

Appendix II

Description of the methodology utilised for the collection, assessment and utilization of the scientific evidence

Introduction

This document describes the systematic methodology utilized for the collection, assessment and use of the scientific evidence at the basis of the International Standards on Drug Use Prevention. The methodology was developed and implemented by UNODC with the inputs of a globally representative group of 85 scientists and policy makers in the field of drug prevention, a smaller group of which also volunteered to provide more in-depth support to its conceptualisation.

Members of this Group of Experts were identified through an inclusive process aiming at involving as many recognised individuals in the field of drug prevention as possible, and ensuring that all the sub-fields of prevention research were part of the expert group. The Member States of the United Nations were requested to nominate experts, as well as other international organisations (EMCDDA, CICAD, WHO, ILO) and recognised institutions (NIDA) and civil societies organisations (CCSA, Mentor) in the field. The list of participants is acknowledged in the main text of the Standards.

The methodology allows the Standards to be based on scientific-evidence, whilst taking into account the limits of the evidence, the resources available to the process of development of the Standards, the practical nature of the Standards that are aiming to inform policymakers, rather than fellow scientists. In particular, this methodology strives to provide a transparent picture of the strength of evidence that is available to support different interventions and policies with regard to their efficacy and effectiveness in general, as well as in different geographical, socio-economical and cultural settings, and their characteristics.

To this end, interventions and policies were included and described in the Standards on the basis of a hierarchy of study designs, and on the basis of assessing the methodological quality of these studies, as described in the following sections.

Evidence-Based Practice

Since the early 1990s there has been a growing movement in health, education, and other behavioural service fields toward the delivery of services/ practices whose impact on positive outcomes are grounded in science and research. The movement is defined by the term 'evidence-based' that is assigned to practices, programs, or interventions. There are a variety of definitions of the term 'evidence-based'. The Evidence-Based Practice Institute of the University of Washington's definition encompasses the common elements:

“Evidence Based Practice (EBP) is the use of systematic decision-making processes or provision of services which have been shown, through available scientific evidence, to consistently improve measurable client outcomes. Instead of tradition, gut reaction or single observations as the basis for making decisions, EBP relies on data collected through experimental research and accounts for individual client characteristics and clinician expertise.” (Evidence Based Practice Institute, 2012; <http://depts.washington.edu/ebpi/>)

Several groups have established criteria for the scientific basis for evidence-based practices or programmes¹. In general there are great similarities across the criteria, with groupings of evidence into “best” or “excellent” down to “good” or “promising”. It is in the lower range of categories of evidence where there are the greatest disparities.

¹ E.g. the National Registry of Evidence-based Programs and Practices (NREPP) of the Substance Abuse and Mental Health Services Administration (SAMSHA) in the USA, Blueprints for Violence Prevention, a project of the Center for the Study and Prevention of Violence at the University of Colorado, and, in the medical field, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system, etc.

Efficacy and Effectiveness

The science of evaluation of social and behavioural interventions, as well as most clinical practices, sets out standards for the conduct of quality research. These standards apply to research design and statistical methodologies in order to establish a causal link between exposure to the intervention and the outcomes of interest. Such standards require: an evaluation design using randomization if possible or alternative design strategies for addressing confounding if not; clearly articulated research objectives; a theory-based logic or conceptual model that shows how the intervention and intervention components are associated with the short-, intermediate, and long-term outcomes of interest, so that participation or exposure to the intervention can be related to the outcomes of interest and not to other, external influences; and that there is a link between the components of the intervention derived from the model and the outcomes. Another aspect of a quality evaluation study includes an examination of what other factors associated with the target group or the environment in which the intervention was delivered modify or temper the relationship between exposure to the intervention and its outcomes.

Clearly, not all interventions impact all of those exposed to them. Variations in outcomes often are noted within the exposed populations and this information along with the findings from mediation analyses (analysing the role of different intervention components) serve to assist intervention developers to modify or enhance their program.

