgesics or painkillers like those shown on the card.

SHOW CARD NO. 7.

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?
SHOW CARD NO. 7. 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, Roxicodone, 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?
SHOW CARD NO. 7. 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, Roxicodone, 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, Contin)
AN4.10		Other:

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

	Years old

AN6. When was the <u>first time</u> you took an analgesic <u>without a prescription</u>?

	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. How easy or difficult do you think it would be to obtain an analgesic without a prescription if you wanted to?

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

MARIJUANA (MA): Now, I am going to ask you some questions about using other substances. We'll begin with marijuana.

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

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

	1. Yes	(skip to MA6**)**
	2. No	

MA5. In the past, how often did you use marijuana?

	9. Only once	(skip to MA13**)**
	10. Several times during a year	(skip to MA13**)**
	11. Several times for several years	(skip to MA13**)**
	12. Several times a month, for a year	(skip to MA13**)**
	13. Several times a month for several years	(skip to MA13**)**
	14. Every day, for one year	(skip to MA13**)**
	15. Every day for several years	(skip top MA13**)**

MA6. How often in the past 12 months have you smoked 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

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

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

	1. Yes	
	2. No	(skip to MA13**)**

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

	1. Only once
	2. Several times in the past 30 days
	3. Several times a week
	4. Every day

MA10. Thinking about the past 30 days, how many marijuana cigarettes (joints or spliffs) do you smoke on a typical day when you are using? (if less than 1, write 1)

	Number of joints

MA11. As far as you know, how much does a marijuana cigarette or joint (spliff) cost?

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

MA12. How much did you spend on marijuana in the past 30 days?

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

MA13. 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, etc.					
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 drug-dependant persons who commit non-violent crimes to enter court-supervized treatment programs instead of prisons.					
d) Allow marijuana cultivation for scientific research					

MA14. Some countries have permitted marijuana for recreational use. Have you heard about this?

Yes	
No	(skip to MA16)
Don't know/no opinion	(skip to MA16)

MA15. Do you agree or disagree with the laws in those countries that permit the recreational use of marijuana?

Yes
No
Don't know/no opinion

MA16. Would you agree with having a law allowing marijuana for recreational use in [INSERT NAME OF COUNTRY HERE]?

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

HASHISH (HA): Now, I would like to ask you some questions about the use of hashish.

HA1. Have you ever used hashish?

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

HA2. How old were you when you used hashish <u>for the first time</u>?

	Years old

HA3. When was the <u>first time</u> 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 <u>past 12 months</u>?

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

HA5. How often have you used hashish in the past 12 months?

	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): Now, I would like to ask you some questions about cocaine.

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, how many days did you use cocaine?

	Number of days (from 0 to 30)

CO8. How many hits/grams/ounces of cocaine do you use in a month? Use best guess to estimate

	Number of (indicate measure used) ________________

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) Countries should edit the name of the substance and the questions.

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 past 12 months?

	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 past 30 days?

	Yes	
	1. 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 opinio**

Notes:

- ✓ **EACH COUNTRY SHOULD USE THE NAME GIVEN TO THIS SUBSTANCE LOCALLY**
- ✓ **IF THE COUNTRY DECIDES NOT TO INCLUDE THE MODULE ON CRACK, THEN IN QUESTIONS PB1, PB4, and PB6, THE INSTRUCTION "SKIP TO CR1" SHOULD BE CHANGED TO READ "SKIP TO EX1."**

CRACK (CR)

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

	1. Yes	
	2. No	(skip to EX1**)**

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 EX1**)**

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 EX1**)**

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

	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 money did you spend on crack in the past 30 days?

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

NOTE: EACH COUNTRY SHOULD USE THE TERM MOST COMMONLY USED.

MDMA – ECSTASY (EX): Going on to another type of substance,

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 first time 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 past 12 months?

	1. Yes	
	2. No	(skip to LS1)

EX5. How often did you use ecstasy 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

EXT6. Have you used ecstasy in the past 30 days?

	1. Yes	
	2. No	(skip to LS1)

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

	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

LSD (LS): Let's talk about LSD:

LS1. Have you ever used LSD?

	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). Now I am going to ask you some questions about inhalant use. I am going to show you this card. SHOW CARD NO. 8.

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 first time you used inhalants?

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

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

	1. Yes	
	2. No	(skip to PO1)

IN5. How often did you use inhalants 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

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

	1. Yes	
	2. No	(skip to PO1)

IN7. And those times you used inhalants in the past 30 days 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 Interne
	10. Other Which? ………………

POPPERS (PO)

PO1. Have you ever used poppers?

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

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

	Years old

PO3. When was the <u>first time</u> 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 <u>past 12 months</u>?

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

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 <u>past 30 days</u>?

	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 IN8 ON INHALANTS AND SHOULD ALSO BE ADDED TO THE LIST ON CARD 8.**
- ✓ **COUNTRIES MAY ADD ADDITIONAL QUESTIONS FOR SPECIFIC SUBSTANCES.**
- ✓ **COUNTRIES MAY ADD A COMPLETE MODULE FOR A SPECIFIC SUBSTANCE THAT NEEDS SPECIAL ATTENTION.**

OTHER DRUGS (OD). To finish with this topic, we're now going to ask you some questions about using other substances.

When was the last time you used any of these substances?	Never used (1)	More than 1 month ago but less than 1 year ago (2)	More than 1 year ago (3)	In the past month (4)
OD1. Methamphetamine: Meth, ice, crystal	1	2	3	4
OD2. Amphetamine: fet, speed				
OD3. Heroin: Paste, "H," white powder, skag, and tar).	1	2	3	4
OD4. Ketamine: Keta, vitamin K, super K, CK o Calvin Klein, Mary Kay or María-K.	1	2	3	4
OD5. Opium	1	2	3	4
OD6. GHB: Liquid X, liquid ecstasy, Georgia homeboy, Oop, Gamma-oh, grievous bodily harm, Mils, "G," Liquid G, Fantasía.	1	2	3	4
OD7. Synthetic cannabinoids: Synthetic marijuana, Spice, K2, Joker, Black Mamba, Kush, or Kronic.	1	2	3	4
OD8. Synthetic Cathinones: bath salts	1	2	3	4
OD9. Aminoindanes: MDAI gold	1	2	3	4
OD10. Phencyclidine: PCP, angel dust, embalming fluid, hog, killer weed, love boat, ozone, peace pill, super weed, rocket fuel, estrafalaria.	1	2	3	4
OD11. Phenethylamines (Europa, 4-FMP, RDJ, 4-MMA, Methyl-MA, 2C-C-NBOMe, Bomb, N-Bomb N, 251, Nexus, 2C-E, and Blue mystic)	1	2	3	4
OD12. Piperazines: BZP, mCPP, A2, Legal X, and Pep X.	1	2	3	4
OD13. 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.	1	2	3	4
OD14. Anabolic steroids	1	2	3	4
OD15. Caffeine products: caffeine pills, energy drinks, powdered caffeine	1	2	3	4
OD16. Lean, purple drank o sizzurp (Sprite mix, cough syrup, and pastilles)	1	2	3	4

NOTE: COUNTRIES MAY DELETE SOME OF THESE OR INCLUDE OTHERS.

MODULE 3: PERCEIVED RISK (PR) AND FACTORS ASSOCIATED (FA) WITH DRUG USE. Moving on to another topic,

What kind of risk do you think someone runs if he or she does any of the following things? **MARK THE ANSWER WITH AN X IN ONE OF THE BOXES FOR EACH LINE**	No risk 1	Slight risk 2	Moderate risk 3	High risk 4	Don't know 9
PR1. Smoking one or more packs of cigarettes a day	1	2	3	4	9
PR2. Drinking five or more drinks of alcohol a day	1	2	3	4	9
PR3. Taking medication without a doctor's prescription frequently	1	2	3	4	9
PR4. Inhaling or sniffing glue, paint, varnish, deodorants, gasoline, or similar once	1	2	3	4	9
PR5. Often inhaling glue, paint, varnish, deodorants, gasoline, petrol, or similar	1	2	3	4	9
PR6. Smoking marijuana several times a month	1	2	3	4	9
PR7. Smoking marijuana often	1	2	3	4	9
PR8. Smoking cocaine base paste or crack once	1	2	3	4	9
PR9. Smoking cocaine base paste or crack often	1	2	3	4	9
PR10. Snorting cocaine sometimes	1	2	3	4	9
PR11. Snorting cocaine often	1	2	3	4	9
PR12. Inhaling poppers once	1	2	3	4	9
PR13. Inhaling poppers often	1	2	3	4	9
PR14. Using ecstasy sometimes	1	2	3	4	9
PR15. Using ecstasy often	1	2	3	4	9

Notes: Countries may add other substances or adapt the name of the substance for PR8 and PR9

PR16. Did you ever feel curious about trying a drug? *(marijuana, cocaine, cocaine base paste, ecstasy, or similar)*

	1. Yes
	2. No
	3. Perhaps
	4. Doesn't apply, have already used
	9. Don't know/no opinion

PR17. If you had the chance, would you try drugs? *(marijuana, cocaine, cocaine base paste, ecstasy, or similar)*

	1. Yes
	2. No
	3. Perhaps
	4. Doesn't apply, have already used
	9. Don't know/no opinion

ONLY FOR THOSE WHO SAID THEY HAD USED ALCOHOL OR OTHER DRUGS IN THE PAST 12 MONTHS. If not used, mark <u>NEVER</u> in FA1, and skip to FA21. Otherwise, begin with FA2.

During the past 12 months, how often did you experience the following situations as a result of using alcohol or other drugs?	Never 1	Rarely 2	From time to time 3	Fairly often 4	Very often 5
FA1. Never used alcohol or other drugs **(skip to FA21)**					
FA2. Performing poorly on an important test or project	1	2	3	4	5
FA3. Getting into trouble with the police	1	2	3	4	5
FA4. Getting involved in an intense argument or fight	1	2	3	4	5
FA5. Memory loss	1	2	3	4	5
FA6. Having domestic problems	1	2	3	4	5
FA7. Being taken advantage of sexually or having someone take liberties with you	1	2	3	4	5
FA8. Taking sexual advantage of someone or taking liberties with another person	1	2	3	4	5
FA9. Trying in vain to stop drinking or stop taking illicit drugs	1	2	3	4	5
FA10. Self-harm (cutting)	1	2	3	4	5
FA11. Thinking seriously about committing suicide	1	2	3	4	5

STILL THINKING ONLY ABOUT THE PAST 12 MONTHS	No 1	Yes, for alcohol only 2	Yes, for drugs only 3	Yes, for alcohol and drugs 4	Don’t know/no opinion 9
FA12. ¿Have you been arrested for drinking alcohol or using a drug?	1	2	3	4	9
FA13. Have you been arrested for possession or trafficking a drug?	1	2	3	4	9
FA14. Have you missed work, university, or school because you felt ill because of using alcohol or another drug?	1	2	3	4	9
F16.Were you fired from your job or dropped out of university or school because of problems related either directly or indirectly to alcohol or other drugs?	1	2	3	4	9
FA17. Have you gone to a hospital, clinic, or doctor’s office because of alcohol or other drug use problems?	1	2	3	4	9
FA18. Have you been involved in a traffic accident or an accident at home or at work that may have been due, directly or indirectly, to alcohol or other drug use?	1	2	3	4	9
FA19. Have you gone into debt or had to sell something important so that you could buy alcohol or another drug that you needed?	1	2	3	4	9
FA20. Did you have to leave home, face a break-up with your spouse or partner, or separate from your loved ones because of alcohol or another drug?	1	2	3	4	9

INTERVIEWER: ASK QUESTIONS FA21 to FA23 ONLY OF PEOPLE AGED 18 OR OLDER

STILL THINKING ONLY ABOUT THE PAST 12 MONTHS	No 1	Yes, for alcohol 2	Yes, for drugs 3	Yes, for alcohol and drugs 4	Don’t know/no opinion 9

FA21. Were you the victim of a robbery or assault committed by someone who appeared to be under the influence of alcohol or another drug?	1	2	3	4	9
FA22. Have you been intimidated, threatened, beaten up, pushed, or kicked by someone who was under the influence of alcohol or another drug?	1	2	3	4	9
FA23. Have you been sexually intimidated or forced to perform a degrading sexual act by someone who was under the influence of alcohol or another drug?	1	2	3	4	9

INTERVIEWER: ALL RESPONDENTS SHOULD ANSWER QUESTIONS FA24 TO FA27.

FA24. Does anyone at home, as far as you know, have the habit of smoking every day?
Not including yourself

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

FA25. At home, as far as you know, does anyone usually drink alcohol every day, even if it's just one glass?
Not including yourself

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

As far as you know, does anyone in your house use any of these drugs?
If you do, do not answer

	FA26. MARIJUANA			FA27. COCAINE			FA28. COCAINE BASE PASTE*
	1. Yes			1. Yes			1. Yes
	2. No			2. No			2. No
	9. Don't know/no opinion			9. Don't know/no opinion			9. Don't know/no opinion

As far as you know, do any of your close friends or the people you see a lot use any of these drugs?

	FA29. MARIJUANA			FA30. COCAINE			FA31. COCAINE BASE PASTE*
	1. Yes			1. Yes			1. Yes
	2. No			2. No			2. No
	9. Don't know/no opinion			9. Don't know/no opinion			9. Don't know/no opinion

*Note: In FA28 and FA31, use the name of the substance commonly used (cocaine base paste, crack)

MODULE 4: INFORMATION ON TREATMENT (TT) OPTIONAL

TT1. Have you ever received any type of treatment for alcohol or drug use?
(do not include treatment for stopping smoking)

	1. Yes
	2. No **(skip to TD1)**
	3. I have never used alcohol or other drugs **(skip to TD1)**

TT2. In the past 12 months, have you received treatment for alcohol or drug use?

	1. Yes **(Answer TT3 and TT4 and skip to TD1)**
	2. No **(skip to TT5)**

TT3. Was the treatment you received for alcohol use, drug use, or both?

	1. For alcohol only
	2. For another drug, only
	3. For alcohol and other drug(s)

TT4. Thinking about the last time you were in treatment to stop using alcohol or another drug. ADAPT TO OPTIONS AVAILABLE IN THE COUNTRY.

	1. In-patient in a rehabilitation center or therapeutic community
	2. Outpatient health clinic or rehabilitation center
	3. Private clinic
	4. Self-help groups like Alcoholics Anonymous (AA)
	5. Other

(skip to TD1)

TT6. How long were you in treatment this last time for treatment for drugs or alcohol?

	Years		**Months**

TT5. Over the past 12 months, have you felt the need for help or treatment to reduce or stop drinking alcohol?

	1. Yes
	2. No
	3. Not applicable, did not use **alcohol** in the past 12 months

TT6. Over the past 12 months, have you felt the need for help or treatment to reduce or stop using another drug?

	1. Yes
	2. No
	3. Not applicable, did not use **other drugs** in the past 12 months

MODULE 5: INFORMATION ON DRUG TRAFFICKING/DEALING (TD)

As far as you know, how many of the following things are happening in your neighborhood? *SHOW CARD NO. 9*	A great deal 1	Some 2	Little 3	None 4	Don't know/no opinion 9
TD1. Drug trafficking/dealing	1	2	3	4	9
TD2. Breaking and entering homes	1	2	3	4	9
TD3. Graffiti, broken street lights, or things of that kind	1	2	3	4	9
TD4. Drug use in public places (in the street or parks)	1	2	3	4	9
TD5. Armed robbery or mugging in the street	1	2	3	4	9
TD6. Young people standing around or loitering at street corners or on the block	1	2	3	4	9
TD7. Shootouts and violence with firearms	1	2	3	4	9

TD8. How would you describe drug use in your neighborhood?

	1. Serious
	2. Slight
	3. There is none
	9. Don't know/no opinion

TD9. Do you know whether there are places, streets, alleys, or other areas in your neighborhood where drugs are sold?

