PROJECT CT.00.EP.14

TECHNICAL IMPLEMENTATION AND UPDATE OF THE EUROPEAN UNION DATABANK ON NATIONAL POPULATION SURVEYS ON DRUG USE AND CARRYING OUT A JOINT ANALYSIS OF DATA COLLECTED

FINAL REPORT

(revised version)

Reported by:

RUUD BLESS
(Quinx Research)

in collaboration with:

JULIAN VICENTE, RICHARD HARTNOLL, NORBERT FROST (EMCDDA)

HILARY BEEDHAM, The Data Archive, University of Essex
(Creation of the Databank)

DIRK KORF (Bonger Institute of Criminology, University of Amsterdam)
LUDWIG KRAUS (Institut für Therapieforschung, Munich)
(Joint Analysis)

December 2002

Please use the following citation:

European Monitoring Centre for Drugs and Drug Addiction. Technical Implementation and Update of the European Union Databank on National Population Surveys on Drug Use and Carrying Out a Joint Analysis of Data Collected. EMCDDA project CT.00.EP.14, Lisbon, EMCDDA, December 2002.

Contact Details

European Monitoring Centre for Drugs and Drug Addiction
Rua Cruz de Santa Apollónia 23/25
1100, Lisboa
Portugal

Further copies of this report can be obtained from the EMCDDA at the above address.

CONTENTS

PREFACE BY THE EMCDDA

Since 1996 the EMCDDA has been developing several epidemiological indicators to assess the drug situation in the EU. Five of these indicators have been considered as “Key Indicators” (prevalence of drug use among the general population, prevalence of problem drug use, drug treatment demand, drug-related deaths and drug-related infectious diseases), although others may be developed and adopted in the future.

The key indicators have been developed with the assistance of contractors and groups of experts form all Member States. Guidelines for the five key indicators were endorsed by the Scientific Committee of the EMCDDA in December 2000, and adopted by the Management Board of the Centre in September 2001. The guidelines adopted by the Management Board reflect the state of the art at that moment, but it is expected that they will continue to be improved.

The indicator “Prevalence and patterns of drug use among the general population” is based on population surveys of representative samples of the general population. The guidelines for this indicator include a set of common core items (the European Model Questionnaire –EMQ-) to include in national questionnaires or to report information from national surveys and general methodological recommendations. The guidelines were developed through several consecutive projects and are compiled in a Handbook that can be downloaded from the EMCDDA website[1]

In addition to the development of guidelines for the key indicator, two consecutive EMCDDA projects have outlined and started a system for collecting and harmonising existing national surveys on drugs, following the European Model Questionnaire (EMQ). This system has been called “European Union Databank on National Population Surveys on Drug Use” (NPSD-EU or “EMCDDA Databank on Surveys”). The mentioned two projects were the CT.99.EP.08 and the CT.00.EP.14, which can be considered as two consecutive phases of the same developmental process.

The report presented here (Project CT.00.EP.14) describes the progress obtained until August 2002 regarding the implementation of the Databank, including the harmonisation procedures of the national databases available at that moment. The report also presents two cross-national joint analyses carried out on the basis of harmonised database created during the project (age of first cannabis use and gender gap in drug use).

During the initial developmental phase of the Databank, the data and documentation were located at the UK Data Archive (British national data archive for social sciences). Once this phase was completed, data and documentation were transferred to the EMCDDA premises. Further development of the Databank will continue at the EMCDDA, with incorporation of new national databases as they are made available by countries. This change on location does not affect neither procedures for data depositing nor the existing Licence Agreements between data owners and the EMCDDA.

Most data depositors chose to use the standard Licence Agreement developed within the project, based on the experience of the UK Data Archive. One country decided to proceed through exchange of letters, as a more simple procedure from an administrative point of view. Standard Licence Agreements are preferable as they make more explicit all the conditions and rights of each party but, if necessary, other forms of agreement could be considered.

The minutes of the expert meeting of May 2002 have been included as an annex to the report. During this meeting progress on the key indicator and on databank an joint analysis were reviewed, and minor modifications of the EMQ agreed.

Julian Vicente
EMCDDA Project Manager
December 2002

[1] http://www.emcdda.org/situation/themes/drug_use_general_population.shtml

INTRODUCTION

This Final Report accounts for the achievements of the agreed objectives of project CT.00.EP.14. The four chapters of the report correspond to the four objectives of the project.

The first objective of the project is the technical implementation of the European Databank of National Population Surveys on Drug Use (NPSD-EU), which concept had been developed in project CT.99.EP.08 “Creation of a European Databank on Population Surveys on Drug Use and Joint Analysis of Data Collected”. It was originally intended to embed this databank in a type of computer based archive environment that is commonly applied in the social sciences for storing, accessing and retrieving of survey data. For this reason a trial database had been developed and installed by the end of 2001 at The Data Archive of the University of Essex (United Kingdom), awaiting a decision by EMCDDA about its future. In the meantime EMCDDA has decided to incorporate the NPSD-EU databank into a general EMCDDA-based archiving system that also contains data on other indicators of drug use and drug problems. As this system will be developed and managed by EMCDDA, the previously elaborated structure of the NPSD-EU has become obsolete. The EMCDDA-based system is still under construction. The final report of CT.00.EP.14 therefore only contains an adapted description of the structural elements of the prevalence databank and the content of the databank realised and delivered by August 2002 under the terms of contract.

The second objective of the project is a joint analysis of the data collected during the creation of the databank. Several proposals for joint analyses were presented in the final report of project CT.99.EP.08. The discussions about these proposals resulted in a two studies. The first, “Analysis of Age of First Cannabis Use in Germany, Greece And Spain” has been carried out by a research team coordinated by Ludwig Kraus, the second study, “Drug Use and the Narrowing Gender Gap”, has been carried out by a research team coordinated by Dirk Korf. Both are presented as separate papers within this report.

The third objective of project CT.00.EP.14 is the re-assessment of the core items of the European Model Questionnaire (EMQ) and the recommendations about data collection methodology, which were presented in the final report of project CT.99.EP.08. The re-assessment resulted in two proposals for amendment of the EMQ - adding “age of first use” as a core item for all drugs covered by the EMQ and different categories for last month frequency of drug use – and in an elaboration of good practise in sampling procedures. The new text on sampling has already been incorporated into the final report of project CT.99.EP.08 B.

The fourth objective of the project concerns the contractor’s assistance to EMCDDA in the implementation of the key indicator on general population prevalence. This report presents a summary overview of the assistance activities.

Ruud Bless

Project coordinator

Quinx Research
In de Poldermolen 3
1115 GR Duivendrecht
The Netherlands

IMPLEMENTATION OF THE NPSD-EU DATABANK

1.1. STRUCTURAL ELEMENTS

Introduction

This section gives an overview of the structural elements of the NPSD-EU. These elements had first been elaborated in project CT.99.EP.08 in the framework an existing computer based archiving system for social research data. As in the meantime the decision has been made to incorporate the NPSD-EU in a dedicated database of indicator data at EMCDDA, the previous descriptions needed to be adapted. As mentioned in the Preface, the revision presented below does not include the new database structure as this is still under construction and was not included in the tasks of project CT.00.EP.14.

Aims

The NPSD-EU is a database of data based on surveys about the prevalence of drugs among the general population, created by EMCDDA as an instrument to,

- Collect existing national survey data;
- Harmonise existing survey data with regard to common core items and variables to improve comparability;
- Support the implementation by EU Member States of the key indicator “drug prevalence among the general population”;
- Promote the harmonisation of future drug prevalence surveys with regard to common core items and categories as well as common methods and standards for data collection;
- Stimulate and facilitate scientific research on the extent, patterns and trends of the use of illicit drugs in the European Union.

Concept

The concept of the NPSD-EU is based on the following principles:

- Acquisition of survey data on the basis of formal License Agreements with data owners to safeguard copyrights and other interests of the data owners and the researchers involved in the surveys;
- Access to data for secondary analysis on the basis of User Agreements that prevent misuse of data and infringements of copyrights;
- Storage and preservation of data with high quality techniques to ensure platform independent availability for future use;
- Creation of data subsets and joined datasets of comparable core item derived from deposited survey data.
- Active involvement of the creators of the deposited data and the Expert Group on Population Surveys in the management and further development of the NPSD-EU.

Datasets and documents

The NPSD-EU will contain the following datasets and documents.

Datasets:

- **Survey data**: original data from prevalence surveys in EU Member States, deposited on the basis of License Agreements between EMCDDA and the data owners;
- **Derived datasets**: subsets of deposited survey data, created by EMCDDA, containing harmonised core items and categories;
- **Joined datasets**: datasets created by EMCDDA by joining together derived datasets based on different survey datasets;
- **Aggregated data**: aggregated prevalence data by gender, age groups and drug types, corresponding to the standard table formats which are annually reported to EMCDDA by the REITOX Focal Points.

Documents:

- **Survey questionnaires**: questionnaires pertaining to the deposited survey data;
- **Survey documentation**: technical reports and other documentation pertaining to the deposited survey data, providing information about sampling frames, sampling methods, weighting procedures, interview modes and other operational aspects of the survey and the construction and content of the survey datasets;
- **Harmonisation documentation**: information about the procedures and programming routines applied in the creation of harmonised datasets.

Operational procedures

Data depositing

Data owners will be regularly invited to deposit prevalence survey data and documentation in the NPSD-EU. Depositing will be based on a *License Agreement* between the depositor and EMCDDA. In the agreement the depositor specifies the conditions of access and use of the data. The License Agreement ensures that the copyrights of the original data rest with the depositor and that the depositors acquire a copyright jointly with EMCDDA in the datasets that are created within the NPSD-EU from original survey data.

The License Agreement includes a Deposit Form that specifies the data to be deposited. A model of the agreement was included in the final report of project CT.99.EP.08.

Data owners can deposit complete survey datasets or partial datasets. Partial datasets might exclude survey data that are not related to the core items of the European Model Questionnaire or not relevant for the further development of these core items. The documentation of partial datasets should however give a full account of the survey questionnaire and survey processes.

Cataloguing and storage of datasets and documentation

EMCDDA will continue to develop in consultation with the Expert Group on Population Surveys appropriate procedures, methods and technical facilities for the storage, cataloguing and management of the datasets and supporting documentation of the NPSD-EU. These procedures will be consulted with those countries that deposit data from their national surveys.

The procedures proposed in the Final Report of project CT.99.EP.08 were based on the practice of UK Data Archive, based on the software and facilities developed at that institution acting as British national data archive for social sciences. The procedures that will be developed by the EMCDDA will be simpler and adapted to the specific needs of the NPSD-EU.

Creating derived and joined datasets

Derived and joined datasets containing common core items and categories will be created by EMCDDA from the deposited survey data. The permission to do so is explicitly stated in the License Agreement with the depositors. Derived datasets will include as far as possible the core items of the EMQE. Joined datasets are combinations or derived datasets based on survey data of different years and/or countries.

The specific manipulations required to create derived datasets will depend on the characteristics of each individual survey, considering the general guidelines outlined in the next section, and will be elaborated for each deposit separately in consultation with the research team that has carried out the original survey in order to guarantee the integrity of data.

The manipulations on the survey data deposited by August 2002 have been carried out by the contractor of CT.00.EP.14. The guidelines for harmonisation are based on this experience.

Data entry of aggregated data

EMCDDA has developed a format for entering aggregated data from standard indicator tables of prevalence survey data, which are reported by the REITOX Focal Points, into a dedicated database developed in MS Access. This data entry format has been evaluated and successfully tested in the framework of project CT.00.EP.14.

In principle aggregated data can also be generated from deposited datasets. For the time being however, the aggregated data in the NPSD-EU will be entered separately as data depositing is a voluntary decision of data owners, whereas the reporting of aggregate prevalence information is regulated in the relations between EMCDDA and the REITOX Focal Points. As a consequence the database of aggregated data relates to more surveys than have been deposited in the NPSD-EU.

Access to the NPSD-EU

Access to data

The access to the actual data in the NPSD-EU will be governed by the License Agreement between EMCDDA and the original data depositors. Unless data depositors allow access to anyone, data will only be accessible for registered users who have acquired the necessary permission and signed the *User Agreement* conforming to the conditions specified in the License Agreement pertaining to the data. EMCDDA will develop appropriate modalities to implement accessibility whenever granted. This can vary from data on transferable media to on-line electronic access.

