DESIGNING STUDIES TO EVALUATE AND VALIDATE SCALES AND INDICATORS OF PROBLEM DRUG USE

COMPONENT 1 CONSOLIDATION OF THE NATIONAL DRUGS OBSERVATORIES
WG 1.4 STUDIES TO EVALUATE AND VALIDATE SCALES AND INDICATORS OF PROBLEMATIC DRUG USE

This project is funded by
the European Union

PROGRAMA DE COOPERACIÓN ENTRE AMÉRICA LATINA, EL CARIBE Y LA UNIÓN EUROPEA EN POLÍTICAS SOBRE DROGAS
COOPERATION PROGRAMME BETWEEN LATIN AMERICA, THE CARIBBEAN AND THE EUROPEAN UNION ON DRUGS POLICIES

Proyecto financiado por la Unión Europea

Designing Studies to Evaluate and Validate Scales and Indicators of Problematic Drug Use

Cooperation Programme between Latin America, the Caribbean and the European Union on Drugs Policies (COPOLAD II)

Credits

This document has been developed within the framework of the Cooperation Programme between Latin America, the Caribbean and the European Union on Drugs Policies (COPOLAD II); Component 1: Institutional Strengthening of National Drug Observatories; Activity 1.4: Design of studies to evaluate and validate scales and indicators of "Problematic Drug Use".

LEADING COUNTRIES

Spanish-speaking group: Chile
English-speaking group: Jamaica
European reference countries: Romania and Cyprus

EXPERT GROUP

Coordination: Francisco Cumsille (Chile)
Editor: Francisco Cumsille (Chile), Álvaro Castillo-Carniglia (Chile), Carlos Ibañez (Chile), Milica Georgescu (Romania) and Colette Cunningham-Myrie (Jamaica).

Review: Julián Vicente (EMCDDA), Marya Hynes (OID/CICAD/OEA), Pernell Clarke (OID/CICAD/OEA), Novie Younger-Coleman (Jamaica), María Elena Alvarado (Chile), José Marín (Chile), Esteban Pizarro Muñoz (Chile), Nicolás Rodríguez (Chile), Ioanna Yiasemi (Cyprus)), Ruxanda Iliescu (Rumania), Roberta Caixeta (PAHO/WHO), Carolina Chavez (PAHO/WHO), Luis Alfonzo (PAHO/WHO) and Graciela Ahumada (Coordinator of Component 1 of COPOLAD II).

Edition

Execution and Coordination Body (ECB) of the Cooperation Programme between Latin America, the Caribbean and the European Union on Drugs Policies (COPOLAD II), FIIAPP.

Acknowledgements

The ECB would like to thank those who have participated in the Expert Group; and to specially thank the *European Monitoring Centre for Drugs and Drug Addiction* (EMCDDA) and the Inter-American Drug Abuse Control Commission (CICAD-OAS).

C. Beatriz de Bobadilla, 18. 28040 Madrid, Spain. February 2019.

1. Introduction

As part of the collaborative effort between the **Cooperation Programme between Latin America, the Caribbean and the European Union on Drugs Policies (COPOLAD)** and the National Drug Commissions (NDC) in Latin America and the Caribbean, this program is addressing Problematic Drug Use as a research priority. The national drug observatories (NDOs), which are the research and analysis office within the NDCs reviewed the existing scientific literature how to measure Problematic Drug Use through national, population-based, epidemiological surveys, through the use of scales and other scientific tools.

There are several limitations to the use of scales and tools to measure substance use disorders (SUDs) in population surveys, also referred as problematic drug use, abuse or dependence, which will be defined in more detail bellow. The primary issue is the validity of scales and instruments when they are used in drug surveys. This refers to whether those scales and instruments are measuring what they are meant to measure in terms of content and prediction The validity of a scale or instrument lies in its ability to distinguish between individuals who have a characteristic under study (e.g. dependence on some substance), from those who do not have it.

In order to confirm validity, it is essential to determine whether positive and negative cases are being correctly classified. The correctness or incorrectness of the classification will always be in reference to a defined gold standard, for example, by a clinical interview conducted by an expert.

For a valid measurement of dependence or problematic use to be taken in a survey, the survey questions must evaluate all the relevant dimensions of problematic use or dependence. These questions should be well designed, the interviewee must clearly understand what is being asked, and the conditions under which the interview is conducted must favor candid responses. The later, should be standardized thus it can be replicated under the same conditions in different studies.

Therefore, the first step is to assess if the questions are being well formulated in order to measure the criteria that they are supposed to measure. Then, it is important

to know whether the questions are understood by the target population. And finally, it depends on those who answer the questions to do so truthfully, and this depends on the conditions in which interviews are conducted.

Therefore, a potential weakness in the application of scales on SUDs in national surveys is the unknown validity of the instruments under those circumstances. Most of the scales and instruments for assessing SUDs in the general population were designed for a face-to-face interview modality performed by trained interviewers. Questions designed for face-to-face interview allow for an opportunity for the subject to ask for clarification. However, in the case of secondary school population studies, the surveys are self-administered in the classroom and in groups, which may impact the validity of responses. Furthermore, we are applying the same questions to adolescents aged 12 and 18, who may have different levels of understanding.

Most school population surveys do not seek to estimate diagnostic categories such as such as abuse or dependence, but do include instruments that assess levels or risky behaviors, such as binge drinking, risk of dependence on marijuana through the CAST scale (Cannabis Abuse Screening Test)(1), or problematic consumption, abuse and dependence of alcohol and other substances, CRAFFT[1] (Substance Abuse Screening Test Among Adolescents)(2).

The problem with these indicators, similar to those used in the general population, is that we do not know their validity under the current conditions of application. Efforts to assess the psychometric characteristics of these instruments, for example, in terms of internal consistency, factorial structure or predictive capacity are undoubtedly important and necessary, but it is also necessary to assess the predictive validity of these instruments. In this sense, knowing the sensitivity, specificity and predictive values of the instruments, under conditions of self-administration and for the entire age range included in those surveys (usually from

[1] https://www.masspartnership.com/pdf/CRAFFTScreeningTool.pdf

8th to 12th grade or 13 to 18 years of age) is key if it is decided to maintain and to deepen the measurement of drug use disorders in the school population.

Another consideration in the use of the results derived from these scales and instruments has to do with the interpretation of phenomena with a lower prevalence than that used when estimating the sample size. For example, in the year 2016 in a given country in the general population study, 0.4% of the population met criteria for cocaine abuse or dependence (problematic use) and 0.2% for coca paste. With so few cases, opportunities to learn more about this population— in terms of its distribution by sex, age, region or any other characteristic of interest — are very limited. In that same year, the percentage of women with Problematic Drug Use of coca paste was lower than 0.1%. On the other hand, the absolute error for lifetime prevalence of illicit drug use in the effective sample was 1.5%; that is, we are talking about phenomena with a prevalence many times lower than the sample error with which the survey was designed.

With this background in mind, it is of paramount importance to carry out studies that assess the validity of the instruments and scales used to estimate SUDs in national drug survey. In addition, it is important for NDOs to incorporate confidence intervals or standard errors as a regular practice, so that people who use this data are able to assess the quality of these statistics, or alternatively, to provide well-founded critique when the data is not used appropriately.

In the following sections we will focus on how the countries of the community of Latin American and Caribbean States (CELAC) have been evaluating what has been called problem substance use, describing the criteria use to create that category. We will also discuss the diagnostic criteria that are currently used in the scientific literature.

2. Objectives

The general objective of this document is to produce methodological designs necessary for the CELAC NDOs and other users to **measure Problematic Drug**

Use in their populations, in order to generate valid information for public policy

The specific objectives are:

1.- To know the criteria and instruments that countries are using to estimate the proportion of past-year users of marijuana and alcohol who meet the criteria for problematic use.

2.- To describe and analyze the new criteria and instruments developed by international bodies that are now available.

3.- To propose methodologies that allow comparison of criteria and instruments, current and new, under a variety of conditions; on the one hand taking as reference a gold standard from a structured clinical interview, and on the other hand comparing the different diagnostic criteria between them under real conditions of application of a drug survey.

3. Basic definitions

3.1. Problematic use

The concept *Problematic use*, has been widely used by CELAC National Drug Observatories in their surveys; however, is not incorporated into the International Classifications of Diseases (ICD). This presents some issues as the term appears to be interpreted in different ways making it difficult to compare data from other regions of the world that use internationally agreed terminologies such as Substance Use Disorders (SUDs).

According to the World Health Organization, *problem drinking* refers to a pattern of alcohol consumption that causes problems, whether individual or collective, health or social. This term has been used since the mid-1960s in a more general sense to avoid referring to the concept of *alcoholism* as a disease.

In certain contexts, problematic drinking has been used as a synonym for the concept of alcohol dependence in its early or less severe phases.

Other formulations used to avoid stigma are *drinking-related problems* and *drinking problems*. Some experts have used the term *problematic alcohol use* to cover another related concept: a use that has the potential to cause problems (more or less equivalent to *high-risk use*).[2]

3.2. Problematic Use vs. Substance use disorders (SUD)

Disorders arising from substance use, like the vast majority of diagnoses in psychiatry, have no objective criteria for assessment, nor a biological basis to confirm a clinical diagnosis. Clinical evaluation in psychiatry has historically been associated with the idiosyncrasies of the theoretical postulates adopted by the examiner and open to significant subjectivity(3) This results in a lack of reliability in diagnostics, a fact that has been repeatedly noted in the scientific literature. A classic example that reported marked discrepancies between psychiatric diagnoses in the United States and England(4) served as a warning to review these diagnostic disparities.

A systematic effort has been made to establish internationally common criteria for the diagnosis and classification of mental illness, improving the reliability and validity of psychiatric diagnoses, examples range from the Feighner Diagnostic Criteria in the early 70's(5) to the ICD 11(6) and the fifth edition of the Diagnostic and Statistical Manual (DSM-5). Efforts to create a reliable diagnostic system were primarily aimed at developing glossaries that made it easier for professionals to agree on the concepts used; and specify operational inclusion/exclusion criteria for the various disorders.

The term ***Problematic use*****,** widely used in epidemiological study reports in the Americas, is usually an indicator composed of two diagnostic categories identified in ICD manuals. These are the ICD-10 (8) *Dependence Syndrome* and *Harmful*

---

[2] Lexicon of Alcohol and Drug Terms. Edited by the World Health Organization in 1994

Substance Use, and the categories of Substance Dependence and Substance Abuse from the DSM-IV (7). However, as already mentioned, most countries in the Americas use different classification categories to construct *Problematic Drug Use* indicator. As an example, they use the definition and diagnostic criteria of abuse taken from DSM-IV and that of dependence from ICD-10 to study and report prevalence of *Problematic Drug Use*.

This report suggests reconciling the terminology used in the reports of drug observatories in the region with the diagnostic categories used internationally. This will avoid confusion of concepts, and avoid difficulties in communicating results between countries.

Therefore, we recommend adopting the definition of ***Substance Use Disorders (SUD)*****,** present in one of the two main diagnostic systems in use internationally. The recently published versions of the ICD in its 11th revision of the World Health Organization(6) and the DSM in its 5th version published by the American Psychiatric Association (APA)(8) have significantly modified the conceptualization and ordering of SUDs. Next, we will review the definition of Substance Use Disorders in ICD-11 and DSM-5.

3.3. Substance use disorders or addictive behaviours in ICD-11

Changes to this group of disorders in the 11th revision of the International Classification of Diseases include a higher specification of different harmful patterns of substance use, which may be continuous or episodic and recurrent. These changes present a new category to denote unique episodes of harmful use and a simplification of diagnostic guidelines for substance dependence. Diagnostic concepts and categories of dependence and harmful use have been maintained; however, dependence has three simplified diagnostic criteria instead of six in ICD-10. Individuals must meet **at least** two criteria for a classification of dependence.

<u>Diagnostic Categories of SUDs in ICD-11:</u>

- Single episode of harmful substance use
- Harmful pattern of substance use

- Substance dependence

For the purposes of the objectives of this document we will only focus on the most relevant diagnoses, i.e., *harmful pattern of substance use* and *substance dependence*, which are described below.

3.3.1. Harmful pattern of substance use

It is a pattern of use of a substance that has caused harm to a person's physical or mental health or has resulted in some behaviour that can harm the health of others. The pattern of use of the substance is evident over a period of at least 12 months if use is episodic or at least one month if continuous. Harm to the individual's health occurs due to one or more of the following:

1) intoxication-related behaviour;

2) direct or secondary toxic effects on organs and body systems; or

3) harmful route of administration. Harm to the health of others includes any form of physical harm, including trauma, or mental disorder that is directly attributable to behaviour related to poisoning by the substance to the person to which the diagnosis of harmful pattern of use of the substance applies.

3.3.2. Substance dependence

Substance dependence is a disorder of the regulation of the use of substances arising from the repeated or continuous use of the substance. The characteristic trait is a strong internal impulse to use substances, manifested as an inability to control the use, increasing the priority given to substance use over other activities and the persistence of the use despite the harms or negative consequences. These experiences are often accompanied by a subjective sense of need or desire to use the substance. The physiological characteristics of dependence may also be present, including tolerance to the effects of the substance, withdrawal symptoms after cessation or reduction of substance use, or repeated use of the substance or pharmacologically similar substances to prevent or alleviate withdrawal symptoms. The characteristics of dependence are usually evident over a period of at least 12

months, but diagnosis can be made if the use of the substance is continuous (daily or almost daily) for at least 1 month.

Diagnostic Guidelines:

1. At some point in the previous 12 months or on an ongoing basis two or more of the following domains have been present: Loss of control over substance use: in relation to its onset, quantity, circumstances or completion of the consumption, often but not necessarily accompanied by a subjective sense of intense desire or need to use the substance.
2. Substance use becomes a growing priority in life, surpassing other interests, daily activities, responsibilities, personal or health care. Substance use has an increasingly central role in a person's life and relegates other areas of life to the periphery. Substance use often continues despite the occurrence of problems.
3. Physiological characteristics (indicating a neuroadaptation to the substance) manifesting through (i) tolerance, (ii) withdrawal symptoms after cessation or reduction of substance use or (iii) repeated use of the substance (or a pharmacologically similar substance) to prevent or alleviate withdrawal symptoms.

3.4. Substance-Related Disorders and Addictive Disorders according to DSM-5

Substance-related disorders are divided into two groups: substance-induced disorders and SUD. The great change introduced in this manual is the removal of the categories *Dependence* and *Abuse* to be integrated into the new SUD category that orders the disorder according to its level of severity according to the number of diagnostic criteria identified.

These criteria investigate a problematic mode of consumption of the substance that causes a clinically significant deterioration or discomfort and manifests itself **by at least 2 of the following events within** 12 months:

1. The substance is often used in larger amounts or for a longer period of time than intended.
2. There are persistent attempts or failed efforts to cut down or control the use of the substance.
3. A great deal of time is spent on the activities needed to obtain the substance, consume the substance or recover from the effects.
4. Cravings or a strong desire or urge to use the substance.
5. Recurring consumption of the substance leading to non-compliance with fundamental duties at work, school or home.
6. Continued substance use despite having persistent or recurrent social or interpersonal problems, caused or exacerbated by the effects of the substance.
7. Important social, occupational or recreative activities are given up or reduced because of substance use.
8. Recurrent consumption of the substance in situations in which it is physically hazardous.
9. Substance use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance.
10. Tolerance, as defined by either of the following:
    a. Markedly increased amounts of the substance in order to achieve intoxication or desired effect
    b. Markedly diminished effect with continued use of the same amount
11. Withdrawal, as manifested by either of the following:
    a. The characteristic withdrawal syndrome for the substance.
    b. The same (or a closely related) substance is taken to avoid withdrawal symptoms.

The number of diagnostic criteria identified allows to classify the disorder into:

- No disorder: 0–1 symptom present.
- Mild: 2–3 symptoms present.
- Moderate: 4–5 symptoms present.

- Severe: 6+ symptoms present.