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

TD10. How easy or difficult would it be for you to obtain any of the following drugs? **(Put an X in the correct column for each drug**)	It would be easy 1	It would be difficult 2	It would not be able to obtain 3	Don't know 4
1. Marijuana	1	2	3	4
2. Cocaine	1	2	3	4
3. Cocaine base paste	1	2	3	4
4. Ecstasy	1	2	3	4
5. LSD	1	2	3	4
6. Heroin	1	2	3	4

TD11. When was the last time you were offered any of these drugs, either to try or to buy? **(***Mark an X for the response for each drug)*	During the past 30 days 1	More than 1 month ago but less than 1 year ago 2	More than 1 year ago 3	Have never been offered 4
1. Marijuana	1	2	3	4
2. Cocaine	1	2	3	4
3. Cocaine base paste	1	2	3	4
4. Ecstasy	1	2	3	4
5. LSD	1	2	3	4
6. Heroin	1	2	3	4

Note: In QUESTIONS TD10 and TD11, each country should name the substance used in the country for response No. 3.

MODULE 6: DRUG PREVENTION AND CONTROL MEASURES (PC)

Do you agree with the following? ***SHOW CARD NO. 10***	Fully agree 1	Agree 2	Disagree 3	Very much disagree 4	Don't know/no opinion 9
PC1. Most young people smoke marijuana	1	2	3	4	9
PC2. An addict can never recover	1	2	3	4	9
PC3. Drug use and drug trafficking are the main causes of crime in this country	1	2	3	4	9
PC4. Prevention of drug use should be obligatory in all schools	1	2	3	4	9
PC5. Drug dealing should not be punished with prison sentences	1	2	3	4	9
PC6. The police are easily corrupted by drug traffickers or dealers	1	2	3	4	9
PC7. Marijuana should be legal for everyone over the age of 18	1	2	3	4	9
PC8. A drug test should be required to hold public office	1	2	3	4	9
PC9. Drugs have made criminals more violent	1	2	3	4	9
PC10. People who use drugs should be left alone: after all, it's up to each person to decide whether to do drugs or not	1	2	3	4	9

PC11. Do you think that drug use in this country has increased, stayed the same, or decreased in recent years?

	1. Has increased
	2. Has stayed the same
	3. Has decreased
	9. Don't know/no opinion

PC12. Do you think that in the coming years, the drug problem is going to get worse, remain the same, or decrease?

	1. Will increase
	2. Will stay the same
	3. Will decrease
	9. Don't know/no opinion

Do you think that the drug issue is one of the core concerns of the following authorities?

PC13. The city or town authorities		**PC14. Parliament or Congress**		**PC15. Central government authorities**	
	1. Yes		1. Yes		1. Yes
	2. No		**2.** No		2. No
	9. Don't know/no opinion		**9.** Don't know/no opinion		9. Don't know/no opinion

Note: In QUESTIONS PC13 and PC14, each country should adjust the questions to its circumstances.

PC16. This card shows some possible measures authorities could take to address the drug problem. In your opinion, which ones would be most effective addressing the drug problem in your city or town? (MARK A MAXIMUM OF 3). ***SHOW CARD NO. 11***

	1. Increase punishment for drug traffickers/dealers
	2. Punish drug users
	3. Carry out prevention programs in the schools
	4. Conduct prevention campaigns in the media
	5. Increase funding for the police
	6. Increase funding for addiction treatment
	9. Don't know/no opinion

PC17. Do you know of any drug prevention program(s) in your neighborhood or town/city?

	1. Yes
	2. No **(skip to PC19)**
	9. Don't know/no opinion **(skip to PC19)**

PC18. How would you rate that drug prevention program?

	1. Good
	2. Middling/so so
	3. Poor
	9. Don't know/no opinion

PC19. Over the past 12 months, have you talked to your teenage children about the dangers and problems associated with drug and alcohol use?

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

PC20. In the past 12 months, have you attended a meeting in your children's school(s) where they discussed drug and alcohol prevention?

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

PC21. Do you know what the (Name of National Drug Agency) is?

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

PC22. And do you know what that agency does?

INTERVIEWER: A CORRECT RESPONSE WILL LINK THE NATIONAL DRUG AGENCY WITH PREVENTION/TREATMENT OF DRUG USE AND/OR CONTROL OF DRUG TRAFFICKING OR DEALING

	1. Correct response
	2. Incorrect response
	9. Don't know/no opinion

Now I am going to ask you some questions about alcohol:

Do you agree or not with the following measures? ***SHOW CARD NO. 12***	Fully agree	Agree 2	Disagree 3	Very much disagree 4	Don't know/no opinion 9
PC23. Increase taxes on alcoholic beverages	1	2	3	4	9
PC24. Reduce the number of places that sell alcoholic beverages	1	2	3	4	9
PC25. Raise the age limit for buying alcohol	1	2	3	4	9
PC26. Increase penalties for people who drink and drive (drive when drunk)	1	2	3	4	9
PC27. End sponsorship of sports events or teams by alcoholic beverage companies	1	2	3	4	9
PC28. Reduce the hours when alcohol can be sold	1	2	3	4	9

And now I would like to ask you some questions about illicit drugs such as marijuana, cocaine, and cocaine base paste:

Do you agree or not with the following measures? ***SHOW CARD NO. 13***	Fully agree 1	Agree 2	Disagree 3	Very much disagree 4	Don't know/no opinion 9
PC29. Do not prosecute first offenders provided they agree to enter a treatment program	1	2	3	4	9
PC30. Give the same sentence to people who deal in marijuana as those who are dealing in cocaine or crack	1	2	3	4	9
PC31. Authorize screening of students at the entrance to their school to see whether they have drugs in their possession	1	2	3	4	9

MODULE 7: INFORMATION ON EMPLOYMENT (PL)
ONLY FOR THOSE WHO GAVE RESPONSES 1, 2, or 3 TO QUESTION 1 OF MODULE 1

PL1. What is your position in your current job? ***SHOW CARD NO. 14***

	1.	Businessman, manager of a large business
	2.	Employer/owner of a medium-sized business
	3.	Employer/owner of a small business
	4.	Self-employed
	5.	Professional in the private sector
	6.	Government professional
	7.	Private sector employee
	8.	Private sector worker
	9.	Public or municipal employee
	10.	Public or municipal worker
	11.	Domestic service
	12.	Work in a family business without pay
	99.	Don't know/no opinion

PL2. How large is the company or business where you work? Include all of the offices or branches of your company in the country. (READ responses)

	1.	1- 9 people
	2.	10-49 people
	3.	50-199 people
	4.	200-999 people
	5.	1,000 or more
	9.	Don't know/no opinion

PL3. Thinking about the past 12 months, were you unemployed at any time?

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

PL4. In the past 12 months, were you involved in any type of workplace accident?

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

PL5. In the past 30 days, how many full workdays did you miss because of illness or an accident?

	Number of days

Enter 99 if responds Don't know/no opinion

PL6. Thinking about the past 3 years, how many employers have you had?

INTERVIEWER: IF HE/SHE HAS BEEN SELF-EMPLOYED DURING THIS PERIOD, THE RESPONDENT SHOULD COUNT HIMSELF AS AN EMPLOYER.

	1. Number of employers
	2. Not applicable
	9. Don't know/no opinion

PL7. Still thinking about the past 3 years, were you dismissed (fired) from any job during that time? *If the respondent voluntarily left the job or finished a temporary job, this is not dismissal.*

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

PL8. Now I am going to ask you some questions about your current job.
Can you tell me whether, in your current job, there are any special rules about the use of alcohol or other drugs by employees?

INTERVIEWER: IF THE PERSON HAS MORE THAN ONE JOB, HE/SHE SHOULD ANSWER CONCERNING THE SAME JOB FOR ALL JOB-RELATED QUESTIONS.

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

PL9. Have you been given any information about alcohol or drug use prevention in your current job?

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

PL10. In your current job, is there any kind of program to help an employee who has a problem with the use of alcohol or other drugs?

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

HOUSEHOLD SURVEY SHOW CARDS

CARD NO. 1 (question 2 MODULE 1)
ADAPT FOR EACH COUNTRY

1. Senior government official, civil servant, senior management of a company. Senior officer in the armed forces. Owner of a large business or owner of a large farm.
2. Professional, scientist, university professor, armed forces officer, owner of medium-sized business (6-25 employees), owner of medium-sized farm.
3. Technician, teacher, small business owner, middle management, administrative or service personnel, small farmer.
4. Office staff in business, industry, services, civil service or similar.
5. Small business owner, small farmer.
6. Skilled worker (plumber, hairdresser, airconditioning technician, etc.)
7. Laborer, farm worker.
8. Apprentice
9. Armed forces enlisted personnel, security personnel.
10. Domestic service /street vendor.
11. Retired on an old age pension, or disability pension or similar
99. Don't know/no opinion

CARD NO. 2 (question 5 MODULO 1)
USE LOCAL CURRENCY

1. Less than USD200
2. $201-$400
3. $401-$600
4. $601-800
5. $801-$1,000
6. $1,001-1,500
7. $1,501-$2,000
8. $2,000- $2,500
9. $2,501-$3,000
10. $3,001-3,500
11. $3,501-$4,000
12. $4,001-5,000
13. More than $5,000

CARD NO. 3 (questions 6 and 7 MODULE 1)
ADAPT FOR EACH COUNTRY

1. Never attended school
2. Did not complete primary school
3. Completed primary school
4. Did not complete secondary school
5. Completed secondary school
6. Technical/vocational school
7. Higher education/did not complete university
8. Higher education/completed university
9. Graduate/post-graduate

CARD NO. 4 (QUESTIONS for AUDIT AU2 and AU3 MODULE 2)
TYPES OF LIQUOR SHOULD BE ADAPTED FOR EACH COUNTRY

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

CARD No. 5 (LIST OF TRANQUILIZERS FOR QUESTIONS TR1 TO TR9 MODULE 2)
ADAPT FOR EACH COUNTRY

Alprazolam (Alprazolam, Intensol, Xanax and Xanax XR)
Clonazepam (Klonopin)
Diazepam (Diastat AcuDial, Diazepam, Intensol, Diastat, Valium)
Flunitrazepam (Rohipnol)
Chlordiazepoxide (Klopoxid, Libritabs, Librium, Monthural, Multum, Novapam, Risolid, Silibrin, Sonimen, Tropium, Zetran)

CARD NO. 6 (LIST OF STIMULANTS FOR QUESTIONS ES1 TO ES9 MODULE 2)
ADAPT FOR EACH COUNTRY

Methylphenidate (Ritalin, Concerta)
Phenmetrazine (Preludin)
Amphetamine (Adderall, Adderall XR, Mydayis, Evekeo, Zenzedi, Dexedrina)
Dextroamphetamine (Dexedrine, DextroStat)
Pemoline (Cylert)
Other:

CARD NO. 7 (LIST OF ANALGESICS FOR QUESTIONS AN1 TO AN9 MODULE 2)
ADAPT FOR EACH COUNTRY

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, MS Contin)

CARD NO. 8 (LIST OF SUBSTANCES FOR QUESTIONS IN1 TO IN8 ON INHALANTS MODULE 3)
ADAPT FOR EACH COUNTRY

Glue
Paint
Varnish
Deodorant
Ether
Toluene
Gasoline/benzene/petrol
Paraffin

CARD NO. 9 (questions TD1 to TD7 MODULE 5)

1. Drug dealing/trafficking
2. Breaking and entering homes
3. Scribbling graffiti on the walls, damage to street lighting, or things of that kind
4. Taking drugs in a public place such as the street or squares or parks
5. Armed robbery or mugging in the street
6. Young people standing around or loitering at street corners
7. Shootouts and violence with firearms

CARD NO. 10 (questions PC1 to PC10 MODULE 6)

1. Most young people smoke marijuana
2. An addict can never recover
3. Drug use and drug trafficking are the main causes of crime in this country
4. Prevention of drug use should be obligatory in all schools
5. Drug dealing should not be punished with prison sentences
6. The police are easily corrupted by drug traffickers or dealers
7. Marijuana should be legal for everyone over the age of 18
8. A drug test should be required in order to hold public office
9. Drugs have made criminals more violent
10. People who use drugs should be left alone: after all, it's up to each person to decide whether to do drugs or not.

CARD NO. 11 (question PC16 MODULE 6)

1. Increase punishment for drug traffickers/dealers
2. Punish drug users
3. Carry out prevention programs in the schools
4. Conduct prevention campaigns in the media
5. Increase funding for the police
6. Increase funding for treatment of addicts

CARD NO. 12 (questions PC23 to PC28 MODULE 6)

1. Increase taxes on alcoholic beverages
2. Reduce the number of places that sell alcoholic beverages
3. Raise the age limit for buying alcohol
4. Increase penalties for people who drink-drive (drive when drunk)
5. End sponsorship of sports events or teams by alcoholic beverage companies
6. Reduce the hours when alcohol can be sold

CARD NO. 13 (questions PC29 to PC33 MODULE 6)

1. Do not prosecute first offenders provided they agree to enter a treatment program
2. Give the same sentence to people who deal marijuana as those who are dealing in cocaine or crack
3. Allow the use of cannabis for therapeutic purposes
4. Authorize screening of students at the entrance to their school to see whether they have drugs in their possession
5. Penalize the possession and use of marijuana even in small quantities and even if for personal use

CARD NO. 14 (question PL1 MODULE 7)

1.	Businessman, manager of a large business
2.	Employer/owner of a medium-sized business
3.	Employer/owner of a small business
4.	Self-employed
5.	Professional in the private sector
6.	Government professional
7.	Private sector employee
8.	Private sector worker
9.	Public or municipal employee
10.	Public or municipal worker
11.	Domestic service
12.	Work in a family business without pay

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*, *marital status, socio-economic level, and geographical distribution,* if the study has coverage at that level, 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**			
12-17			
18-34			
35 and older			
TOTAL	**n**	**N**	**100**

*Other variables may be included, such as marital status, socio-economic level
** Another classification could be used, such as 12-17, 18-24, 25-34, 35-44, 45 and older

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 people aged 12-65 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 5,000 cases (corrected for possible rejections or other contingences.) Based on this sample size, the blocks are selected and then the households, and finally the individuals. It is possible that the household selected no longer exists, the map has not been updated, no-one is at home when the interviewer calls, or that the individual selected does not agree to be interviewed, etc. Therefore, even though provision may have been made for certain contingencies, the number of successful interviews may be less than planned—4,800, 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 the interview. It is important that those directing the study, that is, the professional staff of the National 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 4,700 valid cases.

c) **Expansion factor**: as noted above and as explained more fully in Annex 2, there are different sampling designs. A household survey in particular corresponds to what is called "complex sampling." This involves different stages in the process of selecting the individuals 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 person becomes 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 individuals 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 (4,700 in the example here) will produce the total number of individuals in the target population (people aged 12-65), shown as **N** in the table above.

6.2 ESTIMATING INDICATORS ON SUBSTANCE USE

The second area of analysis is the substance use prevalence and incidence estimates. Note that studies based on samples only allow for estimates of what occurs in the population under investigation and do not represent an exact, error-free determination of drug use in the population. Therefore, some uncertainty or error will be associated with these estimates. Researchers estimate the degree of error through the sample data. When reporting, researchers should explicitly state the degree of error in the analysis.

The estimated error is calculated as the ***standard error***, which in turn is used to calculate the ***Confidence Intervals***. The unknown population indicators that we are trying to estimate are termed **parameters.** We, therefore, have three concepts in the estimation process:

1. The **parameter** to be estimated (for example, the past month prevalence of alcohol use in the country's population aged 12-65),

2.- The **estimate** of the parameter according to 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 1,000,000 people aged 12-65, and that the effective sample is 2,000 individuals. By correctly using the expansion factors and the sample design (further details in Annex 2,) the results of the study show that 40% of the individuals reported that they had used an alcoholic beverage in the past month, with a standard error of 2%. Thus, the prevalence of alcohol use in the past month is 40%, with a 95% Confidence Interval of between 36.08% and 43.92%. The 95% Confidence Interval to estimate a proportion (expressed as a percentage) of a population is expressed as:

$$\mathbf{p-1{,}96*ee(p)\ ;\ \ p+1{,}96*ee(p)} \qquad (1)$$

where:

- 1.96 is the value of the normal distribution of an estimate with a confidence level of 95%, and
- ee(p) is the **standard error** of the estimate,
- **p-1,96*ee(p)** as the ***lower limit*** **and p+1,96*ee(p)** is the ***upper limit*** of the 95% Confidence Interval, and
- **d=1,96*ee(p)** is called the **precision** of the estimate.