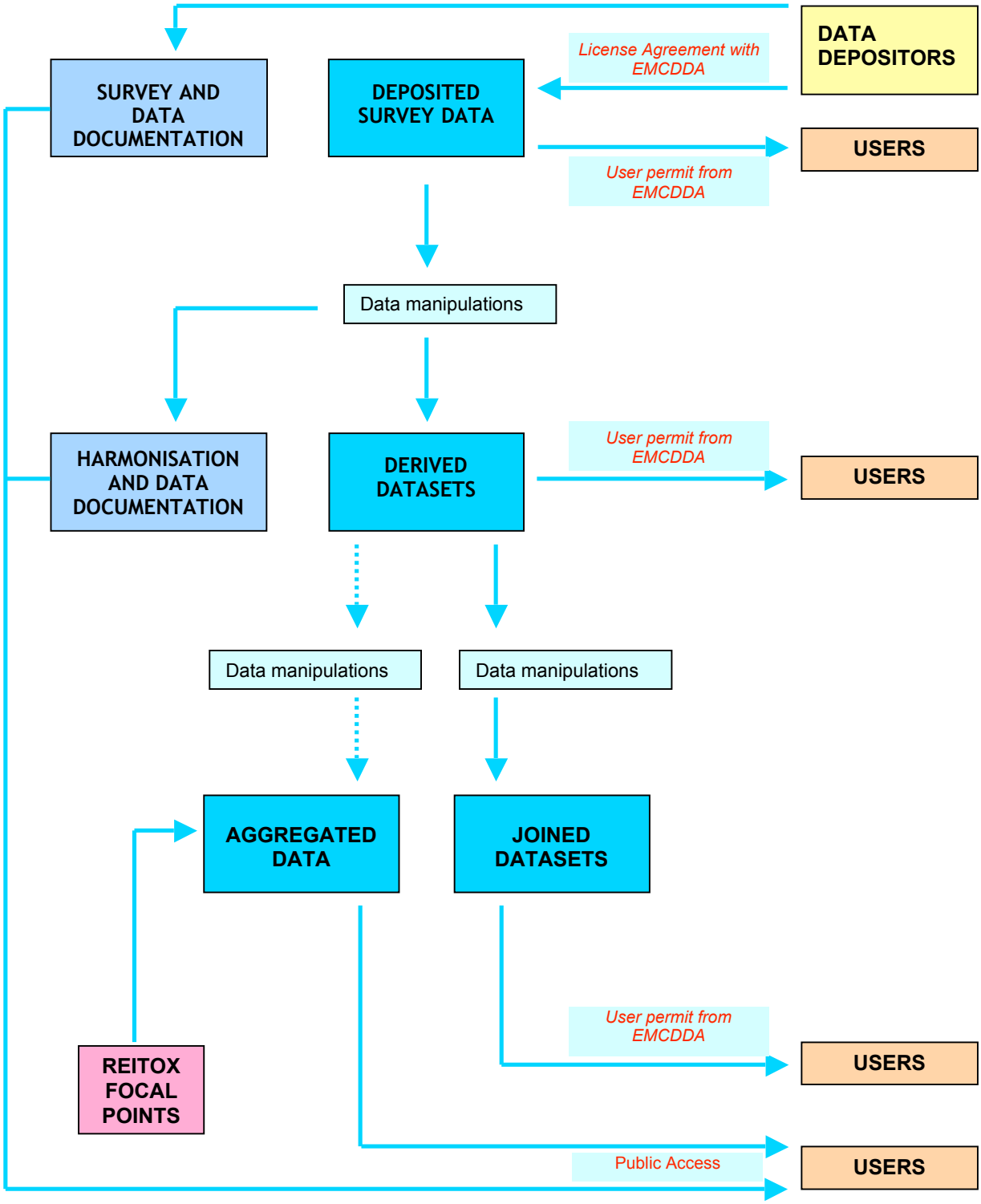
Access to documentation

Documentation relating to the deposited survey data will accessible on the website of EMCDDA. In principle there will be public access to this documentation. EMCDDA will develop appropriate facilities to search within the documentation.

Languages in the NPSD-EU

As a general rule the NPSD-EU will store all documentation supplied by depositors in their original language and all information created from these originals by EMCDDA in English. This means that survey data, questionnaires and documentation will be held in the original languages or the translation provided by the depositor, whereas derived, joined and aggregated data and documentation will be available in English.

Summary model of the NPSD-EU



1.2. CONTENT REALISED AND DELIVERED (August 2002)

Introduction

In this section we first give an overview of the process of data acquisition and the survey data and documentation that have been deposited by August 2002. We then describe the process of creating derived datasets and present an overview of the harmonised and joined datasets. Exhaustive descriptions of the survey data deposited and the derived and joined datasets are included in the electronic files (SPSS system files) pertaining to these datasets which have been delivered by the contractor to EMCDDA. The documentation –questionnaires, technical reports - related to the survey data has been delivered as paper copies or in electronic format as far as available and is not included in this final report.

Data acquisition

The acquisition of data has proven to be a time consuming and sometimes frustrating process, which caused much delay in the achievement of the project objectives.

- The optimistic time plan of project CT.00.EP.14 was based on the positive experience of data collection in the framework of the preceding project CT.97.EP.08 “Co-ordination of an Expert Working Group to develop instruments and guidelines to improve quality and comparability of general population surveys on drugs in the EU”. However, it was not fully taken into account that this acquisition regarded a single specific use of the data based on a gentlemen’s agreement with the researchers that created the datasets and participated in the Expert Group, whereas the deposit of data in the NPSD-EU requires formal procedures, the involvement of formal data owners and relates to non-specific multi-use of the data in the future.
  Although no one disputes that a legal cover of data depositing is a necessary safeguard for the rights of data owners, the proper use of data and a sustainable databank, the formal process required much time and effort to explain the concept and purposes of the databank to third parties, in particular where data owners are not related to the field of drugs or the National Focal Points.
- Depositing data in a databank requires detailed documentation of data and surveys and is more demanding than the traditional exchange of data for single use between cooperating researchers. The process of depositing calls for additional work by the data owners and producers in retrieving data from internal archives, searching and selecting documentation and providing basic descriptions of the deliverables. Moreover, in the case of past surveys the necessary documentation is not always completely anymore or not available in a ready-to-use electronic format.
- For obvious reasons survey data are not made available to others before they have been exploited by the researchers that developed the survey. In general data are not available for deposit within one year after the completion of the survey field work, but in some cases this period of exclusive use is extended. By the closing of the project in August 2002 we had not been able to acquire data of surveys carried out after 1999. The exception is the German survey of 2000, but the formal deposit procedures were still not completed in August 2002. This survey was used in the joint analysis on the grounds that the German national expert for the EMCDDA indicator (and responsible for the German survey) was member of the project analysis team.
- Some data owners resist the idea of depositing data in the NPSD-EU. In the case of one country, the reason is that the data owner considers the concept of a drugs databank premature as long as drug prevalence surveys across Europe are not based on a comparable methodology (in the sense of a relatively identical methodology across countries). In other cases however the arguments against depositing could not be clarified.

By the closing of the project data of eleven surveys from four countries have been deposited. In nine potential cases the acquisition efforts failed. The table below gives an overview of the state of affairs by the end of 2001.

Table 1: Overview of the data acquisition process

Austria	No national surveys available yet.
Belgium	No national surveys available yet.
Denmark	The first national general population survey that includes relevant prevalence data has been executed in 2000; the data were not yet available. It is expected however that they will be deposited upon request.
Finland	In principle there are three relevant surveys (1994, 1996, 1998). The data owner (Statistics Finland), carried out a consultation with its data protection committee that took considerable time. In the meeting of the Expert Group of may 2002 it was informed that the committee had given a favourable opinion to data deposing. This conclusion can be regarded as positive for the NPSD-EU concept. It is expected that actual deposit will be done in the near future.
France	In principle there are two relevant surveys (1995, 1999). The data owner (CFES) did not agree to deposit the data, despite interventions of the French Focal Point that co-owns parts of the survey data. It remains unclear if this position will change in the future.
Germany	The data of the prevalence surveys of 1995 and 1997 have been deposited. The formalities regarding the deposit of the 2000 survey were not concluded before the closing of the project, but as we could use the 2000 data for the joint analyses (see previous page), the 2000 data have already been included in the harmonisation process. It is expected that future surveys will be deposited upon request.
Greece	Survey data of 1993 and 1998 have been deposited. Although the 1993 survey was limited to the Athens metropolitan region, it is be included in the NPSD-EU as it was the precursor of the national survey of 1998 and the 1998 survey allows a substantial extraction of Athens data for comparison with the 1993 data. It is expected that future surveys will also be deposited upon request.
Ireland	There has been a national survey with some prevalence questions in 1998 and a trial drug prevalence survey in 1999. The Irish Focal Point considered the quality of the 1999 survey disputable and not fit for deposit, whereas the scope and context of the 1998 survey is not compatible with the requirements of the NPSD-EU. It is expected that future surveys will be deposited upon request.
Italy	No national surveys available yet. A first national survey has been carried out in 2001. It is expected that this survey will be deposited in due time.
Luxembourg	There has been a national survey in 1998, but with a limited sample size and not compatible with the demands of the NPSD-EU. A deposit is not pursued.
Netherlands	Relevant surveys have been carried out in 1997-98 and 2000. The 2000 data were not yet available. Although the data owner (CEDRO) has indicated that there are no in principle objections to deposit, discussions about the license conditions did not result in a positive decision.
Portugal	No national surveys available yet. A first national survey has been carried out in 2001 and it is expected that this survey will be deposited in due time.
Spain	The survey data of 1995, 1997 and 1999 have been deposited. Spain has also formally deposited national (non-ESPAD) school survey data; these are not processed in the framework of the current project. It is expected that future sweeps of the Spanish surveys will also be deposited upon request.
Sweden	There are relevant surveys of 1995, 1997 and 1999. The data owner (Ministry of Public Health) has decided however that the Swedish data will not be deposited in the NPSD-EU as long as European prevalence surveys are not comparable in methods and standards (in the sense of similar) of data collection.
United Kingdom	Relevant survey data of 1994, 1996, 1998 (British Crime Survey) covering England and Wales have been deposited. It is expected that future sweeps will also be deposited upon request.

Survey data and documentation deposited

The table below lists the surveys that have been deposited and processed in the NPSD-EU by the end of August 2002. Full details of the surveys can be found in the corresponding electronic data files and related documentation.

Table 2: Deposited survey data and documentation

Country	*Name*	*Year*	*Net response*	*Documentation*
England-and-Wales	British Crime Survey	1994	11693 [1]	Technical Report, incl. Questionnaire
		1996	11244 [1]	Technical Report, incl. Questionnaire
		1998	10294 [1]	Technical Report, incl. Questionnaire
Germany	Repräsentativerhebung zum Konsum und Missbrauch Illegalen Drogen, usw.)	1995	7833	Technical Report, incl. Questionnaire [2]
		1997	8020	Technical Report, incl. Questionnaire [2]
		2000	8157	Technical Report, incl. Questionnaire [2 *]
Greece	Population survey on the use of licit and illicit substances in Greater Athens	1993	2103	Part of the questionnaire [3]
	National population survey on the use of licit and illicit substances	1998	3752	Questionnaire [4]
Spain	Encuesta Domiciliaria de Consumo de Drogas	1995	9984	Technical Report, incl. Questionnaire [5]
		1997	12515	Technical Report, incl. Questionnaire [5]
		1999	12488	Technical Report, incl. Questionnaire [5]

1 Figures refer to respondents aged 16-59. Respondents of 60 years and older had not been asked to complete the drug questions of the British Crime Survey, but they are included in the deposited datasets.

2 Documents only available in German; variables and labels included in the survey data file also in German (*) For the time being, 2000 survey was included for the purposes of the joint analysis.

3 Only part of the questionnaire related to EMQ available, in English; variables and labels included in the survey data file in English. Translations provided by the Greek Focal Point.

4 Survey questionnaire in Greek; variables and labels included in the survey data file in English. Translation provided by the Greek Focal Point.

5 Documents only available in Spanish; variables and labels included in the survey data file also in Spanish.

Creation of derived datasets

Introduction

Derived datasets are created from original survey data by constructing common core items (variables) and categories (values) on those items on the basis of a selection of the original data items and data categories. This construction process we call “harmonisation”. The harmonisation of survey data is based on the following principles:

- Survey variables and values are constructed attributes of or related to respondents, based on the answers that the respondents give to questions in the survey questionnaire.
- Different questions can result in the construction of comparable variables if the phrasing of the questions are the same, if the exact phrasing is not considered to have much effect on the answer, or if combinations of different questions can be transformed into a common variable.
- Different answer categories can result in the construction of comparable categories if the wording of the categories are the same, if the exact wording of a category does not affect its semantic meaning, or if different categories can be combined into comparable categories.