4. Criteria currently used by CELAC countries

In order to understand the criteria for measuring *problematic use* in the general population and school population studies currently used by CELAC countries, an *Ad Hoc* questionnaire was sent (see questionnaire in Annex 1) to those in charge of National Drug Observatories (NDO) in the Spanish-speaking countries (Argentina, Bolivia, Brazil, Chile, Colombia, Dominican Republic, El Salvador, Guatemala, Honduras, Mexico, Panama, Peru, and Uruguay) and the English-speaking ones (Antigua and Barbuda, Bahamas, Barbados, Belize, Dominica, Grenada, Guyana, Haiti, Jamaica, St. Kitts and Nevis, St. Vincent and the Grenadines, Saint Lucia, Suriname and Trinidad and Tobago).

Spanish speaking countries 87% of the countries report that among the main objectives of institutions conducting population surveys on substance use is the collection of data on use; 75% refer their main targets include monitoring and surveillance of drug use as well as education and prevention and; 63% report the development of investigation to raise awareness on evidence-based interventions and policies on this matter.

In the case of the Caribbean countries, from the seven countries that responded to questionnaire, only one referred to the instruments that were referred to in the questionnaire, while two other countries point out that the same instruments have been used by other institutions, beyond the country's Drug Commission.

4.1 Instruments used in the general population

In relation to the main tools used in the general population, the results of this survey in Spanish-speaking countries show that 88% use a set of questions in their surveys

to assess diagnostic criteria for marijuana and alcohol dependence from ICD-10 and 75% use a set of questions that evaluates DSM-IV diagnostic criteria for abuse.[3]

Using the DSM-IV criteria for abuse and ICD-10 for dependence, among people who claim to have used the relevant substance during the last year, a person is classified as having *problematic use* if they meet at least one of these two conditions; Table 1 represents this situation:

Table 1: Classification for *problematic use* by combining DSM-IV and ICD-10

Abuse DSM-IV	ICD-10 Dependence	
	Yes	No
Yes	a	c
No	b	d

Where ***a+b*** represent cases that according to ICD-10 are classified on the condition of dependence, ***c*** represents cases that according to DSM-IV meet the condition of abuse (but not dependence according to ICD-10), and ***a+b+c*** is the total number of people classified with problematic use of the specific substance. Thus, for example, if ***N* is the total population** (expanded sample) and ***M*** represents the total number of people who, for example, claim to have smoked marijuana in the last year, the prevalence (in percentage) of marijuana use in the last year is $\boldsymbol{P = \frac{M}{N} * 100}$ and the percentage of people classified in the *problematic use* condition is given by $\boldsymbol{PU = \frac{a+b+c}{M} * 100}$

Another tool that has been used in general population surveys to early assess risky and harmful alcohol consumption is the AUDIT *(Alcohol Use Disorders Identification Test[4])(9, 10)* It is a 10-question instrument that has been validated for ICD-10

[3] With the exception of experience in a country for ICD-10 and DSM-IV diagnostic criteria, there are no validation experiences of the constructs being measured.

[4] In Spanish: https://www.who.int/substance_abuse/activities/en/AUDITmanualSpanish.pdf

alcohol use disorder diagnoses; each question is evaluated with scores from 0 to 4 so the total sum of the AUDIT varies from 0 to 40 points. There are different classification criteria which are presented in Annex 2 together with the corresponding questionnaire. Eighty-eight per cent of Spanish-speaking countries report using it for the definition of alcohol dependence and 50% to define harmful or hazardous use. Regarding English-speaking countries, only 1 of the seven countries that submitted information report having used it.

4.2 Instruments used in the school population

In the case ***of the adolescent population,*** beyond the fact that any use of psychoactive substances carries a specific risk in this population, several countries refer having used self-administered surveys in students (usually between eighth and tenth grade, i.e. between 13 and 17 years of age), to know the prevalence of problematic drug use in this population.

Sixty-three % of Spanish-speaking countries have ever used the CAST (Cannabis Abuse Screening Test) instrument to assess risky marijuana use, as have most Caribbean countries (11 countries in this region included the CAST in studies carried out between 2013 and 2014). This instrument applies to students who report having used marijuana in the last year. The scale contains 6 questions with answers with ordinal categories going from *never* to *very often*, which are dichotomized (with values 0 and 1 each) and whose sum ranges from a score of 0 to 6. Based on the score in most countries that use this scale, students are classified into three categories according to the following Table 2:

Table 2: Ranking by CAST scale score

Score	Classification
0-1	No risk
2-3	Low risk
4-6	High risk

Some countries use other cut-off points to classify risk levels: *low* (1-2 points), *moderate* (3 points), and *high* (4-6 points).

Another instrument used at some point in 38% of countries is the CRAFFT research tool to assess the risk of problematic use, abuse or dependence of alcohol and other drugs in adolescents. However, this instrument has not been used consistently in school population surveys.

5. Review of scales and criteria used to measure *problematic use*.

In addition to the progress brought by international classifications of diseases for the generation of a common language and terminology, it is also very important to avoid discrepancies in the diagnostic elaboration process. To this end, psychiatric interviews with different levels of structuring have been developed based on international disease classifications.

A structured psychiatric interview is a methodology that includes a script, which specifies the questions the interviewer must ask exactly as they are written; also may provide a selection of response alternative for the interviewee. These questionnaires may indicate questions to skip questions depending on the response given by the subject or depending on certain pre-defined characteristics the subject may have.

Similarly, in **semi-structured questionnaire,** questions should also be read literally, however, it allows the interviewer to add questions to specify symptoms and signs based on their clinical interpretation of the interviewee's answers.

Despite criticism that structured interviews reduce the complexity of psychopathological events and limit the expression of individual psychological issues, diagnostic reliability is significantly reduced when structured or semi-structured psychiatric interviews are not used. These findings justify the use of structured interviews as a method of choice for understanding the prevalence of SUDs.

In this context there have been developed structured psychiatric interviews aiming at both exploring the general psychiatric pathology of the adult, as well as specific types of problems (e.g. SUDs) or specific population categories (child population, elderly people, etc.).

Next, we will briefly introduce those which, even though making a diagnosis of a wide variety of psychiatric disorders, turn out to be more relevant to the study of the prevalence of SUD:

- Structured Clinical Interview for DSM-IV for Axis I disorders (SCID-I)(11, 12)
- Composite International Diagnostic Interview (CIDI)(13): Diagnosis with ICD-10 criteria, which also allows DSM-IV diagnostics.
- Schedules for Clinical Assessment in Neuropsychiatry (SCAN)(14): Diagnostics with ICD-10 criteria.
- Present State Examination last revision (PSE-10)(15): Symptoms that appeared in the last month are assessed.
- Diagnostic Interview Schedule (DIS)(16): Gathers diagnostic information into three systems (Feighner, DRC, DSM-III).
- Psychiatric Research Interview for Substance and Mental Disorders (PRISM)(17): Semi-structured interview to detect other mental disorders in people with SUDs (dual pathology).
- Mini International Neuropsychiatric Interview (MINI)(18):The MINI is a short-term structured diagnostic interview that explores the main psychiatric disorders of Axis I of DSM-IV and ICD-10.

In addition to the diagnosis of SUDs, we mention some interviews designed for the evaluation and screening of alcohol and other drug use:

- Addiction Severity Index (ASI 6)(19, 20): Is a semi structured interview that assesses the degree of severity of substance use in seven areas: general medical status, employment and support, alcohol use, drug use, legal status, family and social status, and psychiatric status.
- EuropASI. European Addiction Severity Index is based on the 5th version of the ASI (21)
- Opiate Addiction Treatment Index (OTI)(22): Specially developed to provide standard measures of opiate treatment. It is structured in 6 areas: drug use, risky behaviors for HIV, social functioning, criminal activity, health status and

psychological situation (consisting of the 28 items of the General Health Questionnaire, GHQ-28)(23). The OTI has advantage over the ASI in that takes approximately 20 to 30 minutes to complete, while the ASI takes around 45-60 minutes.

- Maudsley Addiction Profile (MAP)(24, 25): The MAP is a structured interview for treatment outcome research. It covers sixty-item over four areas: substance use, health risk behaviors, physical and psychological health, and personal/social functioning. Regarding application time it is shorter than the previous ones, from 10 to 15 minutes. The Maudsley was designed for adults.
- Alcohol-Related Disorder identification test (AUDIT) (9, 26): Simple screening test to detect problems from alcohol consumption through ten questions, developed by the World Health Organization.
- Screening for Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)(27, 28): to detect problems with alcohol, tobacco and other drug use through eight questions, developed by the World Health Organization.

Some epidemiological studies on general psychiatric disorders have been conducted in the community since the 1980s. They studied the prevalence rates of some mental disorders, including SUDs in countries such as Brazil, Chile, Colombia, Mexico, Peru and Puerto Rico(29) through the use of structured or semi-structured interviews based on the diagnostic criteria of international classifications. Among the instruments used are**:** The Present State Exploration (PSE)(30), the Diagnostic Interview Schedule (DIS) (31) and the Composite International Diagnostic Interview (CIDI)(13).

In the region there is a reasonable development of competencies for the realization of epidemiological studies of mental health disorders mainly using the CIDI. Drug Use Disorders show the highest sensitivity with 80% (95% CI,34.94%-100%) and specificity with 98.46% (CI 95%, 94.7%-100%), of all psychiatric disorders evaluated through CIDI, when compared to those with SCID-I(32).

Due to its widespread international use, its good performance in the detection of SUDs in epidemiological studies, its ability to deliver psychiatric diagnoses using both DSM and ICD criteria, the proposed gold standard for conducting studies to detect Problematic Drug Use/Drug Use Disorders, is the CIDI in its 3rd version.

Training for implementing the CIDI

WHO's WMH-CIDI (World Mental Heath-CIDI) standard training programme includes 30 hours of home-based pre-study, with a series of manuals and supporting CD-ROMs, followed by 3 to 5 days of in-person training by CIDI trainers (depending on the CIDI WHO-Approved Training and Referral Centre in charge of the training). An evaluation is carried out at the end of the in-person training, and only those who pass this assessment will obtain a WMH-CIDI certification. Those who fail to pass the training exam can attend a second training session for free to retake the test after reviewing the materials again.

6. Study designs to select SUD classification instruments

As mentioned above, the objective of this document is to have a methodology that allows one to measure SUD in population surveys; this section proposes different strategies for identifying people with a SUD. Basically, these strategies, some of which have been used in the past and others are more recent, are options on which a decision must be made in the future.

With the purpose of identifying the best classification criteria for SUDs, two research designs are proposed, which should be assessed through pilot studies in some countries of the region:

- The first design is to compare the different classification criteria (those that have been used so far and the new ones) with a gold standard obtained through a structured clinical interview conducted by professionals trained for these purposes.
- The second design do not consider a gold standard, and compares different classification criteria (those that have been used so far in CELAC countries

and the latest disease classification system, DSM-5 an ICD-11) through surveys that simulate the conditions for the classification of cases.

Based on these two options, the following methodologies are proposed for countries to evaluate the different classification criteria for SUDs. Both proposed designs are aimed at analyzing SUD due: alcohol and marijuana.

6.1. Validation through comparison with a gold standard.

To apply a gold standard to validate these instruments, the following four elements are required:

- Development of the validation instrument
- Selection of the gold standard
- Select the indicator(s) that will allow the identification of the best criterion
- Study design and sampling strategy.

The first element concerns the construction of an **instrument** (questionnaire), which can be applied by a trained (non-clinical) pollster, containing the questions out of which the criteria for shaping diagnostic categories can be derived from the following disease classification manuals: DSM-IV, ICD-10, DSM-5 and ICD-11. In the case of alcohol, the inclusion of the questions contained on the AUDIT scale should also be considered.

A comprehensive review was carried out on the items included in the national general population drug surveys of several countries, as well as those proposed by the Inter-American Drugs Observatory of CICAD/OAS for this type of study. As mentioned above, the survey questionnaire includes 5 items to determine the presence of substance abuse based on DSM-IV criteria, and 10 items to determine the presence of symptoms of dependence according to ICD-10 criteria. These 15 questions cover almost all of the criteria contained in ICD-10, ICD-11, DSM-IV and DSM-5 to measure harmful use, abuse, dependence and SUDs. However, some of the currently used items omit aspects that are described within the criteria, so we suggest making the following two changes:

- First change: item from original country survey: *Have you taken/used [name of drug] even though you intended not to?*

  Suggested modification: *Have you tried to control, reduce or stop using [name of drug], and failed to do it?*

  This question explicitly explores failed attempts to control or abandon use, which is a criterion included in all classification systems. One could even include "on more than one occasion" to explain that failed attempts to control or abandon use are repeated.

- Second change: the original item of the country survey: *Have you ended up taking/using [name of drug] in larger quantities than you expected?*

  Suggested modification: *Have you ended up taking/consuming [name of drug] in larger amounts or for longer than you planned to do?*

With this change we also include that the criterion considers use for longer than initially thought, rather than in larger quantities alone.

In addition, it is necessary to add the following question (question16) to investigate potential harm to third-parties as a consequence of substance use (necessary to determine harmful consumption according to ICD-11): *Has any family member or person close to you been physically or psychologically harmed due to your use of [name of drug]*?.

In short, the following 16 questions are proposed for 13 criteria which are presented in Table 3. It is important to remember that when "[name of drug]" is mentioned, for the purposes of this document we only refer to marijuana or alcohol.

Table 3: SUD criteria, questions and their relevance to different classification systems.

Criteria	DSM-IV	DSM-5	ICD-10	ICD-11	Questions. Thinking in the last 12 months:
1.-Withdrawal	D[1]	UD	D	D	1. Did you ever have times when you use [DRUG] to keep from having problems like these? *Anxiety/restlessness/irritability; stress/depression; nausea/vomiting; concentration problems; tremors; see, hear or feel non-existent things; fatigue/drowsiness/weakness; tachycardia; insomnia*
					2. Did you ever have times when you stopped, cut down, or went without using [DRUG] and then experienced withdrawal symptoms?
2.-Tolerance	D	UD	D		3. Did you ever need larger amount of [DRUG] to get the same effect you used to get?
					4. Did you ever find that you could no longer get high on the amount you used to use?
3.-Giving up of activities due to use	D	UD	D	D	5. Did you ever have a time when you gave up or greatly reduced important activities because of your [DRUG] use – like sports, work, or seeing friends and family?
4.- Time spent	D	UD			6. Did you ever have several days or more when you spent so much time using or recovering from the effects of [DRUG] that you had little time for anything else?
5.-Physical and psychological problems due to use	D	UD	D		7. Did you ever continue to use [DRUG] when you knew you had a serious physical or emotional problem that might have been caused by or made worse by using?
6.- Neglect of roles	A	UD	HU	HU	8. Was there ever a time when your use of [DRUG] frequently interfered with your work or responsibilities at school, on a job, or at home?
7.-Harmful use	A	UD	HU	HU	9. Were there times when you were often under the influence of [DRUG] in situations where you could have gotten hurt, for example when riding a bicycle, driving, operating a machine, or anything else?
			-		10. Did you ever get into physical fights while using [DRUG] or right after using?
8.-Legal problems	A	-	HU	HU	11. Were you arrested or stopped by the police more than once because of driving under the influence of [DRUG] or because of your behavior while you were under the influence of [DRUG]?
9.-Repeated attempts to abandon or control use	D	UD	D	-	12. Were there times when you tried to stop or cut down on your use [DRUG] and found that you were not able to do so?
10.-Use in larger quantities or for longer	D	UD		D	13. Were there ever times when you used [DRUG] more frequently or for more days in a row than you intended?
11.-Intense desire (Craving)	-	UD	D		14. Was there ever a time when you often had such a strong desire to use [DRUG] that you couldn't stop using or found it difficult to think of anything else?
12.-Social/inter-personal problems	A	UD	HU	HU	15. Was there ever a time when your use of [DRUG] caused arguments or other serious or repeated problems with your family, friends, neighbors, or co-workers?
13.-Damage to third parties	-	-	-		16. Was there ever a time when you could say that your using [DRUG] has negatively affected other people?

[1]Only applies to alcohol, not for marijuana;

Nomenclature: **D=**Dependence; **A=**Abuse; **UD=**Use Disorder; **HU**=Harmful use

Diagnostic categorization according to the different classifications can be found in Annex 2 along with the questionnaire.