Thus, a Confidence Interval may be expressed as:

$$\mathbf{p-d\ ;\ \ p+d} \qquad (2)$$

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	40	36.08 – 43.92

As well as reporting the overall prevalence (40% 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, using dummy data:

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

Variables		Prevalence (%)	95% CI
Sex	Male	45	40.7 – 49.3
	Female	35	31.5 – 49.3
Age	12-17	20	14.5 – 25.5
	18-34	50	44.7 – 55.3
	35 and older	45	39.1 – 50.9
Total		**40**	**36.1 – 43.9**

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 marital status, or else according to the perception of 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 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 know 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 perception of 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, which means:

- Determining the *geographical areas* (according to the size of cities/towns, urban/rural area), and

- Determining the *age group* (in accordance with SIDUC, these are individuals aged 12-65.)

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

Sampling frame: As we said in Section 3.3, the best way of reaching/finding people is by means of a **sampling of households** in the geographical areas defined. This leads us to define the **sampling frame** as the list of ***households*** in the geographical areas defined. The households are grouped into what we have called "blocks." These clusters must be mapped using the most up-to-date maps available. This information is usually available in the country's statistics institute, or in agencies that conduct household surveys. This ***up-to-date sampling frame is the basis of a research study*** such as that described in this Protocol.

Having defined the target population and developed the corresponding sampling frame, the means of obtaining the sample(s) will depend on the representativeness that is desired for the study. If the study has only ***national representation***, then there will be only **one sample**, which will be selected directly from the **sampling frame**.

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

The same procedure is used when estimates for a smaller geographical level are planned. If a country has 20 regions or departments, and it is decided to estimate indicators at that level, then the earlier sampling frame must be divided into 20 sampling frames, each containing information on the households in the cities in the respective region or department. A random sample of households should be obtained for each of these, ***large enough to produce estimates with acceptable errors.***

Regardless of the level of representativeness, as we saw earlier in Section 3.3, the sampling process (nationally or regionally) based on the sampling frame of blocks and households is generally done in three stages:

- ***First, a random selection of blocks*** in the geographical areas defined,
- ***Then a random selection of households*** in the blocks selected in the previous phase,
- ***Lastly, random selection of a person*** between the ages of 12 and 65 in each of the households selected in the second phase.

Random samples will thus be obtained of persons aged 12-65 ***for each of the geographical areas to be represented*** in the study and for which it is possible to obtain estimates with acceptable errors.

It is important to remember that generally speaking, blocks do not consist of the same number of households. Thus, for example, if there are three apartment buildings in one block, each building having 10 floors, with six apartments on each floor (i.e., six households), there will be 180 households in these buildings alone. The blocks must therefore be divided up according to their size, that is, according to the number of households in each block. This may mean adding a new condition when the primary selection units, i.e., the blocks, are being chosen.

Generating a priori the sampling frame according to a particular criterion is what is known as pre-stratification. However, when analyzing the survey data, it becomes necessary to generate subgroups in order to look at specific characteristics, such as males and females, or specific age groups, and so on. In these cases, we talk about post-stratification. The main difference between pre-stratification and post-stratification is that in the former, the size of the sample of each subgroup (stratum) is determined from the study design. On the other hand, the sizes of the samples obtained from a post-stratification may be insufficient to obtain robust estimates at the desired levels. Thus, for example, if the specific objectives called for the need for estimates by age group (12-17, 18-34 and 35 and older), let us suppose that 15% of the target population is in the 12-17 age group. It may therefore be the case that the sample size is not sufficient to obtain some estimates for this group with acceptable sampling errors.

It should be borne in mind that the discussion of representativeness, sample sizes and precision of the estimates is a discussion that forms part of the ***planning of the study***, and that the decisions taken about them, including levels of representativeness, will have a direct impact on the size of the study sample and therefore on the complexity of administration of the study and of course on its cost.

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

- The first point to be considered is the level of **representativeness** of the sample, that is, whether estimates are needed only at the national level or at a particular division of the country determined a priori during the planning of the study. For example, if estimates are wanted only at the national level, the sample size will be smaller than if estimates are also needed at the level of regions or provinces within the country.
- A second point is related to the **sampling or design method** that will be used.
- The third factor looks at the **variability** of the population with regard to the indicator considered to be the most important. For example, we might assume that the indicator of the greatest interest is the **prevalence of past year use of an illicit substance**, or the **past month prevalence of alcohol use**. Different magnitudes of the estimator will produce different sample sizes.
- A fourth point has to do with the **precision** of the estimates, which accounts for the width of the Confidence Interval or also the **coefficient of variation** associated with the estimate of interest. Greater precision of the estimates will require larger sample sizes.
- Lastly, the fifth factor is related to the desired **level of confidence** in the estimate. Again, the greater level of confidence of the estimate, the larger the sample size will be.

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

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

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

Another item mentioned above is the **variability** found in the population with respect to the variable under study. Thus, for example, if the variable is *use of an illicit drug in the past 12 months*, it may take in different forms in a country: in the population of one particular country, the prevalence (in %) for that variable may be 1%, while in another country, it may be 10% or 50%. If we call that prevalence P, then the variability will be given by the product of P and its complement, that is Q=100-P, or P*Q. A value of P=50% means that the population is divided into two equal parts, which means maximum variation/variability. This variation decreases to the extent that P moves away from 50%. The size of the sample is directly proportionate to variability, which means that ***the more variation, the larger the required sample size will be*.**

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

In summary, there are many important points to be considered when determining the size of the sample during the planning of the study, as we have described. One in particular is the sampling design: as stated earlier, the sampling design in a household survey carried out by means of a three-stage sampling of households as discussed is what is called ***complex sampling***, which is very different from the design known as simple random sampling, which would involve a sample frame based on people aged 12-65, rather than one based on households.

In order to explain how to determine the size of the sample, we go back to the theory of a simple random sample. When estimating a proportion (P) in simple random sampling, the variance of that estimate[11] is given by:

$$\boxed{\boldsymbol{V_{mas}(p) = \frac{P(1-P)}{n}}} \qquad (3)$$

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

Unlike simple random sampling, as stated earlier, the household study is based on **complex sampling**, because of its three-phase design. This means that the variance defined in equation (3) is not that of a complex sample; however, it is possible to make changes to this variance in order to reflect the sampling design that is used.

To do this, we must return to the drawing of the sample frame and the selection in the first phase of the sampling. The households that make up the sampling frame are grouped into clusters called "blocks," and therefore the people in these blocks share a common geographical space. Therefore, within these *clusters,* the units of analysis may not be totally independent of each other. In other words, the variable being studied, in this case, *substance use*, may be impacted by the overall conditions in the specific geographical area occupied by the clusters, which means that the responses of the people in them are not necessarily independent of each other. This situation means that there is a certain degree of correlation, known as ***intra-class or intra-cluster correlation coefficient*** (ICC), which is called **ρ**. In the case of a simple random sample, this coefficient[12] would equal zero.

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

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

[11] Throughout the annexes a sufficiently large population size is assumed and therefore the correction factor is not considered for the population.

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

where m=number of clusters and ρ=ICC.

In the case of a simple random sample, where the units are independent of each other, ρ=0 and thus deff=1. ***Deff in other cases will be greater than 1.***

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

$$\boxed{V_{mc}(p) = \frac{\mathbf{p(1-p)}}{\mathbf{n}} * ED} \quad (5)$$

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

$$\boxed{V_{mc}(p) = V_{mas}(p) * ED} \quad (6)$$

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

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

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

For this purpose, let us again consider the definition of precision that we saw above:

d=1.96*ee(p)

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

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

such that the size of the sample is defined as:

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

To solve this equation, it is necessary to:

i. Determine the **confidence level** of the estimate; in other words, determine the value of ***z***. Based on the approximation of the normal distribution associated with the sample distribution of a proportion, and given that 95% Confidence Intervals are those 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 used instead of the approximation to the normal distribution, particularly when working with statistical packages. One of these distributions is a binomial distribution.

ii. Determine a value for ***p***, that is, the most important indicator to be estimated in the study, such as, for example, prevalence of the use of an illicit drug in the past year. There are some options available for approximating this value: if the country has already conducted an equivalent study, then the prevalence from that study may be used. If the country does not have any previous equivalent studies, there might be some local studies that could be of help, or indicators could be used from countries with similar characteristics, or else regional or subregional averages given in the Reports on Drug Use prepared by the OID/CICAD/OAS. The worse option would be to resort to the criterion of "maximum variation", which occurs when **p** (prevalence) is 50%. This is the worst option since it is highly unlikely (highly undesirable) that the prevalence of the use of an illicit substance in the general population would have reached such a figure. In short, in order to properly identify this value, what is needed is local information. It also requires considerable judgment on the part of the researchers in the National Observatory on Drugs. But a value must be determined.

iii. The third component of formula (9) is value ***d***, or precision**.** This value determines the width of the Confidence Interval, and is closely related to the prevalence discussed in the point above—and in fact, it can be determined on the basis of the p-value. For example, to declare that **d** will not exceed 10% of **p**. That is, if we assume that p=10%, then d=1% such that the Confidence Interval is between 9% and 11%. However, if p is estimated at 20% then d=2 and the CI would be 18% to 22%. On the other hand, if it is determined that d shall not exceed 20% of the p-value, then d=2% in the first case and d=4% in the second.

iv. The last item at work in calculating the sample size is the **design effect**. Again, it is necessary here to have recourse to other sources of information, whether earlier studies in the country or similar studies in other countries. As discussed above, the value of deff is more than 1, and may have different values. Once again, determining the possible value of deff requires judgment and know-how.

The preceding formula for the size of the sample in equation (9) could be modified by correcting it for two factors: on the one hand, the ***non-response rate***, that is, the percentage of original households in which it was not possible to conduct the interview, whether because the household refused, or because it did not have any people in the age group defined, etc. The other factor that may have an influence is the ***average number of households within a building***, understood as the total number of household in divided by the total number of buildings. If the non-response rate was 0 and the average number of homes/households per housing area was 1, then expression (9) above does not change.

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

Here, the **question here is the size of the sample the study should have.** By way of example, let us focus on the response in the line highlighted in yellow, which allows us to analyze the contents of the la table. This line assumes that the prevalence should be around 10% and that the design effect is 3. Then, if:

- ✓ The desired precision is 1%, which means that the Confidence Interval will be between 9% and 11%, the size required under these conditions should be 10,372 persons.
- ✓ On the other hand, if the desired precision is 2%, the sample size should be 2,593 persons. In this case, the Confidence Interval will be 8% to 12%.
- ✓ So, if the precision is 5%, that is, an interval of between 5% and 15%, the study should be conducted on a total of 415 persons, which would result in a loss of a fair amount of precision.
- ✓ Lastly, with a precision of 10%, the number of cases required is 104, but the interval will be too wide (0% to 20%) and therefore, not very informative.

Table A1.1: Sample sizes for combinations of prevalence (p), precision (d) and design effect (deff)

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

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

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

It should be stressed that the sample size decided on is that value that satisfies the requirements for estimates at the national level. This means that for information disaggregated by region, for example, the precision of those estimates on the basis that sample size will be less than that obtained at the national level. This means that if the study was planned with the goal of obtaining estimates at the local level (regions, for example), as many sizes of samples should be determined as the number of areas that have been defined—which can be adjusted as a function of the size of the population in each of them.

All of these considerations should be taken into account when taking the decision about the size of the sample.

Another important point for decision concerns **the errors of the estimate** mentioned earlier. Generally speaking, there are two ways of looking at these errors: on the one hand, we have **absolute error (AE)** of the estimate, which is the same value we defined earlier as precision, that is, **1.96*ee**[13] for estimates at 95% confidence (using the approximation to a normal distribution). On the other hand, we have **relative error (RE)**, which is the *quotient of the absolute error and prevalence*, and is usually expressed as a percentage. Thus:

$$EA = 1{,}96 * ee \qquad and \qquad ER = \frac{EA}{p} * 100 \tag{10}$$

Looking again at the yellow line in the previous table, where p=10%, the table below shows the relative errors associated with different values for the absolute error (AE):

Table A1.2: Relative errors (RE) for a prevalence of p=10% with different absolute errors (AE). All values are expressed in percentages.

AE	1	2	5	10
RE	10%	20%	50%	100%

We see in Table A1.2 that for a prevalence of 10%, if the absolute error (expressed by the precision) is 5% (i.e., a Confidence Interval of 10% ± 5%, or between 5% and 15%), the relative error will be 50%, which is excessively high. Remember that the size of the sample in this case was 415, as seen in Table A1.1. There are no rules about acceptable values of relative error, but it is suggested that it be no higher than 30%. This places a new condition on determining on the size of the sample for the study, and is something to be considered in the process.

It will be appreciated that determining the sample size in a population study, and selecting the housing units (households) included in the sample is no small thing. It must consider all of the elements discussed earlier. Therefore, starting with the planning of the study, the staff who will be responsible for determining all of these issues related to the sample, who must have training in statistics and particularly in sampling, must be on board. These staff may be attached to the country's statistical office, or to universities or companies that work in the field. Their work is essential to

[13] ee refers to standard error.

the success of the study. It should focus on the following issues based on the most up-to-date sampling frame possible (not an exclusive list):

- ✓ Determine **the size(s) of the sample(s)** on the basis of the study objectives, the conditions described and the information available, and on that basis, determine the number of persons aged 12-65 that will be included in the sample,
- ✓ Select **the sample**, i.e., select the blocks and the housing units in each of the blocks selected,
- ✓ Once the fieldwork has been completed, **calculate the expansion factors** for each individual in the sample, values needed for the subsequent statistical analyses, an issue that will be discussed in more detail in the next annex.

ANNEX 2: STATISTICAL ANALYSIS

This Annex will go into some detail about the ***statistical analysis*** of the results of the household survey, conducted by means of a multistage 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,[14] STATA,[15] SAS,[16] and R.[17] 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

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.

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

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:

a) Cleaning the database

This means essentially deciding on certain criteria for exclusion, that is, to decide on the reasons why certain questionnaires should be excluded from the database. The National Observatory on Drugs must decide on these criteria in advance. Some typical criteria for exclusion should be:

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

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

b) Generating new variables in the database

What do we mean by a variable? Basically, a ***variable*** is a characteristic that we want to study in the subjects of the study, in this case, people in the country aged 12-65. The variables to be included in the study derive directly from the study objectives.

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

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

Regarding "first use", we are interesting 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 month

	3. More than 1 month ago but less than 1 year ago
	4. More than 1 year ago

The second answer enables us to find out who used the substance for the first time in the "past month," while affirmative responses 2 or 3 provide information on first use in the "past year." This particular situation is examined in more detail in the section on statistical analysis.

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).**[18] 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 is 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.)

[18] 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

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 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.

c) 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.

Let us take an example based on a **three-stage** sampling based on a geographic stratification of the country, with K being the number of geographic areas or strata:

First stage:

- ✓ Let us say that M_k is the total number of blocks in geographical area k, and m_k is the number of blocks to be selected in area k, and
- ✓ $P_{k(j)}$ is the probability of selecting block j within geographical area k (k=1,2….K), then:

$$P_{k(j)} = \frac{m_k}{M_k} \qquad k=1, 2, \ldots\ldots K \tag{11}$$

Second stage:

- ✓ Let L_{jk} be the total number of households in block j of area k, and let l_{jk} be the number of households to be selected in this block, and
- ✓ Let $P_{k(i/j)}$ be the probability of selecting household i within block j in geographical area k, then:

$$P_{k(i/j)} = \frac{l_{jk}}{L_{jk}} \qquad j=1, 2, \ldots\ldots J_k \tag{12}$$

Third stage:

- ✓ If n_{ijk} represents the number of eligible people in the selected household, then the probability of selecting one person in particular will be the inverse of that number, i.e., $1/n_{ijk}$. Thus, if in a selected household there are 2 people aged 12-65, the probability of selection for each will be 0.5. If there are 4 eligible people, the probability will be 0.25, and if there is only one eligible person, the probability will be 1.