In the continuum of evaluation research, studies demonstrating that the intervention as designed has a positive impact under the controlled research conditions are called efficacy studies. Once the intervention impact has been demonstrated under these controlled circumstances, the next stage of the evaluation is to determine the extent to which the impact is sustained in 'real world' delivery settings. These next stage evaluation studies are generally referred to as effectiveness studies.

The majority of evaluations showing the positive impact of preventive interventions presented here have been conducted within controlled experimental conditions rather than in "real world" conditions and are therefore mostly efficacy studies. Environmental or policy prevention strategies have mostly been evaluated in 'real world' conditions as controlling exposure is quite challenging if possible at all, and they would thus fall mostly in to the effectiveness studies.

Outcomes of evidence-based prevention

Interventions and policies were included in the Standards if they have demonstrated efficacy or effectiveness. The outcomes of interest were the elimination or reduction of the use of illicit drugs, alcohol and tobacco in a follow-up at least one year after exposure to the intervention. At least two and a majority of primary studies should have reported positive effects in this respect, and no studies should have reported iatrogenic effects on important outcomes. Impact on mediating variables was considered only in the case of interventions and policies targeting young children (see below).

In fact, a number of interventions and policies target children well before the age of onset of substance use (infancy or primary school years). Some of these have been evaluated in long term follow up studies showing effects in terms of preventing drug or substance use in adolescence or adulthood. However, some of these interventions and policies have not been evaluated through long-term follow up studies, and thus data on their effectiveness on preventing future substance use is not yet available. Moreover, data on their impact on important outcomes that have been shown in the scientific literature to be associated with the onset of substance use is available (mediating variables).

Therefore, interventions and policies targeting young children and showing an impact on outcomes strongly linked in the scientific literature to the onset of substance use were also included in the Standards, although the strength of the evidence was classified as one step weaker. In order to be included, two primary studies needed to report a positive effect at least one year after intervention delivery in terms of at least two mediating variables. Mediating variables were identified on the basis of consensus of the Group of Experts and are listed in Annex I. No relative weight was identified and assigned to the variables.

Collection and screening of the scientific evidence

To try and reduce to the maximum extent possible the risk of publication bias, a multiple research strategy was followed. First and foremost, the members of the Group of Experts on the Prevention Standards were requested to provide all relevant evidence published in scientific journals or in official reports in any language. No inclusion/ exclusion criteria were set as to the date of the publication.

In general, the Experts were requested to provide systematic reviews and meta-analysis. Moreover, with regard to interventions or policies that are well researched, they were requested to provide what they considered key studies. Finally, with regard to intervention or policies or with regard to implementation in geographical areas that are not well researched, they were requested to provide any available study. As all the evidence was identified by a group of leading international experts, this is seen to be the first source of reliability assuring that the evidence base includes all the most important studies and that the studies are of sufficient quality.

In addition to this, other sources of quality scientific literature were consulted, as follows:

- ✓ All references included in the EMCDDA Best Practice Portal were included in the process;
- ✓ The Cochrane and the Campbell libraries were searched for reviews related to the prevention of drug use;
- ✓ References from a review of reviews that was being undertaken by Liverpool John Moores University were kindly shared with UNODC and were also included in the process
- ✓ References included in the selected studies received by the Group of Experts were also cross-checked.

A total of 584 references were received and were screened for relevance to the process. To be included in the process of assessment of the evidence, studies needed to report the impact of an intervention or a policy with regard to the prevention of drug alcohol or tobacco use after the intervention on any population (256). Originally, the criteria requested for outcomes to be assessed at least one year after the intervention. While many reviews did not clearly report on this and were included anyway, Annex V details this information for each study that was eventually accepted as part of the evidence base and, in case there is not sufficient information with regard to these issues, this is indicated in the main text.