It is important to note that the combined cells correspond to a domain (1 or more criteria) and those domains correspond to a set of symptoms. For example, ICD-11 considers 3 domains to assess substance dependence, one of which is the impact of the substance on some physiological aspects of the individual (indicative of neuroadaptation to the substance) containing criteria 1 and 2, corresponding to tolerance and withdrawal; these are analysed separately in the other classification systems.

Finally, the criterion of withdrawal according to DSM-IV in the context of this document should only be considered for alcohol dependence but not for marijuana dependence. This criterion was included for marijuana dependence in DSM-5.

The AUDIT instrument has been used to classify alcohol use disorders. The AUDIT questionnaire can be found in in Annex 1, as well as its classification pattern.

For the second element, regarding the decision on the **gold standard** and its operationalization, we propose the use of CIDI (Composite International Diagnostic Interview) standardized interview applied by a trained clinical professional. Through this interview, for each study subject, classifications of SUD should be obtained by two criteria:

a) **DSM-5**. By this diagnostic criterion, persons who have been using a specific substance during the last year are classified into 1 of the 4 following categories according to SUD: ***without, mild, moderate or severe disorder*****.** The last three categories constitute the condition of ***SUD.***
b) **ICD-11**. Unlike the above criterion, ICD-11 classifies people in **one of** the following conditions: ***no disorder, harmful pattern of use, presence of symptoms of dependence***.

Using this same interview structure, we can populate a database with the answers for each of the questions that make up the different criteria used by the clinical

professional; and of additional analyses and even use other classification criteria such as DSM-IV and ICD-10.

Thirdly, the indicator(s) should be defined to allow for the best criterion to be decided and included in the questionnaires administered through surveys to estimate the proportion and number of people with alcohol and marijuana use disorder. These indicators should be related to the ability of an instrument to identify both cases with substance disorder and substance disorder, i.e. ***Positive and Negative Predictive Values, PPV and NPV,*** **respectively,** which correspond to two conditional probabilities defined as:

PPV- Probability that a case that has been classified as with a SUD in the interview through the survey (+), actually has a disorder (D) according to the clinical interview (gold standard), i.e. **PPV** = P(D|+).

NPV- Probability that a case that has been classified as without a SUD in the interview through the survey (-), does not have a disorder according to the interview (gold standard), i.e., **NPV**=P(no D|-).

It is important to keep in mind that the instrument used **in the survey necessarily** implies the existence of questions derived from some specific diagnostic criterion (e.g. DSM-5 or ICD 11). On the basis of this, people are classified in one of the two possible conditions, just as DSM-IV and ICD 10 are used today to determine whether or not a person is considered in the condition of *problematic use*.

However, in order to determine these indicators, there are **at least two methodological strategies** that can be developed.

1. The first option is to assume that we select a **simple random sample** **of *n* people who declare, as an example, having used marijuana during the last year;** these people are subjected to a clinical interview and also to the proposed face-to-face questionnaire. In both cases, each subject in the sample is classified as positive if a SUD is detected or negative in case of the contrary. The following Table 4, shows the configuration of the n cases according to two criteria:

Table 4: Classification of cases according to clinical interview and proposed survey in a sample of n cases.[5]

Clinical interview	Proposed survey		Total
	+	-	
D	a	b	a+b
no D	c	d	c+d
Total	a+c	b+d	n

This way, and in this specific case in which a random sample of marijuana users is used, the PPV and NPV correspond to the following expressions (which are usually multiplied by 100 to show them in %):

$$\boldsymbol{PPV = \frac{a}{a+c}} \qquad \boldsymbol{NPV = \frac{d}{b+d}} \qquad \textbf{(1)}$$

The higher (closest to the maximum of 1 or 100%) are these values, the best the classification procedure used in the proposed survey will be**.**

In addition to the two indicators above, which we reiterate are the most relevant, it is possible to determine four other indicators:

Prevalence (P): we will understand as prevalence of SUDs (at the sample level), the **proportion of people in the sample** among those who claim to have used the substance in the last year, **who were classified in that condition through the clinical interview** (*gold standard*). Its expression is (also usually shown in percentage):

$$\boldsymbol{P = \frac{a+b}{n}} \qquad \textbf{(2)}$$

Sensitivity (S): corresponds to the probability that a person with *SUD* (determined through clinical interview) will be classified in that condition by some criterion, in this

[5] In this table and hereafter, we will use the symbol "D" and "no D" to refer to cases with or without disorders detected in the clinical interview, and "+" and "-" for cases with and without disorders classified in the face-to-face interview.

case by the question proposed in the face to face interview, i.e., S=P(+|D), and its calculation formula in this case is:

$$\boldsymbol{S = \frac{a}{a + b}} \quad \textbf{(3)}$$

Specificity (Sp): corresponds to the probability that a person *without SUD* (determined through the clinical interview) will be classified in that condition by such criterion, i.e., Sp=P(- | no D) and its calculation formula in this case is:

$$\boldsymbol{Sp = \frac{d}{c + d}} \quad \textbf{(4)}$$

Proportion of cases with correct classification (PCC): corresponds to the proportion of people (with or without SUD according to clinical criteria) who were correctly classified by the face to face interview.

Its expression is as follows:

$$\boldsymbol{PCC = \frac{a + d}{n}} \quad \textbf{(5)}$$

Suppose a sample of the 1,000 people who claim to have used marijuana in the past year, who are voluntarily interviewed by a clinical expert, and then by an interviewer using a questionnaire (in face-to-face version) that contains the questions needed to classify according to some predetermined criteria; Table 5 below yields the results of this exercise:

Table 5: classification of cases according to clinical interview and proposed instrument in a sample of 1.000 cases.

Clinical interview	Proposed instrument		Total
	+	-	
D	150	50	200
no D	70	730	800
Total	220	780	1,000

According to the above, of the1,000 people of the sample (last-year marijuana users), 200 of them were classified with marijuana use disorder according to the clinical interview, and of these 150 were classified in the same condition in the face-

to-face survey. Hence, the results of indicators defined above are presented in the following, Table 6:

Table 6: Estimates of validation measures under random sampling

Prevalence $P = \frac{200}{1,000} = 0.2\ (20.0\%)$	**Positive Predictive Value=PPV=P(D|+)** $PPV = \frac{150}{220} = 0.682\ (68.2\%)$
Sensitivity=S=P(+|D) $S = \frac{150}{200} = 0.75\ (75.0\%)$	**Negative Predictive Value=NPV=P(-|no D)** $NPV = \frac{730}{780} = 0.936\ (93.6\%)$
Specificity= Sp= P(-|no D) $Sp = \frac{730}{800} = 0.913\ (91.3\%)$	**Correctly Classified Cases** $PCC = \frac{150 + 730}{1,000} = 0.88\ (88.0\%)$

That is, if a certain criterion is used by a proposed interview (as used today in general population surveys), 75% of cases (with disorder according to the clinical interview) are correctly classified according to the proposed survey (Sensitivity), on the other hand, 91.3% of "non-cases" (no disorder), were classified as such in the proposed survey.

However, of those classified as positive in the proposed survey, i.e. 220 people, 150 of them actually had marijuana use disorder according to clinical judgment (PPV). On the other hand, of the 780 cases that were classified as negative in the survey, 730 of them did not have marijuana use disorder according to the gold standard (NPV). Hence the PPVs and NPVs are respectively 68.2% and 93.6%. In addition, 88% of all cases were correctly classified in the survey through the proposed interview.

This strategy (random sampling of last-year marijuana or alcohol users) requires defining a particular population group in which the study will be conducted, for example, adults attending primary health care services, college students, etc. The

downside is determining the initial number of people who need to be interviewed to achieve a sufficient number of people who meet the criteria of ever using marijuana or alcohol in the past year. Suppose a scenario in a country where, for example, the prevalence of last year marijuana use in populations aged 18 to 64 is 10%. That is, if 1,000 people are interviewed in primary health care services, on average 100 of them are expected to have used marijuana in the last year, so they are eligible for clinical interview and survey administration; this number may be insufficient to achieve robust estimates. The above scenario requires having too large logistics and would take too long, with the costs involved. Probably a strategy of this nature could be more appropriate to study indicators in the case of alcohol, where the prevalence of use in the last year is much higher than that of marijuana. In addition, it is assumed that people attending primary health care have a similar behaviour to the general population from which the estimate of 10% comes from, and that in addition all of them will consent to participate in a study of this nature.

2. Given the practical limitations of the previous strategy, we will discuss a **second option**. The fundamental change lies in the search for *cases*, that is, of people with a SUD (being alcohol or marijuana the drugs subject of this work). This involves resorting to some setting where the **population group has a high probability of satisfying the condition sought**, i.e. SUD. It is probable that the most appropriate settings for this are accredited treatment centres in the country.

Following this logic, let´s assume that we select a predetermined number (n_1) of people known to have a SUD (marijuana or alcohol as the case may be), and on the other hand, we select another number of people (n_2) who, having used the substance over the past year, do not meet the criteria for being diagnosed with a SUD.

People from both groups are subjected to the clinical evaluation and the proposed interview and, by some specific predefined criteria (DSM-5, ICD-11 or other) contained in the questionnaire of the interview, are classified into one of two groups: with and without SUD. The following, Table 7, shows this situation:

Table 7: Case classification according to survey of samples with and without substance disorder

Substance disorder	Proposed instrument		Total
	+	-	
Yes (D)	a	b	n_1
No (no D)	c	d	n_2
Total	a+c	b+d	n_1+n_2

As mentioned in the first sampling **strategy,** the equations for **determining** PPVs and NPVs are based on a random sample of people who have used some substance in the last year, marijuana for example, and those calculation formulas are valid only in that specific case. However, if we opted for the strategy we are looking at, these indicators must be determined through a different path (using the Bayes theorem). An important difference from the previous strategy is that in this one, the total number of people with a disorder according to clinical criterion (n 1) and without disorder (n_2) are set *a priori*, (they may even be the same size) so the prevalence of SUD, P, (which in theory would be equal to $n_1/(n_1+n_2)$) cannot be deduced as in the sample model above.

In strategies such as this, the corresponding probabilities (Prob. in the following formulas) PPVs and NPVs are determined by three other indicators: by the prevalence of SUDs in the population of last-year users for the substance being analysed (**P**), by the sensitivity (**S**) and the specificity (**Sp)** of the method used in the face-to-face interview. The expressions are as follows:

$$\boldsymbol{PPV = Prob(D \mid +) = \frac{S * P}{S * P + (1 - Sp) * (1 - P)}} \quad \textbf{(6)}$$

and

$$\boldsymbol{NPV = Prob(no\ D \mid -) = \frac{Sp * (1 - P)}{(1 - S) * P + Sp * (1 - P)}} \quad \textbf{(7)}$$

Where, from the table above, **S** and **Sp** are determined according to the following expressions:

$$S = \frac{a}{n_1} \qquad Sp = \frac{d}{n_2} \tag{8}$$

In the case of prevalence **P,** i.e. **the proportion of people who, having used a substance in the last year, have a disorder for the use of that substance, should be estimated by external sources (e.g. , surveys in the same country or in other surveys with similar characteristics, consultations with experts, or other ways),** since, as previously explained, it cannot be deduced through a methodological strategy of this type. **This is essential; to use a strategy such as the one described, it is necessary to have a reliable estimation of P**.

Example: Estimation of validation in a study with participants in treatment centres.

Let´s suppose a study to analyze the predictive capacity of the DSM-5 to determine the proportion of marijuana users with the condition of use disorder for that substance. From one (or several depending on the country) treatment centre, the first 100 cases admitted and clinically diagnosed with SUD are selected. From another population, for example primary care, we select another 100 people who have used marijuana in the last year, undergo a clinical interview and do not have marijuana use disorder. In this way, the 200 people (assuming everyone consents to it) are interviewed by a pollster as it would have been done in a general population study (i.e. by a layman, not a professional on specific subjects), and are classified in one of two categories: with disorder (either mild, moderate or severe under DSM-5 or harmful pattern of use or dependence under ICD-11) or without marijuana use disorder.

The following, Table 8 presents the results of this simulated study:

Table 8: Case classification according to survey of samples with and without disorder

Substance disorder	Face-to-face		Total
	+	-	
Cases (D)	95	5	100
Controls (no D)	15	85	100
Total	110	90	200

As mentioned above, to determine PPVs and NPVs it is necessary to have an estimate of the prevalence P, that is, of the proportion of people with marijuana use disorder among users in the last year. Suppose that in that country in a recent general survey with face-to-face interviews, 15% of people who reported having used marijuana in the last year were classified as problematic use by combining DSM-IV (abuse) and ICD-10 (dependence), and that is the best information available in the country. Therefore, that value is used as the best estimate of P.

According to the above, the measures of Sensitivity (S), Specificity (Sp), Positive Predictive Value (PPV) and Negative Predictive Value (NPV) are shown in Table 9, respectively:

Table 9: Estimates of validation measures from cases in treatment centres and public health services

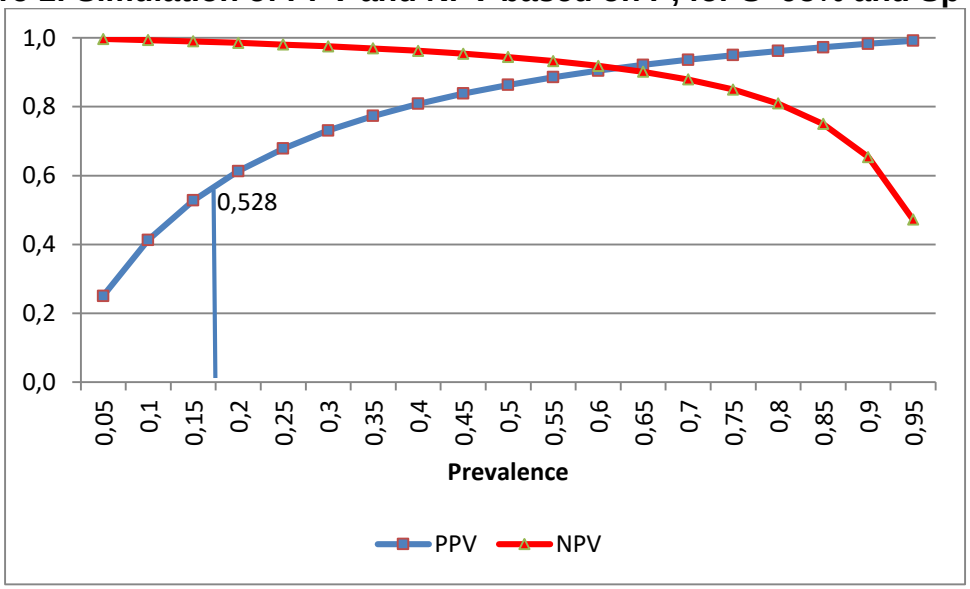
$\boldsymbol{S = P(+\mid D) = \frac{95}{100} = 0.95\ (95\%)}$	**PPV=P(D|+)** $\boldsymbol{PPV = \frac{0.95 * 0.15}{0.95 * 0.15 + (1 - 0.85) * (1 - 0.15)} = 0.528\ (52.8\%)}$
$\boldsymbol{Sp = P(-\mid no\ D) = \frac{85}{100} = 0.85\ (85\%)}$	**NPV=P(no D|-)** $\boldsymbol{NPV = \frac{0.85 + (1 - 0.15)}{(1 - 0.95) * 0.15 + 0.85 * (1 - 0.15)} = 0.99\ (99\%)}$

The PPV estimate shows that out of every 100 cases classified with marijuana use disorder in the face-to-face interview, only 53 of them (52.8%) really present such a condition. In contrast, out of every 100 people classified without disorder in the interview, 99 of them (99%) actually don't have that condition.

It is important to note that even though sensitivity and specificity have fairly acceptable values, 95% and 85% respectively, the positive predictive value is not very desirable, only 52.8%. Basically, this situation is related to the prevalence value of the problem being studied.