Looking at expressions (11) and (12) above and the probability in stage three, the final probability of selecting one person in household i (i=1,2….n_{jk}) in block j in geographical area k is:

$$P_{ijk} = \frac{m_k}{M_k} * \frac{l_{jk}}{L_{jk}} * \frac{1}{n_{ijk}} \qquad i=1, 2, \ldots.\ n_{jk} \tag{13}$$

The expansion factor for each person in the sample is defined as the inverse of the probability of selection as determined in expression (13), that is:

$$f_{ijk} = n_{ijk} * \frac{M_k}{m_k} * \frac{L_{jk}}{l_{jk}} \tag{14}$$

where:
- subindex k (k=1…K) represents the geographical areas or **strata**,
- subindex j (j=1…J_k) represents the **blocks** within the strata,
- subindex i (i=1…n_{jk}) represents the **households** within the blocks in the strata.

Having determined the expansion factor for each individual in the sample, the next step is to incorporate it into the database as a new variable.

N.B.: The example above is simply an example of what is probably the simplest situation when plans go forward without mishaps, but there are some assumptions that need to be revised in each case.

For example, as we saw earlier, the number of households in different blocks may be very different and this may lead us to divide up the blocks according to the number of households. This may introduce a new factor to be considered in the sampling process and therefore in the expansion factor. Also, some of the expressions above may be modified by the **non-responses**, i.e., by those households selected for the sample but where the interview was not able to be conducted for whatever reason. There are other reasons why this process may become more complex.

This means that expression (14) for the expansion factor might be modified, and these changes should be taken into account by those responsible for drawing up the sample design and by those in charge of calculating those factors.

As stated earlier, the variable **expansion factor** is a fundamental part of what we have called the second database, because it ensures that ***unbiased estimates can be obtained for the parameters*** defined in the study objectives.

3. Statistical analysis

With these conditions, we can move to the analysis itself. This will require an analysis plan that basically covers three areas that complement each other but 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, socio-economic level, or geographical areas if pertinent. 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, socio-economic level or geographical area.

To better explain the analyses that will be described below, the table below shows the variables in a database containing 10,707 cases. We then give a list with a small part of the database (only 23 cases and 12 variables), which we will use for subsequent examples.

Table A2.0. Description of variables in a database of 10,707 cases

Name of Variable	Description
id	Identification of the case
estrato	Geographical area (region of the country)
upm	Primary selection unit, i.e., the blocks within the stratum
P99	When was the first time you tried alcohol? 1=past month 2=more than 1 month ago less than 1 year ago 3=more than 1 year ago, missing=never used
sex	Sex of respondent, 1=Male, 2=Female
age	Age of respondent
Riesgo_mar	Perceived risk of occasional use of marijuana, 1=no risk 2=slight risk 3=moderate risk 4=high risk
pbeb_month	Past month alcohol use, 1=Yes, 0=No
Ibeb_ano	First use in past year, 1=Yes, 0=No
Ibeb_month	First use in past month, 1=Yes, 0=No
suma_audit	Sum of AUDIT score; Missing if PBEB_ANO=0
fac_mar	Easy to obtain marijuana, 1=Yes, 0=No, 9=Don't know
pmar_ano	Marijuana use in past year, 1=Yes, 0=No
fexp	Expansion factor

id	estrato	upm	p99	sexo	edad	riesgo_mar	pbeb_mes	ibeb_ano	ibeb_mes	suma_aud	fac_mar	pmar_ano	fexp
1	1	1	3	2	35	1	0			3	1	0	2597,52
2	1	1	3	1	43	2	1			5	0	0	2383,8
3	1	1	3	1	42	3	1			4	1	0	4603,2
6	1	2	3	2	31	2	1			12	0	0	748,02
7	1	2	3	1	38	4	1			2	1	0	287,7
8	1	2	3	2	30	1	0			4	0	0	501,42
9	1	2	3	1	18	4	0				1	0	1315,2
18	1	3		2	43	4	0	0	0		0	0	706,92
19	1	3		2	22	1	0	0	0		1	0	328,8
20	1	3	3	2	44	4	1			5	1	1	1413,84
21	1	4	2	1	17	2	0	1		4	0	0	600,06
22	1	4	3	1	23	2	1			2	1	0	945,3
5507	4	1	3	2	47	1	0			1	1	0	1701,54
5508	4	2	3	2	28	2	0				0	0	320,58
5509	4	2	3	2	27	1	0				1	0	476,76
5512	4	2	3	1	19	3	1			20	1	1	912,42
5513	4	2	1	2	46	4	0	1	1		1	0	123,3
5514	4	2	3	1	45	4	0			1	0	0	723,36
5515	4	3	3	1	53	3	1			2	0	0	583,62
5521	4	3	3	1	43	3	1			13	0	0	180,84
5522	4	3		2	26	3	0	0	0		1	0	131,52
5523	4	3	3	1	46	3	1			0	1	0	583,62
5524	4	4		2	46	1	0	0	0		0	0	706,92

What we see in these 23 cases is that they come from two geographical areas or strata, namely, **strata 1 and 4.** There are 12 cases in stratum 1 (ID from 1 to 22). The first three cases (ID=1, 2 and 3) belong to the same *primary selection unit,* UPM=1, that is, these three persons belong to households located in the same block. The first is a 35 year old female, the second is a male aged 43, and the third is also a male aged 42. In UPM=2 of stratum 1, there are four persons, and so on. The last column represents the individual expansion factors.

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_month) 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 ***perception of 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 for/by 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 22,876, which corresponds to the population represented by this sub-sample of 23 persons. We see in the database that there are 11 males and 12 females. If we do **not weight** for expansion factors, we would conclude that the distribution is 47.8% males and 52.2% 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 2.383,8; 4.603,2; 287,7;……583,62. The sum of these expansion factors is 13.119,12.

In the case of females, the expansion factors are 2.597,52; 748,02; 501,42; …706,92. The sum of these values is 9.757,14. The percentages of males and females are then calculated, resulting in 53.4% and 46.6% respectively, which are quite different from the values that would have been obtained if not weighted for the pertinent factors.

In the same way, we can estimate the percentage of people who perceive the frequent use of marijuana as being of ***high risk***. Of the 23 cases in the sample, 17 said that it is high risk: if **not weighted** for the expansion factors, it might be concluded that **73.9%** perceive frequent use of marijuana as being high risk. However, when **weighting** for the expansion factors, i.e., considering the sum of the factors of these 17 cases, we have a total of 16,144, which means **7.6%** of the population represented, as opposed to the 73.9% obtained without weighting.

Formally speaking, let us recall that f_{ijk} is the expansion factor of the individual selected in household i of block j of area or stratum k.

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:

- ✓ B**inary 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 **average**, **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_month 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*[19] as:

$$\boldsymbol{p} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}\sum_{i=1}^{n_{jk}} \boldsymbol{x_{ijk}} * \boldsymbol{f_{ijk}}}{\sum_{k=1}^{K}\sum_{j=1}^{J_k}\sum_{i=1}^{n_{jk}} \boldsymbol{f_{ijk}}} \tag{15}$$

where **X** represents the variable pbeb_month, that is, **X**=pbeb_month and

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

$\boldsymbol{f}_{ijk}$ = *expansion factor of the person selected in household i of block j of area or stratum k*

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.

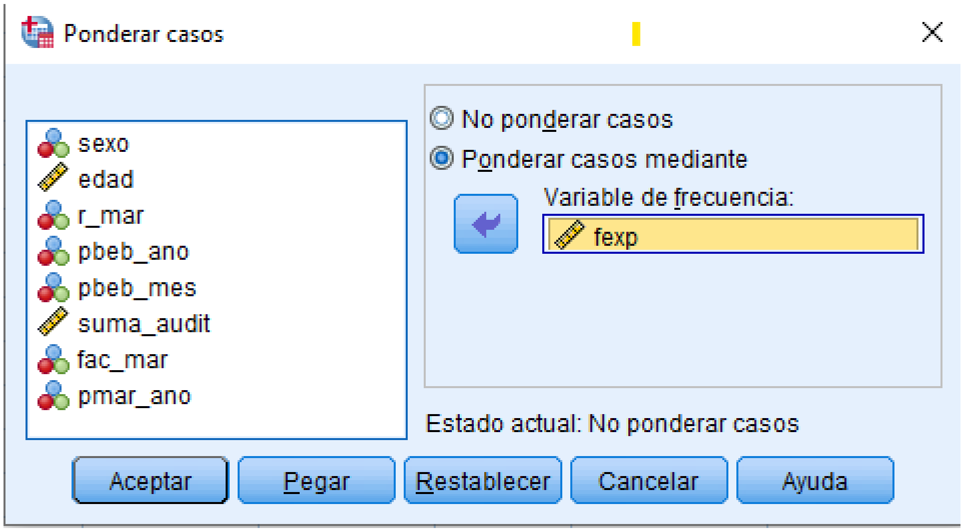
Formula (15) operates in the same way. The result is as follows, where x̅ represents the **average** age in the weighted sample:

$$\bar{X} = \frac{35 * 2.597{,}52 + 43 * 2.383{,}8 + \cdots + 46 * 706{,}92}{2.597{,}52 + 2.383{,}8 + \cdots + 706{,}92} = \frac{856.351{,}38}{22.341{,}96} = 38.3 \text{ years}$$

Henceforth, we shall use the SPSS statistical software to illustrate the calculations of interest, and will go back to the complete database of 10,707 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, ***fexp*** in this case, and press *Done*.

[19] 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.



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 and then *Done*. The result for the variable sex is given in Table A2.1, which shows us that the population represented is 10,078,213 individuals, with 48.4% male and 51.6% female.

Table A2.1. Distribution by sex of the population represented

		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	Male	4882524	48.4	48.4	48.4
	Female	5195689	51.6	5.6	100.0
	Total	10078213	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 A2.2, 38.9% of the population represented said they had used alcohol in the past month (value 1 in the variable pbeb_month indicates a positive case.) This value is the point 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_month)

pbeb_month		Frequency	Percentage	Valid percentage	Cumulative percentage
Valid	.00	6154470	61.1	6.1	61.1
	1.00	3923743	38.9	38.9	100.0

Total	10078213	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 A2.3.

Table A2.3: Description of age

	N	Minimum	Maximum	Average
age	10078213	12.00	65.00	35.6010
N valid (por lista)	10078213			

This table shows that the ages of the people in the sample range from 12 to 65, with an average age of 35.6.

However, the average 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 A2.3_1: Percentiles de age

	Percentiles						
	5	10	25	50	75	90	95
Age	14.00	17.00	23.00	35.00	48.00	57.00	60.00

We can therefore say that 50% of the population represented is aged at the most 35 (50th percentile), 25% are 23 years old or less (25th percentile), and 25% are aged 48 or more (75th percentile). We also know that 5% are aged 14 or less (5th percentile), and that 10% of the population represented are aged 57 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 use the previous database of 10,707 cases, which represent a population of 10,078,213. First, using SPSS, we shall determine the average age de las personas in three different situations.

a.- ***Without weighting for the expansion factors, i.e., assuming a simple random sample of 10,707 persons aged 12-65.***

Table A2.4: Description of average age using simple random sampling

	N	Average	
	Statistical	Statistical	Standard error
Age	10707	36.40	.142
N valid (list)	10707		

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

Table A2.5: Description of average age with expansion factors

	N	Average	
	Statistical	Statistical	Standard error
age	10078213	35.60	.005
N valid (list)	10078213		

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

Table A2.6: Description of average age using a complex sampling design (with expansion factors)

	Estimate	Standard error	95% Confidence Interval	
			Lower	Upper
Median age	35.60	.324	34.97	36.24

We see that in the first case (Table A2.4), the average age is 36.40. However, in the other two cases, (Tables A2.5 and A2.6) it is 35.60. The first estimate is **incorrect since it does not take into account the weighting factors**. The second and third estimates of the average age are **correct**. The difference between these two latter cases lies in the form in which the standard errors are calculated. In the second, the standard error is very small (ee=0.005) 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=0.324, which is much higher than the previous one. **This is the correct standard error** with a 95% Confidence Interval for the average age, with limits of 34.97 and 36.24 years of age.

The ***design effect*** can also be obtained directly together with the previous estimate:

Table A2.7: Estimate of the design effect

	Estimate	Standard error	95% Confidence Interval		Design effect
			Lower	Upper	
Median age	35.60	.324	34.97	36.24	5.168

As we saw in the previous example, the average of the age variable can be obtained by again using expression (15) 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 (15).

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	Cumulative percentage
Never used	1040915	10.3	10.3	10.3
In the past month	92360	.9	.9	11.2
More than 1 month ago but less than 1 year ago	253990	2.5	2.5	13.8
More than 1 year ago	8690948	86.2	86.2	100.0
Total	10078213	100.0	100.0	

From the above, we can see that using expression (15), 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 a 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	6154470	61.1	61.1	61.1
Yes	3923743	38.9	38.9	100.0
Total	10078213	100.0	100.0	

In the case of prevalence (Table A2.9), the result is direct: 3923743/10078213=0.389 (38.9%); that is, almost 39% of the population represented said they had used an alcoholic beverage during the past month.

- **To estimate past year and past month Incidence** (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, 1,040,915+92,360+253,990 = 1,387,265 persons. Excluded are the 8,690,948 individuals (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:

$$Incidence\ alcohol\ past\ year = \frac{92.360 + 253.990}{10.078.213 - 8.690.948} = \frac{346.350}{1.387.265} = 0.25\ (25.0\%)$$

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, 92,360 people. 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:

$$Incidence\ alcohol\ past\ month = \frac{92.360}{10.078.213 - 8.690.948 - 253.990} = \frac{92.360}{1.133.275} = 0.081\ (8.1\%)$$

Formalmente, expression (15) 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 that 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 (15) 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:

$$Past\ year\ incidence = \frac{B + C}{N - D} \qquad (16)$$

and

$$Past\ month\ incidence = \frac{B}{N - C - D} \tag{17}$$

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, with expressions (16) and (17).

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	Cumulative percentage
Never used	1040915	10.3	10.3	10.3
In the past month	92360	.9	.9	11.2
More than 1 month ago but less than 1 year ago	253990	2.5	2.5	13.8
More than 1 year ago	8690948	86.2	86.2	100.0
Total	10078213	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" (8,690,948 cases, 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	0 (No)	1040915	10.3	75.0	75.0
	1 (Yes)	346350	3.4	25.0	100.0
	Total	1387265	13.8	100.0	
Missing values		8690948	86.2		
Total		10078213	100.0		

The frequency shown as 1 in the table above (*Used for the first time during the past year*), i.e., 346,350 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*" in Table A2.13, that is, 92,360+253,990.

The column "Valid percentage" shows incidence in the past year, 25.0%, which was what was obtained earlier.

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

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	0 (No)	1040915	10.3	91.9	91.9
	1 (Yes)	92360	.9	8.1	100.0
	Total	1133275	11.2	100.0	
Missing	values	8944938	88.8		
Total		10078213	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, 92,360 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, 8.1%.

Using this approach, namely, constructing two new variables for past year incidence and past month incidence, it should be relatively easy and straightforward 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 "use of substance," then:

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

Using formula (15), we see that the estimator of the prevalence of substance use for the period is:

$$\boldsymbol{p} = \frac{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \sum_{i=1}^{n_{jk}} \boldsymbol{x_{ijk} * f_{ijk}}}{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \sum_{i=1}^{n_{jk}} \boldsymbol{f_{ijk}}} \qquad (18)$$

where the subindices i,j,k identify the person selected from household i, in block j in stratum h; 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 <u>prevalence</u> of "lifetime use", "past year use", and "past month use", in addition to the <u>incidence of use</u> "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 (18) (i.e., with expansion factors) is 50%. If we also know that the error of the estimate (what we have called "precision") 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 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 subsample of 10,707 cases. The sample is taken from 6 geographical areas (known as strata) from which the total number of cases were chosen.

For the purposes of the examples that are given below, we shall use the same database with the variables described in Table A2.0.