Studies reporting impact in terms of treatment of drug, alcohol or tobacco dependence, as well as studies reporting impact only in terms of prevention of the health and social consequences of drug, alcohol or tobacco use (e.g. prevention of crashes due to intoxicated driving) were not included (60). Studies reporting impact on mediating variables (e.g. improved parenting) were included only if they were targeting children during middle childhood and younger (31). Epidemiological studies (i.e. investigating the link of certain individual or environmental factors to the onset of substance use) and/or studies exploring important general issues with regard to the prevention of drug use and substance

abuse were considered as an important part of the context of the Standards (268). Flow chart 1 summarises this phase of the process.

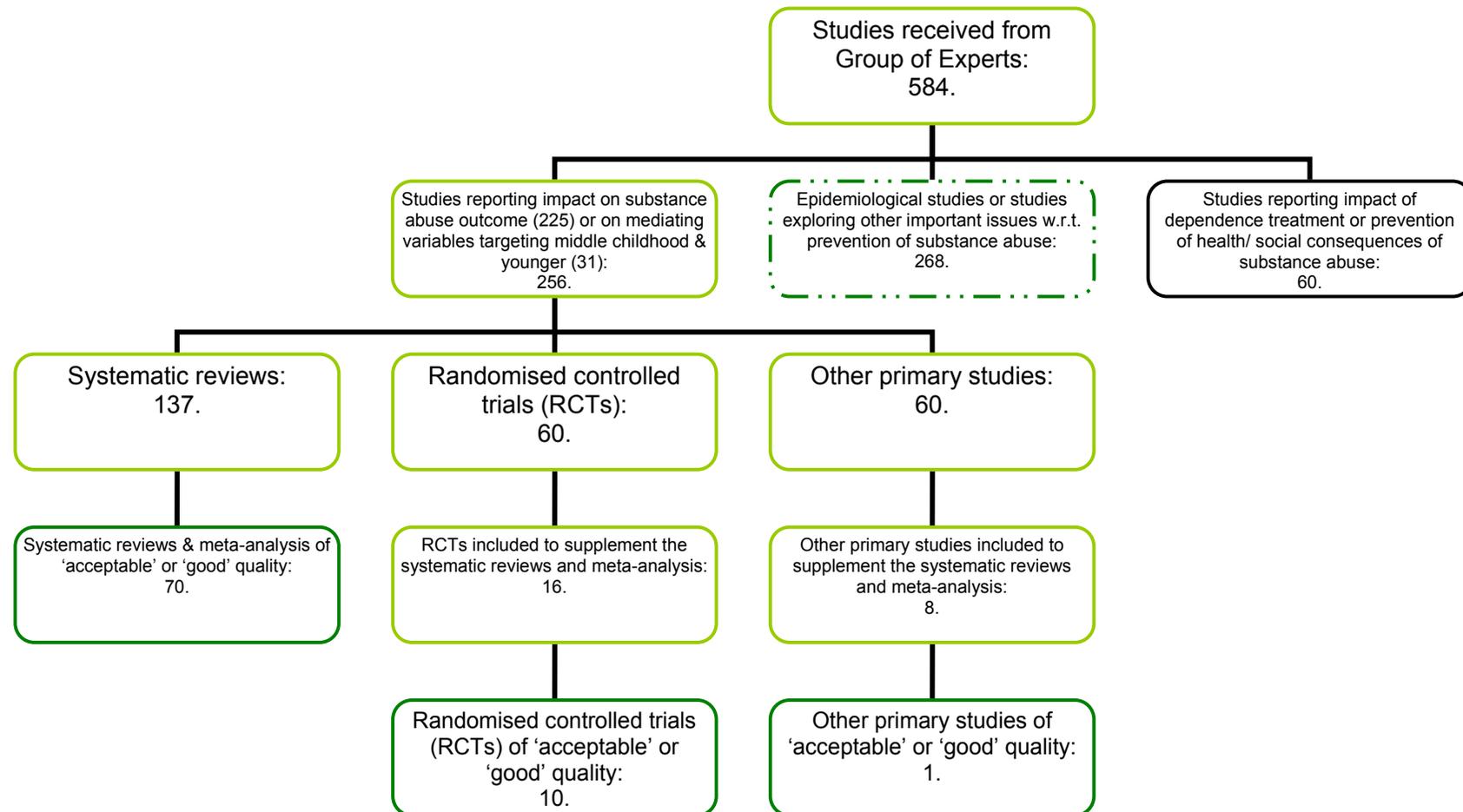
Assessment of the scientific evidence

Introduction