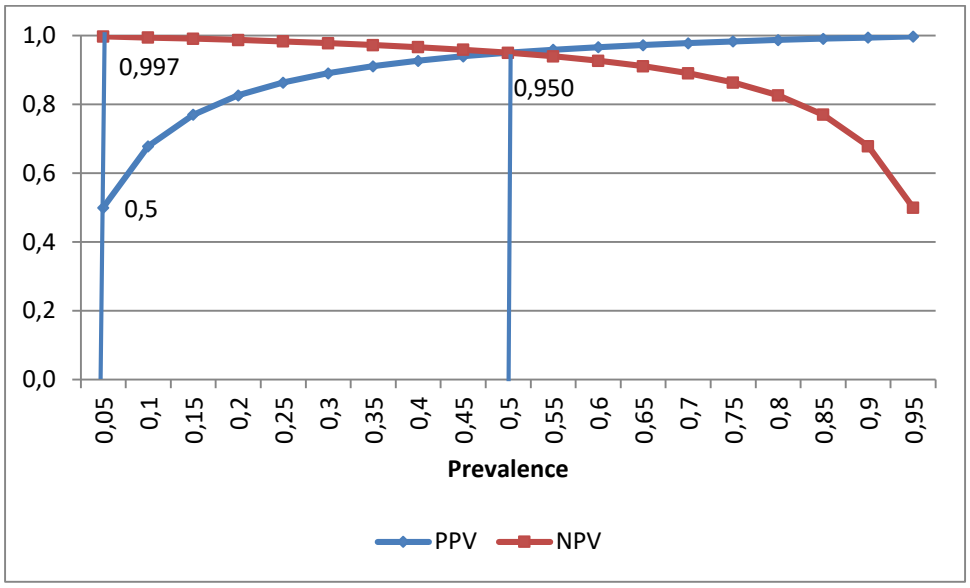
The following graph, using a simulation, presents the positive and negative predictive values that would be obtained based on the P prevalence, for fixed values of S=0.95 and Sp=0.85.

Figure 1: Simulation of PPV and NPV based on P, for S=95% and Sp=85%



Another simulated example is presented in the following graphic Figure 2, now with S=Sp=95%.

Figure 2: PPV and NPV simulation based on P, for S=95% and Sp=95%



Note that even though the values for sensitivity and specificity are quite high, 95% in each case, the PPV is low when the prevalence of the phenomenon studied is low. In fact, in the graph above we observe that if prevalence is 0.05 (5%), then the PPV will be 0.5 (50%), that is, of every 2 cases classified as *with disorder* in the face-to-face interview, only 1 of them would actually have the same result in the clinical interview. In other words, we might be overestimating the amount of cases with SUD through face-to-face interviewing. On the contrary, in this same case (P=5%) the NPV is 0.997 (99.7%), i.e. virtually all cases classified as *without SUD* through face-to-face interviewing would test negative through the clinical interview. On the other hand, if the prevalence was 0.5 (50%), then both the PPV and NPV would be 95%.

To conclude this section, let's go back to a key subject: how to solve ***the sampling issues*** in both strategies?

However, before analyzing the strategies we have mentioned, it is necessary to bear in mind some considerations when a country makes the decision to conduct a pilot study to assess the predictive properties of the criteria discussed in this document:

- What information does the country have regarding *problematic use* of marijuana and/or alcohol?
- When and how was this study conducted? Can its results still be considered up-to-date?

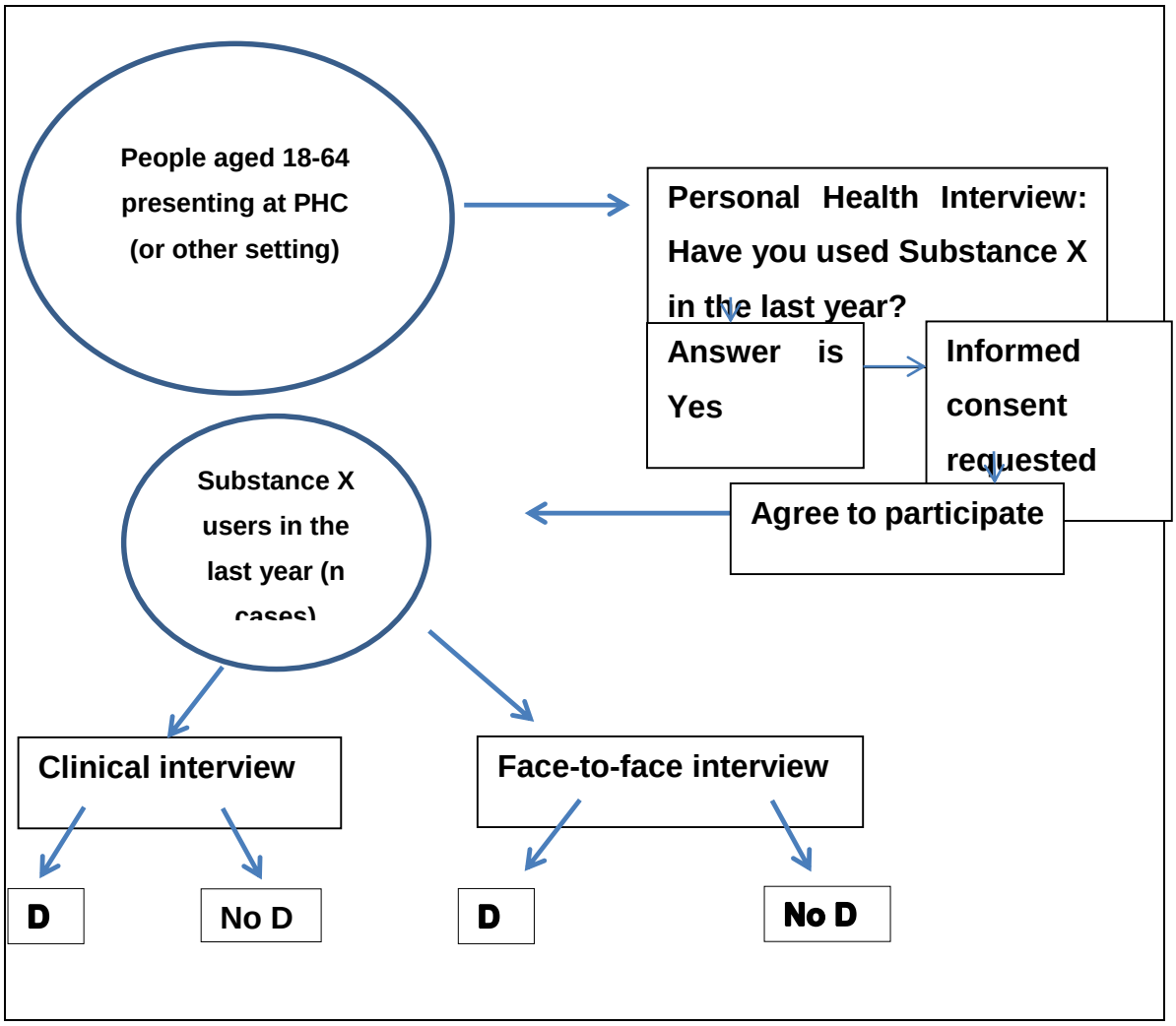
- What were the estimates for last year's rate of use of the substance(s) under study (alcohol and/or marijuana)?
- What were the estimates regarding the percentage of people with problematic use for such substances among people using them in the last year, i.e. the of prevalence estimate(s)?
- Which criteria were used to estimate these prevalence?

The answers to these questions (in addition to those related to human and financial resources) are crucial in deciding whether the country is in a position to conduct such a study, and if the answer is positive, decide on the methodological strategy that best fits its reality. Now let's look at what these strategies consist of:

In the **first case** (i.e. a random sample of last year users) the population group in which the study will be carried out must be defined. An alternative is to resort to people who visit primary health care services (PHC), in the age range used in general population surveys, i.e. 18 to 64 years (although surveys consider people aged 12 to 64 , the 12-17 age group should be treated independently). The scheme is as follows (see Figure 3):

- For a specified and necessary period of time to access a sufficient number of persons, in those who meet the age criterion, they are asked whether or not they have used a particular substance in the last 12 months. This must be done by a person from the health service team.
- In cases with a positive response, they are informed about the study and asked for their consent to participate in both the clinical interview by a trained professional, and the face-to-face survey (simulating what is done in a general population survey). Appropriate conditions must be created for both activities, and since they are performed by different people, a code for the participant must be generated, for example, initials of first and last names plus year of birth. This is for the purpose of combining the information obtained by both procedures and thus generating the table as described above (see Table 8).

Figure 3: Flow chart for scale validation using a study design based on random sampling.



According to the above flow chart, Figure 3, the following table represents the results obtained through this strategy.

Table 10: classification of cases according to clinical interview and survey in a sample of n cases.

Clinical interview	Face-to-face		Total
	+	-	
D	a	b	a+b
no D	c	d	c+d
Total	a+c	b+d	n

The key question now is about n, i.e. the sample size, in this case, the number of people who claim to have used a substance (marijuana or alcohol) during the last year, and who agree to participate in the study. Since we are in a scheme based on simple random sampling (with the variation that the entire population is not available to obtain the corresponding sample, but that people are accessed depending on how they are demanding attention in the system, in such a way that the first n persons who agree to participate constitute the sample for the study, under the assumption that they adequately represent the group of people who demand attention in such service), we can resort to the traditional sample size formula for these cases, i.e.,

$$\boldsymbol{n = \frac{Z^2 * P * (100 - P)}{d^2}} \qquad \textbf{(9)}$$

where **Z** is the value for the normal distribution corresponding to a predetermined confidence level, **P** is the prevalence of the SUD among last year users, **Q** = 100–P, and finally **d** represents the desired error in the estimate of the P prevalence (note that values in the formula are expressed as a percentage).

So, for example, if we want an estimate with a confidence level of 95%, then Z=1.96, and an error not exceeding 2% assuming that the prevalence may be close to 10% (P), then

$$\boldsymbol{n = \frac{1.96^2 * 10 * (100 - 10)}{2^2} = 864}$$

If the objective was to analyze the predictive capacity of instruments to estimate the proportion of people with alcohol use disorder among last-year alcohol users, and assuming that the prevalence of alcohol use in that period is 50%, it will then be needed to consult twice as many people (i.e. 1,728), to detect 864 who have consumed alcohol in the last year, assuming that they all consent or are in a position to participate. However, if instead of alcohol we want to study the classification criteria for marijuana, and if we assume that the prevalence of marijuana use in the last year in the population aged 18 to 64 in that country is 10%, it will be necessary to interview ten times the size 8,640 people. The latter probably requires too large logistics, which would make it very difficult to implement. It is important to note that

we have assumed a prevalence of marijuana use of 10%, which is quite high for the vast majority of countries in the region of the Americas, where that figure may not exceed 5%, which would require starting the process with more than 17,000 people in these countries. In short, in this modality the complex part is to access an appropriate number of people with SUD according to clinical criteria.

The **second strategy,** as already noted, is to have two independent samples:

- A first sample obtained directly from one or more accredited treatment centres in the country, corresponding to *newly admitted cases* (in a given time) and diagnosed with SUD (marijuana or alcohol) by the professionals of the centre(s). All these cases are invited to do the structured clinical interview and also the face-to-face interview by a pollster. This resolves the difficulty mentioned above in terms of having a sufficient number of cases. This provides a first table like the following, Table 11:

Table 11: Classification of cases <u>with</u> and <u>without</u> disorder in a sample of Treatment Centres according to clinical and face-to-face interview

Clinical interview	Face-to-face		Total
	+	-	
D	a_1	b_1	n_{11}
no D	c_1	d_1	n_{12}
Total	a_1+c_1	b_1+d_1	$n_{1.}$

What should we expect from this table? In first place we must remember that the total number of cases studied have been diagnosed with SUD (the one under analysis) by the health team of the respective centre, so the number of cases classified with no disorders in the structured clinical interview (no D) should be marginal, and ideally zero, this is n_{12} equals 0 or close to 0. In this extreme case the table above would look as presented in table 12:

Table 12: Error-free classification of cases <u>with</u> and <u>without</u> disorder in sample Treatment Centres according to clinical and face-to-face interview

Clinical interview	Face-to-face interview		Total
	+	-	
D	a_1	b_1	n_{11}
no D	0	0	0
Total	a_1	b_1	n_{11}

This way what we have called the cases group (n_{11} of the tables above) would be getting set up.

- Once the cases have been defined, we must determine the way to obtain the group we have previously called control, i.e. those who, having used the substance under study over the past year, do not have a SUD according to the structured clinical interview. In this case it is possible to use the scheme presented above, i.e. people visiting primary health care services. Since the group that is needed are those without SUD, it probably won't be as complex to access a predetermined number of them with this strategy (as it was to complete a necessary number of cases).

For example, if we need to access a sample of 200 people who, having used marijuana in the last year, *do not have a disorder for the use of that substance*, it is very feasible that we will have to do a clinical interview with about 220 last-year marijuana users (assuming that about 10% of them may have a disorder and the remaining 90% have no marijuana use disorder). Likewise, a significant number of people should be interviewed (by PHC service officials) regarding the use of the substance in the last year, but much lower than described in the first strategy. This is represented in Table 13:

Table 13: Classification of cases with and without disorder according to clinical and face-to-face interview in PHC patients

Clinical interview	Face-to-face		Total
	+	-	
D	a_2	b_2	n_{21}
no D	c_2	d_2	n_{22}
Total	a_2+c_2	b_2+d_2	$n_{2.}$

where $n_{2.}$ represents the total number of people visiting PHC who report using the substance in the past year, and n_{21} and n_{22} represent people diagnosed with or

without use disorder for that drug. What is expected is that the number of people with a disorder in the clinical interview (D) will be much lower than the number without disorder (no D), i.e. it is expected that the greatest contribution of this strategy will be precisely in the latter group, which is what is missing to supplement cases from treatment centres. Thus, the following table, Table 14, presents both tables where the subgroups with the largest number of people are highlighted in bold to form the ***cases*** and ***controls*** in this strategy, but there could also be crossed inputs, for example, n_{21} are cases obtained from PHC and n_{12} are controls from treatment centres.

Table 14: Classification of cases with and without disorder according to clinical interview and face-to-face survey in sample from Treatment Centres and PHC

Clinical interview	Face-to-face					
	Treatment Centre			PHC		
	+	-	Total	+	-	Total
D	a_1	b_1	n_{11}	a_2	b_2	n_{21}
no D	c_1	d_1	n_{12}	c_2	d_2	n_{22}
Total	a_1+c_1	b_1+d_1	$n_{1.}$	a_2+c_2	b_2+d_2	$n_{2.}$

Overlaying both tables in one, we get the following representation in Table 15:

Table 15: Classification of cases with and without disorder according to clinical interview and face-to-face survey combining samples from treatment centres and PHC visitors

Clinical interview	Face-to-face		Total
	+	-	
Cases (D)	a	b	n_1
Controls (no D)	c	d	n_2
Total	a+c	b+d	n_1+n_2

Based on the information obtained by these two pathways, and properly organized as presented in the table above, it is then possible to determine the PPV and NPV, in addition to the sensitivity S (and specificity Sp) by means of formulas (6), (7) and (8) described previously for each of the proposed criteria (DSM-IV and V, ICD-10 and 11, in addition to AUDIT for alcohol).

6.2. Comparison between instruments

The strategy described in the previous point focuses on comparing different instruments with a clinical diagnosis made by a trained professional, and as mentioned, choosing the criterion that best suits a country's reality is defined by its predictive capacity, i.e. by the criterion that leads to a better combination of PPV and NPV.

In addition to the above, and on the basis of the information obtained, it is also possible to compare the different instruments with each other, by means of the agreements obtained in the classification of SUD; indeed, if we consider only information from the face-to-face survey**,** it will be possible to compare, for example, the criteria based on the DSM, or those based on the ICD.

Let´s take from Table 15 the classification obtained by the face-to-face survey, i.e. with disorder (+) or without disorder (-) by DSM-5 and ICD-11, which can be represented in Table 16 below.