We first need to give a brief description of the sample for the two demographic variables that are in the reduced 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 minors and young people, for example. ***For this, we shall construct a new variable for age, using the following age categories: 12-17, 18-34 and 35-65. Of course, these are not the only categories possible, but they are the ones that will be used in the following examples, since they are the categories that the OID/CICAD uses to prepare its regional reports. As we shall see later, these categories will need to be revised/reviewed for purposes of analysis.***

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

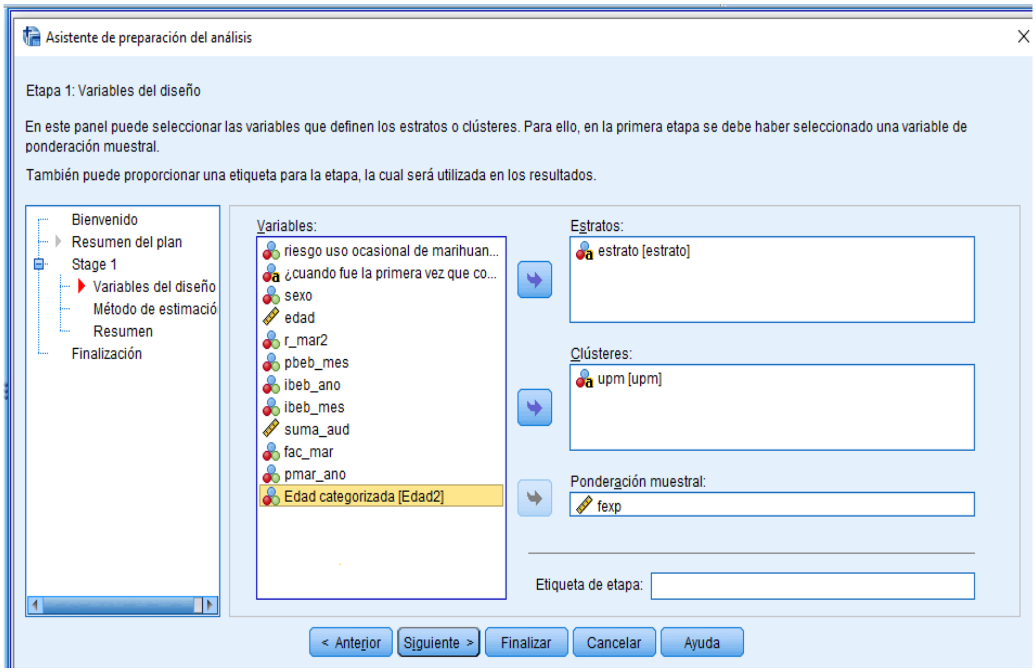
Variables	Sample size	Population represented	%
Sex			
Male	4,263	4,882,524	48.4
Female	6,444	5,195,689	51.6
Age			
12-17	1,139	1,276,969	12.6
18-34	4,025	3,724,071	37.0
35-65	5,543	5,077,173	50.4
TOTAL	10,707	10,078,213	100

As shown in the table, in the expanded sample there are nearly 52% females and around 37% are people aged 18-34.

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

- ✓ The sample design must be determined first; basically, in this specific example, we indicate the **strata** (geographical areas) and the **clusters** (primary selection units, i.e., blocks), 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 variables corresponding to the ***Strata***, 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***.

✓ Then we use the SPSS "sentences" 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_month**, which is 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	Standard error	95% Confidence Interval	
				Inferior	Superior
Media	pbeb_month	.3893	.00941	.3709	.4078

The prevalence of the use of alcohol in the past month is 38.93%, 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 37.09% and 40.78%. The error of the estimate and the Confidence Interval are correct and are properly computed on the basis of the sample design used.

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

Table A2.18: Design effect on indicator of prevalence of past month alcohol use

		Estimate	Standard error	95% Confidence Interval		Design effect
				Lower	Upper	
Media	pbeb_mes	.3893	.00941	.3709	.4078	3.995

As shown above, the Design Effect is almost 4, i.e., the variance of the model used is four times the variance obtained if a simple random sampling had been used. ***This is important information to be borne in mind when determining the sample size for similar studies in the future.***

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

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

$$EA = \frac{Ls - Li}{2} \quad \text{and} \quad ER = \frac{EA}{p} * 100 \tag{19}$$

In the example:

$$EA = \frac{40,78-37,09}{2} = 1,85\% \text{ and } ER = \frac{1,85}{38,93} * 100 = 4,7\%$$

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 95% Confidence Intervals for the categories sex and age group. The results are given in the table below:

Table A2.19: Prevalence of past month alcohol use and 95% Confidence Intervals (CI), by sex and age groups

Variables	Prevalence (%)	Standard error (%)	CI (%)
Sex			
Male	48.67	1.35	46.02 – 51.31
Female	29.78	1.13	27.57 – 32.00
Age			
12-17	22.30	1.63	19.10 – 25.49
18-34	51.24	1.42	48.45 – 54.03
35-65	34.09	1.25	31.63 – 36.54
TOTAL	38.93	0.94	37.09 – 40.78

The first point to notice in the above table is that the standard error for the total sample is the lowest (smallest) of all, and the reason is that it is the estimate that considers the total sample: the larger the size of the sample, the smaller the error of the estimate (for similar prevalence rates.) When estimates are made at the level of sections of the sample, the precision of those estimates will be less; thus, for example, in people aged 12-17, the width of the Confidence Interval (upper limit minus lower limit) is almost 6 percentage points, almost double the result at the national level.

Looking at the precise estimates of prevalence, we see that the prevalence of past month alcohol use is higher among males than among females, 48.7% and 29.8% respectively. We also see that the highest prevalence rate is among the 18-34 year age group, at 51.2%, more than double that of the first age group, and very much higher than in the last age group.

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 prevalence of alcohol use at the level of the population on the basis of the results obtained from the samples. For example, are there differences between males and females? Between 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 10% to 15%, and for the other subgroup it is 18% to 24%, 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 13% to 18%, then there is an overlap of the two intervals, and the conclusion would therefore be that there are no statistically significant differences between the groups. A third situation might occur if there is a slight or very small overlap of the intervals, for example 10% to 15% and 14.9% to 16%, or 10% to 15% and 15.1% to16%. 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 probability terms.

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*.

The 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 show that: p_1=48.67% and p_2=29.78%.

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 A2.20:

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

	Chi-squared	F corrected	gl1	gl2	Sig.
Pearson	401,161	129,320	1	1238	0,000
Odds ratio	403,490	130,071	1	1238	0,000

The above table gives two results for the Pearson's chi-squared statistic: the first column of results of Table A2.20 (chi-squared=401,161) does not consider the sample design, whereas the second column (***F corrected=129,320***) 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 129,320 (Q=129,320) with an associated probability (p-value) of less than 0.0001. If the significance level is defined as 5%, that is, $\alpha=0.05$, given that the p-value<0.0001, the null hypothesis is rejected, that is to say, there is evidence for stating that at the level of the population, there are differences between males and females in the prevalence of past month use of alcohol. In other words, *there are 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 also present at the level of the population, and it is therefore assumed that the difference is "not by chance" but is explained by sex.

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

H_0: P_1=P_2=P_3 that is, prevalence of past month alcohol use is the same for the three age groups, and

H_1: at least two indicators are different.

Again using SPSS, we obtain the following results:

Table A2.21: Test of independence for comparison of past month alcohol use by age category

	Chi-squared	Sig.
Pearson	463,447	0,000
Odds ratio	470,453	0,000

According to the results in Table A2.19, the prevalences in the sample of past month alcohol use by age group are (n %): p_1=22.30%, p_2=51.24% and p_3=34.09%.

In Table A2.21 we see that Pearson's chi-square value according to the conditions of the sampling design is **Q=89,996** with p<0.0001. 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 three 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 A2.19 suggest that there may be statistically significant differences between the three groups.

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**.
    - o 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**.
    - o 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**.
    - o 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 population aged 12-65, we have:

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

On the other hand:

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

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[20] 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):

Odds ratio: epidemiological studies on prevalence of drug use—the category to which a household study of people aged 12-65 belongs—are cross-sectional studies conducted at a specific moment in time, in which past and present events are

[20] There are other measures of association depending on the study design, particularly longitudinal studies.

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.

The hypothesis behind a question of this nature is that for those respondents 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:

$$RD = \frac{P_1/Q_1}{P_0/Q_0} = \frac{P_1 * Q_0}{P_0 * Q_1} \qquad (20)$$

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

$$P = P(Y = 1/X = x) = \frac{e^{\beta_0+\beta_1 * X}}{1 + e^{\beta_0+\beta_1 * X}} \tag{21}$$

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:

$$P_1 = P(Y = 1 \;/\; X = 1) = \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}} \tag{22}$$

and

$$Q_1 = 1 - P_1 = P(Y = 0 \;/\; X = 1) = 1 - \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta_1}} = \frac{1}{1 + e^{\beta_0+\beta_1}} \tag{23}$$

From (21) and (22) we have:

$$P_1/Q_1 = e^{\beta_0+\beta_1} \tag{24}$$

If X=0 then:

$$P_0 = P(Y = 1 \;/\; X = 0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \tag{25}$$

and

$$Q_0 = 1 - P_0 = P(Y = 0 \;/\; X = 0) = 1 - \frac{e^{\beta_0}}{1 + e^{\beta_0}} = \frac{1}{1 + e^{\beta_0}} \tag{26}$$

From (25) and (26) we have:

$$P_0/Q_0 = e^{\beta_0} \tag{27}$$

Lastly, going back to expression (20) for the odds ratio (OR), we have:

$$RD = \frac{P_1/Q_1}{P_0/Q_0} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1} \qquad (28)$$

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 (P1) is equal to the prevalence among females (P_2),[21] 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 A2.22: 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)	0.0532	0.054	-0.053	0.159	0.987	0.324
[sex=,00]	0.8044	0.071	0.665	0.944	11.284	0.000
[sex=1,00]	0.000[a]	.	.	.	.	.

The most important result in Table A2.22 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, $\hat{\beta}_1$=0.804 with p<0.001 and a 95% Confidence Interval of between 0.665 and 0.944. Given that the p-value is very low, the conclusion is that there are statistically significant differences between the past month prevalence rates of alcohol use between males and females.

[21] 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.

The estimator of the odds ratio is:

$$R\widehat{D} = e^{\hat{\beta}1} = \exp(\hat{\beta}_1) = \exp(0{,}8044) = 2{,}235$$

With a 95% Confidence Interval of 1.944 to 2.571, and given that the null value 1 is not within the limits of the CI, we reach the same conclusion as above. The same results in SPSS are shown in Table A2.23, which gives the estimate of the odds ratio, and the corresponding Confidence Interval:

Table A2.23: Estimate of odds ratio for past month alcohol use and sex based on logistic regression model

Odds ratio	95% Confidence Interval	
	Lower	Upper
2.235	1.944	2.571

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. If the prevalence among males is higher than among females, 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 X. In some situations, the X cannot be modified, but in others, it is feasible to do so.

Example 1: X may represents the variables sex, age group, region or area. These are variables do not change. The data show a statistically significant difference in alcohol use by sex. This information is important because statistically significant differences found between males and females could be useful to target interventions, but variables male and female do not change.

Example 2: X represents the *perceived risk of substance use.* . In this example, the variable *perceived risk* can in fact be modified. The data show that as the perception of risk rises, the probability of substance use decreases. This is information because it implies that policies aimed at increasing the number of individuals who perceive that substance use is highly risky, the probability of use.

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

Table A2.24: Prevalence of past year marijuana use and 95% Confidence Intervals (CI), by sex and age group

Variables	Prevalence (%)	Standard error (%)	CI (%)
Sex			
Male	6.15	0.51	5.23 – 7.23
Female	1.77	0.31	1.26 – 2.48
Age2*			
12-17	5.50	0.86	4.04 – 7.45
18-34	6.80	0.62	5.69 – 8.12
35-65	1.36	0.23	0.98 – 1.89
TOTAL	3.90	0.31	3.34 – 4.54

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

Comparing the prevalence by sex, i.e., prevalence among males with the prevalence among females, we see that the Confidence Intervals for males and females do not overlap, indicating a statistical difference between the two groups.

Let us now examine the hypothesis that both prevalence are equal when we apply a logistic regression model. Again, using SPSS, we obtain the following results (Tables A2.25 and A2.26):

Table A2.25: Estimate of parameters 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)	2..724	0.088	2.552	2.897	30,979	0,000
Male vs female	1.290	0.188	0.921	1.660	6,859	0,000

Table A2.26: 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	3.635	2.513	5.257

We see from this table that the $\hat{RD}$=3,635 with p<0.0001 (see previous table,) less than α=0.05 and with a Confidence Interval of between 2.51 and 5.26 (does not include null value 1), which leads to the conclusion that there is a statistically significant difference between the past year prevalences of marijuana use among males and females.

In the two preceding examples, the independent variable (sex) had two possible reponses. Let us now look at the case where this variable has more than two levels or categories, for example, the variable riesgo_mar, 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 A2.27: Prevalence of past year marijuana use and 95% Confidence Intervals, by perceived risk of occasional use of marijuana

Perceived risk of occasional use of marijuana	Prevalence (%)	Standard error (%)	CI (%)
No risk	15.82	2.70	10.52 – 21.12
Slight risk	10.89	1.34	8.27 – 13.51
Moderate risk	8.65	1.20	6.31 – 11.00
High risk	1.15	0.16	0.84 – 1.45
TOTAL	3.90	0.31	3.34 – 4.54

As shown in the above table, the prevalences of marijuana use decline to the extent that the perceived risk of occasional use of the substance increases, going from nearly 16% of those who perceive no risk to a little more than 1% of those who see high risk. This is clearly a trend. The Confidence Intervals do not suggest a difference between the first three categories, but do show differences between each of them vis-à-vis high risk. 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, 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 $F=70,392$ with $p<0.0001$.

Table A2.28: Estimate of parameters of the logistic regression model for past year marijuana use and perceived risk

Origin	gl1	gl2	Wald F	Sig.
(Corrected model)	3,000	1236,000	70,392	0,000
(Intersection)	1,000	1238,000	1047,054	0,000
riesgo_mar	3,000	1236,000	70,392	0,000

Table A2.29 gives the odds ratios for comparison of each category with perceived high risk.

Table A2.29: 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	16.187	10.067	26.030
Slight risk vs. High risk	10.527	7.207	15.375
Moderate risk vs. High risk	8.161	5.478	12.159

As shown in the table, all the odds ratios are associated with Confidence Intervals that do not contain the null value, and therefore the three categories show prevalences of marijuana use that are significantly higher than the group with the lowest prevalence, namely, "High risk." Despite this, it is not possible to determine differences between the age groups with higher prevalence rates. The results of a second model are given in Table A2.30, but now with "no risk" as the reference category; here, we see that there are no statistically significant differences in the prevalences of past year marijuana use between those who perceive "no risk" and "slight risk" (CI between 0.41 and 1.03); however, there are statistically significant differences between the "no risk" group and the "moderate risk" group (CI between 0.301 and 0.845).

Table A2.30: Estimate of odds ratio for past year marijuana use and perception of risk based on logistic regression model with "slight risk" as reference

Comparison	Odds ratio	95% Confidence Interval	
		Lower	Upper
Slight risk vs. No risk	,650	,410	1,030
Moderate risk vs. No risk	,504	,301	,845
High risk vs. No risk	,062	,038	,099

In Tables A2.25 to A2.30, 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 (21) on a logistic regression model:

$$P = P(Y = 1/X = x) = \frac{e^{\beta_0+\beta_1 * X}}{1 + e^{\beta_0+\beta_1 * X}} \tag{29}$$

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:

$$Q = 1 - P = \frac{1}{1 + e^{\beta_0+\beta_1 * X}} \tag{30}$$

Formula (29) 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 <u>*natural logarithm of the quotient of P and Q*</u>, (which is called the logit of P) and which gives the following ***linear relationship***:

$$\ln(P/Q) = \beta_0 + \beta_1 * x \tag{31}$$

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 20 and 21, or between people aged between 50 and 51. 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 A2.31: Estimated of 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)	-1,543	,181	-1.897	-1.189	-8,543	,000
age	-,054	,006	-.066	-,043	-9,386	,000

In Table A2.31, we see that the estimate of the coefficient of age is negative (-0.054) with a value of p<0.0001, which shows that there is a statistically significant reduction in the *ln (P/Q)* as the age increases. We conclude from this that the probability of the use of marijuana in the past year, i.e., P, also diminishes as the age increases. However, looking again at Table A2.24, we see that the prevalence rate increases in the 12-17 and 18-34 age groups, from 5.5% to 6.8% and then decreases in the third group to 1.36%, which would seem to contradict the result in Table A2.31. This may well be related to the assumption discussed above that there was a linear relationship, which, in this case, is not supported by the data.