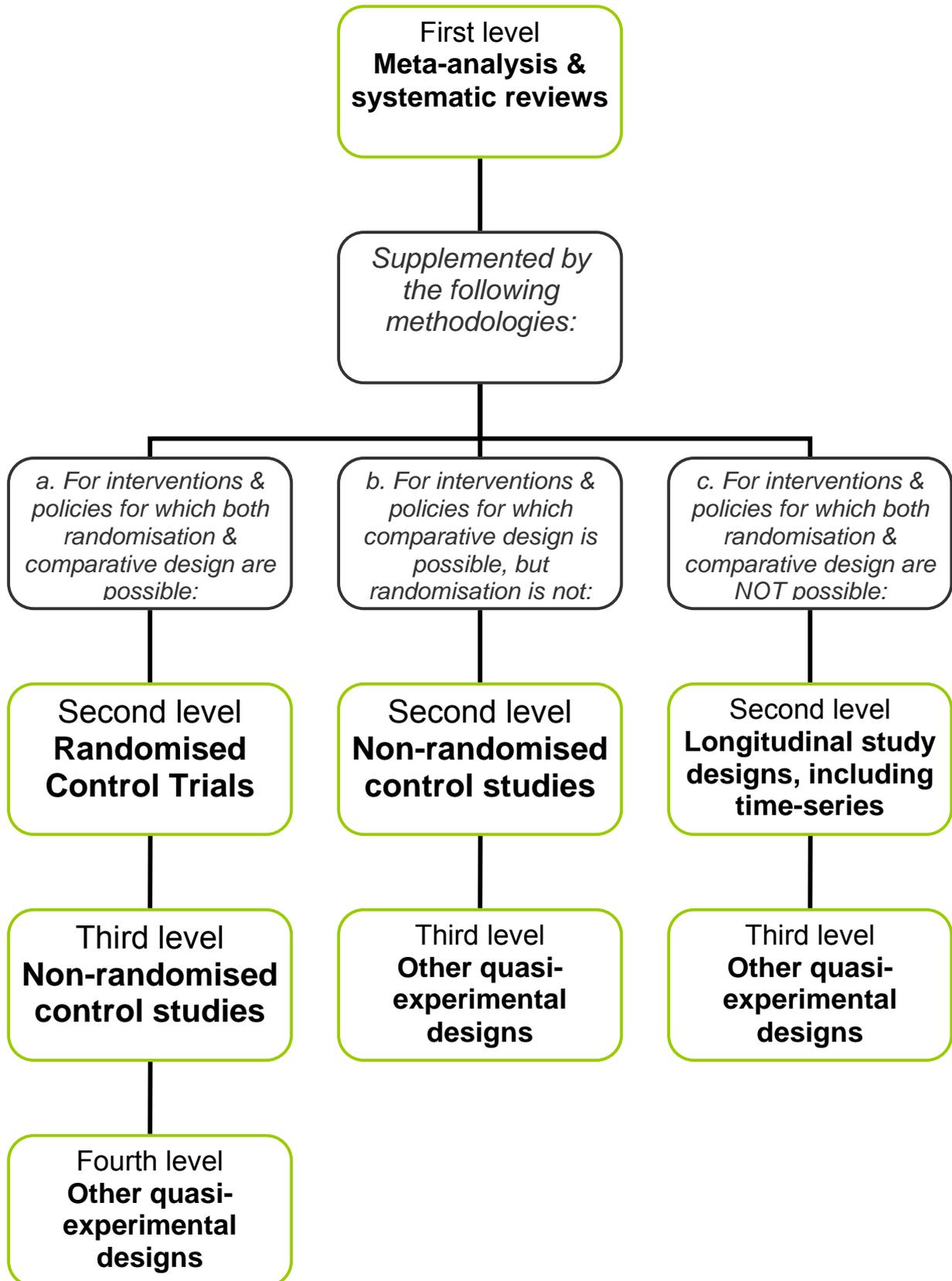
Evidence was reviewed according to a hierarchy of study designs, as represented in Flow Chart 2. The first level of the hierarchy included systematic reviews and meta-analysis, but, in case of a gap in the available evidence at this first level, this was supplemented with experimental or quasi-experimental primary studies (including randomised control trial, non-randomised control studies, time-series, etc.) at a second, third and fourth level, as described below. This process allowed the methodology to be both transparent and systematic, while substantially reducing the number of studies to be examined. The quality of all studies included in the process was assessed. The combination of the kind and the quality of studies supporting interventions and policies became the basis for the indication of the strength of evidence of effectiveness included in the main text of the Standards.