It should be noted that data harmonisation only means the creation of comparable variables and categories from different survey data. It does not imply that the output figures of harmonised datasets are also comparable between different surveys, as the magnitude of these figures might have been affected by differences in sampling procedure, interview mode, questionnaire context or the cultural context in which the survey has been carried out.

Implementation

The actual creation of derived datasets in the framework of project CT.00.EP.14 has been performed in a few consecutive steps. All data manipulations were performed using SPSS 10.0 and the resulting datasets are stored on the report CD as SPSS system files.

Step 1

At first all items and categories that we expected to create from the original data were added with compute statements to the original survey datasets with a uniform value of 7777 for all cases. The new items comprise the core items of the EMQ, but also other items that are we considered relevant for the future development of the EMQ or for future analysis of survey data. The value 7777 is an arbitrarily chosen code and indicates that the variable has not (yet) any valid categories.

The procedure above is only implemented for efficiency reasons. By adding all variables that we attempted to harmonise to every original datasets, no matter if this could be effectuated or not, we did not have to declare in the next steps variable and value labels for each dataset separately. Also creating derived datasets which at least in name contain all variables and values allows future users of the datasets to develop standard syntaxes for data manipulation and statistical procedures which can run on each file. If code 7777 is a priori declared as missing, such standard syntaxes will not produce errors but will simply exclude the not really existing variables as all cases have a missing value.

To allow an easy control of the data transformations in the harmonisation process we made undone any missing value declaration in the original data and included weight factors were turned off.

Step 2

Within the extended datasets the intended harmonised variables and values have been actually created by compute, recode and other transformation statements on the existing items and categories in the survey data. When applicable standard codes were assigned for values that in future use of the data might be declared as missing, according to the list below, and inconsistencies in the responses have been corrected as far as possible.

Step 2 resulted in extended survey datasets, which contain both the original data and the new harmonised variables and categories. On the report CD these datasets have the term “ADD-ON” in the filename. These “add-on” files can be used for alternative data manipulations or the creation of new harmonised items if such might be needed in the future.

Step 3

According to the procedures above a question that for a particular case has been skipped due to a preceding filter question results in a variable with code 8888. In the source data the value is usually left blank at data entry and specified as “system missing” in the SPSS file. Treating such cases as missing is however not correct: not the answer is missing, but the question has not been asked or should not have been asked.

For the prevalence questions which are considered the core items of the datasets, we have applied a reinterpretation of these skipped questions on the basis of the answers given to the preceding filter questions. For example, if life-time prevalence LTP=2 (“no”) and last-12-month prevalence LYP=8888 (“question skipped”), LYP is recode into LYP=2, etc. Obviously, this reinterpretation has not been implemented when the original survey questionnaire does not require to skip questions, as is the case in the Spanish surveys.

In step 3 we also calculated prevalence measures regarding “any” illicit drug. In the transformations the prevalence of “any illicit drug” is for each survey calculated as “any of the drugs mentioned in the survey”. This implies that the concept can have a different meaning across survey datasets as each survey has it’s own set of drugs covered.

At the end of step 3 the original survey data were removed from the extended add-on file, resulting in a dataset that only contains harmonised variables and values, and the applicable weight factor is turned on again. On the report CD the resulting datasets have the term “HARMONY” in the filename. As mentioned above the “harmon” files can also contain non-existing harmonised variables with the uniform value 7777 assigned in step 1. For convenience we have also added for each survey dataset a files without these obsolete variables; on the report CD these have the term “HARMON” in the filename.

The transformation syntaxes applied to each deposited dataset are also stored on the report CD.

In Annex 1 we give an overview of all harmonised variables and values and their correspondence to the EMQ. The overview includes the amendments to the EMQ adopted by the Expert Group on Population Surveys in May 2002.

General rules applied in the harmonisation process

Reduction to ordinal scales
In order to allow and improve a match between the EMQ and the derived datasets we have in several cases reduced the information content of the original source data to simple ordinal scales (e.g. low - medium - high). The scale values might correspond to different cut-off points or combinations of categories in the original datasets.

Missing values
In many cases we were unable to assess how missing values have been interpreted and managed in the original surveys. Even if questionnaires distinguish between categories like “don’t know”, “no answer” or “refuse to answer”, the distinctions are not always kept in the survey dataset and we don’t know how and to what extent data entry errors had been corrected or replaced by declared missing values. Many survey datasets also do not discriminate between real item non-response, i.e. the respondent gives no (valid) answer, and item non-response caused by preceding filter questions, i.e. the respondent was not supposed to answer.

As far as possible the following codes for “missing values” were applied, whereby it should be noted however that in the derived datasets no values are actually declared as missing values.

Table 3. Missing value codes

Value	***Label***	**Description**
7777	*Does not exist*	Assigned to all cases of a harmonised variable, which cannot be constructed from the source data. The variable does not actually exist. Variables with this uniform value appear only in the intermediate “add-on” and “harmon” working files of the harmonisation process. For analyses the value 7777 should not be declared as missing.
8888	*Question skipped*	Assigned to cases of variables that have been skipped due to the

		responses on one or more preceding filter questions. The value is first created as a replacement of all original “system missing” values and in the harmonisation process to label skipped questions. In step 3 of the harmonisation process 8888 values of prevalence variables have been reinterpreted consistent with the preceding filter questions.
8889	*Question skipped due to cluster skip of a set of questions*	This code applies when a whole section of the questionnaire, containing several questions, has been skipped on the basis of a preceding filter question. In this project it applies only to the British Crime Survey, where respondents could skip all drug questions if they refused self-completion of the drug section. Code 8889 is not re-interpreted or recoded in the final harmonised files.
9997	*Cannot be assessed*	Assigned to constructed variables which value could not be assessed by the logical operations on existing survey variables.
9998	*Refused to answer*	Some surveys discriminate for some questions between “no answer” or “don’t know” and “refuse to answer”. As far as the source data allow this distinction is retained in the harmonisation process.
9999	*No valid answer*	Answers not corresponding to one of the valid categories of a variable and not corresponding to any of the “missing” codes above. For lack of underlying data this item non-response is not further differentiated.

Corrections for inconsistencies
Respondents are not always consistent in their answers to questions and people who construct datasets from completed questionnaires can also make errors. In both cases we may end up with inconsistent data and in general we don’t know how and to what extent the deposited data have been cleaned or corrected for inconsistencies. Pen-and-paper interview modes will result in more inconsistent data than computer aided modes, where the software can prevent inconsistencies by accepting only selected codes and guiding the interviewer or respondent through the questionnaire.
In general there are two methods to correct for inconsistencies.

- Inconsistent answers are considered as the result of human errors in the assignment of values to a variable. If so, the “real” value is unknown and the value should be considered as “missing”.
- Inconsistent answers originate from the fact that not everyone might understand the structure or instructions of the questionnaire, which can imply that people tend to skip questions, which at first sight do not seem to relate to them. If so, inconsistencies can sometimes be corrected by a logical reinterpretation of the answers on preceding or following questions.

In the harmonisation process we have chosen for the second approach with regard to all prevalence questions and the first method in other cases. The general rules that we followed are listed below:

IF LMP_… = 1 (yes) → LYP_… = 1 (yes)

IF LYP_… = 1 (yes) → LTP_… = 1 (yes)

IF AGE_.. = > 1 and < 97 → LTP_… = 1 (yes)

IF LTP_... = 9998 (refused) and LYP_… ≠ 1 → LYP_… = 9998 (refused)

IF LYP_... = 9998 (refused) and LMP_… ≠ 1 → LMP_… = 9998 (refused)

IF LTF_.. = > 0 and < 9997 → LTP_… = 1 (yes)

IF LMF_.. = > 0 and < 31 → LMP_… = 1 (yes)

The application and logical order of the statements depends on the structure of the questionnaire concerned and was adapted accordingly. For the exact transformations pertaining to each dataset one should consult the SPSS syntaxes on the report CD.

Although the total number of corrections for inconsistencies is often not very high, the effect on prevalence and continuation rates for individual drugs can be relatively substantial when we deal with very low figures in the general population.

Weight factors
Derived datasets are stored as weighted datasets and all cases include the weight factor pertaining to the survey source data. There is no common practice regarding the total figures that application of weight factors should return. In some deposited datasets weighting returns population totals, in others weighting returns the totals of the net response. As we prefer the second method original weight factors have sometimes been adapted so that the weighted files return the same totals and the non-weighted files would do.

It should be noticed that joined datasets, based on weighted derived datasets, can only be used for year-by-year or country-by-country comparisons. Aggregating across survey years or countries would require to calculate new weight factors based on the aggregated populations.

Case numbers
The original case identifying numbers of each dataset are preserved in the resulting derived datasets as the variable CAScyy, where cc corresponds to a two-digit country code and yy to the year of survey.

Alcohol
Some surveys measure prevalence for different types of alcohol separately, without a combined category “any alcohol”. In such cases the prevalence of (any) alcohol has been assessed as the prevalence of either alcoholic drink covered.

If frequencies of drinking are assessed separately as number of times or days drinking each different type of alcohol, frequency of drinking any alcohol has been defined as the maximum frequency of any of the listed drinks. It should be noted that this might have resulted in an underestimation of the real frequency of alcohol use.

Illicit drugs
Some surveys include other drug types than specified in the EMQ and some surveys include drug classes that comprise several drug types. In the harmonisation we have retained the differentiations of the original surveys. As a result the derived harmonised datasets have more specifications by drugs than the EMQ.

Any drug
Although a question about the use of any drug is not part of the EMQ, we did include the item in the harmonisation process because in some countries a question about the use of any illicit drug might be used as a primary filter (though not in the datasets deposited at this moment) and the concept of “any drug” can facilitate the investigation of multi-drug use. In the assessment of “any drug” we did not include methadone and “other opiates” as these may refer to licit use of drugs prescribed by a doctor.

Introductive question
The EMQ includes for each drug type the introductory question “Do you personally know someone who takes (a specific type) drug”. This introductory question is not used in any of the deposited prevalence surveys. Instead we have added the more common introduction “Have you ever heard of (substance)”, which appears in the surveys from Greece and England and Wales. In the case of the British Crime Surveys (England and Wales) this question is also used as a filter for the prevalence questions.

Pharmaceuticals
In the EMQ pharmaceutical drugs are included as the combination “sedatives and/or tranquillisers”. The combined category does not exist in the deposited datasets. Instead we have included common variables for sedatives and tranquillisers separately. Even if these substances are covered in the original surveys, the harmonisation attempt cannot always be considered satisfactory as the concept and the measurement of the use of sedatives and tranquillisers are not always comparable between different surveys.

Opinions
Questions about opinions are only incorporated in the derived datasets if the original surveys contained questions that read (almost) the same as those proposed in the EMQ.

Respondent attributes

Basic attributes like age and gender are easy to harmonise across surveys but for other attributes the harmonisation should only be considered as a tentative exercise as concepts, definitions and actual practice differ considerably between countries, which prevents reliable transformation into common variables and values. In many cases even the limited categories of the EMQ could not be constructed for all cases of the original survey data, which results in relatively large numbers of cases where the value of the variables had to be set to “unknown” or “cannot be assessed”.

Harmonised derived datasets

All deposited survey datasets have been entered into the harmonisation process. In Annex 2 we present a summary overview of the results of the exercises for each dataset, indicating the level of comparability (good or moderate) across different datasets. When comparability is moderate, one should always consult the original questionnaires and datasets as well as the exact transformations that have been applied.

2. JOINT ANALYSIS

Introduction and acknowledgements

The final report of project CT.99.EP.08 presented several suggestions for a joint analysis of the data to be deposited in the NPSD-EU. The implementation of some of these proposals was included in the tasks of project CT.00.EP.14. One important reason for combining the implementation of the databank with analytical exercises was that this can demonstrate the potential of the databank as a resource for epidemiological research which in turn might result in more support for and depositing of survey data in the NPSD-EU.

The proposals for joint analysis have been discussed with Expert Group on Population Surveys and in consultations between EMCDDA and the contractor two separate studies have been chosen for implementation under the terms of contract of project CT.00.EP.14.

The study "*Analysis of Age of First Cannabis Use in Germany, Greece and Spain*" has been carried out by Ludwig Kraus and Rita Augustin of the Institut für Therapieforschung, Munich, Germany, and the study "*Drug use and the narrowing gender gap*" has been carried out by Dirk Korf and Annemieke Benschop of the Bonger Institute of Criminology at the University of Amsterdam, the Netherlands. Both studies are reported in the next sections.