What is the solution in a case like this? To group the ages into ranges that make sense for our particular study. For example, in Table A2.24 we had three age groups:

- Age2=1 if the age is between 12 and 17
- Age2=2 if the age is between 18 and 34
- Age2=3 if the age is between 35 and 65.

It is possible that substance use is not homogeneous within the 18-34 and 35-65 age groups, and it may be that a good option would be to create subdivisions in those two groups, so as to go from 3 categories to 5 groups. We will therefore construct a new variable, which we will call Age3, as follows:

- Age3=1 if the age is between 12 and 17
- Age3=2 if the age is between 18 and 24
- Age3=3 if the age is between 25 and 34
- Age3=4 if the age is between 35 and 44
- Age3=5 if the age is between 45 and 65.

The following table gives these prevalences and respective Confidence Intervals for each of these new age groups.

Table A2.32: Prevalence of past year marijuana use and 95% Confidence Intervals, by perceived risk of occasional marijuana use

Age	Prevalence (%)	Standard error (%)	CI (%)
12 – 17	5.50	0.86	4.04 – 7.45

18 – 24	9.32	1.07	7.43 – 11.63
25 – 34	4.85	0.65	3.72 – 6.29
35 – 44	2.03	0.40	1.37 – 3.00
45 – 65	0.93	0.21	0.60 – 1.45
TOTAL	3.90	0.31	3.34 – 4.54

What the table shows is that there is an increase up to the age of 24, and then a steady decline, which reinforces the idea that there is no linear trend. In these cases, it is advisable to adjust a logistic regression model using the 45-65 year age group as the reference, since it has the lowest prevalence rate. The results of this analysis are shown below in terms of odds ratios.

Table A2.33: Estimated odds ratio in logistic regression model for past year marijuana use and age groups (reference 45-65 year age group)

Comparison	Odds ratio	95% Confidence Interval	
		Lower	Upper
12-17 vs. 45-65	6.180	3.593	10.631
18-24 vs. 45-65	10.916	6.504	18.320
25-34 vs. 45-65	5.408	3.233	9.048
35-44 vs. 45-65	2.201	1.318	3.675

As expected, there are statistically significant differences in the prevalence rates for each group vis-à-vis the reference group, but we can say nothing about the other comparisons. Below we have a new model but now with the 18-24 age group as the reference, since it is the group with the highest prevalence rate.

Table A2.34: Estimated odds ratio in logistic regression model for past year marijuana use and age groups (reference 18-24 years)

Comparison	Odds ratio	95% Confidence Interval	
		Lower	Upper
12-17 vs. 18-24	0.566	0.391	0.820
25-34 vs. 18-24	0.495	0.347	0.708
35-44 vs. 18-24	0.202	0.126	0.324
45-65 vs. 18-24	0.092	0.055	0.154

The odds ratios are less than 1 since the reference group has the highest prevalence rate, and in all groups, the prevalence rates are statistically lower than that of the 18-24 age group (both of the limits of the Confidence Intervals are below the null value of 1). It is important here to point out and to reinforce the need to resort to this type of analysis to compare groups rather than taking decisions on the basis of the Confidence Intervals of the prevalences of each group, is what is happening with the first two age groups. Indeed, looking again at the results shown in Table A2.32, and comparing the

Confidence Intervals for the 12-17 and 18-24 age groups, we see that there is a slight overlap of the intervals, which might suggest that there are no differences between the prevalences of the two groups. However, according to Table A2.34, the Confidence Interval of the odds ratio in the comparison between these groups does not include the null value, the CI is between 0.391 and 0.820, the conclusion being that there is evidence to say that there are statistically significant differences at the population level between the prevalence rates of these two age groups.

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 (31) 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. There is no doubt that the results in Table A2.32 were of great help in deciding to adjust a model with the variable age as a continuous variable did not appear to be a good decision.

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 in the analyses. It was concluded from the above analyses (Tables A2.25 and A2.26) that the prevalence 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 there were a higher proportion of males in the age groups with a higher rate of marijuana use, 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:

$$\boxed{P = P(Y/\,Sexo, Edad) = \frac{e^{\beta_0+\beta_1 * Sexo+\beta_2 * Edad}}{1+e^{\beta_0+\beta_1 * Sexo+\beta_2 Edad}}} \qquad (32)$$

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 prevalences 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 A2.35: Estimated odds ratios 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	3.501	2.403	5.101

The odds ratio shown in Table A2.35, that is, $\hat{RD}$=3.501 is called an **age-adjusted** odds ratio, which is different from that given in Table A2.26 where $\hat{RD}$=3.635, which is the **crude** or **unadjusted** odds ratio. The important point is that the adjusted model examines the comparison between prevalences 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 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 subcategories "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."

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:

$$P(Y = 1 \,/\, X = 1) \; > P(Y = 1/X = 0) \qquad (33)$$

In other words, if the factor is present (X=1), the likelihood of substance use (Y=1) is ***greater*** than when the factor is not present (X=0).

If **Z** is another factor associated with substance use, we will say that **Z** is a ***protective factor*** for **Y** if:

$$P(Y = 1 \,/\, Z = 1) \; < P(Y = 1/Z = 0) \qquad (34)$$

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 be focused 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 1.15% of those who perceive occasional use of marijuana to be "high risk," to a prevalence of 15.82% 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 A2.36: 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	1.15

Not high risk	10.79
TOTAL	3.90

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 1.15%, almost 10 times lower than those who perceived a lower risk, 10.79%.

We need now to know what proportion of cases in the expanded sample fall into each of these two categories. Having a prevalence rate of 10.79% associated with 60% of the population represented is very different than having it correspond to 5% of that population. Table A2.37 describes this distribution:

Table A2.37: Distribution of perception of risk of occasional use of marijuana

Perception of risk of occasional marijuana use	N	(%)
High risk	7,205,775	71.5
Not high risk	2,872,438	28.5
TOTAL	10,078,213	100.0

Some 28.5% of the population represented *do not perceive* “high risk” of occasional use of marijuana.

It may be concluded that if there are positive changes in the variable analyzed, in other words, if the distribution described in Table A2.36 increases the percentage of cases in the “high risk” category, 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, which can be used to target the prevention strategies appropriately for this specific population.

3.4 Types of substance users

Thus far, we have discussed in some detail the five basic indicators of substance use, which are a fundamental part of the objectives of a study of this nature—namely, the three indicators on prevalence of use, ***lifetime***, ***past year*** and ***past month***, and the two indicators on new users, namely ***incidence in the past year*** and ***incidence in the past month***.

In paragraph 4.2 of Section 4 of the Protocol on *operational definitions*, we said that it was necessary to look further into the population subgroups in order to describe other aspects of drug use, such as ***frequency and intensity of use***, beyond the indicators on substance use already discussed. In this section, we shall look at this question for two substances, tobacco and marijuana.

3.4.1. Tobacco

The additional indicators on **tobacco** are based on the world tobacco survey, known as the Global Adult Tobacco Survey (GATS)[22] developed by the World Health Organization and the U.S. Centers for Disease Control and Prevention (CDC). In Section 5 of the questionnaire of this Protocol, the GATS questions are **TA6**, **TA7** and **TA8**, combined with those questions that examine prevalence of lifetime, past year and past month use.

We have four indicators:

a) **Prevalence of current tobacco smokers**: The percentage of respondents that currently smoke tobacco. Para ello se considera la question **TA6** on the questionnaire, y se define:

[22] https://www.who.int/tobacco/surveillance/en_tfi_tqs.pdf

N_1 = Total number of weighted cases in which **TA6=1 or 2**.

Answer 1 (TA6=1) is those who say that they smoke "**every day**" and answer 2 (TA6=2) is those who say they smoke "**Only some days,**" of those who stated that they had used in the past month (current smokers). Answer 3 to this question is for people who answer "**Don't know/no opinion.**" Then we define the ***prevalence of current smokers*** as:

$$\boldsymbol{Prevalence\ current\ tobacco\ smokers} = \frac{\boldsymbol{N_1}}{\boldsymbol{Total\ of\ the\ expanded\ sample}} * 100$$

b) <u>Prevalence of current *daily* tobacco smokers</u>. This second indicator is designed to estimate the percentage of respondents who currently smoke tobacco **every day**. Here, we take into account only those cases in which **TA6=1.** Then, let

N_2 = Total weighted cases where **TA6=1.**

The prevalence of current daily tobacco smokers is then defined as:

$$\boldsymbol{Prevalence\ current\ daily\ tobacco\ smokers} = \frac{\boldsymbol{N_2}}{\boldsymbol{Total\ of\ the\ expanded\ sample}} * 100$$

c) <u>Percentage of former daily tobacco smokers who currently do not smoke tobacco</u> <u>(of all those interviewed.)</u> This indicator and the next are designed to estimate those persons who **smoked every day in the past** and who **currently do not smoke**. Here we have two groups of people:

- On the one hand are those who *smoked once in their lives ("ever smoked")* (TA1=1) *but who have not smoked in the past 12 months* (TA4 answers 2 or 9); in these cases, former daily smokers can be found from question **TA8** answer **1**;

  Let **N_3 =** Total weighted cases in which **(TA8=1).**

- The second group consists of those people who *stated that they had smoked in the past year* (TA4=1) *but who had not smoked in the past month* (TA5 answers 2 and 9) and who also *said that they had smoked every day* (TA7=1); these cases are found by combining questions **TA5=2 or TA5=9** (did not smoke in the past month) with question **TA7** answer **1**.

  Let **N_4 =** Total weighted cases in which **(TA5=2 or 9 and TA7=1).**

De lo anterior se define:

$$\mathbf{N_5 = N_3 + N_4}$$

Then N_5 is the ***number of people who in the past had smoked every day and who do not currently smoke tobacco***.

The basis for this indicator if the <u>total number of persons in the study</u>.

The percentage of former daily smokers is then defined as:

$$\boldsymbol{Percentage\ of\ former\ daily\ tobacco\ smokers = \frac{N_5}{Total\ of\ the\ expanded\ sample} * 100}$$

d) <u>Proportion of former daily smokers of tobacco who currently do not smoke</u> (of those who smoked every day in the past.) Unlike the previous indicator which referred to the total of the expanded sample, in this case, we want to find out the percentage of former daily smokers who currently do not smoke *as a proportion of former daily smokers* (including those who current smoke "only some days.")

Those who had smoked every day in the past, independently of whether they currently smoke, are those cases in which **TA7=1 or TA8=1**. The sum of the two is the denominator of this indicator. Let

$\mathbf{N_6}$ **=** Total weighted cases in which **(TA7=1).**

Again, let $\mathbf{N_3}$ **=** Total weighted cases in which **(TA8=1),**

Then $\mathbf{N_7 = N_6 + N_3}$

$$\boldsymbol{Proportion\ of\ former\ daily\ tobacco\ smokers = \frac{N_5}{N_7}}$$

Note the difference between N_5 and N_7. The former is the number of persons who in the past had smoked every day <u>but who do not currently smoke</u>. The latter, on the other hand, are those who had smoked every day in the past <u>independently of whether they currently smoke (or not)</u>.

In all of the above indicators, cases where the response to the various questions is "Don't know/no opinion" should be excluded from the denominators.

3.4.2. Marijuana

For the case of marijuana, we shall use questions **MA1 to MA10** on the questionnaire as the reference. Let us recall:

- questions **MA1, MA4 and MA8** allow us to estimate the prevalence of lifetime, past year and past month marijuana use respectively,
- and question **MA3** allows us to estimate the incidence in the past year and past month.

In addition:

- question **MA5** let us analyze the frequency of use among those who said they had *used in their lives ("ever used")* but who *had not done so in the past year*,
- question **MA6** is also related to the frequency of marijuana use but refers to those who said they had used in the past year, and
- questions **MA9** and **MA10** are designed to study the frequency of use and amounted used by *those who had used in the past month*.

It is possible on this basis to develop other profiles of marijuana users by means of the following indicators based on specific subgroups of users:

a) Users in the past month:

i. ***Daily use of marijuana**:* this is directly related to those cases that gave answer **4** to question **MA9**, that is, those who *stated they had used marijuana every day during the month prior to the study*. It is possible to develop two new indicators. Let us define:

$\mathbf{M_1}$ = Total weighted cases in which **MA9=4**, then:

$$\boldsymbol{Prevalence\ daily\ use = \frac{M_1}{Total\ of\ the\ expanded\ sample} * 100}$$

$$\boldsymbol{Proportion\ of\ daily\ use = \frac{M_1}{Expanded\ cases\ of\ past\ month\ users}}$$

Both indicators have the same numerator ($\mathbf{M_1}$) but a different denominator: the first refers to the total number of cases in the expanded sample, and the second represents a proportion of daily users entre those who said they had used marijuana in the past month. Thus, the first indicator answers the question about what percentage of people in the population uses marijuana every day. By contrast, the second question is an attempt to find out what proportion of past month users smoke marijuana every day.

ii. Both of the two previous indicators can in turn give rise to two more indicators dealing with the **quantity** of cigarettes or spliffs used on a typical day of marijuana use, which can be obtained from question **MA10**. It might be interesting in particular to classify daily **users** according to whether they smoke **one cigarette** a day, or **two or more cigarettes a day**.

iii. Question **MA9** also offers other options that might be interesting to study. On the one hand, we have those people who stated that they had used marijuana during the past month but who had only done so “Only once” or “Several times in the past 30 days,” i.e., those who gave answers **1 or 2** to this question. These cases could be called ***occasional active users***, and we could define two new indicators. If:

M_2= Total weighted cases where **MA9=1 or MA9=2,** then:

$$\textit{Prevalence occasional active users} = \frac{M_2}{\textit{Total of the expanded sample}} * 100$$

$$\textit{Percentage occasional active users} = \frac{M_2}{\textit{Expanded cases of past month users}} * 100$$

The difference between the two indicators lies in the denominator: the first is a percentage of the total population represented in the study, while the second is a proportion of past month users.

b) **Users in the past year:**

This group is one of the main indicators of marijuana use, the ***prevalence of use in the past year***, (also referred to as **recent users**.) Let:

M_3= Total weighted cases in which **MA4=1**,

that is to say, those who said they had used marijuana in the past year. We know that:

$$\textit{Past year prevalence} = \frac{M_3}{\textit{Total of the expanded sample}} * 100$$

In this group, we can examine the frequency of marijuana use in the past year, and for this, we can look at question **MA6** (frequency of use in the past 12 months)**,** and develop other indicators of interest based on frequency of use. For example, if we define:

M_4 = Total weighted cases in which **MA6=1 or MA6=2,**

That is, cases who said they had used marijuana in the past year "only once" or "several times in the past 12 months", we can define the following two indicators:

$$\boldsymbol{Past\ year\ prevalence\ of\ occasional\ users} = \frac{\boldsymbol{M_4}}{\boldsymbol{Total\ of\ the\ expanded\ sample}} * \mathbf{100}$$

and

$$\boldsymbol{Proportion\ of\ past\ year\ occasional\ users} = \frac{\boldsymbol{M_4}}{\boldsymbol{M_3}}$$

This second indicator gives us the proportion of people who occasionally used marijuana, among those who used during the past year.

c) **Former daily users**

Former daily users of marijuana may come from different groups. On the one hand, we have those who used marijuana in their lifetime ("ever used") and at that time used daily, but who have not used in the past year. Let:

M_5 = Total cases in which MA5=6 or MA5=7,

In addition, we have those persons who stated that they had used marijuana every day in the past year (MA6=5) but who do not currently use marijuana (MA8=2). Thus:

M_6 = Total cases where MA6=5 and MA8=2,

Based on the above, we can determine the percentage of persons who in the past used marijuana every day but who currently do not use.

$$\boldsymbol{Percentage\ of\ former\ daily\ users} = \frac{\boldsymbol{M_5 + M_6}}{\boldsymbol{Total\ of\ the\ expanded\ sample}} * \mathbf{100}$$

We can use the same numerator but referring to those who said they had ever used marijuana (lifetime,) i.e., those who gave the answer **1** to question **MA1**.

$$\textbf{\textit{Proportion of former users}} = \frac{\boldsymbol{M_5 + M_6}}{\boldsymbol{Total\ lifetime\ users}}$$

The indicators shown above are a sample of several other indicators that could be defined in accordance with the specific objectives of the study. These indicators could also be of interest for other substances considered in the questionnaire.