From a practical point of view, the assessment was undertaken by two staff members of UNODC. They developed and piloted the rating tools using a number of studies. The provisional ratings were compared and discussed to ensure high inter-rater reliability in applying the criteria, and the tools were revised to make it easier to use. They then proceeded to rate all studies independently. Cases of disagreement were discussed and resolved, if necessary, with the input of a third staff member. Moreover, the table detailing the assessment of each study included in the process (whether it was rated 'good', 'acceptable' or 'not acceptable') was shared with the Group of Experts together with the draft of the Standards.

Flowchart 1: Summary of screening on studies received



Flowchart 2: Hierarchy of study designs for the inclusion of interventions and policies in the Standards



Review of systematic reviews and meta-analysis

The first step was the assessment of the quality of systematic reviews and meta-analysis. Recognised methodologies such as those adopted by the Cochrane² and Campbell³ Collaborations and the Community Guide⁴ were rated as 'good', while others were rated as "good", "acceptable", or "not-acceptable", utilising an instrument adapted from the standards of the Cochrane Reviews. The instrument is attached as Annex II and reviewed the following issues:

- A. Clear, transparent and sufficient inclusion criteria for study selection;
- B. Transparent, broad and diverse methods for literature search;
- C. Sufficient detail on included studies concerning methodology, participants, intervention characteristics and findings;
- D. Documentation and quality of data analysis and interpretation.
- E. The quality of the data analysis and interpretation (multiple assessors used in assessing the quality of the studies, clearly reported results, sufficiently similar results, reported and elaborated reasons for variations in results, description of missing data, assessed and reported possibility of bias, double counting of primary data avoided).

The result of the quality assessment is attached as Annex IV to this document. There were 137 systematic reviews and 70 were rated 'good' or 'acceptable'. A 'not acceptable score' was mainly due to a lack of detail in the reporting either in relation to the search strategy or concerning the included studies.

Assessment of primary studies

This first part of the process was supplemented with the results of other primary studies that were included if:

1. They covered an intervention or policy for which no 'acceptable' or 'good' review was found at all;
2. They reported impact on drug use on an intervention or policy for which only 'acceptable' or 'good' reviews reporting impact on alcohol or tobacco or mediators were found;
3. They reported impact on drug use on an intervention or policy for which 'acceptable' or 'good' reviews reporting impact on drug use were found AND they were published AFTER the data collection of the last 'acceptable' or 'good' review.
4. They study the cost-effectiveness of an intervention or policy.

² <http://handbook.cochrane.org/>.

³ <http://www.campbellcollaboration.org/>

⁴ <http://www.thecommunityguide.org/index.html>

5. They reported impact on the implementation of an intervention or a policy in a country other than the USA, Canada, Europe, Australia and New Zealand.