The contractor and the researchers wish to acknowledge the contribution of others to the concept and realisation of these studies. They are:

- The researchers responsible for the original survey data that have been used in the studies: Manina Terzidou (Greek surveys), Jacinto Rodriguez-Osuna (Spanish surveys) and Malcolm Ramsay (British Crime Surveys for England and Wales)[1];
- The project leaders of CT.00.EP.14 at EMCDDA: Julian Vicente and Richard Hartnoll;
- The members of the Expert Group on Population Surveys, listed on the cover of this report.

[1] Ludwig Kraus was also the researcher responsible for the German surveys.

2.1. ANALYSIS OF AGE OF FIRST CANNABIS USE IN GERMANY, GREECE AND SPAIN

LUDWIG KRAUS, RITA AUGUSTIN
Institut für Therapieforschung, Munich, Germany

Introduction

Based on results from General Population Surveys (GPS) illicit drug use and in particular the use of cannabis is well documented in the majority of the member states of the European Union (EU). Trend data from these cross-sectional surveys observed a general increase in both lifetime and 12-month prevalence of cannabis use in the general population and particularly among the younger generations in recent years (EMCDDA, 2001, Hibell et al., 2000). Prevalence rates, however, are generally reported by age groups and thus represent the age distribution of the population at a given time. Although these cross-sectional data are feasible to study cohort effects, adequate analyses are seldom carried out. Since most surveys contain a question on the age of first use of a particular substance, cohort effects of drug experience can be analysed by using survival analysis (Singer & Willet, 1993; Singer & Willet, 1991). This technique allows the concurrent analysis of incidence and prevalence. By analysing the event history of onset of cannabis use for specific cohorts two research questions can be tackled: (1) at what age do different birth cohorts commence the use of cannabis and are there differences between cohorts in different cultures, and (2) how did the epidemic of drug use develop in different countries with respect to different generations.

Apart from research interests in cohort-specific cannabis prevalence onset of cannabis use has become a crucial issue. Longitudinal studies point at early onset of cannabis use in men to be associated with a higher risk for the consumption of other illicit drugs (Yamaguchi & Kandel, 1984). Research into transitions between licit and illicit drugs and between different illicit substances were carried out in order to find evidence for the gateway theory (Kandel & Faust, 1975). Support for the theory comes from the finding that adolescents are unlikely to experiment with marijuana without prior experimentation with licit drugs (Kandel, Yamaguchi & Chen, 1992). Furthermore, prior experiences with legal drugs was significantly associated with the early development of cannabis consumption in teenagers (Höfler et al., 1999).

Other research focuses on the relation between age of onset and drug-related harm later in life. Robins and Przybeck (1985) found early onset of drug consumption to be related to a higher risk for drug and alcohol abuse and dependence and antisocial personality disorder. While this was also found for the relation between early onset of regular alcohol use and later experience of abuse and dependency (Kraus et al., 2000), this relation could not be found in a longitudinal study among pure cannabis users. The association between early onset and cannabis-related disorders was not significant (Schumann et al., 2000).

While prior research suggests that patterns of cannabis use are rather unstable, more likely to be experimental (Kandal & Logan, 1984), and involve a high rate of cessation (Johnston, O'Malley & Bachmann, 1992; Silbereisen, 1997), in a longitudinal study Perkonigg et al. (1999) found relatively low remission rates among regular and repeated users, and similarly more continued use among more experienced users. The authors conclude that patterns of cannabis use are rather stable and that higher baseline use is associated with a higher probability of continued and regular use .

Thus, age of onset seems to matter when it comes to progression into repeated, regular and heavy use. Furthermore, the earlier cannabis use starts, the more prone an individual becomes to other risk factors. From a Swiss study Müller and Gmel (2002) report that an early onset was associated both with a higher prevalence of other illicit drugs (e.g. cocaine, ecstasy, hallucinogens), and an early onset

into the use of these drugs. Similarly, cessation of cannabis use was less frequent in early onset users.

Studies on cohort-specific changes in age of onset are scarce. Based on cross-sectional data from several surveys in the general population Kraus, Bauernfeind and Herbst (1998) observed a sharp increase in lifetime experience of illegal drug use in Western Germany in the 1990s after prevalence increased only slightly in the period between 1970 and 1990. The observed increase of prevalence in younger age groups could not be explained with an age-shift of onset of drug use. The risk functions of first drug experience showed an almost proportional increase over time, and their maxima remained approximately constant at the ages 16 to18. Using also survival analysis, a study in Switzerland could show that onset into the use of cannabis has decreased from 17 years in 1992 to approximately 16 years in 1997 (Müller & Gmel, 2002).

Research on age of onset may thus help to determine when and how many of a certain cohort commence with drug use, whether age of onset of drug use has changed over cohort and thus over time, and - from a cultural perspective - whether differences in these patterns can be observed when cohort-specific data on age of onset are compared across countries.

Method

Samples

Germany
In Germany, nation-wide representative surveys on the use of licit and illicit drugs were conducted in 1995, 1997 and 2000. All three surveys are representative of the German-speaking general population aged 18-59 years. A self-administered questionnaire was filled out at home and returned by post. In each survey about 8000 respondents were surveyed. There was no oversampling of certain subgroups.

In 1995 and 1997, a stratified, 3-stage, random sampling design was utilised. In the first stage constituencies were stratified according to region and selected at random. Within selected constituencies points on the city map were chosen at random. Starting from these points every third household was selected. In the last stage the respondent chosen in each household was the individual with the most recent birthday. Employees of the survey firm that was assigned to carry out the field work personally approached subjects at their home, and invited them to participate in a survey on “life and health issues” that had been commissioned by the German Federal Ministry of Health. In both years, the response rate amounted to 65%.

In 2000, sampling design changed to a two-stage design. In the first stage communities were sampled. In the second stage a systematic sample was drawn, using the list of inhabitants of the selected communities. The questionnaire was mailed to the selected addresses and returned by the respondents. The response rate declined to 45%.

Greece
In 1993 about 2000 12-64-year-old inhabitants of Athens were surveyed. Five years later, the survey was extended to all of Greece and the sample size amounted to about 3750. The sample was a multistage stratified probability sample. The four initial strata, defined by degree of urbanization, were Greater Athens, Thessaloniki, other urban areas, and the semi-urban and rural areas together. In the first stage, 15 towns and 30 villages were selected randomly from the third and fourth strata, respectively. In the second stage, city blocks were sampled at random in the first two strata and from the selected towns and villages in the other two strata. In the third stage selected blocks were visited to record all existing households. Household members were then stratified by age (12-17, 18-24, 25-35 and 36-64 years) and sex. In the fourth stage the stratified sample was selected through systematic sampling. The two younger age groups were oversampled.

Substitutions were made from the beginning of the survey in case of change of address, hospitalization, permanent absence, repeated refusals, mental retardation or death. All substitutions were made with that member of the same sex and age group living in the same block whose serial

number was closest to that of the person initially selected, subject to no other member of his or her household having already been selected or used for a substitution.

Spain
Survey data for Spain were available for 1995, 1997 and 1999. Sampling procedures of the 1995 survey, however, differed substantially from the two other surveys, data were thus taken only from the surveys conducted in -1997 and 1999. In both surveys the target population were the 15-65-year-olds in Spain including the municipalities of Ceuta and Melilla on the Moroccan coast. The respondents from these cities were omitted from the present analysis. The sample sizes amounted to about 12,500 in both years. Both smaller Autonomous Communities and the age group 15-39 were oversampled. The surveys were conducted as face-to-face interviews at interviewees' homes with a self-administered questionnaire on the consumption of legal and illegal drugs.

The samples are probabilistic, multi-staged and stratified within each cluster section. The selection of the primary units (municipalities) and the secondary units (sections) – territorial units of approximately 2000 inhabitants – followed a selection with a probability proportional to size design (PPS). The selected sections and the number of interviews to be performed in each section were handed to the fieldwork agency as the interviewer had to perform the next stages of selection. The interviewer had to draw a route through the whole section passing all buildings or blocks of buildings. Households were selected by means of a systematic sampling along the routes. Finally, individuals were selected from households by means of randomly numbered tables.

Analyses

Analysing the age of first cannabis use, a particular missing-data-mechanism has to be considered: First cannabis use of a respondent may occur after the survey and thus age of first cannabis use is unknown. These cases are called "censored". Restriction of the analysis to the non-censored cases leads to systematically downward biased results. Thus, survival analysis techniques were used, which take censored cases into account.

In each country, 5-year-birth cohorts were built. Applying a discrete Cox regression model (Lawless, 1982), it was tested if the responses on age of first cannabis use of the same birth cohort differ significantly between the surveys in the same country. In the case of insignificant differences, 5-year-birth cohorts from different surveys were combined. The same was applied to one-year-birth cohorts. To account for the complex sample design, the Cox regression model was calculated with the program SUDAAN 7.5.2 (Shah, Barnwell & Bieler, 1997).

Age of first cannabis use of the different birth cohorts was analysed using the life table method. Hazard rates estimate the risk of first cannabis use at a certain age. In contrast to the "usual" use of survival analysis techniques, in the context of onset of cannabis use not the proportion of "survivors" (i.e. those who have not yet taken cannabis) but the proportion of "deaths" is of interest. Thus, instead of the survivor function, the empirical distribution function was calculated, which gives the cannabis lifetime prevalence rates at certain ages. As the life table procedure of SPSS ignores case weights, the hazard rate as weighted mean of an indicator variable was calculated. This variable was given the value 1 if first cannabis use occurred at a certain age and 0 if this event occurred at a higher age. If first cannabis use took place at an earlier age the indicator variable was set to a missing value. Of course, for censored respondents all indicator variables are zero as first cannabis use can only happen in the future. On the basis of the hazard rates, all concepts of survival analysis – survivor function, empirical distribution function, cumulative hazard rates – can be calculated (Lawless, 1982).

To analyze cohort effects 5-year-cohorts (1978-82, 1973-77,…,1933-37) were built and empirical distribution functions and hazard rates of age of first cannabis use were compared. To study the historical development of cannabis experience empirical distribution functions and hazard rate for selected one-year cohorts were calculated according to calendar year. Since for most birth cohorts the risk of first cannabis use is close to zero at age 25 or higher ages hazard rates are shown only up to age 25 or age 30.

Results

GERMANY

Distribution of age of first cannabis use

For the five-year-cohorts, no significant differences between the 1995 survey, the 1997 survey, and the 2000 survey were found. Thus, data from all three surveys were pooled.

Figure 1 shows the empirical distribution function for the birth cohorts 1978-82, 1973-77,…,1938-42 in Germany. While in the six youngest cohorts the first use of cannabis commenced at age 12, in older cohorts a later onset is observed. The curves increase rapidly between 15 and 30 years and stabilise after this point. The curves are steeper the younger the respondents are, which implies a nearly monotonic ordering of the prevalence rates: The lowest prevalence rate (1.5%) is observed for the oldest birth cohort and the highest prevalence rate (33%) for the second youngest birth cohort. While the prevalence rate of the youngest birth cohort (30%) is close to the highest prevalence rate, only 25% of the birth cohort 1968-72, which is the third oldest birth cohort, have ever tried cannabis.

The hazard rates are only given up to age 30 as they are close to zero for older ages. Except for the youngest birth cohort and those born in 1963-1967, the hazard rates show a common shape with a maximum at age 18 (Figure 2). The peak of the youngest cohort may, however, shift to the right when forthcoming surveys can be included. In the 1997 and 2000 survey participants of this cohort were aged 18-19 years and 18-22 years, respectively, thus respondents with first cannabis use in young adulthood may be censored. Apart from this statistical artefact, there is a clear tendency to higher hazard rates for younger cohorts at each age.