ANNEX 3: INSTRUCTIONS FOR THE FIELDWORK

This Annex deals with two topics that are related to implementation of the study: first, we give instructions about the work of the interviewer when administering the questionnaire to the people selected, and we also make some recommendations about organizing and conducting the fieldwork on the ground.

1. INSTRUCTIONS FOR COMPLETING THE QUESTIONNAIRE

The questionnaire is in two parts: the first is the cover sheet, and the second part is the actual questionnaire, which is divided into seven modules, as described in Section 5.

1.1. COVER SHEET

- The cover sheet given in Section 5 of this Protocol consists of the following four parts:
    1. ***Geographical identification*** of the household to be interviewed (use maps as a backup,) and the ***letter for random selection*** of the person to be selected.
    2. ***Kish Grid*** for choosing the person to be interviewed.
    3. Identification of the ***Person Selected*** in the sample.
    4. The ***Field Log*** showing the dates and month of the visits to the household, the supervisor's record and review of the questionnaire.
- The Supervisor will be responsible for filling out **Part 1** of the cover sheet (Geographical identification) before distributing the questionnaires for the fieldwork (Field Assistant or interviewers.) This Part 1 gives the letter for the random selection of the person in the household.

1.1.1.6

1.1.1.7 GEOGRAPHICAL IDENTIFICATION

Region/ Department or province	Town/ District	Area code	Code of household	Number of questionnaire	Letter for random selection
					D

In this case, the letter **D** indicates that only this column of the Kish Grid should be used, and the remaining columns disregarded.

- **Part 2** of the cover sheet is the Kish Grid, which will be used to determine which person in the household should be interviewed.
  The Kish Grid printed in the questionnaire is, in fact, 18 grids (one per column, labelled A to R). For each questionnaire, only the column corresponding to a la box "**Letter for random selection,**" as shown in the example below (column **D is shaded**).

KISH GRID FOR RANDOM SELECTION

Name	Age	Number	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
			1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
			1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3
			1	1	1	2	2	1	2	2	2	3	3	3	4	3	3	4	4	4
			1	1	1	2	2	3	2	2	3	3	4	4	3	4	4	5	5	5
			1	2	1	2	2	2	3	3	3	4	4	4	5	5	5	6	5	6
			1	1	2	1	2	3	4	3	4	4	5	4	5	6	7	6	7	7
			1	1	2	2	3	4	4	3	4	5	6	5	5	6	7	7	8	8
			1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9
			1	1	2	2	3	3	4	5	4	6	7	6	8	8	9	9	10	10
			1	2	3	4	2	4	5	5	6	6	7	7	8	10	8	9	10	11
			1	2	3	4	2	4	5	6	5	7	8	7	9	10	9	11	10	12

The random selection is done as follows:

I. In the first column, write in the name of each person living in the household ***aged 12-65***, beginning with the ***oldest, in descending order of age***. Write down only those who are eligible.

II. In the second column, write down the ***age*** of each person (age at last birthday.)

III. In the third column, beginning with the oldest, **assign a number to each person, starting with 1.**

IV. The last number in the "Number" column indicates the ***row*** of the Kish Grid that should be used.

As an example: supposing that the eligible members of the household are those shown in the following table: following instructions, numbers have been assigned ***only to those persons aged 12 to 65*** (6 persons). Those who are outside this age range are not included in the list. The oldest person, Mary, is assigned number 1 and so on up to the last eligible person, Carla, who is assigned number **SIX** in the "Number" column. The **SIXTH** row of the grid should be used, and where it intersects with the letter **D**, we find the number **2**. The person numbered **2** in the "Number" column is the person selected for the sample, in this case, **Joseph, aged 43**.

Name	Age	Number	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Mary	62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Joseph	43	2	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
Charles	25	3	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3

Anne	22	4	**1**	**1**	**1**	**2**	**2**	**1**	**2**	**2**	**2**	**3**	**3**	**3**	**4**	**3**	**3**	**4**	**4**	**4**
Julie	19	5	**1**	**1**	**1**	**2**	**2**	**3**	**2**	**2**	**3**	**3**	**4**	**4**	**3**	**4**	**4**	**5**	**5**	**5**
Carla	12	6	**1**	**2**	**1**	**2**	**2**	**2**	**3**	**3**	**3**	**4**	**4**	**4**	**5**	**5**	**5**	**6**	**5**	**6**
			1	**1**	**2**	**1**	**2**	**3**	**4**	**3**	**4**	**4**	**5**	**4**	**5**	**6**	**7**	**6**	**7**	**7**
			1	**1**	**2**	**2**	**3**	**4**	**4**	**3**	**4**	**5**	**6**	**5**	**5**	**6**	**7**	**7**	**8**	**8**
			1	**1**	**2**	**2**	**3**	**3**	**4**	**4**	**5**	**5**	**6**	**6**	**7**	**7**	**8**	**8**	**9**	**9**
			1	**1**	**2**	**2**	**3**	**3**	**4**	**5**	**4**	**6**	**7**	**6**	**8**	**8**	**9**	**9**	**10**	**10**
			1	**2**	**3**	**4**	**2**	**4**	**5**	**5**	**6**	**6**	**7**	**7**	**8**	**10**	**8**	**9**	**10**	**11**
			1	**2**	**3**	**4**	**2**	**4**	**5**	**6**	**5**	**7**	**8**	**7**	**9**	**10**	**9**	**11**	**10**	**12**

Before beginning the selection, the interviewer should introduce himself or herself to the person who answers the door of the household selected.

INTRODUCTION

Good morning, good afternoon
We are conducting a study in thousands of households in the country on issues related to public health. Our goal is to obtain information that will help guide actions that will help solve public health problems in our country. Your cooperation in answering this survey will be very useful for this effort. Your answers will be totally confidential and anonymous. They will be used only for statistical purposes.

- **Part 3** of the cover sheet consists of information on the ***person selected***. The interviewer will write in the sex and age of the person interviewed (age at last birthday). As to the question ***Sex***, the interviewer should enter one of the two options by observation, without actually asking the question. The interviewer should then confirm whether the age of the person selected is the age given in the Kish Grid, and if not, should correct it. To do so, the interviewer will ask: *How old are you*? And write the answer in the box provided.

1.1.1.7.1.1

PERSON SELECTED

Sex of respondent		**Age of respondent**	
	1. Male		**Age at last birthday**
	2. Female		

- **Part 4** is the ***Field Log***, where the date and time of each visit is entered, as well as the outcome or result (interview conducted, the person to be interviewed is absent, date agreed on for a new visit, refusal, etc.)

  The Field Assistant or Supervisor will enter the final outcome of the process (questionnaire supervised and sent for data entry or discarded due to absence, or refusal, or failure to find the household, etc.)

Example:
1.1.1.8
1.1.1.9 3- FIELD LOG

Date and time of VISIT	Interviewer Code	Outcome	Interviewer signature
14/09/2003- 10 a.m	**E10**	**Residents absent**	1.1.1.9.1 *John Smith*
14/9/2003 13 a.m	**E10**	**Person chosen was busy. Interview rescheduled for 15/09 at 11.30 a.m.**	**1.1.1.9.2** *John Smith*
15/09/2003 11.30 a.m	**E8**	**Interview conducted**	**1.1.1.9.3** *Mary Jones*
SUPERVISION	**Supervisor code**	**Result**	**Supervisor signature**
Field Assistant	**A1**	✓ **Supervised**	
Supervisor	**A**	✓ **Supervised. Send for data entry**	

1.2. GENERAL INFORMATION ABOUT THE QUESTIONS ON THE QUESTIONNAIRE

MODULE 1: GENERAL INFORMATION

QUESTIONS 1 and 2. For question No. 1, if the respondent **does not have** a job that involves full-time, part time or occasional work (answer 4 and following,) skip directly to question No. 3. Question No. 2 should be answered by those who gave responses 1 to 3 to question 1. Show Card No. 1.

QUESTIONS 3 and 4. Question 3 asks about the household's main source of income. In the event the person interviewed is unsure about question 4, the *Head of Household* should be that member of the household who brings in the highest amount of income in question 3. If the person interviewed is the person considered as Head of Household, enter that information on question 4.

QUESTION 5. Ask the respondent about the household's monthly income in an average month, and write it down in the pertinent income range. In the event the person interviewed is a minor, ask him or her to try to obtain the information from an adult present in the household. The answers should be given in local currency. Show Card No.1.

QUESTIONS 6 and 7. Education systems vary from country to country, and therefore the answers should be adjusted to reflect the situation in the country. Ask the question as it appears on the questionnaire, and show Card No.1. If the respondent is the Head of Household, do not ask question 7, but enter the same answer as for question 6.

QUESTIONS 8 and 9. Question 8 asks about the respondent's type of health insurance. In question 9, the respondent should indicate both his/her marital status and whether or not he/she lives with a partner.

The country may wish to add to this module questions related to nationality, ethnicity or religious affiliation.

MODULE 2: SUBSTANCE USE

This module is divided into groups of questions about each substance examined. The instruction "skip" is specified in each group of questions, according to whether the respondent had or had not used the substance in a particular time period. The first question asks about use of the substance in the respondent's lifetime ("ever used"; if the answer is NO, then skip to the next substance, with the exception of some substances for which there are other questions of general interest, even though the respondent had never used. The instructions about "skipping" must be followed to avoid any inconsistency in the responses.

QUESTIONS TA1 to TA8. These questions ask about current and past use of tobacco.

QUESTIONS TA1 to TA8. These questions ask about the current and past use of electronic cigarettes (e-cigarettes.) These devices can be used with three different products: **nicotine, marijuana or with flavorings.** Question TA1 asks about the use of any of these alternatives. If the answer is NO, the interviewer ends this segment and skips to the next substance. Otherwise, the interviewer will continue to ask about the product(s) used with e-cigarettes. A respondent may answer that he or she uses one or more products.

QUESTIONS AL1 to AL9 and AU1 to AU10. The first questions are a set of questions on alcoholic beverages. Questions AU1 to AU10 correspond to the Alcohol Use Disorders Identification Test (AUDIT), which can help determine signs of problem use of alcohol and dependence on alcohol.

QUESTIONS TR1 to TR9. These are a series of questions about the use of tranquilizers. Card No. 5 should be shown here. Questions are asked about tranquilizer use with or without a medical prescription, but the emphasis is on the latter. For question TR6, several answers are possible, i.e., answering about more than one drug.

QUESTIONS ES1 to ES9. These are a series of questions about the use of prescription stimulants. Card No. 6 should be shown here. Questions are asked about stimulant use with or without a medical prescription, but the emphasis is on the latter. For question ES6, several answers are possible, i.e., may refer to more than one drug.

QUESTIONS AN1 to AN9. These are a series of questions about the use of prescription analgesics. Card No. 7 should be shown. Questions are asked about use of analgesics with or without a medical prescription, but the emphasis is on the latter. For question AN6, several answers are possible, i.e., may refer to more than one drug.

QUESTIONS MA1 to MA12. This set of questions asks about lifetime us of marijuana (ever used,) age of first use, and most recent and current use of marijuana, frequency of use, the quantity used per month, price, and amount of money spent on marijuana use per month.

QUESTIONS HA1 to HA6. This set of questions asks about the use of hashish, and the frequency of use.

QUESTIONS CO1 to CO10. Questions about lifetime use, age of first use, and most recent and current use of cocaine hydrochloride, the frequency of use, the amount used per month, price of the drug and amount of money spent on use per month.

QUESTIONS PB1 to PB10. These are a set of questions about lifetime use, age of first use and most recent and current use of cocaine base paste, the frequency of use, amount used per month, price of the drug, and the amount of money spent on use per month.

QUESTIONS CR1 to CR10. This section has questions about lifetime use, age of first use, and recent and current use of crack, frequency of use, amount used per month, price of the drug, and amount of money spent on use per month.

QUESTIONS EX1 to EX10. These are questions about lifetime use, age of first use, and most recent and current use of MDMA or ecstasy, frequency of use, amount used per month, price of the drug, and amount of money spent on use per month.

QUESTIONS LS1 to LS6. Questions about lifetime use, age of first use, and most recent and current use of LSD, frequency of use in the past year.

QUESTIONS IN1 to IN8. Questions about lifetime use, age of first use, and most recent and current use of inhalants, frequency of use in the past year.

QUESTIONS PO1 to PO6. Questions about lifetime use, age of first use, and most recent and current use of poppers, frequency of use in the past year.

QUESTIONS OD1 to OD16. These questions ask about the use of other substances, with each drug on a new line.

MODULE 3: PERCEPTION OF RISK AND FACTORS ASSOCIATED WITH DRUG USE

This module has two parts that are central to the study. It asks about the perceived risk of occasional or frequent use of a series of substances, which are identified in questions **PR1 to PR17**.

The second part has questions about factors associated with drug use. Only respondents who ***said they had used alcohol or other drugs in the past 12 months*** should answer these questions (questions **FA1 to FA20**.)

Then we have questions **FA21 to FA23**, which are asked only of respondents **aged 18 and older.**

Lastly, questions **FA24 to FA27** are asked of all respondents.

MODULE 4: INFORMATION ON TREATMENT

QUESTIONS TT1 to TT6. This module is designed to find out whether the respondents have received any time of treatment for their use of alcohol or other drugs, or whether they have felt the need for treatment.

MODULE 5: INFORMATION ON DRUG TRAFFICKING/DEALING QUESTIONS TD1 to TD11.

This module asks about different issues related to drug trafficking or dealing. Questions **TD1 to TD9** are concerned with security and safety issues; question **TD10** is about ease of access to different substances, and lastly, question **TD11** refers to drugs offered to respondents.

MODULE 6: DRUG PREVENTION AND CONTROL MEASURES

QUESTIONS PC1 to PC36. This module consists of questions about the respondents' opinions about different drug-related issues. Card No. 11 should be shown when asking question PC16; a maximum of three answers is permitted. Card No. 12 should be shown in connection with questions PC23 to PC28, and Card No. 13 for questions PC29 to PC33.

MODULE 7: INFORMATION ON EMPLOYMENT

This module applies only to those who gave answers 1, 2 or 3 to question 1 of module 1, i.e., those who said that they are working. This module contains questions PL1 to PL10.