Table 16: Classification of marijuana use disorder by DSM-5 and ICD-11

DSM-5	ICD-11		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	N

From the table above it is possible to determine different indicators, namely:

- Proportion of cases classified as with disorder according to DSM-5 (+) which are also classified as with disorder (+) according to ICD-11, i.e. a/(a+b).
- Proportion of cases classified as with disorder (+) according to ICD-11 which are also classified as with disorder (+) DSM-5, i.e. a/(a+c).
- Proportion of cases classified as without disorder according to DSM-5 (-) which are also classified as without disorder according to ICD-11 (-), i.e. d/(c+d).

- Proportion of cases classified as without disorder according to ICD-11 (-) which are also classified as without disorder according to DSM-5 (-), i.e. d/(b+d).

On the other hand, it is necessary to determine the proportion of agreeing cases (PAC) and a statistical indicator to analyze that proportion. The PAC was defined in formula (5) as

$$\boldsymbol{PAC = \frac{a+d}{n}}$$

and the statistical indicator corresponds to Cohen's Kappa coefficient[6] defined as follows:

$$\boldsymbol{K = \frac{Agreeing\ observations - Agreeing\ observations\ attributed\ to\ chance}{Total\ of\ observations - Agreeing\ observations\ attributed\ to\ chance}}$$

Where according to the table above:

- Agreeing observations = a+d
- Total of observations = n

With regard to the third element of the equation, i.e. the *number of agreeing observations attributed to chance,* the hypothesis that both classification criteria are really independent of each other must be assumed as true. Therefore, for example, from a probabilistic perspective, the likelihood that an individual is classified through both criteria with SUD (+) is equal to the product of the individual probabilities, i.e.,

$$\boldsymbol{\left[\frac{a+b}{n}\right] * \left[\frac{a+c}{n}\right]}$$

[6] Cohen J, A coefficient of agreement for nominal scales. Educ. Psychol. Meas, 20 (37-46), 1960.

so that the number of randomly classified cases presenting a disorder (+) is equal to n times the previous amount, i.e.:

$$\boldsymbol{n} * \left[\frac{\boldsymbol{a}+\boldsymbol{b}}{\boldsymbol{n}}\right] * \left[\frac{\boldsymbol{a}+\boldsymbol{c}}{\boldsymbol{n}}\right] = \frac{(\boldsymbol{a}+\boldsymbol{b}) * (\boldsymbol{a}+\boldsymbol{c})}{\boldsymbol{n}} = \boldsymbol{m_1}$$

The number of randomly consistent cases in terms of case classification in SUD (-) is determined in the same way.

Let's call these cases m_2 where:

$$\boldsymbol{m_2} = \boldsymbol{n} * \left[\frac{\boldsymbol{b}+\boldsymbol{d}}{\boldsymbol{n}}\right] * \left[\frac{\boldsymbol{c}+\boldsymbol{d}}{\boldsymbol{n}}\right] = \frac{(\boldsymbol{b}+\boldsymbol{d}) * (\boldsymbol{c}+\boldsymbol{d})}{\boldsymbol{n}}$$

According to the above, the Kappa coefficient takes the following expression:

$$\boldsymbol{K} = \frac{(\boldsymbol{a}+\boldsymbol{d}) - (\boldsymbol{m_1}+\boldsymbol{m_2})}{\boldsymbol{n} - (\boldsymbol{m_1}+\boldsymbol{m_2})}$$

Example: measuring the quantification of agreement between instruments

Consider the following example where 500 people who have used marijuana in the past year are interviewed using two previous criteria, and their results were as follows:

Table 17: Classification of marijuana use disorder using DSM-5 and ICD-11 in simulated example

DSM-5	ICD-11		Total
	+	-	
+	50	10	60
-	20	420	440
Total	70	430	500

It follows from the above that the proportion of agreeing cases is:

$$\boldsymbol{PAC} = \frac{\boldsymbol{50+420}}{\boldsymbol{500}} = \boldsymbol{0.94}$$

In addition

$$m_1 = \frac{(a+b)*(a+c)}{n} = \frac{60*70}{500} = 8.4$$

$$m_2 = \frac{(b+d)*(c+d)}{n} = \frac{440*430}{500} = 378.4$$

for which the kappa index is:

$$K = \frac{(a+d)-(m_1+m_2)}{n-(m_1+m_2)} = \frac{(50+420)-(8.4+378.4)}{500-(8.4+378.4)} = \frac{83.2}{113.2} = 0.735$$

There are different interpretations of the Kappa value, though not very different from each other. One of these is shown in Table 18 below:

Table 18: Interpretation of the Kappa Index

K Value	Interpretation
Under 0	Match less than simply obtained by chance
0.01-0.20	Poor
0.21-0.40	Slight
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

Similarly, it is possible to compare, for example, dependence using ICD-10 with dependence using ICD-11, or other comparisons. For more details on these comparisons see the article by Lago et al (33).

In the event that a country could not carry out studies based on a clinical gold standard**,** such as the one described in section 6.1, it is possible to perform face-to-face surveys using the same questionnaire in a group like patients, college students, or others. For these purposes, the sample aspects described above should be taken into account.

7. General considerations for studies in adolescent populations

One of the main concerns of countries is the onset of substance use at an early age. This has led to measuring this type of behavior in the school population for early interventions.

As already mentioned in point 3.2, any level of substance use in the adolescent population is considered to be risky to health. Having said that, it is necessary to distinguish between different types of consumption including frequency, intensity, pattern of use and risk behaviors associated with the use of substances. In this context, the countries of the region have been using the various instruments already described (Binge drinking, CAST). The team responsible for this document believes that no further innovations are needed in terms of these instruments, but strongly recommends deepening of the questions and protocols in the indicated direction, that is frequency, intensity, pattern of use, as well as risky behaviours associated with substance use (risky sexual behavior, driving under the effects of alcohol). In the event that the questionnaire in use in a country does not include such questions, these should be incorporated into future studies.

It is important to reiterate the need to move forward in the depth of analyses on the basis of information already available, beyond the description of phenomena, through the generation of hypotheses that aim to better understand the problems associated with substance use.

Annex 1: Questionnaire for National Drug Observatories for Diagnosis

A part of efforts by the **Cooperation Programme beetwen Latin America, the Caribbean and the European Union on Drugs Policies (COPOLAD)** to provide guidance on the validation of instruments used to determine problematic drug use, you are being invited to participate in this survey. We wish to learn how you have used any or all of the following tools in the past. The tools of interest are the following:

- ICD 10 (WHO) - Diagnostic Criteria
- DSM IV (American Psychiatric Association) – Diagnostic Criteria
- AUDIT – Alcohol Use Disorders Identification Test (WHO)
- ASSIST – Alcohol, Smoking and Substance Screening Test(WHO)
- CUPIT – The Cannabis Use Problems Identification Test. (Jan Bashford, Ross Flett, Jan Copeland).
- CIDI – Composite International Diagnostic Interview(WHO)
- CAST – Cannabis Abuse Screening Test. (Austin, Beck and Legleye)
- CRAFFT – Substance Abuse Screening Test among Adolescents (Knight JR, Shrier LA, Bravender TD, Farrell M, Vander Bilt J, Shaffer HJ)
- SDS – Severity of Dependence Scale (Ludwing Kraus)
- PUM – Problem Use of Marijuana (Janusz Sieroslawski)

This questionnaire will gather data on whether or not you have used the tools, the index your study measured when using the tool as well as the target groups for your study. Any work done to validate the tool should also be documented.

Date: _____

Time: _____

Interviewer: _____

Section 1: Background of Organization

1. What is the name of your organization?
   _____.
2. State the country in which your organisation is located _____
3. How would you most correctly classify your organization? (Tick one.)

□ University or Research Centre
□ Health Institution
□ International Agency
□ NGO or Social Organisation
□ Drug Observatory
□ Institute of Statistics and Censuses
□ Private Consulting Agency
□ Professional Organisation
□ Government Organisation
□ Other (please specify) _____
□ Don't know [88]
□ No Response [99]

4. The primary function of the organization is to facilitate (select one)

□Local policy development[1] □Regional policy development[2]
□ Don't know [88] □ No Response [99]

5. In what year did your organisation start operating? _____
6. What type of work is done by your organisation? (Tick all that apply.)

□ Monitoring and surveillance of drug use[1]
□ Treatment for drug abuse[2]
□ Education and drug abuse prevention[3]
□ Control of drug use/abuse (security)[4]
□ Drug use data gathering[5]
□ Funding drug use studies[6]
□ Research using drug awareness interventions[7]
□ Policy development based on drug awareness interventions[8]
□ Other (please specify) _____[9]
□ Don't know [88]
□ No Response [99]

Section 2: Instruments used and Indices measured

7. Place a tick in the spaces provided to indicate whether you have ever used any of the following tools to gather data and write in the space provided the year of last use in a population survey.

Measurement Tool	Ever Used (Yes[1], No[0])	Year last Used
ICD 10 (WHO) - Diagnostic Criteria		
DSM IV (American Psychiatric Association) – Diagnostic Criteria		
AUDIT – Alcohol Use Disorders Identification Test(WHO)		
ASSIST – Alcohol, Smoking and Substance Screening Test(WHO)		
CUPIT – The Cannabis Use Problems Identification Test. (Jan Bashford, Ross Flett, Jan Copeland).		
CIDI – Composite International Diagnostic Interview(WHO)		
CAST – Cannabis Abuse Screening Test. (Austin, Beck and Legleye)		
CRAFFT – Substance Abuse Screening Test among adolescent. (Knight JR, Shrier LA, Bravender TD, Farrell M, Vander Bilt J, Shaffer HJ)		
SDS – Severity of Dependence Scale (Ludwing Kraus)		
PUM – Problem Use of Marijuana (Janusz Sieroslawski)		

(If you have used none of the instruments above please STOP HERE and return the questionnaire.)

8. Please place a tick in the space provided to indicate the types of changes (if any) that were made to the instrument before use. (Mark all that apply for each tool. (Yes[1], No[0]))

Name	Used Original Document	Changed Language of questionnaire	Changed The Questionnaire items	Changed Response options	Made Other Change (Please Specify)
1. ICD 10 (WHO)					
2. DSM IV					
3. AUDIT					
4. ASSIST					
5. CUPIT					
6. CIDI					
7. CAST					
8. CRAFFT					
9. SDS					
10. PUM					

9. Indicators measured in last study:

a. Place a tick in the spaces provided to select the indicators measured in your last study using the instruments listed below. (Mark all that apply for each instrument. (Yes[1], No[0]))

Name Of Tool	Addiction To Cannabis	Harmful Alcohol Use	Alcohol Dependence	Other (Please Specify)
1. ICD 10 (WHO)				
2. DSM IV				
3. AUDIT				
4. ASSIST				
5. CUPIT				
6. CIDI				
7. CAST				
8. CRAFFT				
9. SDS				
10. PUM				

b. Place a tick in the spaces provided to select the indicators measured in your last study through the instruments listed below. (Mark all that apply for each tool. (Yes[1], No[0]))

Name Of Tool	Addiction To Cocaine	Addiction To Tranquilizers	Addiction To Amphetamines	Addition To Analgesics	Other (Please Specify)
1. ICD 10 (WHO)					
2. DSM IV					
3. AUDIT					
4. ASSIST					
5. CUPIT					
6. CIDI					
7. CAST					
8. CRAFFT					
9. SDS					
10. PUM					

10.- Targeted population:

a.- For each tool listed below, place a tick in the space provided to indicate the label that best describes the population that was targeted for the last study that was conducted. (Mark all that apply for each tool. (Yes[1], No[0]))

Name Of Tool	Students In Schools	Adolescents	Young Adults (18-35)	Pregnant Women	General Population	Other (Please Specify)
1. ICD 10 (WHO)						
2. DSM IV						
3. AUDIT						
4. ASSIST						
5. CUPIT						
6. CIDI						
7. CAST						
8. CRAFFT						
9. SDS						
10. PUM						

b.- For each tool listed below, place a tick in the space provided to indicate the label that best describes the at-risk group that was targeted for the last study that was conducted. (Mark all that apply for each tool. (Yes[1], No[0]))

Name Of Tool	Men Who Have Sex With Men	Alcohol Addicts	Prisoners	Treatment Centre Clients	Commercial Sex Workers	Unattached Youth	Other (Please Specify)
1. ICD 10 (WHO)							
2. DSM IV							
3. AUDIT							
4. ASSIST							
5. CUPIT							
6. CIDI							
7. CAST							
8. CRAFFT							
9. SDS							
10. PUM							

11. Please place a tick in the space provided to indicate, for each instrument, the age group targeted in your last survey in which the instrument was used. (Mark all that apply for each tool. (Yes[1], No[0]))

Name of tool	10-15 years	12-65 years	15-19 years	15-74 years	18-24 years	18-35 years	35-74 years	50 years & over	65 years & over
1. ICD 10 (WHO)									
2. DSM IV									
3. AUDIT									
4. ASSIST									
5. CUPIT									
6. CIDI									
7. CAST									
8. CRAFFT									
9. SDS									
10. PUM									

12. For the last study in which you used this tool, please write the number represented by the tool (1. ICD; 2. DSM IV; 3. AUDIT; 4. ASSIST; 5. CUPIT; 6. CIDI; 7. CAST; 8. CRAFFT; 9. SDS; 10. PUM), and place a tick in the appropriate box to indicate the types of prevalence estimates obtained. (Mark all that apply for each tool. (Yes[1], No[0]))

Outcome measure	Tool #	Age-specific estimates	Sex-specific estimates	Estimates within other demographic categories	Within other socioeconomic categories	Other (Please specify)
a. Addiction to cannabis, harmful alcohol use, addiction to prescription drugs, addiction to cocaine/illegal drugs						
b. Alcohol dependence						
c. Alcohol use						
d. Amount/duration of drug and alcohol (D&A) use						
e. Craving D& A						
f. Frequency of substance use score						
g. At-risk drinking						
h. Lifetime substance use						
i. Prevalence of alcohol dependence						
j. Psychological dependence						
k. Severity of dependence						
l. Substance abuse						
m. Substance dependence						
n. Substance involvement score						
o. Substance risk score						
p. Substance use treatment						

13. For each instrument that was used in your last study, name the form of problematic drug use that was measured and indicate the cut-off point used to define problematic drug use.

Name of tool	Form of Problematic Drug Use	Cut-point used to define problematic drug use
1. ICD 10 (WHO)		
2. DSM IV		
3. AUDIT		
4. ASSIST		
5. CUPIT		
6. CIDI		
7. CAST		
8. CRAFFT		
9. SDS		
10. PUM		

14. Have you ever done a study that gathered data on incidence of problematic drug use?

□Yes[1] □No[0]

(If response to item 14 is No, please go to item 16.)

15. Please tick the outcomes for which incidence was measured. (Tick all that apply)

(1) Addiction to cannabis [1]
(2) Harmful alcohol use [2]
(3) Alcohol dependence [3]
(4) Addiction to prescription drugs[4]
(5) Addiction to cocaine[4]
(6) Addiction to other illegal drugs [5]
(7) Other (please specify)_____[6]

16. Did you measure the quality of the data you gathered for any of your studies?

□Yes[1] □No[0]

(If No, skip to item 19).

17. Place a tick beside the questionnaire name if you assessed the quality of the data gathered using this tool. (Tick all that apply)

(1) ICD 10 (WHO)	(5) CIDI	(9) CUPIT
(2) DSM IV	(6) CAST	(10) PUM
(3) AUDIT	(7) CRAFFT	
(4) ASSIST	(8) SDS	

18. State the method(s) used to assess data quality.

Name of tool	Method used to assess data quality
1. ICD 10 (WHO)	
2. DSM IV	
3. AUDIT	
4. ASSIST	
5. CUPIT	
6. CIDI	
7. CAST	
8. CRAFFT	
9. SDS	
10. PUM	

19. On a scale of 1 to 5, for each tool, indicate, with a tick in the space provided, how you would rank the quality of the data that was gathered using the tool.

Name of tool	Very Poor (1)	Below Average (2)	Average (3)	Above Average (4)	Excellent (5)
1. ICD 10 (WHO)					
2. DSM IV					
3. AUDIT					
4. ASSIST					
5. CUPIT					
6. CIDI					
7. CAST					
8. CRAFFT					
9. SDS					
10. PUM					

20. Were any steps taken to determine the validity of the data gathered using any of the tools used for your most recent study?