Figure 1. Empirical distribution function for age of first cannabis use: Germany, 5-year-cohorts

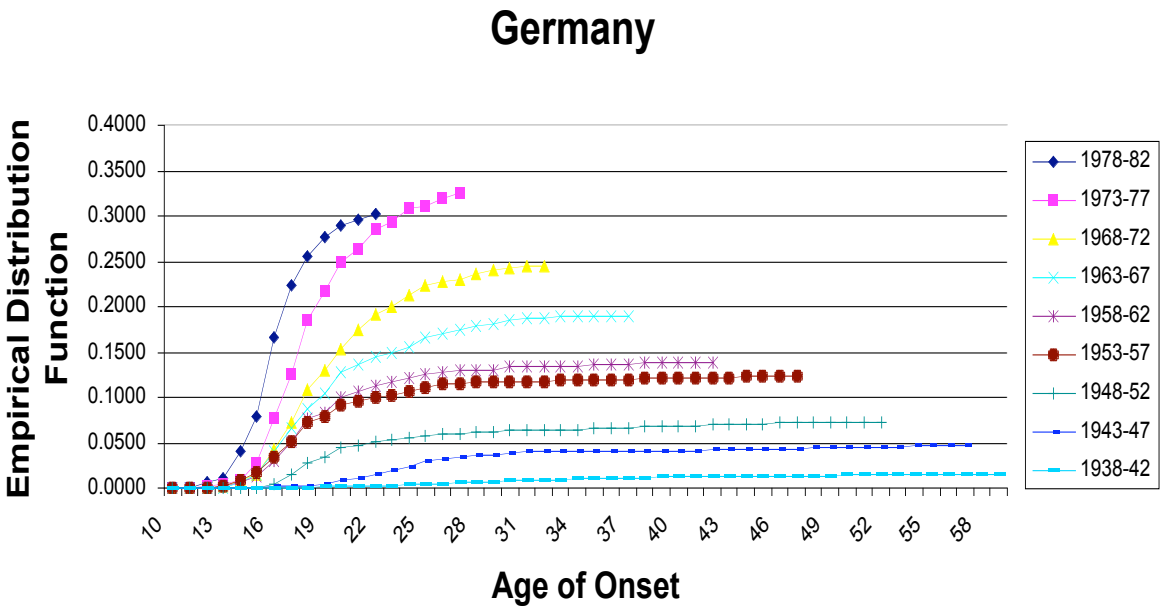
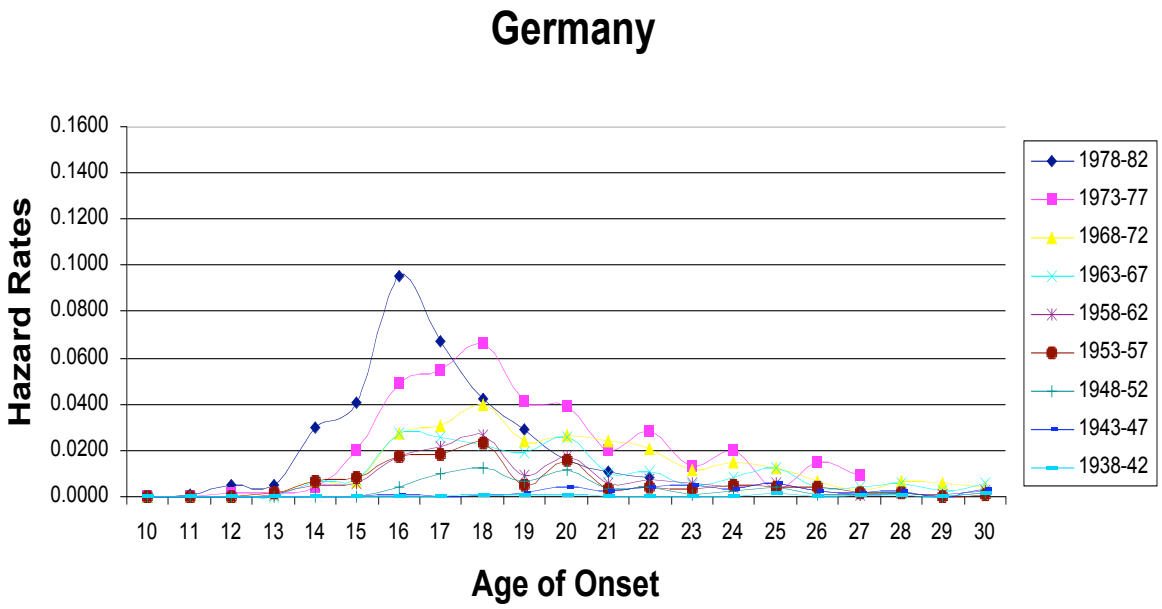


Figure 2. Hazard rates for age of first cannabis use: Germany, 5-year-cohorts



Historical development

To analyse the historical development of cannabis experience empirical distribution functions and hazard rates for selected one-year cohorts were calculated according to calendar year. Again, because of non-significant differences between survey responses on age of onset all three surveys were pooled.

Figure 3 shows an increase in steepness of the empirical distribution function across births cohorts. While the difference between adjacent birth cohorts is moderate for the birth cohorts 1945 to 1965 the slope of the empirical distribution function increases constantly for the three following birth cohorts (1970 ,1995, 1980). In terms of calendar years, Figure 3 points at a moderate development of the drug epidemic between its start in the late sixties and the beginning of the nineties and a steep increase of prevalence afterwards. The same can be seen from Figure 4.

Figure 3. Empirical distribution function for year of first cannabis use: Germany, one-year-cohorts

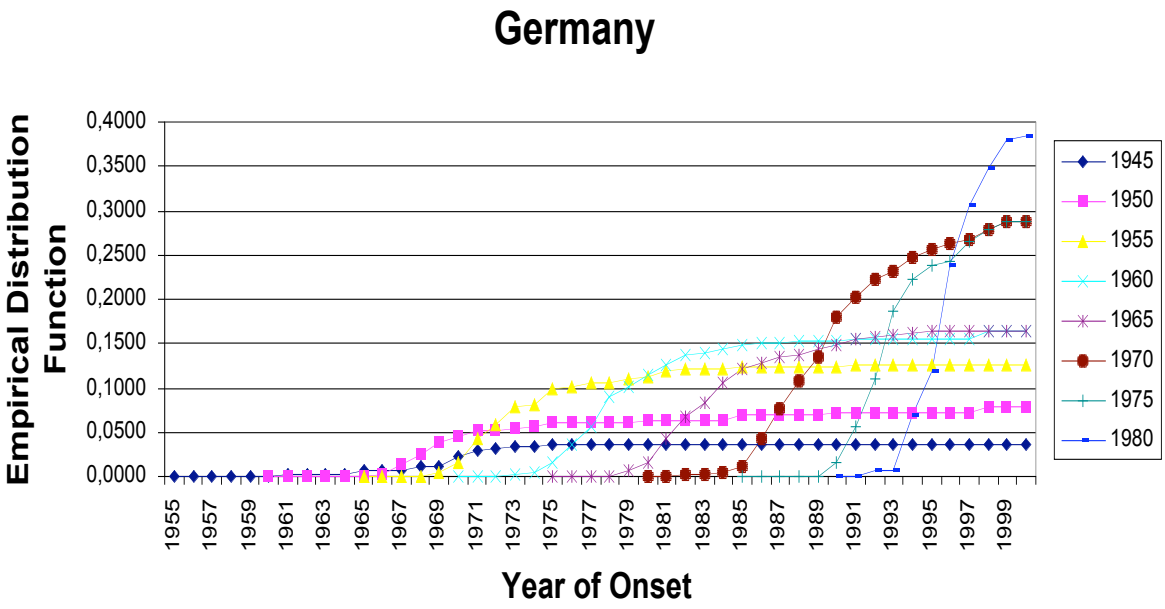
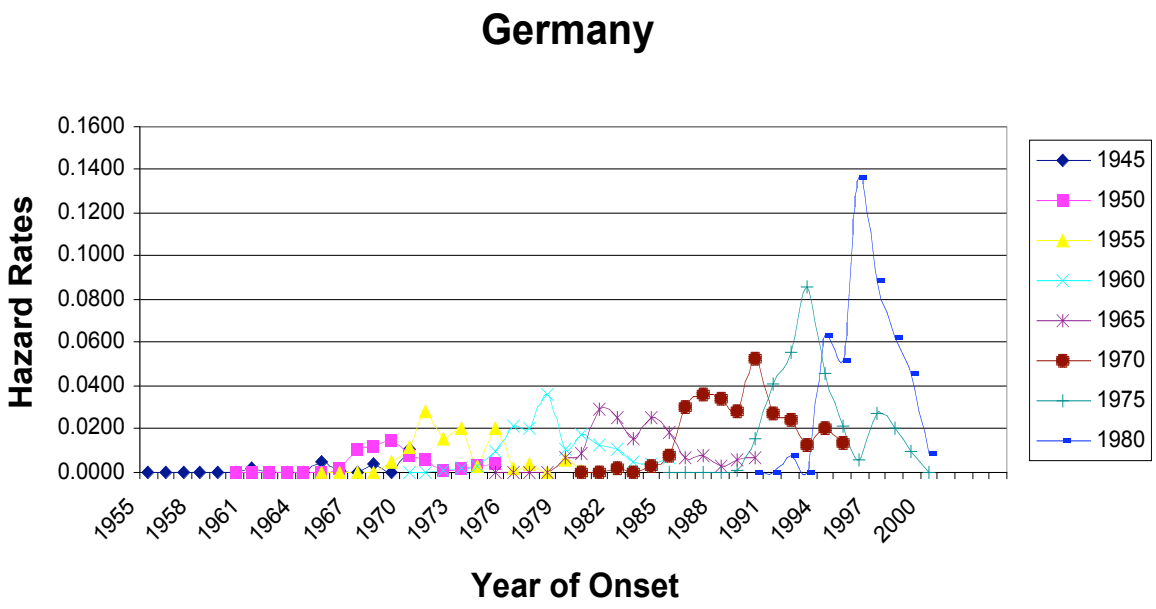


Figure 4. Hazard rates for year of first cannabis use: Germany, one-year-cohorts



Historical development in Eastern- and Western Germany

Figure 1 to Figure 4 show the results for Germany. The access to illicit drugs was, however, rather different in both parts of Germany. While in Western Germany illicit drugs have been available since the late 60s there existed no drug market in Eastern Germany before the re-unification of Germany in 1990 (Kirschner, 1997). Different surveys in Eastern Germany reported a rapidly increasing lifetime prevalence of cannabis among young respondents during the first half of the 90s (Kirschner, 1997; BZgA, 2001; Kraus & Augustin, 2001). Analysing age of first use of illicit drugs in Western Germany, Kraus and colleagues (1998) reported also an increase in lifetime prevalence of illicit drugs in the beginning of the nineties in the former Federal Republic of Germany. To account for the different historical situation we analysed Eastern and Western Germany separately. To increase the sample sizes we also included surveys from the beginning of the nineties with the same sampling design as in 2000.

Figure 5 and Figure 6 depict the empirical distribution functions and the hazard rates of first cannabis use for seven different birth cohorts in western Germany. The prevalence rates of those born in 1955, 1960 and 1965 clearly exceed the prevalence rates of those born in 1945 and 1950 while the two youngest birth cohorts (1970, 1975) exhibit even higher prevalence rates. These changes happened in the late 60s and in the beginning of the 90s respectively.

As in eastern Germany cannabis use was uncommon before 1990 single respondents with cannabis experience lead to significant differences between the surveys. Since these significant differences are rather statistical artefacts and do not reflect systematic differences between the surveys we pooled the surveys and calculated empirical distribution function and hazard rates of first cannabis use for the respondents born in 1970 and 1975.

Compared to the respondents of the same age in Western Germany, those born in 1970 in Eastern Germany were at a higher risk of first cannabis use in the second half of the 90s. Their lifetime prevalence amounts to 14%, which is half the prevalence rate of the Western German respondents born in 1970 (Figures 5 and 7). The empirical distribution functions of Eastern and Western German respondents born in 1975, however, differ only slightly.