1.3. <u>INFORMATION ON DRUGS</u>

i. MARIJUANA (Cannabis)

Cannabinoids are substances derived from the cannabis plant. The product obtained from cutting the upper portion of the plant (leaves and stems), and when it is chopped up and rolled like cigarettes, it is called *marijuana*. *Hashish* is derived from the separated resin that is extracted and filtered from the upper part of the plant and the underside of the leaves; *hashish* oil is a distilled concentrate of hashish. Cannabinoids are usually smoked, but can also be used orally and are sometimes mixed into tea or food.

ii. COCAINE

Cocaine, a natural substance produced by the coca plant, is consumed in various forms (e.g., coca leaves, coca paste, cocaine hydrochloride and cocaine alkaloids), whose potency varies due to differing levels of purity. Cocaine is the active ingredient in all of the preparations. Chewing coca leaves is a practice generally confined to the indigenous populations of Central and South America, where coca is grown. The use of cocaine base paste, a crude extract of the coca plant, is found almost exclusively in the cocaine-producing countries of South and Central America, where it is called "*basuco”* or *“pasta base*." The solvents used in the preparation of cocaine base paste often contaminate the paste and can have toxic effects on the central nervous system and other organs when it is smoked. Powdered cocaine hydrochloride is usually “snorted” through the nose (“inhaled"), or dissolved in water and injected intravenously. Common in the United States is the use of crack, a cocaine alkaloid which is extracted from the powder and mixed with sodium bicarbonate and dried or heated into small rocks. Crack differs from other forms of cocaine in that it is easily vaporized and, when inhaled, its effects are very rapid. The clinical syndrome and adverse effects of *crack* are similar to those produced by comparable doses of other cocaine preparations.

iii. HALLUGINOGENS (LSD, mushrooms, Peyote)

Hallucinogens are a set of substances, the most common of which in our hemisphere are LSD, mushrooms and peyote. They cause changes in perception that occur in a state of complete wakefulness, such as subjective intensification of perceptions, depersonalization, derealization, illusions or hallucinations, which appear during or shortly after use of hallucinogens. Psychological and maladaptive behavioral changes also occur, such as marked anxiety or depression, ideas of reference, fear of losing control, paranoid ideation, impaired judgment or social or work performance. There are also a series of symptoms such as dilation of the pupils, tachycardia, sweating, palpitations, fuzzy vision, trembling, lack of coordination.

iv. HEROIN

Heroin is an opiate drug that is synthetic or partly synthetic. In some countries, it is the opiate that is most commonly used and most abused. It is usually used intravenously (injected), but can also be smoked or snorted when the heroin in very pure.

Heroin use causes changes in behavior: euphoria followed by apathy, agitation or psychomotor retardation, changes in thinking and judgment, or social or work impairment. It also causes constricted (“pinpoint”) pupils (miosis at the beginning, constriction of the pupil of the eye, which becomes very small), drowsiness, jabbering, attention impairment, or memory loss.

v. INHALANTS contact cement, gasoline/petrol, acetone, etc.)

These substances include aliphatic and aromatic hydrocarbons that are found in substances such as gasoline (petrol), glue, solvents and spray paint. Halogenated hydrocarbons are less commonly used (they are found in cleaning products, correction fluid, sprays), together with other volatile compounds that contain esters, ketones and glycols. Most of the compounds inhaled are a mixture of a large number of substances that can produce psychoactive effects, and it is often difficult to determine which is the precise substance responsible for the disorder.

A wide variety of methods are used to inhale the intoxicating vapors. Commonly, a rag soaked in the substance is held over the mouth and nose, and the vapors are inhaled. The substance can be placed on a piece of paper or in a plastic bag and is then inhaled from the bag. The substances can also be inhaled directly from their containers, and in the case of aerosols, they can be sprayed directly into the mouth or nose. There have also been cases in which the person heats the compounds to speed up vaporization. Inhalants reach the lungs, blood and CNS very rapidly.

vi. SYNTHETIC DRUGS

Synthetic drugs are a large group of psychoactive substances that have certain characteristics in common. They are manufactured by means of chemical synthesis in clandestine laboratories, and can be made out of pharmaceutical products such as ephedrine and pseudoephedrine that are found in some flu medications. They are used in the illicit drug market to manufacture methamphetamines.

Methamphetamine is the synthetic drug most commonly used. Illicit production of methamphetamine is significant, particularly in North America, which is why it receives more attention in action plans to control synthetic substances in that part of the world. Although methamphetamine does have legitimate medical uses, most methamphetamines are trafficked illegally and produced in clandestine laboratories. Amphetamine, a synthetic stimulant like methamphetamine, is also produced clandestinely using very similar processes and its effects are similar.

Other substances such as MDMA, GHB, Rohypnol and Ketamine are also known as "club drugs" because of their association with rave parties attended by young middle-class people. GHB and Rohypnol are central nervous system depressants. Use of them has been described in certain cases where victims of sexual assault who had taken them were rendered incapable of resisting. Ketamine is a dissociative anesthetic that has become somewhat popular in rave circles.

Other substances may be included in synthetic drugs, such as synthetic opiates, which are less frequently used in Latin America: oxycodone and the dissociative anesthetic phencyclidine (PCP). While not an opiate, we also have the hallucinogen lysergic acid diethylamide (LSD) which has been used in Latin America since the nineteen seventies, but at very low rates of consumption.

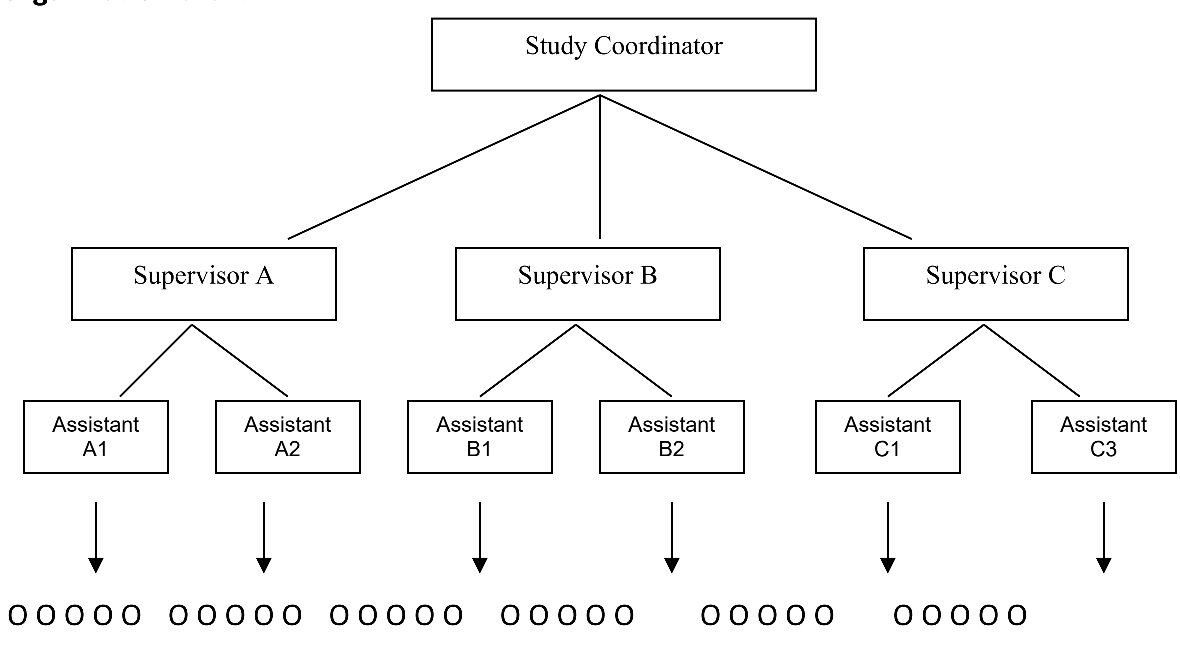
2. INSTRUCTIONS

As explained earlier, the National Observatory on Drugs is responsible for the study. It may delegate the fieldwork to an outside agency that has experience running studies as complex as the one described in this Protocol. Assuming that to be the case, the following sections deal only with the more general aspects of the field work.

2.1. ORGANIZATION OF THE FIELDWORK

The National Observatory on Drugs will be responsible for and will have overall coordination of the study. Execution may be delegated to an outside agency, who will designate the Study Coordinator. Below we give a typical organization chart for the fieldwork. In addition to the Study Coordinator, the team consists of supervisors, assistants and interviewers. The degree of complexity of the organization of the workload and the specific number of people performing each function will depend on the size of the study, both geographically and in terms of the number of people making up the sample.

Organization chart



Ref - . O = Interviewer

2.2. INSTRUCTIONS FOR THE STUDY COORDINATOR

2.2.1. OVERALL RESPONSIBILITIES OF THE STUDY COORDINATOR

The Study Coordinator is responsible, in coordination with the National Observatory on Drugs, for overall direction of the study, and control and supervision of all of the procedures involved The Supervisors, Assistants, Interviewers and data entry personnel must report in detail to the Study Coordinator during collection and processing of the information.

2.2.2. SPECIFIC TASKS OF THE STUDY COORDINATOR

- Have direct supervision of the work of the Supervisor and data entry personnel.
- Make sure that all of the technical and methodological procedures are followed, as well as instructions given by the National Observatory on Drugs during the course of the study.
- The Study Coordinator will receive the materials and send them to their destinations.
- Determine and directly assign the work to the team of Supervisors.
- Is responsible for keeping to the schedule at all stages.
- Must monitor expenditures in accordance with the assigned budget, to ensure that the study fully meets the goals agreed on with the National Observatory on Drugs.
- Prepare and check the materials for the fieldwork.

2.3. INSTRUCTIONS FOR THE SUPERVISORS

2.3.1 OVERALL RESPONSIBILITIES OF THE SUPERVISORS

The Supervisors are responsible for the fieldwork and have direct responsibility for the Field Assistants and the Interviewers. The Supervisors answer directly to the Study Coordinator**.**

2.3.2. SPECIFIC TASKS OF THE SUPERVISORS

- Have direct supervision of the work of the Field Assistants and the Interviewers.
- Make sure that all of the technical and methodological procedures are followed, as well as instructions given by the Study Coordinator during the course of the study.

- Coordinate and supervise the fieldwork assigning responsibilities and workloads to the Assistants and Interviewers.
- Receive and distribute the work materials to the Field Assistants and the Interviewers.
- Receive the questionnaires, supervise that the procedure is correct, and send them for processing.

2.4. INSTRUCTIONS FOR THE FIELD ASSISTANTS

2.4.1. OVERALL RESPONSIBILITIES OF THE FIELD ASSISTANTS

The Field Assistants must help their Supervisor by distributing and collecting the materials, and by helping with second (and third) interview visits if necessary.

2.4.2. SPECIFIC TASKS OF THE FIELD ASSISTANTS

- Help distribute the work and materials to the Interviewers.
- Receive the surveys and perform an initial review of the forms.
- Supervise the work of the Interviewers during any re-interviews.
- Make sure that all of the technical and methodological procedures are followed, as well as instructions given by the Study Coordinator during the course of the study.

2.5 TASKS AND IMPORTANCE OF THE INTERVIEWER

2.5.1. Overview of the Interviewer functions and preferred profile

The Interviewer is the person responsible for obtaining the information by administering the questionnaire and following all the details of the procedures and forms described in the ***Questionnaire Instruction Booklet.***

The interviewer plays a very important role in achieving the study objectives, and the success of the study will depend to a large extent on his/her capacity, sense of responsibility, and engagement with the work.

Special personal attributes are needed, such as empathy, getting along with people, flexibility to adapt to different situations, whether individual or community, as well as a sense of responsibility and discretion, given that he/she will have access to confidential information.

The interviewer answers to the Supervisor, who will assign the work, either directly or via the interviewer's assistants.

2.5.2. Tasks and responsibilities of the Interviewer

From an ethical standpoint, the Interviewer must:

- Take on the job responsibly and with dedication
- Behave correctly at all times, and should be well-groomed.
- Be fully conversant with the present manual and Questionnaire Instruction Booklet.
- Participate actively in all of the scheduled training sessions.
- Notify his/her immediate supervisor and, if pertinent, the Coordinator General, of any irregularities or practices that deviate from the Research Plan.
- Resign from the job if for any reason, once he/she is in the field, the Interviewer finds it impossible to carry out the work according to the guidelines.
- Respect the confidentiality of the information supplied by the respondents, and of the information gathered during the study process.

The Interviewer's obligations overall are:

- Locate the households assigned for interviewing.
- Make every effort to conduct the interview, and re-visit the household at least **three times** if the residents are not at home or, if the person in the household selected according to the Kish Grid is not at home. The Interviewer should obtain additional information in the neighborhood if needed for greater efficiency.
- In the event of a refusal, the Interviewer should, without pressuring, politely stress how important it is that the person participate in the study. If the person continues to refuse, say that another member of the research team may come again and ask the person to participate in the interview.
- Having secured the interview, inform the respondent about the purpose of the study and how it will be carried out.
- Complete the form (including the cover sheet) accurately, following the instructions and without skipping any questions or changing their order.

2.5.3. Specific recommendations for conducting the interview

- The Interviewer should introduce himself/herself by name, identify the organizations that are conducting the study, assure the respondent that the answers are confidential, explain how the household was selected, and explain the purpose of the study.
- The interviews should be conducted as far as possible without distractions or interruptions by other people. The interviewer should note under "Observations" on the cover sheet any unusual situation or interruptions that might affect the quality of the responses.
- The interviewer must protect the anonymity of the person interviewed and must deliver the materials only to the person designated by the Supervisor.
- Insist on the interview. If the person to be interviewed refuses or cannot be located, the interviewer should go back to the household up to three times more, on different days and at different times of day, before giving up on the survey. It will be very helpful to ask the neighbors or other members of the household when it will be easiest to find the members of the household, or the person selected.

2.5.4. Interviewer's behavior during administration of the questionnaire

1. The interviewer should read out the questions and, when pertinent, also read out the possible answers as they appear on the questionnaire. The interviewer should show the cards as indicated on the questionnaire. The interviewer must not change the order of the questions nor the text of the questions. The interviewer must not:
   - Fail to read the answers as written (must not summarize, for example)
   - Alter or change the questions on the questionnaire.
   - Change the order of the questions.
2. Should not ask questions that ought to have been skipped, and should not fail to ask questions that should have been asked.
3. Interviewer questions to clarify answers, encourage the respondent to answer, or to explain any questions on the questionnaire should not introduce any bias or guide responses in a particular direction. Examples of non-leading questions are: "anything else?" or "Could you repeat that, please?"
4. Should never "interpret" the question for the respondent.
5. Should give clear instructions for any questions that may need them.
6. Should not interrupt the respondent, and should not give any personal opinions.
7. If the respondent does not understand a question, read it again slowly, but do not explain it, except for some questions discussed in detail in the Instruction Booklet where some explanation is permitted.
8. Try to maintain eye contact with the respondent to create un ambiente adecuado para la conversación.
9. Keep the interview flowing smoothly and, if possible, record each answer while asking the next question.
10. Record the responses as discreetly as possible.
11. If a respondent's comments are written down, use the respondent's own words. Do not summarize.
12. If the respondents give answers that are ambiguous, insistir para obtener otras más concretas y que se ajusten a las alternativas descritas en el cuestionario.

13. Do not engage in conversation with the respondent about the subject matter of the survey, since the Interviewer's own observations may color the answers.
14. At the end of the survey, go over the questionnaire again before leaving the respondent to check that all of the questions have been asked and all of the answers have been recorded. It is preferable to ask any questions that were omitted, even if they are out of order, or to ask the respondents to repeat an answer rather than leave some of the points blank.
15. Inform the respondent of the possibility that a Supervisor may visit him/her.

2.5.5. Interviewer materials

- Identification card
- Interviewer instructions
- Questionnaire Instruction Booklet
- Pencil, eraser and pencil sharpener.
- Labelled envelope containing the following:
    - Forms for administration of the survey.
    - Map showing the precise location of the household to be interviewed.
    - Cards to show for questions, where indicated in the questionnaire.

3. TASKS PRIOR TO THE FIELDWORK

The Study Coordinator must perform the following prior to the fieldwork:

- Obtain the materials needed to identify the households selected for the sample provided by the National Statistics Institute or other body responsible for mapping. The Coordinator should obtain a map showing the precise location of the households to be surveyed.
- The Coordinator must participate in the organization of all phases of the research. In addition, the Coordinator must be actively involved in the following:

a. Selection of the research team

1.1.1.9.4 *The Study Coordinator should play an active role in selecting the Supervisors, Field Assistants, Interviewers and Data Entry personnel (if pertinent.) The* *people chosen for each of these tasks must fulfill the job requirements and profiles.*

b. Planning and executing the pilot test, if necessary.

c. Training plan.

All of the research team needs to be trained, and the Study Coordinator, along with the country's National Observatory on Drugs, is responsible for the design of the training and selection of experts to do the training.

Generally speaking, the training should include the following essential points:

- A clear definition of the objectives of the study.
- Determine the responsibilities of the team members.
- Ensure full engagement by all.
- Familiarity with how to use the instrument properly, including the use of the Cards. The interviewer must be in a position to answer possible questions from the respondents.