□Yes[1] □No[0]

(If No, please STOP HERE and return the questionnaire).

21. In the space provided please write a number (or numbers) to indicate the method(s) used to assess the validity of the data gathered using the tool used in your most recent study.

1) Clinical observations
2) Expert's independent clinical evaluation
3) Medical reports
4) Collateral informal reports
5) Follow up assessments
6) Biological indicators
7) Standard diagnostic test
8) Gold standard assessment
9) Questionnaires
10) Drug diaries
11) Other _____(please specify)

Name of tool	Method used to assess validity of data gathered
1. ICD 10 (WHO)	
2. DSM IV	
3. AUDIT	
4. ASSIST	
5. CUPIT	
6. CIDI	
7. CAST	
8. CRAFFT	
9. SDS	
10. PUM	

Thank you for your participation!

Annex 2

QUESTIONNAIRE AND DIAGNOSTIC CATEGORIZATION

MINIMUM DEMOGRAPHY

1. Sex	2. How old are you?
Man 1 Woman 2	|____|____| Years

OTHER QUESTIONS.....

ALCOHOL QUESTIONNAIRE

Have you ever used alcoholic drinks in your life?	How old were you when you first used alcoholic drinks?
Yes 1 Continue No 0 Exits alcohol module	|____|____| Years
When was the first time you used alcoholic drinks?	Have you used alcoholic drinks in the last 12 months?
In the last 30 days 1 More than 30 days ago but less than 12 months ago 2 More than 12 months ago 3	Yes 1 Continue No 0 Go to QXX
Have you used alcoholic drinks in the last 30 days?	And how many days have you used alcohol in the last 30 days?
Yes 1 Continue No 0 Go to QXX Does not answer 9 Go to QXX	|____|____| Days
How many days have you been drunk over the last 30 days?	

|____|____| Days
If use in the last month is reported **In the last month, how many days have you had 5/4 drinks or more for men/women on the same occasion (occasion refers to a period of approximately 2 hours)? NOTE FOR THE QUESTIONNAIRE AND THE INTERVIEWER: IF THE INTERVIEWEE IS MALE, THE QUESTION REFERS TO 5 DRINKS OR MORE; IF IT IS A WOMAN IT REFERS TO 4 DRINKS OR MORE.** ____ **No days (0 – 30)**

In the last 30 days What type of alcoholic drink did you use and how often? **CARD**

Pollster: read each drink	Daily	Weekends	Some days of the week	Never
1) Beer	1	2	3	4
2) Brandy	1	2	3	4
3) Rhum	1	2	3	4
4) Whiskey	1	2	3	4
5) Brandy, Cognac, Vodka, Gin	1	2	3	4
6) Wine	1	2	3	4
7) Industrial or medicinal alcohol mixed with powdered drink or soda (chamber, Chamberlain)	1	2	3	4
8) Other, which? _____	1	2	3	4

AUDIT QUESTIONS TO EVALUATE: (ENC: IF USED ALCHOL IN THE LAST 12 MONTHS ACCORDING TO QXX ASK QUESTIONS FROM QXX TO QXX, IN CASE CONTRARY PASS ON TO QXX)

Deliver Card XX (Equivalence Table)

1 drink	**One bottle or can of beer (333 cc)** **One glass of wine (140 cc)** **A drink of liquor such as brandy, rum, whiskey, tequila, vodka etc. (40 cc)**
1 and a half drinks	**Half a litre of beer (500 cc)**
3 drinks	**A litre of beer**
6 drinks	**One bottle of wine (750 cc)**
8 drinks	**One bottle of wine (1 liter)**
18 drinks	**A bottle or bottle of liquor such as brandy, rum, whiskey, tequila, vodka etc.**

Questions	**Score**				
	0	**1**	**2**	**3**	**4**
1. How often do you have a drink containing alcohol?	Never	Monthly or less	2 to 4 times a month	2 to 3 times a week	4 or more times a week
2. How many drinks containing alcohol do you have on a typical day when you are drinking? USE CARD TO ESTIMATE NUMBER OF DRINKS	1-2	3-4	5-6	7-9	10 or more
3. How often do you have 6 or more drinks on one occasion?	Never	Less than 1 time a month	Monthly	Weekly	Daily or almost daily
4. How often during **the last year** have you found that you were not able to stop drinking once you had started?	Never	Less than 1 time a month	Monthly	Weekly	Daily or almost daily
5. How often during the **last year** have you failed to do what was normally expected from you because of drinking?	Never	Less than 1 time a month	Monthly	Weekly	Daily or almost daily
6. How often during the **last year** have you needed a first drink in the morning to get yourself going after a heavy drinking session?	Never	Less than 1 time a month	Monthly	Weekly	Daily or almost daily
7. How often during the **last year** have you had a feeling of guilt or remorse after drinking?	Never	Less than 1 time a month	Monthly	Weekly	Daily or almost daily
8. How often during the **last year** have you been unable to remember what happened the night before because you had been drinking?	Never	Less than 1 time a month	Monthly	Weekly	Daily or almost daily
9. Have you or someone else been injured as a result of your drinking?	No		Yes, but not in the last year		Yes, in the last year
10. Has a relative or friend or a doctor or another health worker been concerned about your drinking or suggested you cut down?	No		Yes, but not in the last year		Yes, in the last year

DSM-IV, DSM-5, ICD-10 and 11 EVALUATION QUESTIONS

If used alcohol in the last year:

Questions, (Thinking in the last 12 months)	Yes	No
1. Did you ever have times when you took a drink to keep from having problems like these? *Anxiety/restlessness/irritability; stress/depression; nausea/vomiting; concentration problems; tremors; see, hear or feel non-existent things; fatigue/drowsiness/weakness; tachycardia; insomnia*	1	0
2. Did you ever have times when you stopped, cut down, or went without drinking and then experienced withdrawal symptoms?	1	0
3. Did you ever need to drink a larger amount of alcohol to get the same effect you used to get?	1	0
4. Did you ever find that you could no longer get high on the amount you used to drink?	1	0
5. Did you ever have a time when you gave up or greatly reduced important activities because of your drinking – like sports, work, or seeing friends and family?	1	0
6. Did you ever have several days or more when you spent so much time drinking or recovering from the effects of alcohol that you had little time for anything else?	1	0
7. Did you ever continue to drink when you knew you had a serious physical or emotional problem that might have been caused by or made worse by drinking?	1	0
8. Was there ever a time when your drinking frequently interfered with your work or responsibilities at school, on a job, or at home?	1	0
9. Were there times when you were often under the influence of alcohol in situations where you could have gotten hurt, for example when riding a bicycle, driving, operating a machine, or anything else?	1	0
10. Did you ever get into physical fights while using alcohol or right after using?	1	0
11. Were you arrested or stopped by the police more than once because of driving under the influence of alcohol or because of your behavior while you were under the influence of alcohol?	1	0
12. Were there times when you tried to stop or cut down on your drinking and found that you were not able to do so?	1	0
13. Were there ever times when you used drank more frequently or for more days in a row than you intended?	1	0
14. Was there ever a time when you often had such a strong desire to drink that you couldn't stop using or found it difficult to think of anything else?	1	0
15. Was there ever a time when your drinking caused arguments or other serious or repeated problems with your family, friends, neighbors, or co-workers?	1	0
16. Was there ever a time when you could say that your drinking has negatively affected other people?	1	0

MINIMUM QUESTIONS FOR MARIJUANA

Have you ever used Marijuana?

Yes	1	Continue
No	0	Exits marijuana module

How old were you when you first tried marijuana?

_____ years

When was the first time you tried Marijuana?

Over the last 30 days	1
More than 30 days and less than 12 months ago	2
More than 12 months ago	3
NS/NC	9

Have you used Marijuana in the last 12 months?

Yes	1	Continue
No	0	Skip
Does not answer	9	

Think about the last 12 months. How often have you used Marijuana?

One time only	1
A few times over the last 12 months	2
A few times monthly	3
A few times weekly	4
Daily	5
Does not answer	9

Have you used Marijuana in the last 30 days?

Yes	1
No	0
Does not answer	9

DSM-IV, DSM-5, ICD-10 and 11 QUESTIONS TO EVALUATE

If used marijuana in the last year:

Questions (thinking in the last 12 months)	Yes	No
1. Did you ever have times when you use marijuana to keep from having problems like these? *Anxiety/restlessness/irritability; stress/depression; nausea/vomiting; concentration problems; tremors; see, hear or feel non-existent things; fatigue/drowsiness/weakness; tachycardia; insomnia*	1	0
2. Did you ever have times when you stopped, cut down, or went without using marijuana and then experienced withdrawal symptoms?	1	0
3. Did you ever need larger amount of marijuana to get the same effect you used to get?	1	0
4. Did you ever find that you could no longer get high on the amount you used to use?	1	0
5. Did you ever have a time when you gave up or greatly reduced important activities because of your marijuana use – like sports, work, or seeing friends and family?	1	0
6. Did you ever have several days or more when you spent so much time using or recovering from the effects of marijuana that you had little time for anything else?	1	0
7. Did you ever continue to use marijuana when you knew you had a serious physical or emotional problem that might have been caused by or made worse by using?	1	0
8. Was there ever a time when your use of marijuana frequently interfered with your work or responsibilities at school, on a job, or at home?	1	0
9. Were there times when you were often under the influence of marijuana in situations where you could have gotten hurt, for example when riding a bicycle, driving, operating a machine, or anything else?	1	0
10. Did you ever get into physical fights while using marijuana or right after using?	1	0
11. Were you arrested or stopped by the police more than once because of driving under the influence of marijuana or because of your behavior while you were under the influence of marijuana?	1	0
12. Were there times when you tried to stop or cut down on your use marijuana and found that you were not able to do so?	1	0
13. Were there ever times when you used marijuana more frequently or for more days in a row than you intended?	1	0
14. Was there ever a time when you often had such a strong desire to use marijuana that you couldn't stop using or found it difficult to think of anything else?	1	0
15. Was there ever a time when your use of marijuana caused arguments or other serious or repeated problems with your family, friends, neighbors, or co-workers?	1	0
16. Was there ever a time when you could say that your using marijuana has negatively affected other people?	1	0

CATEGORIZATION FOR ALCOHOL USE

1.- BINGE DRINKING.

Binge drinking is measured by a single question that asks about the number of days in which the subject has consumed 5 drinks or more in men, or 4 drinks or more in women, on the same occasion:

Binge drinking is indicated if the subject's response is 1 or more days in the past month.

2.- DSM-IV:

a) Abuse.

In the DSM-IV, questions 8, 9, 10, 11 and 15 evaluate **abuse**. Questions 8, 11 and 15 correspond to specific criteria: each criteria is assigned a value as either 0 (absence of symptom) or 1 (presence of symptom). Questions 9 and 10 correspond to the same criterion, so if at least one symptom is evaluated as present, then the value 1 is assigned.

Four criteria are evaluated in total, each with assigned values of 0 or 1, so the total possible sum of the four criteria has a minimum value of 0 and a maximum value of 4.

Abuse is indicated when sum is equal to or greater than 1.

b) DEPENDENCE (usually not used in national surveys).

Questions 1 through 7, 12 and 13 are measures for dependence. These nine questions correspond to seven criteria. Each question is evaluated as either 0 or 1 indicating absence or presence of the corresponding symptom.

- Positive in question 1 or 2, implies symptom is present, and takes value 1
- Positive in question 3 or 4, implies symptom is present, and takes value 1

However, questions 5, 6, 7, 12 and 13 account for specific criteria, respectively.

Presence and absence of symptoms are assigned a value of 0 or 1. Therefore, the possible sum of these seven criteria has a minimum of 0 and a maximum of 7.

Dependence is indicated when the sum is equal to or greater than 3.

3.- DSM-5: ALCOHOL USE DISORDER

Questions 1 to 10 and 12 to 15 correspond to alcohol use disorder. These 14 questions correspond to 11 criteria: for questions 1 or 2, 3 or 4, 5, 6, 7, 8, 9 or 10, 12, 13, 14 and 15 positive symptoms are assigned a value of 1 and negative symptoms are assigned a value of 0.

The total sum of the 11 criteria will fall between 0 and 11, with the following classification.

Sum	Classification	Use disorder
0-1	Absence	No
2-3	Mild	Yes
4-5	Moderate	
6 and up	Severe	

4.- ICD-10: DEPENDENCE.

In the ICD-10, questions 1 through 7, and 12 to 14 assess dependence. These 10 questions correspond to six criteria. Each question is evaluated as 0 and 1 indicating to absence or presence of the corresponding symptom.

- Positive in question 1 or 2, implies positive criterion, and takes value 1
- Positive in question 3 or 4, implies positive criterion, and takes value 1
- Positive in question 5 or 6, implies positive criterion, and takes value 1
- Positive in question 12 or 13, implies positive criterion, and takes value 1

Questions 7 and 8 account for specific criteria.

The total sum of these six criteria has a minimum of 0 and a maximum of 6.

Dependence is indicated when the sum is equal to or greater than 3

5.- <u>ICD-11</u>:

a) Dependence.

In the ICD-11, questions 1 to 7 and 13 and 14 assess dependence. These nine questions account for three domains, which in turn meet seven criteria (criteria 1 through 5, 10, and 11).

- Positive in question 1, 2, 3 or 4 implies positive domain, and takes value 1
- Positive in question 5, 6 or 7, implies positive domain, and takes value 1
- Positive in question 13 or 14, implies positive domain, and takes value 1

The sum of these three domains has a minimum of 0 and a maximum of 3.

Dependence is indicated when the sum is equal to or greater than 2

b) HARMFUL USE.

Questions 8 to 11, 15 and 16 are considered measures for harmful use (it is the only case where question 16 about harm to third parties is used). These six questions correspond to 4 domains, which in turn respond to five criteria (criteria 6 to 8, 12 and 13 defined in Table 3).

- Positive in question 9 or 10 implies positive domain, and takes value 1
- Positive in question 15 or 16 implies positive domain, and takes value 1

Questions 8 and 11 account for specific criteria.

The sum of these 4 domains has a minimum value of 0 and a maximum of 4.

Harmful use is indicated when the sum is equal to or greater than 1

6.- AUDIT

AUDIT questions 1 through 8 consider 5 answer alternatives, and each is evaluated on a scale from 0 to 4 points. In contrast, questions 9 and 10 have three answer options with values 0, 2, and 4 respectively.

The final sum of the scores of the 10 questions ranges from 0 to 40 points.

There are different classifications associated with AUDIT, depending on the purpose of its application.

Risky use or hazardous use if the score is equal to or greater than 8. However, sensitivity is increased if the cut-off point in women is reduced to 1 point, leaving it at 7.
High-risk or harmful level can be determined if answers of 1 or more are obtained in questions 2 or 3.
Presence or start of dependence can be determined if there is a score of 1 or more in questions 4, 5 or 6, especially if the answers are 3 or 4, that is, with daily or weekly symptoms.

On the other hand, for the purposes of interventions based on risk levels, 4 zones have been defined according to the total audit score.

Level of risk	AUDIT score	Intervention
Zone I	0-7	Alcohol education
Zone II	8-15	Brief intervention
Zone III	16-19	Brief intervention and further monitoring
Zone IV	20-40	Referral to specialist for diagnostic evaluation and treatment

CATEGORIZATION FOR MARIJUANA USE

1.- DSM-IV:

a) ABUSE.

Questions 8, 9, 10, 11 and 15 assess abuse. Questions 8, 11 and 15 correspond to respective criteria and each of them is evaluated as 0 (absence of symptom) or 1 (presence of symptom). Questions 9 and 10 correspond to the same criterion, so if at least one of them corresponds to presence the value 1 is assigned.

In total, four criteria are evaluated, each with values 0 or 1, so that the sum of the four criteria has a minimum value of 0 and a maximum value of 4.

Abuse is indicated when the sum is equal to or greater than 1

b) DEPENDENCE

Questions 3 to 7, 12 and 13 assess dependence. These 7 questions correspond to 6 criteria (as opposed to alcohol which also included the withdrawal criterion). Each question is evaluated as 0 and 1 according to absence or presence of the corresponding symptom.