Figure 5. Empirical distribution function for year of first cannabis use: Western Germany, one-year-cohorts

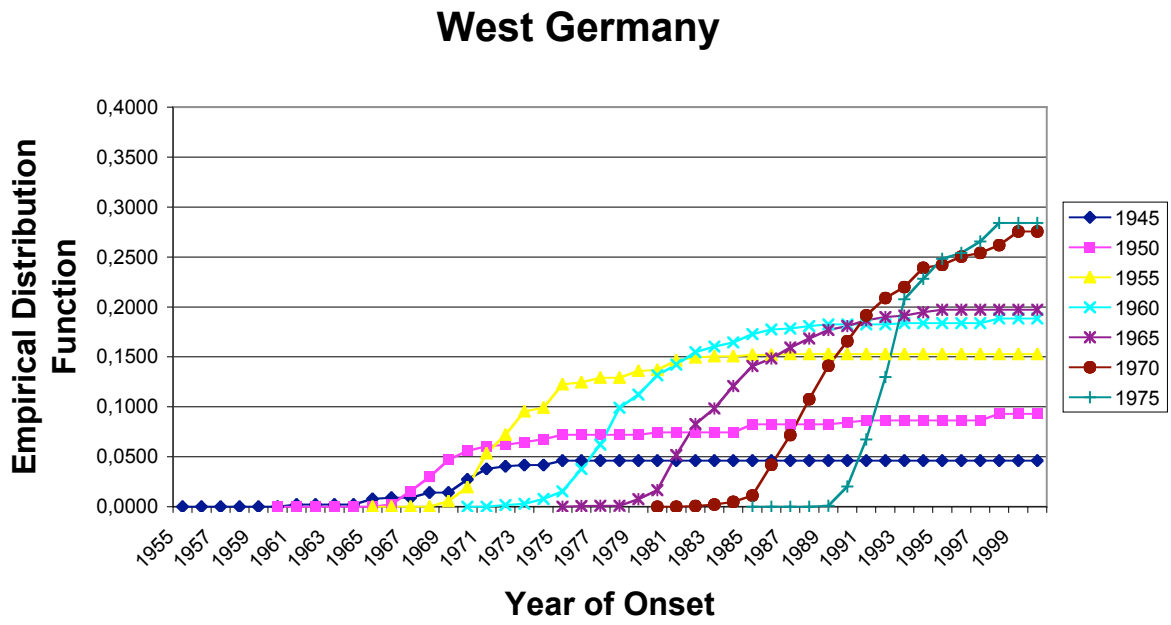


Figure 6. Hazard rates for year of first cannabis use: Western Germany, one-year-cohorts

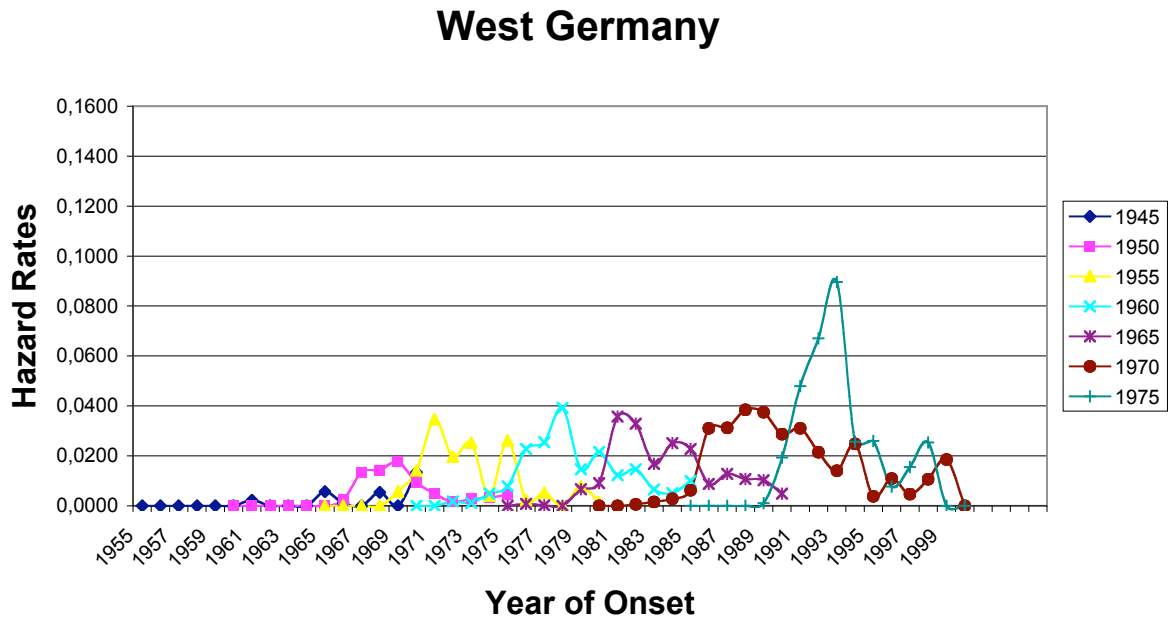
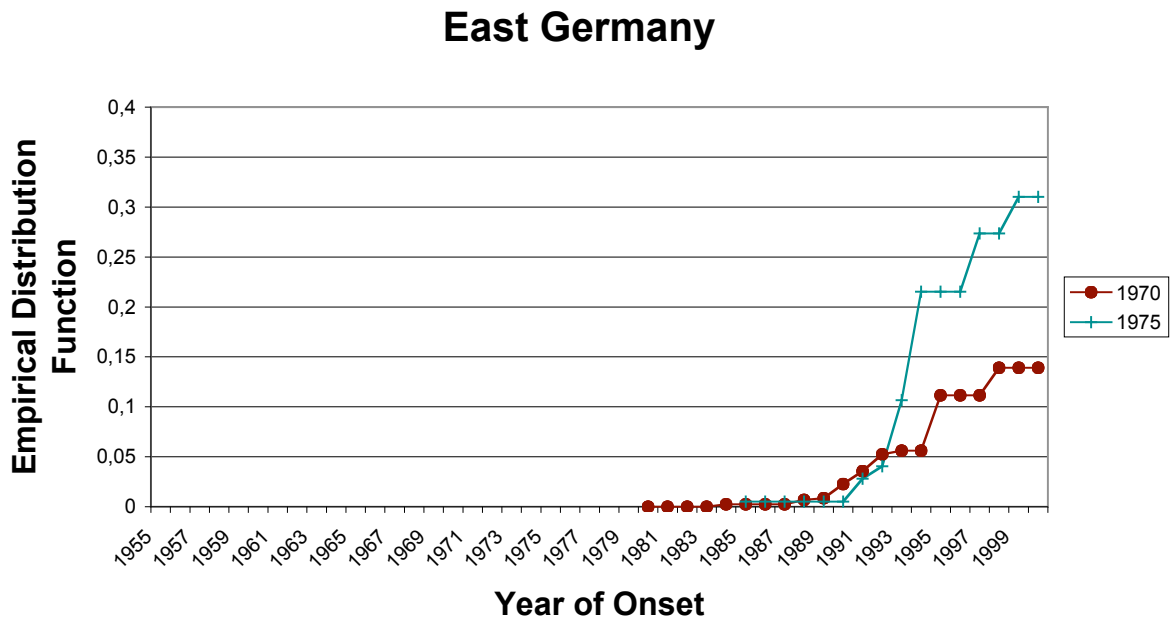


Figure 7. Empirical distribution function for year of first cannabis use: Eastern Germany, one-year-cohorts



GREECE

Distribution of age of first cannabis use

While the 1993 survey covered the Greater Athens area, the 1998 survey was conducted nationwide. In a first step it was tested if in the 1998 full Greek survey the responses on age of onset of those responders living in Athens differ significantly from those not living in Athens. In a second step it was tested if the responses of those cohorts from the 1998 survey with non-significant regional response differences differ significantly from the responses of the same cohort from the 1993 survey. Significant regional differences in responses were found for the three youngest cohorts (1978-82, 1973-77 and 1968-72). Between surveys no significant differences were found, which enabled the pooling of the two surveys for the cohorts 1963-67 up to the oldest cohort born between 1938 and 1942.

Figure 8 shows the empirical distribution functions for the birth cohorts 1978-82, 1973-77,…,1933-37 in Greece. First use of cannabis starts between age 13 (cohorts 1978-82 and 1973-77) and age 20 (cohort 1938-42). The curves increase rapidly until age 30 and level off after that age. There is a tendency to steeper curves for younger cohorts, there are, however, pairs of cohorts with nearly identical curves. This applies to the cohorts 1978-82 and 1973-77, the cohorts 1968-72 and 1963-67, as well as to the cohorts 1958-62 and 1953-57. Prevalence rates range from 1.1% (cohort 1933-37) to 29% (cohort 1973-77).

In accordance with the shift of the starting age of the curves (both empirical distribution functions and hazard rates) towards younger ages also a shift of the peaks of the hazard rates can be observed: While the hazard rates for the cohorts 1973-77, 1968-72 and 1963-67 peak at age 20, for the cohorts 1943-47 and 1938-42 the highest hazard rate is observed at age 25. Furthermore, the maximum of the peaks tends to increase from older to younger birth cohorts: For the cohort 1943-47, the hazard rates achieved a maximum value of 0.0073, i.e. 0.73% of those without cannabis experience until age 25 used cannabis for the first time at age 25. For the cohort 1973-77, however, 5.8% of those under risk of taking cannabis the first time at age 20 experienced first cannabis use at this age. The low hazard rate for the youngest cohort (born 1978-82) at age 20 is based on the reports of only 105 respondents, from which a part may have experienced first cannabis use but after the time of the survey. Thus this hazard rate is not comparable to the hazard rates at age 20 for other cohorts (Figure 9).

Figure 8. Empirical distribution function for age of first cannabis use: Greece, 5-year-cohorts

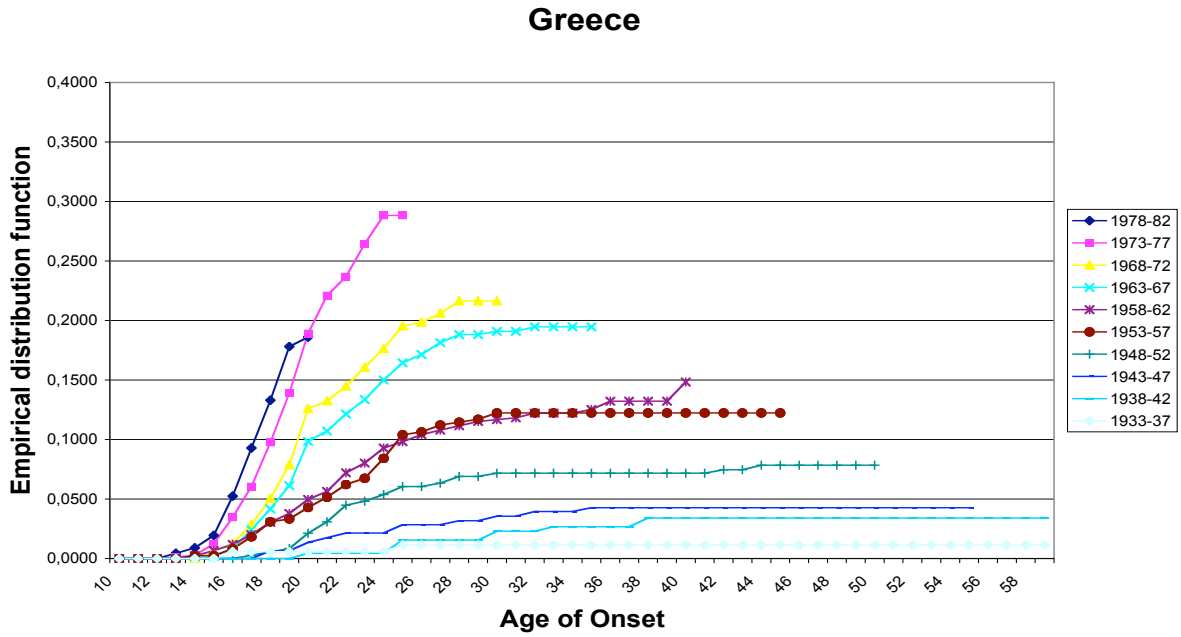
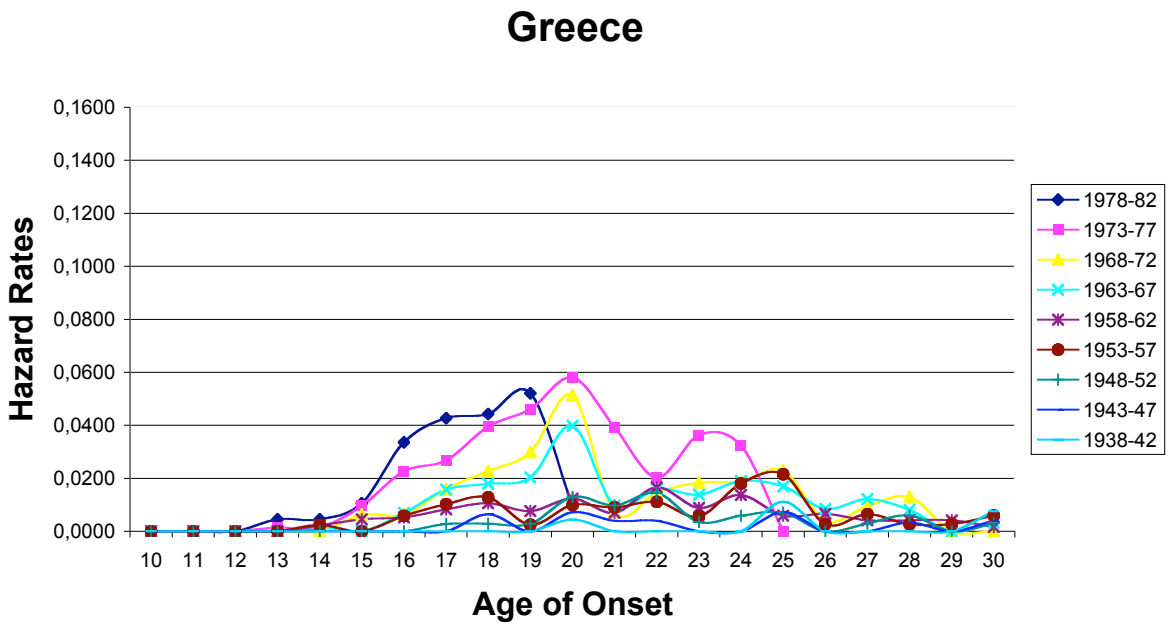


Figure 9. Hazard rates for age of first cannabis use: Greece, 5-year-cohorts



Historical development

Due to the small sample sizes analyses of one-year-cohorts were not conducted. Instead, the analysis of the historical development of cannabis experience is based on the same 5-year cohorts as used in the previous figures 8 and 9. The 5-year-cohorts are thought to represent the middle of the cohort which equals the one-year cohorts used in the corresponding analyses for Germany, i.e. the cohort 1943-47 represents those born in 1945 and the cohort 1978-82 represents those born in 1980.
As can be seen from Figure 10 and 11, cannabis use emerged in Greece during the late 60s and has increased since then. Two prominent changes can be observed: one around the year 1980 and another in the second half of the 80s. These changes correspond to an increase in prevalence of the cohorts 1953-57 and 1963-67 compared to the preceding cohorts.