This step allowed the process to radically reduce the number of primary studies to be assessed and analysed. Sixteen (16) randomized control trials and 8 other primary studies were selected at this stage. Although the Group of Expert is confident that a reliable summary of the available evidence would be generated by this methodology, an important limitation should be noted. There might be cases where a systematic review fails to support a type of intervention, but replicated RCTs of a particular programme within that type provide strong evidence for the particular intervention. In cases, where RCTs were published before or at the same time as the review, their evidence would be lost.

The quality of all primary studies included following this second screening was also assessed. The relevant instruments can be found in Annex II. With regard to the randomised controlled trials, the following criteria, based on those of the Cochrane Drug and Alcohol group (CDAG) (Amato, 2005) were used:

- A. Randomization methods and baseline comparability of groups.
- B. Blinding of participants, personnel and/or outcome assessors.
- C. Amount, nature or handling of incomplete outcome data due to attrition (losses to follow-up) and exclusions.
- D. Other sources of bias, including fidelity of intervention implementation.

With regard to non-randomised control studies, always according to the Cochrane Drug and Alcohol group (CDAG) (Amato, 2005), the instrument utilized was the same as for randomised control studies, with some items rated as 'not applicable'. Finally, although a draft instrument for longitudinal studies (e.g. time-series analysis) had been developed, it was not utilised as no studies of this kind were included following the post-reviews selection.

Ten (10) randomized control trials and 1 other primary study were rated to be 'good' or acceptable. A 'not acceptable' rating was mostly linked to a failure to describe the procedure for random sequence generation.

Data extraction

The studies rated 'acceptable' or 'good' were then coded as to the intervention or policy they were concerned with, the setting where the intervention or policy was implemented, and the age of the target group. An attempt was made to code interventions and policies also according to the gender of the target group.

However, with the obvious exception of interventions and policies targeting pregnant women, in the vast majority of cases the gender of the target group was not specified, nor were results reported by gender. Therefore, this coding was dropped and results of interest were reported together with the other results and presented in the main text of the Standards.

The results included in each study were then summarised, including (where available) the time of follow-up, effect sizes and of characteristics linked to effectiveness. The table summarising the coding and the results of the studies is attached as Annex V.

Inclusion of interventions and policies in the Standards

An intervention or a policy was included in the Standards as an evidence-based strategy if at least one 'acceptable' systematic reviews or meta-analysis reported positive impact with regard to drugs and/or alcohol and/or tobacco use, or, in the case of interventions and policies targeting children during middle childhood and younger, relevant mediating variables. Discrepancies among studies were resolved on the basis of group consensus. In case of no available 'acceptable' or 'good' systematic review or meta-analysis, then demonstrated effectiveness needed to be supported by the results of at least two other primary studies.

Strength of the evidence

The evidence supporting interventions and policies is not all the same. In some cases, there is a lot of evidence of good quality. In others, very limited evidence. In others, there might be evidence, but not of great quality. This information is of crucial importance to policy makers and is summarised in the main text of the Standards. The hierarchy of methodologies and the rating of the quality of the studies described above determined how the strength of the evidence of effectiveness was assessed and presented to policy makers.

The best available evidence ("strong evidence") would be based on studies belonging to the first level of the hierarchy (i.e. systematic reviews and meta-analysis). An intervention or policy that was reported to be effective on the basis of systematic review(s) and/or meta-analysis assessed as 'good' would be

described as being based on 'strong evidence' of effectiveness. Accordingly, if the systematic review(s) and/or meta-analysis were assessed as 'acceptable', the evidence would be characterised as 'good'.

In case of lack of systematic reviews and meta-analysis, the evidence would be based on primary studies with designs belonging to the second level of evidence. The design that deemed acceptable as the second level of evidence was not the same for all interventions or policies, as detailed in Flowchart 1. Consider as an example, the case of an intervention that can be evaluated through a randomised controlled trial (e.g. parenting skills training). In this case, the second level of evidence would be constituted by randomised controlled trials. However, consider, as a second example, an intervention for which it would be impossible to organise a study including a control group (e.g. a nation-wide media campaign). In this case, it was deemed reasonable to consider that the second level of study designs would be longitudinal studies (including time-series).