Dependence through DSM-IV is usually not assessed in national surveys.

- Positive on question 3 or 4, implies positive criterion, and takes the value 1

On the other hand, questions 5, 6, 7, 12 and 13 account for specific criteria.

The sum of these 6 criteria has a minimum of 0 and a maximum of 6.

Dependence is indicated when the sum is equal to or greater than 3

3.- DSM-5: MARIJUANA USE DISORDER.

All questions, except questions 11 and 16, are used to assess marijuana use disorder. The 14 questions respond to 11 criteria:

Questions 1 or 2, 3 or 4, 5, 6, 7, 8, 9 or 10, 12, 13, 14 and 15. Again, for example if questions 1 or 2 are positive, then the criterion is met and value 1 is assigned.

The sum of the 11 criteria will result in a value between 0 and 11, with the following classification.

Sum	Classification	Use disorder
0-1	Absence	No
2-3	Mild	Yes
4-5	Moderate	
6 and up	Severe	

4.- ICD-10: DEPENDENCE.

In this case, consider questions 1 through 7, in addition to 12 to 14. These 10 questions correspond to 6 criteria. Each question is evaluated in 0 and 1 according to absence or presence of the corresponding symptom.

- Positive in question 1 or 2, implies positive criterion, and takes value 1
- Positive in question 3 or 4, implies positive criterion, and takes value 1
- Positive in question 5 or 6, implies positive criterion, and takes value 1
- Positive in question 12 or 13,implies positive criterion, and takes value 1

Questions 7 and 8 account for respective criteria.

The sum of these six criteria has a minimum of 0 and a maximum of 6.

Dependence is indicated when the sum is equal to or greater than 3

5.- ICD-11:

a) Dependence.

Questions 1 to 7, 13 and 14 assess dependence. These 9 questions correspond to three domains, which in turn respond to seven criteria (criteria 1 through 5, 10, and 11 defined in Table 3).

- Positive in question 1, 2, 3 or 4 implies positive domain, and takes value 1
- Positive in question 5, 6 or 7, implies positive domain, and takes value 1
- Positive in question 13 or 14,implies positive domain, and takes value 1

The sum of these three domains has a minimum value of 0 and a maximum of 3.

Dependence is indicated when the sum is equal to or greater than 2

b) HARMFUL USE.

Questions 8 to 11, 15 and 16 are used to *assess harmful use* (it is the only case where the question 16 about harm to third parties is used). These six questions correspond to four domains, which in turn respond to five criteria (criteria 6 to 8, 12 and 13 in Table 3).

- Positive in question 9 or 10 implies positive domain, and takes value 1
- Positive in question 15 or 16 involves positive domain, and takes value 1

Questions 8 and 11 account for specific criteria.

The sum of these four domains has a minimum value of 0 and a maximum of 4.

Harmful Use is indicate when the sum is equal to or greater than 1

Annex 3

Validation of scales: Translation and cultural adaptation of the scale

If you plan to apply one of these instruments in a country different from where it was created, we strongly suggest performing a careful translation and cultural adaptation first. Translation alone is not enough, as there are differences at the linguistic and cultural level that can create confusion, and cause errors in measurement. Scales should be adapted to the language of interest after translation, retro-translation, cognitive interview with users and discussion with expert groups. To do this, we suggest the following steps:

a. **Stage I** - Direct translation into the language of interest: The first stage in adapting a scale is translation. Many recommend that at least two translations of the instrument be performed from the original language to the target language. In this way, translations can be compared and discrepancies can be observed that may reflect more ambiguous words in the original or discrepancies in the translation process. Bilingual translators whose mother tongue is the target language produce the two independent translations. Each translator produces a written report of the translation that they complete. Additional comments are made to highlight challenging phrases or uncertainties. Their reasons for their choices are also summarized in the written report. The contents of the instrument, the questions and response options, and the instructions should be translated. The two translators must have different profiles. One of the translators should be aware of the concepts discussed in the questionnaire being translated. The other translator should not be aware or informed of the concepts being quantified and, preferably, should not have a medical or clinical history. This is called a naive translator, and it is more likely that he or she will detect a different meaning from the original than the first translator.

b. **Stage II** - Synthesis of translations and solution of discrepancies: The two translators and a recording observer meet to synthesize the results of the translations. Working from the original questionnaire as well as the versions

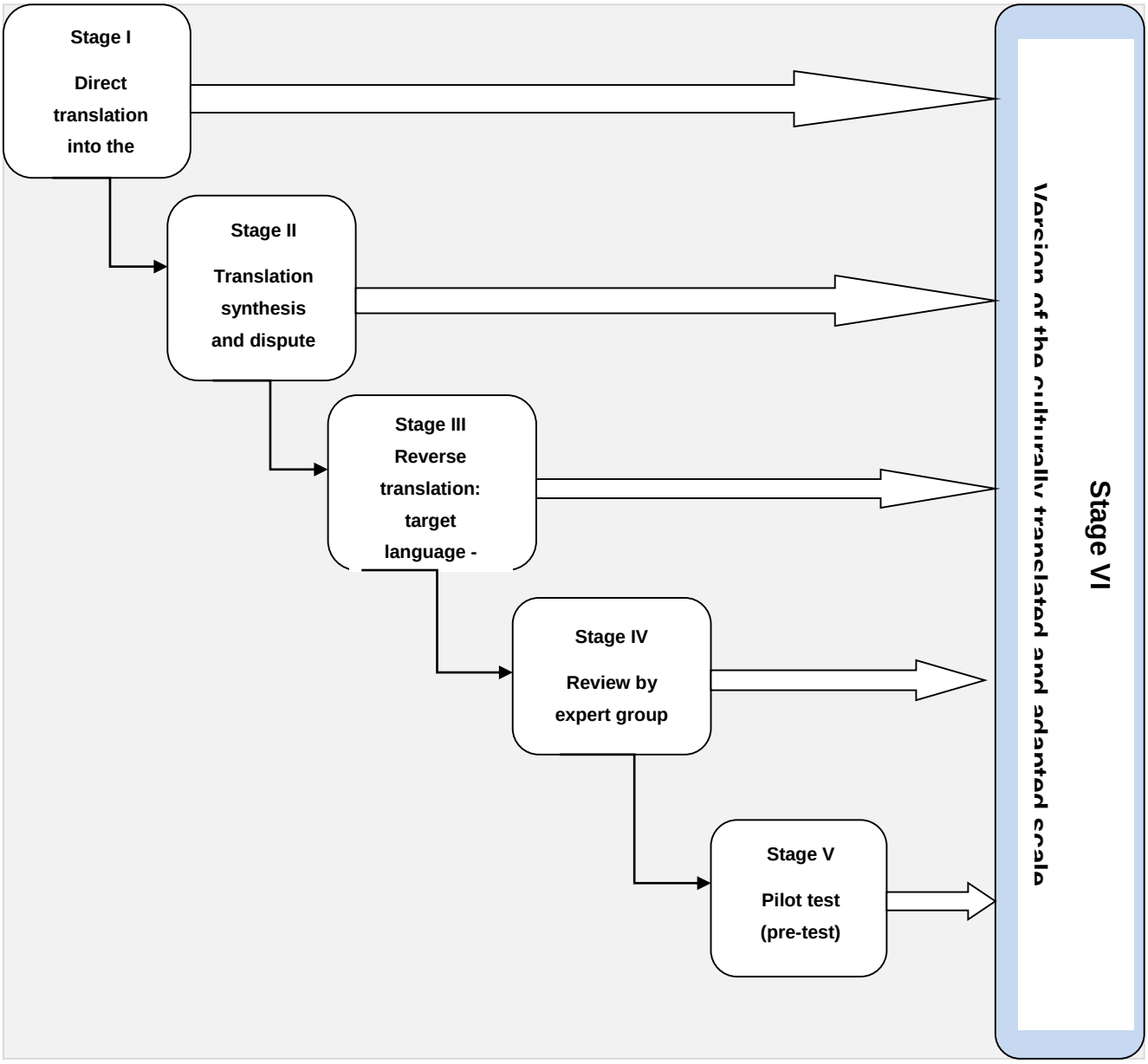
by the first translator and the second translator, the synthesis of these translations is first performed (producing a common translation), with a written report that carefully documents the synthesis process, each of the topics covered, and how they were resolved. Consensus is important, rather than one person solving problems. The next stage is completed with this version of the questionnaire.

c. **Stage III** - Reverse translation: target language - original language: Working from the final version of the questionnaire and completely blind to the original version, the translator then translates the questionnaire back into the original language. This is a validation process to ensure that the translated version reflects the same content of the article as the original versions. This step often increases unclear wording in translations. However, the agreement between the reverse translation and the original source version does not guarantee a successful advanced translation; it simply ensures a consistent translation. Reverse translation is only a type of validity check, highlighting serious inconsistencies or conceptual errors in the translation. Again, two of these retro-translations are considered a minimum. Reverse translation is produced by two people with the source language as their mother tongue. The two translators should not know or be aware of the concepts explored and should preferably have no medical background. The main reasons are to avoid the bias of the information and obtain unexpected meanings of the items in the translated questionnaire, thus increasing the probability of highlighting imperfections.

d. **Stage IV** - Review by Expert Group: The composition of this group is crucial for achieving intercultural equivalence. The minimum composition includes the methodologists, health professionals, language professionals and translators (initial translators and re-translators) involved in the process so far. The original developers of the questionnaire are in close contact with the expert group during this part of the process. The role of the expert group is to consolidate all versions of the questionnaire and develop what would be

considered the pre-final version of the questionnaire for field testing. Therefore, the committee will review all translations and reach consensus on any discrepancies. The material available to the committee includes the original questionnaire and each translation together with the corresponding written reports (which explain the basis of each decision at previous stages). The expert group is making critical decisions; hence, again, full written documentation of the problems and the justification for making a decision must be submitted.

Figure 3: Translation and cultural adaptation of a measurement scale (adapted from Beaton et al)



e. **Stage V** - Pilot Test: The final stage of the adaptation process is the pilot test. This field test of the new questionnaire tries to use the pre-final version in subjects or patients from the target configuration. An appropriate sample of consumers will be selected from a treatment centre, depending of the target population of the study (adolescents, general population). Ideally, between 30

and 40 people should be examined. Each subject completes the questionnaire and interviews each other to prove what they thought each item in the questionnaire and their selected answer meant. Both the meaning of the items and the answers would be explored. This ensures that the adapted version still maintains its equivalence in an applied situation. Response distribution is considered to look for a high proportion of missing elements or individual responses. Apparent and content validity will be assessed through a panel of experts in the treatment of substance use problems that will assess whether, in the context of treatment centres, questions from the original questionnaire appear as relevant, sufficient, reasonable, clear and unambiguous. The analysis of results shall include the assessment of the reliability of the instrument, for which its internal consistency (Cronbach's coefficient) and its test-retest (intraclass correlation coefficient) reliability will be evaluated.

f. **Stage VI** - The final stage of the adaptation process is the submission of all reports and forms to the instrument developer or committee. In turn, they have a means to verify that the recommended steps have been followed, and that the reports seem to reflect this process well. Indeed, it is an audit process, with all steps followed and the necessary reports followed. It is not for this body or committee to modify the content, a reasonable translation is assumed.

Annex 4

GLOSSARY OF TERMS

Appearance validity (logic)	It refers to the degree to which items (questions) on a scale, measure in an apparent or logical way the construct to be measured. To evaluate this property, two groups must be formed, one of experts and the other of subjects that will be measured with the instrument. They both analyze the scale and decide if the questions really seem to measure what is meant. It should be clarified that appearance validity of is not a statistical concept, but depends on the judgement made by the experts on the desirability of the items to evaluate the construct of interest. Moreover, the relevance of this form of validity lies in applicability and, above all, acceptability from the point of view of that who responds and is evaluated with the scale.
Concurrent validity	Concurrent validity seeks to establish the degree of correlation between the results obtained by the evaluation scale and the *criteria* or standard, when both are applied simultaneously. This comparison is performed statistically by Pearson or Spearman correlation coefficients, depending on the characteristics of the data distribution**,** or the type of variable analyzed.
Construct validity	Construct validity ensures that scores resulting from the responses of the instrument can be considered and used as a valid measurement of the phenomenon under study. Thus, this property evaluates the degree to which the instrument adequately reflects the underlying theory of the phenomenon or construct to be measured and consequently, the measure matches that of other instruments that evaluate the same condition. The evaluation of these attributes or constructs requires prior definition of the content of the instrument being validated and the development of a theoretical-conceptual framework that allows the interpretation of the results obtained. Thus, the validity of the construct allows to establish how a measurement of the entity relates in a manner consistent with the hypotheses that are raised to explain the theoretical construct that defines the phenomenon of interest.
Content validity	This property seeks to evaluate whether the different items included in the instrument adequately represent the domains of the construct that is intended to be measured. Content validity is a process in which the

	structure of the scale is determined by ensuring that the scale, through its items, covers all the domains of the entity to be measured, that is, to confirm that the phenomenon is represented appropriately and entirely by its items and domains without leaving any aspect outside of the measurement which means that it encompasses the actual spectrum of the entity, so that inferences arising from the score of the scale are valid within a wide range of circumstances. The procedure for assessing the validity of content involves applying statistical methods such as exploratory factor analysis, this is used to obtain evidence of the underlying dimensions (components) that are present in the instrument and that should correspond, in theory, to the construct to be measured. This seeks to explain the correlations between instrument items from a smaller set of components called domains or factors; in this analysis it is decisive to evaluate the adjustment of the factorial model and the adequacy of the sample and the items evaluated, for which the Barlet and Kaiser-Meyer-Olkin (KMO) sphericity test are used; the latter is considered satisfactory for values greater than 0.7; in addition to rotations, mainly the orthogonal varimax. Globally, the factorial loads or saturations of items (correlation between each item and each factor) are considered optimal if they are equal to or greater than 0.3.
Convergent/divergent validity	This property correlates scores obtained across different scales. If we compare instruments that quantify the same construct and the results between the two measures have significant correlations, they are said to converge, which proves that the scales are conceptually congruent or similar. If, on the other hand, the scale scores that measure with different constructs are compared and obtain low or negative correlations, it means that the scales diverge, indicating non-significant association between the variables, confirming that they measure different constructs; in other words, it would mean that the scale being validated is not specific enough to measure the construct of interest in a given population.
Criterion validity	It sets the degree to which scores obtained from a scale are valid, when compared to a standard or reference pattern (criterion). In this case, the new instrument being evaluated should be compared to an existing scale that

	is widely accepted and has proven to be the best available instrument for measuring the phenomenon of interest. This way the scores obtained with each of the scales are compared in order to assess if there is an adequate correlation between the two.
Cronbach's Alpha Coefficient	It is a coefficient used to measure the reliability of a measurement scale, coined Alpha by Cronbach in 1951, although its origins are found in the works of Hoyt (1941) and Guttman (1945) This method allows to measure the internal consistency of an instrument. It is used in the construction of scales where there are no right or wrong answers, but each interviewee answers the alternative that best represents their way of thinking about the object being asked. A researcher tries to measure a non-directly observable quality (e.g. intelligence) in a population of subjects. To do this, it measures variables that are observable (for example, n responses to a questionnaire or a set of n logical problems) of each of the subjects. Variables are assumed to be related to the unobservable magnitude of interest. In particular, the n variables should make stable and consistent measurements, with a high level of correlation between them. Cronbach's alpha allows to quantify the level of reliability of a measurement scale for the unobservable magnitude constructed from the observed n variables. Cronbach's alpha is not a usual statistic, so it is not accompanied by any p-value that allows to reject the reliability hypothesis on the scale. However, the closer it gets to its maximum value, 1, the greater the reliability of the scale. In addition, in certain contexts and by tacit convention, alpha values greater than 0.7 or 0.8 (depending on the source) are considered enough to ensure the reliability of the scale.
Intraclass correlation coefficient (ICC)	The intraclass correlation coefficient (ICC), originally introduced by Fisher, is a special formulation of Pearson's correlation coefficient (A). This method allows to evaluate the overall agreement between two or more measurements or observation methods based on an analysis of variance model (ANOVA) with repeated measurements. It is defined as the proportion of total variability that is due to the variability of the subjects.
Kappa index	The kappa index is just one of the statistical methods used to evaluate the agreement between 2 or more observers. The Kappa match index (K) is a measure proposed by Cohen in 1960, which is based on