Figure 10. Empirical distribution function for year of first cannabis use: Greece, 5-year-cohorts

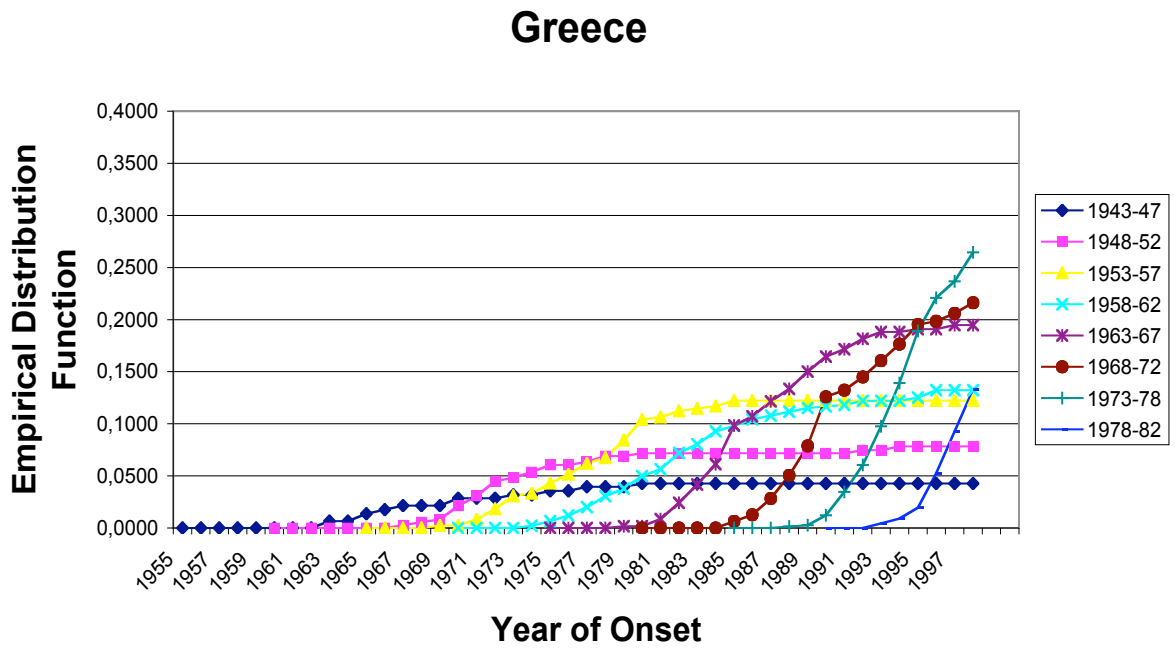
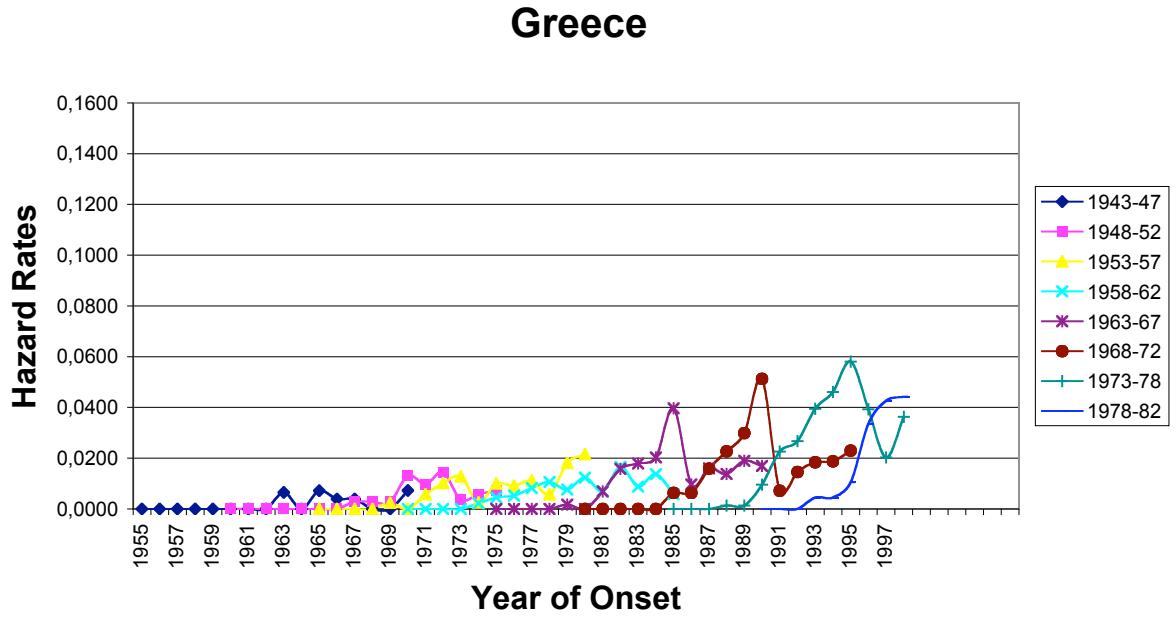


Figure 11. Hazard rates for year of first cannabis use: Greece, 5-year-cohorts



SPAIN

Distribution of age of first cannabis use

Only in the case of three birth cohorts – 1943-47, 1938-42 and 1963-67 – no significant response differences on age of onset between the surveys could be found. In all cohorts with significant

differences between the two surveys the 1997 respondents reported an earlier use of cannabis, resulting in much higher lifetime prevalence rates for the 1997 respondents except for the youngest cohort. These differences may have been caused by differences in sampling design, interview modes or extremely large or small weights for some sampling regions. All these possible causes, however, can be excluded: Neither sampling design nor interview modes differ between the surveys. Moreover, there are no considerable differences between the weights of the surveys and the weighted and non-weighted regional prevalence estimates, which also demonstrates the negligible influence of the weights. On the basis of the data it cannot be decided which survey yields more valid results, thus the results from both surveys are reported.

Figure 12. Empirical distribution function for age of first cannabis use: Spain, 1997, 5-year-cohorts

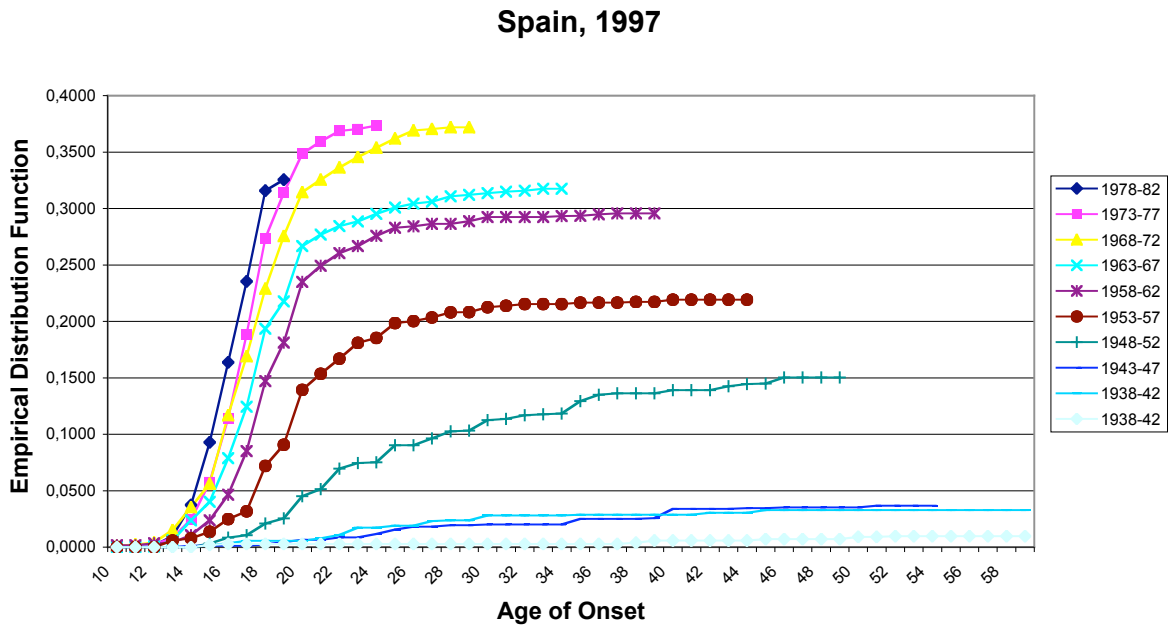
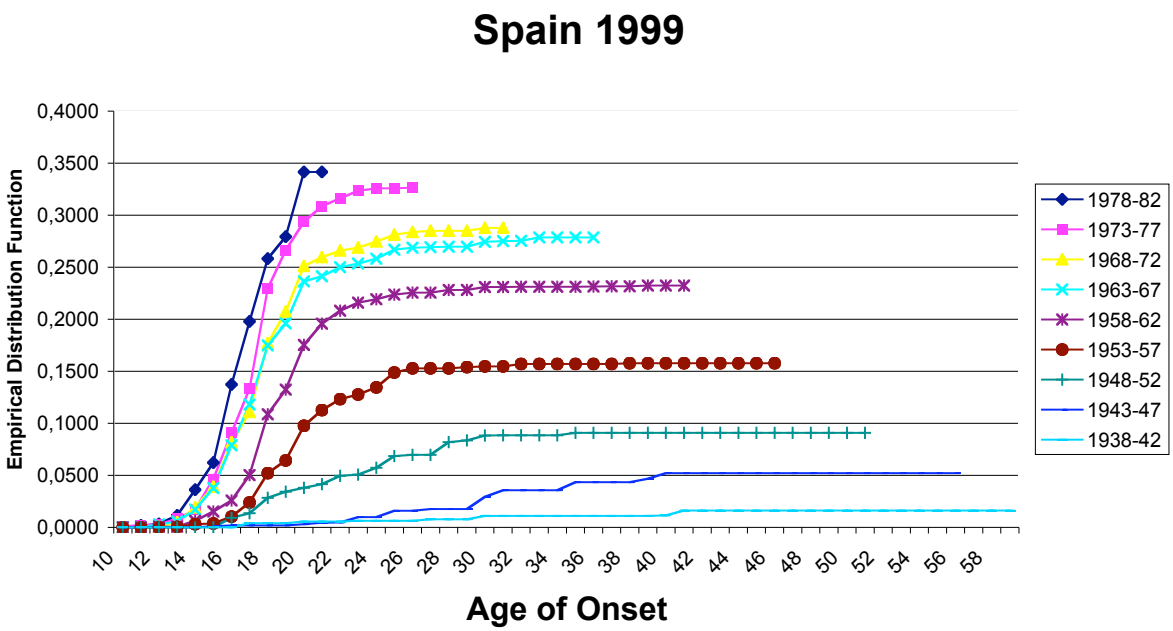


Figure 13. Empirical distribution function for age of first cannabis use: Spain, 1999, 5-year-cohorts



Both Figures 12 and 13 show an increase in lifetime cannabis prevalence from older cohorts to younger cohorts. Cannabis use starts rather early at age 10 to 12. The curves increase rapidly until about age 30 and level off after that age. For the cohorts 1978-82,…,1948-52, the curves in Figures 12 are much steeper than the corresponding curves in Figures 13, resulting in much higher lifetime prevalence rates in the 1997 survey compared to the 1999 survey. While e.g. in 1997 37% of those born in 1973-77 and 15% of those born in 1948-52 reported experience with cannabis the corresponding figures for the 1999 survey are 33% (cohort 1973-77), 29% (cohort 1968-72) and 9% (cohort 1948-52).