Therefore, it could be said that the study designs deemed as acceptable as second level of evidence differed according to the kind of study design that is in principle possible for any given intervention or policy. In this respect, interventions and policies typically fall into three groups, i.e. interventions and policies for which:

1. A randomised controlled trial is possible;
2. A randomised controlled trial is not possible, because randomisation is not feasible or ethical; however, a comparative design is possible;
3. A randomised controlled trial is not possible, because a comparative design is not feasible in the first place (for example in the case of national media campaigns or regulations/policies).

Interventions and policies were assigned to these groups on the basis of consensus of the experts in the evidence working group. The list of interventions and policies grouped accordingly is attached as Annex III to this methodology. For each of these groups, the study designs that were deemed acceptable as second and third level of evidence were identified as described in Flowchart 2.

As in the case of the systematic reviews and the meta-analysis, the study design, its level and quality were combined to provide an indication of the strength of the evidence supporting the indications of effectiveness in the main text of the Standards. Studies based on second level study designs and rated as 'good', would also provide good quality evidence, while those rated as 'acceptable', together with studies based on a third level study design and rated as 'good' would constitute 'promising evidence'. All the rest was not considered in the base of evidence supporting the inclusion (or otherwise) of an intervention or policy in the Standards. Table 1 summarises the criteria for rating the evidence.

Interventions or policies for which the available evidence would not be even rated as ‘promising’ have been briefly described in the Standards in a separate section clearly indicating that at the moment there is no evidence to tell us whether these interventions and policies are effective or not. In a few cases, there is evidence that an intervention and policy is not effective or, worse, has negative effects in terms of substance use. In general, this was found to be the case with characteristics of interventions and policies or with components of certain interventions and policies. Therefore, this information has been reported in the Standards under the relevant intervention and policy.

Table 1
Assessment of evidence

Study design	Quality of study	Assessment of the evidence
First level Meta-analysis and systematic reviews	Good quality	Strong
	Acceptable quality	Good
	Not acceptable	Not included
Second level RCTs/ non-randomised control studies/ time series analysis	Good quality	Good
	Acceptable quality	Promising
	Not acceptable	Not included
Third & fourth level Other research designs	Good quality	Promising
	Acceptable quality	Not included
	Not acceptable	Not included

Transferability

Transferability refers to the evidence that an intervention or policy has been found to be effective in geographical and cultural settings different from those in which the initial assessments were made. In this context, it has to be recognised

that the vast majority of evidence originates from studies conducted in North America, specifically the USA, and a few other countries in North America (Canada), (mostly Western) Europe and Oceania (Australia, New Zealand). That is why the geographical origin of the evidence has been indicated in the main text of the Standards under “Evidence of Effectiveness” for each interventions and policy.

Characteristics of effective interventions and policies

The process described above provided a strong and transparent indication of which interventions and policies are effective in preventing drug use and on the strength of the evidence supporting this statement. In the vast majority of cases, the available evidence did not allow an in-depth analysis of which components or which characteristics of an intervention or a policy ‘are the active ingredient’ or really make the strategy effective. Where available, the results of this analysis were provided.

Further, they were supplemented by indications arising from the other studies provided by the Group of Expert, particularly those reporting the results of mediation analysis, summarised on the basis of expert group consensus. This participatory process allowed the Standards to provide an indication of how interventions and policies should and should not be implemented in order to maximize the chances of their effectiveness. It is very important to note that the resulting indications should not be taken to imply a causal effect between the characteristics of an intervention/policy and its effectiveness. However, they can be taken to provide a description of characteristics that have been found by the Group of Experts to be associated with more effective interventions and policies and, therefore, with a stronger possibility of effectiveness.

Example of application of the methodology

This section briefly describes how the methodology has been applied to one specific intervention: early childhood education.