	comparing the agreement observed in a dataset, with that which could occur by mere chance. If K is zero, this means that the observed match coincides with that which would occur by pure chance. Positive values indicate greater agreement than would be expected by pure chance. If the result was 1, it would be a perfect match. If K has a negative value, it means there is a mismatch. However, it is also necessary to calculate the confidence interval for K, if it includes zero, the conclusion is that the agreement has been by chance.
Negative Likelihood Ratio (NLR) or Negative Probability Ratio (NPR)	It is calculated by dividing the probability of a negative result in the presence of a disease, by the probability of a negative result in the absence of it. Therefore, it is the ratio between the fraction of false negatives (1-sensitivity) and the fraction of true negatives (specificity). The NPR indicates the ratio between the likelihood that a sick individual will get a negative result, relative to the likelihood that a non-sick or healthy individual will get a negative result. In other words, this indicator shows that it is unlikely that in a sick patient the test will be negative, with respect to a healthy patient, with the same negative result. Thus, a positive likelihood ratio greater than 1 indicates that there is a high probability that the individual has the disease, and the higher the value of the ratio obtained, the greater the likelihood of having the disease. Conversely, a negative likelihood ratio of less than 1 will decrease the likelihood that the individual has the disease under study.
Negative Predictive Value (NPV)	On the contrary, the negative predictive value is the probability that an individual with a negative test result, does not present the disease or is healthy.
Positive and Negative Likelihood Ratio (PLR and NLR)	Also called prognostic efficiency index (IEP). This indicator has the characteristic of being a fixed index, because it is used when the diagnostic test has no dichotomous results, but a threshold or breakpoints. Calculating the likelihood ratio is another way or method of assessing the accuracy of a test in the clinical field. It offers the advantage over the other indicators, that it is independent of the prevalence of the disease in a population. The likelihood ratio indicates that a result of a diagnostic test will raise or reduce the likelihood of having the disease, i.e. relative to the previous probability of the disease (prevalence); in other words, it is a ratio and not a proportion. Each diagnostic test is characterized by two ratios of plausibility: the positive

	likelihood ratio or positive probability quotient ratio and the negative likelihood ratio or negative probability quotient, these are described below.
Positive Likelihood Ratio (PLR) or Positive Probability Ratio (PPR)	It is calculated by dividing the probability of a positive outcome in sick patients by the probability of a positive result in healthy individuals. It is, in short, the ratio between the fraction of true positives (sensitivity) and the fraction of false positives (1-specificity) and indicates the disease ratio or the probability of having the disease if the result is positive.
Positive Predictive Value (PPV)	It is the probability that an individual is sick when the test result is positive, in other words, it is a conditioned probability that a patient who is positive for the test, will have the disease.
Predictive validity	Predictive validity assesses the degree to which the new measurement scale is able to predict the score obtained by the gold standard when it is not applied at the same time but at some point in the future. Statistically, this comparison is performed in the same way as in concurrent validity.
Reliability	Reliability is the degree to which an instrument is able to measure without error. It measures the proportion of variation in measurements that is due to the different values that a variable takes and is not the result of systematic (bias) or random (random) error. That is, this property determines the proportion of total variance attributable to true differences that exist between subjects.
Reliability coefficient	It is symbolized by ***rxx*** as it is made up of a correlation between two equivalent measures (also called *test-retest or* RTT) or intra-method reliability coefficient. Its values range from 0 (lack of reliability) to 1 (perfect reliability), it can also be expressed by a percentage. The reliability coefficient is equal to the ratio between the observed variance and the total or true variance: if both types of variance match, the reliability would be equal to

	1 and if there is no match it would be 0, from here some formulas are derived to quantify the reliability of a measurement situation.
ROC Curve	The most commonly used indicator in many contexts is the area under the ROC (Receiver operating characteristic) curve. That allows to evaluate the best cut-off point of a continuous scale that is used to classify people into a binary condition. This is done through determining the sensitivity and specificity (in practice it is its complement) for each cut-off point, and to construct a curve that allows to identify that point for the best combination of those probabilities.
Safety of a Test	The safety of a test is determined by positive predictive (PPV) and negative (NPV) values. These indexes are important for assessing the usefulness of a test, in the clinical field and on an individualized basis, i.e. for each patient, contrary to the information provided by sensitivity and specificity (the latter are not useful in the clinical practice).
Sensitivity	It is defined as the ability of a test to correctly identify those who have the disease. This is equal to the number of subjects with a positive test who have the disease, divided by all subjects who have the disease.
Specificity	It is defined as the ability of a test to identify those who do not have the disease, and is equal to the number of subjects who are negative to the test and who do not have the disease, divided by the number of people who do not have the disease or are healthy.

Bibliography

1. Legleye S, Piontek D, Kraus L. Psychometric properties of the Cannabis Abuse Screening Test (CAST) in a French sample of adolescents. Drug Alcohol Depend. 2011;113(2-3):229-35.

2. Knight JR, Sherritt L, Shrier LA, Harris SK, Chang G. Validity of the CRAFFT substance abuse screening test among adolescent clinic patients. Arch Pediatr Adolesc Med. 2002;156(6):607-14.

3. Vazquez-Barquero JL HCS, Gaite L. La entrevista estructurada en Psiquiatría Rev. Asoc. Esp. Neuropsiq. . 1993;13(44):19-38.

4. Kendell RE, Cooper JE, Gourlay AJ, Copeland JR, Sharpe L, Gurland BJ. Diagnostic criteria of American and British psychiatrists. Arch Gen Psychiatry. 1971;25(2):123-30.

5. Feighner JP, Robins E, Guze SB, Woodruff RA, Jr., Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry. 1972;26(1):57-63.

6. WHO. International Classification of Diseases and Related Health Problems 11th edition (ICD-11), 2018. https://icd.who.int/browse11/l-m/en.

7. APA. Manual Diagnóstico y Estadístico de los Trastornos Mentales DSM-IV-TR.

. Barcelona: American Psychiatric Association; 2002.

8. Kupfer, D. J., Regier, D. A., Arango López, C., Ayuso-Mateos, J. L., Vieta Pascual, E., & Bagney Lifante, A. (2014). DSM-5: Manual diagnóstico y estadístico de los trastornos mentales (5a ed.). American Psychiatric Association.

9. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction. 1993;88(6):791-804.

10. Babor JC, Higgins-Biddle JC, Saunders J, Monteiro MG. AUDIT : the Alcohol Use Disorders Identification Test : guidelines for use in primary health care. World Health Organization; 2001.

11. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. Arch Gen Psychiatry. 1992;49(8):624-9.

12. Williams JB, Gibbon M, First MB, Spitzer RL, Davies M, Borus J, et al. The Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. Arch Gen Psychiatry. 1992;49(8):630-6.

13. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The Composite International Diagnostic Interview. An epidemiologic Instrument suitable

for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry. 1988;45(12):1069-77.

14. Rijnders CA, van den Berg JF, Hodiamont PP, Nienhuis FJ, Furer JW, Mulder J, et al. Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). Soc Psychiatry Psychiatr Epidemiol. 2000;35(8):348-52.

15. Cooper JE, Copeland JR, Brown GW, Harris T, Gourlay AJ. Further studies on interviewer training and inter-rater reliability of the Present State Examination (PSE). Psychol Med. 1977;7(3):517-23.

16. Robins LN, Helzer JE, Ratcliff KS, Seyfried W. Validity of the diagnostic interview schedule, version II: DSM-III diagnoses. Psychol Med. 1982;12(4):855-70.

17. Torrens M, Serrano D, Astals M, Perez-Dominguez G, Martin-Santos R. Diagnosing comorbid psychiatric disorders in substance abusers: validity of the Spanish versions of the Psychiatric Research Interview for Substance and Mental Disorders and the Structured Clinical Interview for DSM-IV. Am J Psychiatry. 2004;161(7):1231-7.

18. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59 Suppl 20:22-33;quiz 4-57.

19. Cacciola JS, Alterman AI, Habing B, McLellan AT. Recent status scores for version 6 of the Addiction Severity Index (ASI-6). Addiction. 2011;106(9):1588-602.

20. McLellan AT, Alterman AI, Cacciola J, Metzger D, O'Brien CP. A new measure of substance abuse treatment. Initial studies of the treatment services review. J Nerv Ment Dis. 1992;180(2):101-10.

21. Bobes J, Bascaran MT, Bobes-Bascaran M, Carballo J, Diaz Mesa E, Saiz P. Valoración de la gravedad de la adicción: Aplicación a la gestión clínica y monitorización de los tratamientos2007.

22. Darke S, Ward J, Hall W, Heather N, Wodak A. Teh Opiate Treatment Index (OTI)Manual. Australia: University of New South Wales; 1991. Contract No.: 11.

23. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med. 1979;9(1):139-45.

24. Marsden J, Gossop M, Stewart D, Best D, Farrell M, Lehmann P, et al. The Maudsley Addiction Profile (MAP): a brief instrument for assessing treatment outcome. Addiction. 1998;93(12):1857-67.

25. Marsden J, Nizzoli U, Corbelli C, Margaron H, Torres M, Prada De Castro I, et al. New European instruments for treatment outcome research: reliability of the maudsley addiction profile and treatment perceptions questionnaire in Italy, Spain and Portugal. Eur Addict Res. 2000;6(3):115-22.

26. Alvarado ME, Garmendia ML, Acuna G, Santis R, Arteaga O. [Assessment of the alcohol use disorders identification test (AUDIT) to detect problem drinkers]. Rev Med Chil. 2009;137(11):1463-8.

27. Soto-Brandt G, Portilla Huidobro R, Huepe Artigas D, Rivera-Rei A, Escobar MJ, Salas Guzman N, et al. [Validity evidence of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) in Chile]. Adicciones. 2014;26(4):291-302.

28. Group WAW. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. Addiction. 2002;97(9):1183-94.

29. Kohn R, Levav I, de Almeida JM, Vicente B, Andrade L, Caraveo-Anduaga JJ, et al. [Mental disorders in Latin America and the Caribbean: a public health priority]. Rev Panam Salud Publica. 2005;18(4-5):229-40.

30. Wing JK, Nixon JM, Mann SA, Leff JP. Reliability of the PSE (ninth edition) used in a population study. Psychol Med. 1977;7(3):505-16.

31. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry. 1981;38(4):381-9.

32. Janca A, Robins LN, Bucholz KK, Early TS, Shayka JJ. Comparison of Composite International Diagnostic Interview and clinical DSM-III-R criteria checklist diagnoses. Acta Psychiatr Scand. 1992;85(6):440-3.

33. Lago L, Bruno R, Degenhardt L. Concordance of ICD-11 and DSM-5 definitions of alcohol and cannabis use disorders: a population survey. Lancet Psychiatry. 2016;3(7):673-84.

COMPLEMENTARY BIBLIOGRAPHY

- Luján-Tangarife, J. and Cardona-Arias, J., Construcción y validación de escalas de medición en salud: revisión de propiedades psicométricas.
- Laura Elisa Montoya Gonzalez, Diana Patricia Restrepo, Bernal Roberto Mejía-Montoya, José Bareño-Silva, Gloria Sierra-Hincapié, Yolanda Torres de Galvis, Daniel Marulanda-Restrepo, Natalia Gómez-Sierra, Silvia Gaviria-Arbeláez. Sensibilidad y especificidad entre la Entrevista Diagnóstica Internacional Compuesta versión 3.0 (World Mental Health, CIDI) con la Evaluación Clínica Estandarizada versión I (SCID-I) en la Encuesta de Salud Mental de la ciudad de Medellín, 2012 Revista Colombiana de Psiquiatría Volume 45, Issue 1, January–March 2016, Pages 22-27.
- Berkanovic E. The Effect of Inadecuate Language Translation on Hispanics Responses to Health Surveys. Am J Public Health 1980; 70:1273-1276.

- Beaton, DE., Bombardier, C., Guillemin, F., Ferraz, MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (PhilaPa 1976) 2000; 25: 3186-3191.
- Carmen R. Wilson VanVoorhis and Betsy L. Morgan, Understanding Power and Rules of Thumb for Determining Sample Sizes, Tutorials in Quantitative Methods for Psychology 2007, vol. 3 (2), p. 43-50.
- Garson, D. G. (2008). Factor Analysis: Statnotes. Retrieved March 22, 2008, from North Carolina State University Public Administration Program.
- Gossop M, Best D, Marsden J, Strang J. Test-retest reliability of the Severity of Dependence Scale. Addiction 1997; 92:353.
- Lamprea, J, Gómez, C. Validez en la evaluación de escalas. Rev. Colomb. Psiquiat. 2007; 36: 340-8.
- Legleye S KBBFRM. Validation of the CAST, a general population Cannabis Abuse Screening Test. Journal of Substance Use 2007; 12:233-42.
- Legleye S., Piontek D., Kraus L., Morand E. et Falissard B. (2013). A validation of the Cannabis Abuse Screening Test (CAST) using a latent class analysis of the DSM-IV among adolescents. International Journal of Methods in Psychiatric Research.
- Martin G, Copeland J, Gates P, Gilmour S. The Severity of Dependence Scale (SDS) in an adolescent population of Cannabis users: Reliability, validity and diagnostic cut-off. Drug Alcohol Depend 2006;83:90-3.
- Obradovic, I., Guide pratique des principaux outils de repérage de l'usage problématique de cannabis chez les adolescents, 2013, OFDT.
- Sánchez, R., Gómez, C. Conceptos básicos sobre la validación de escalas. Rev. Col. Psiquiatría. 1998; 27: 121-30.
- Streiner D, Norman GR. Health Measurement Scales. A Practical Guide to Their Development and Use. Oxford: Oxford University Press, 1995.
- Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: HarperCollins.
- https://www.encorewiki.org/display/~nzhao/The+Minimum+Sample+Size+in+Factor+Analysis

Clauses *ad cautelam,* clarifications and exemptions

COPOLAD is a programme funded by the European Union through the Commission's Directorate-General for International Cooperation and Development (DG DEVCO / EuropeAid).

The opinions or positions expressed in this document are the sole responsability of the authors and editors; in all cases, they do not reflect or represent the views or positions of the COPOLAD Consortium, neither the ones of the European Commission.

Considering that respect for the environment is one of the framework values of COPOLAD, the Consortium is committed to organize its activities taking into account its impact on the environment, particularly CO2 emissions. Therefore, virtual communication techniques are prioritized and the use of recyclable material is recommended along the implementation of the Programme.

COPOLAD Executive & Coordination Body (ECB)
FIIAPP, C. Beatriz de Bobadilla 18 Madrid-28040 (Spain)
+34 911 442 766 @programacopolad /programacopolad
www.copolad.eu

COPOLAD CONSORTIUM

LEADER

FIIAPP COOPERACIÓN ESPAÑOLA

COUNTRIES

SEDRONAR (Argentina) · **SENAD** (Brazil) · **SENDA** (Chile)
MINJUSTICIA (Colombia) · **ICD** (Costa Rica) · **CND** (Cuba)
VLADA (Czech Republic) · **CND** (Dominican Republic) · **MREMH** (Ecuador)
CNA (El Salvador) · **GIZ** (Germany) · **DNII** (Honduras) · **CONADIC** (Mexico)
CONAPRED (Panama) · **DEVIDA** (Peru) · **NBDP** (Poland) · **SICAD** (Portugal)
NAA (Romania) · **DGPNSD** (Spain) · **NDC** (Trinidad and Tobago)
JND (Uruguay) · **ONA** (Venezuela)

EUROPEAN AGENCY

EMCDDA

MULTILATERAL AGENCIES

CICAD – OAS · **PAHO – WHO**

BI-REGIONAL NETWORKS

AIAMP · **IDPC** · **RIOD**

This project is funded by the European Union