The first level of evidence includes systematic reviews. According to Appendix I, early childhood education is covered by only one review (D'Onise et al 2010). According to Annex IV, D'Onise et al 2010 was rated as 'good'.

The second and following levels of evidence include primary studies. However, not all the primary studies listed in Appendix I would be included in the process of assessment of the evidence. As mentioned above, only the following kinds of studies would: 1. studies published after the review, 2. studies from low- and middle-income countries or 3. studies reporting results on drug outcomes where reviews don't. Of the primary studies listed in Appendix I, only the following meets these criteria: Reynolds AJ, Temple JA, Ou SR, Arteaga IA, White BA. School-based early childhood education and age-28 well-being: effects by timing, dosage, and subgroups. *Science* 2011; 333: 360–64. According to Annex IV, this study was assessed to be 'acceptable'.

The content of the studies assessed to be 'good' or 'acceptable' is summarised in Annex V by intervention/ policy. Under 'early childhood education', the following are listed: D'Onise et al 2010 and Jones 2006. Why is Jones 2006 there, and Reynolds et al 2011 is not?

Let us consider the case of Jones 2006. There are many reviews that look at different interventions or policies, and the results are reported separately under each relevant intervention and policy. In Appendix 1, these reviews are typically listed under 'Many settings' and/or 'Many interventions'. According to Annex IV, Jones 2006 was also rated as 'good'.

The case of the primary study Reynolds et al 2011 is different. To determine the level of evidence provided by this study, it is necessary to go back to what kind of intervention this is: is this an intervention for which a randomised controlled trial is possible in principle? Yes, it is (see Annex III). Therefore, the second level of evidence for this kind of intervention is constituted by Randomised Controlled Trials (see Flowchart 2). Reynolds et al 2011 is a matched-group controlled trial (see either Annex IV and V), therefore it is a primary study providing a third level of evidence (see Flowchart 2). Reynolds et al 2011 was assessed to be an 'acceptable' study (see Annex IV). Unfortunately, the evidence provided by an 'acceptable' study of third level is not to be included (see Table 1). That is why the findings of Reynolds et al 2011 are not reported in Annex V, even if the study had been included in the process of assessment.

Therefore, the evidence supporting 'early childhood interventions' is based on the findings of two reviews that were rated as 'good' that report: good results for drugs, mixed results for alcohol, overall positive results for tobacco, and good

results for mediating variables, both in the short and in the long term. This information is therefore summarised in the main text to say that early intervention 'can' prevent the use of drugs and other substances. The reason why it is mentioned that early intervention 'can' prevent is that the reviews reported positive findings that were not perfectly consistent with each other. In the cases where an intervention or a policy is based on studies reporting positive findings that are consistent, the main text reports that the intervention or policy 'prevents'. In contrast, wherever there are reviews reporting inconclusive or mixed findings, the text indicates that the intervention or policy 'may' prevent.

References

- Amato L, Davoli M, Ali R, Auriacombe M, Faggiano F, Farrell M, et al. Cochrane Drugs and Alcohol Group. The Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd, 2005.
- Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
- Baron, R. M., & Kenny, D.A.. (1986). The moderator-mediator variable distinction in social psychology research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51,1173-82.
- Farrington, D.P. (2003). Methodological quality standards on evaluation research. *The Annals of the American Academy of Political and Social Science*, 587, 49-68.
- Flay, B.R., Biglan, A., Boruch, R.F., Gastro, F.G., Gottfredson, D., Kellam, S., Moscicki, E.K., Schinke, S., Valentine, J.C., & Ji, P. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151-159.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, for the GRADE Working Group. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-926
- Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
- McGrath Y, Sumnall H., McVeigh J, Bellis M. Drug use prevention among young people: a review of reviews. 2006. National Institute for Health and Clinical Excellence.

Valentine, J.C., Biglan, A., Boruch, R.F., Castron, F.G., Collins, L.M., Flay, B. R., Kellam, S., Moscicki, E.K., & Schinke, S.P. (2011). Replication in prevention science. *Prevention Science*, 12, 103-117