Figures 14 and 15 show that at each age the risk of cannabis use tends to increase from older cohorts to younger cohorts. Except for the two oldest cohorts, the hazard functions have a common shape with peaks at age 18 or age 20. Most of the figures in Figures 14 exceed, however, the corresponding figures in Figures 15. The biggest differences are found for the ages 15-20. As an example, in 1997 8.4% of the 1973-1977 born without cannabis experience until age 17 reported to have used cannabis the first time at age 17. The corresponding hazard rate calculated with the 1999 data is 4.7%.

Historical development

To illustrate the historical development of cannabis prevalence in Spain one-year-cohorts without significant differences between the two samples were pooled. With two exceptions – the 1975 and the 1980 birth cohorts – it was possible to use the same cohorts as for the German data. Due to significant differences between responses from the 1997 and the 1999 sample the 1975 and the 1980 birth cohorts were substituted by the births cohorts 1974 and 1979, respectively.

Figures 16 and 17 reveal a sharp increase in cannabis prevalence in the second half of the 70s and a smaller one at the beginning of the nineties: The lifetime-prevalence of those born in 1960 and 1965 amounts to about 30%, which is nearly twice the lifetime-prevalence of those born in 1955. Of those born in 1974, 35% had ever used cannabis. The highest hazard rate for the 1960 born (8.9 %, at age 18 in the year 1978) is also nearly twice the highest hazard rate for the 1955 born (4.9 %, at age 20 in the year 1975). The hazard rate for those born in 1974 peaks at age 18 in the year 1992 and is 12.1%. The hazard rate of those born in 1979 lower and peaks at age 17.

Figure 14. Hazard rates for age of first cannabis use: Spain, 1997, 5-year-cohorts

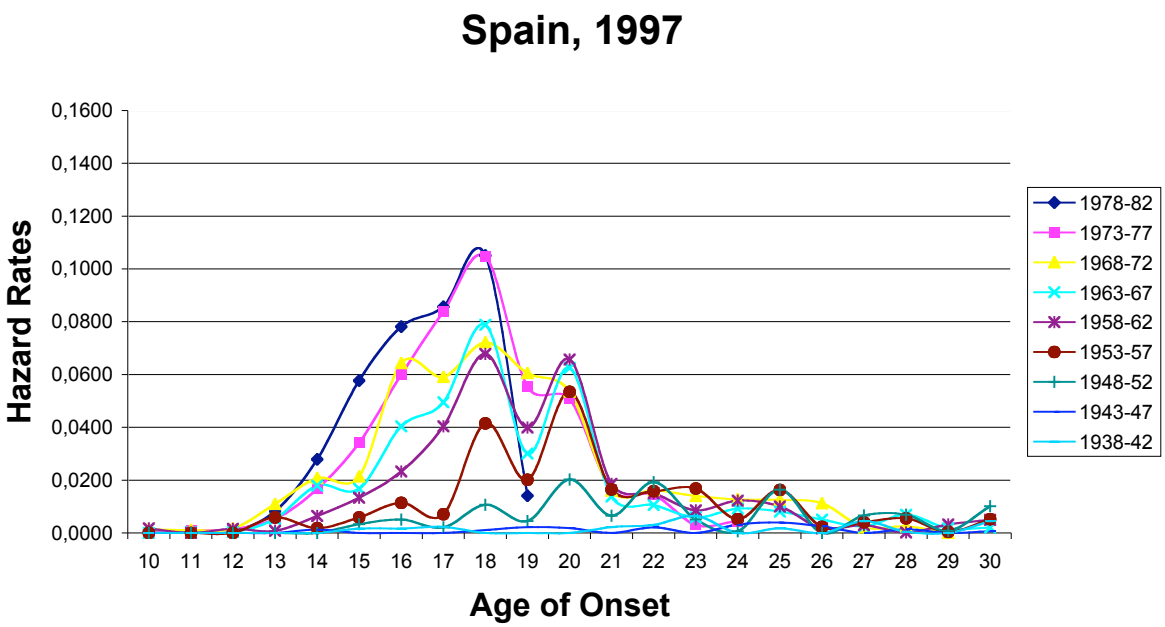


Figure 15. Hazard rates for age of first cannabis use: Spain, 1999, 5-year-cohorts

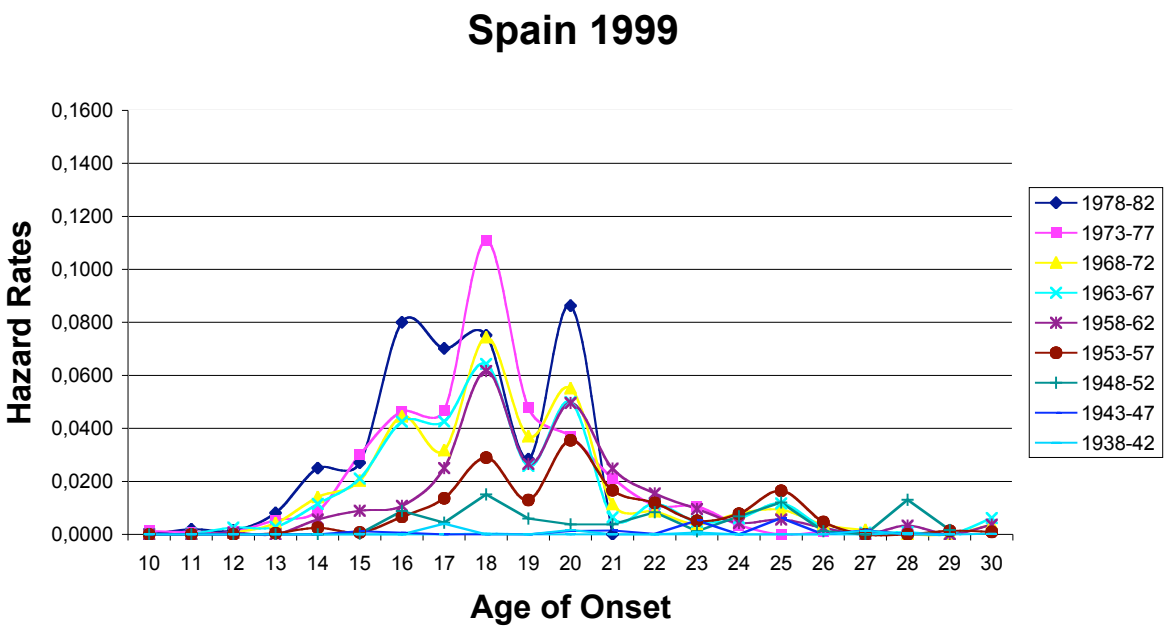


Figure 16. Empirical distribution function for year of first cannabis use: Spain, one-year-cohorts

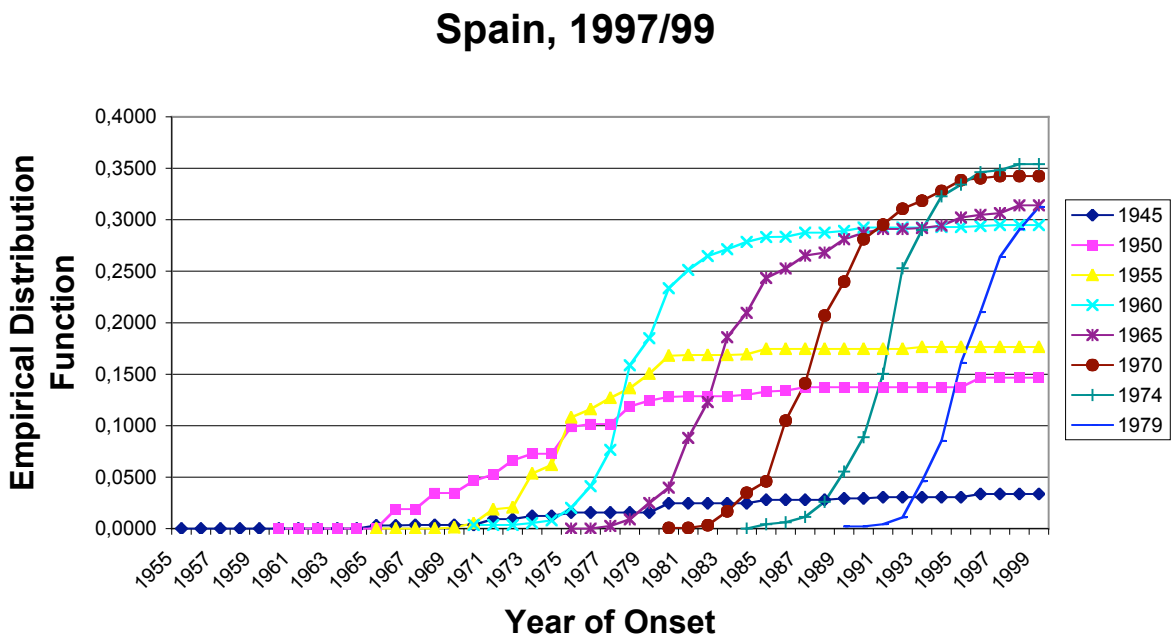
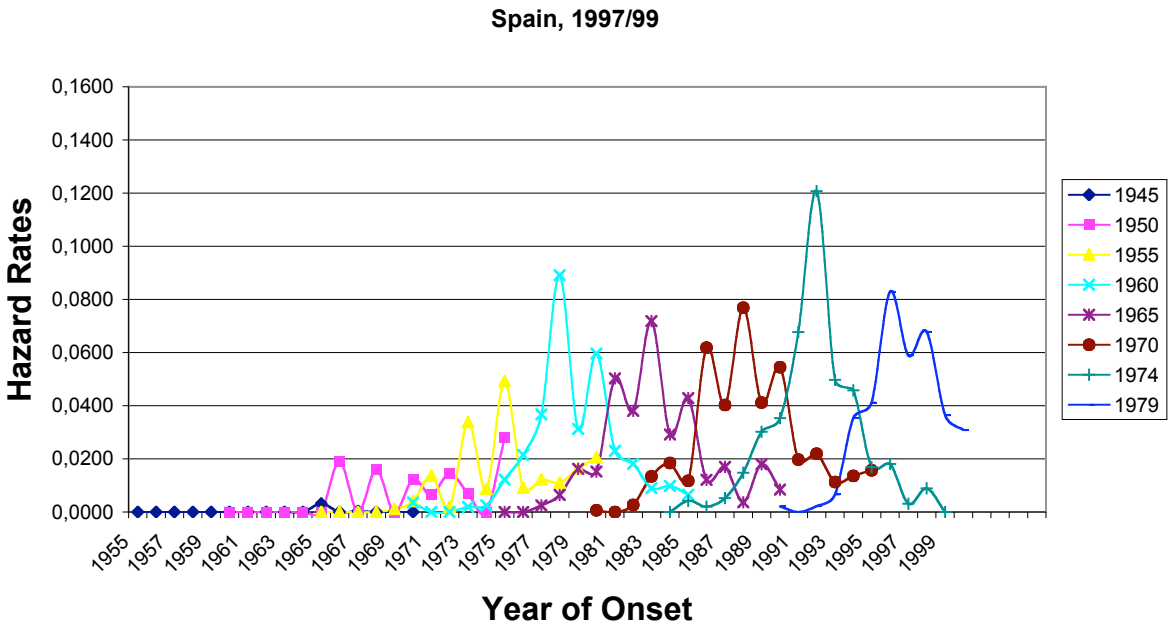


Figure 17. Hazard rates for year of first cannabis use: Spain, one-year-cohorts



Comparison between Countries

In Figures 18 to 23 the empirical distribution functions and hazard rates of the birth cohorts 1943-47, 1963-67 and 1973-77 are compared across countries. Due to the non-significant response differences in the German surveys data from the 1995, 1997 and 2000 survey were pooled. From the Greek survey, the cohort 1973-77 was analysed using only data from 1998, while for the analysis of the two older cohorts data from both surveys were used. While the Spanish data of the cohorts 1943-47 and 1963-1967 could be pooled, the results of the cohort 1973-1977 are shown separately.

The lifetime prevalence rates of the birth cohort 1943-47 are rather similar (Figure 18 and 19). While the German and Greece hazard rates at ages 18-22 i.e. in the mid of the 60s, exceed the Spanish ones, the Spanish respondents report more frequently first cannabis experience at age 35 to 40. This causes a small gap between the German and Greece empirical distribution function and the Spanish ones between ages 20 and 40m (Figure 18).

The cohort 1963-1967 shows a quite different picture. The Spanish hazard rates in the age range 14-20, i.e. in the first half of the 80s, are clearly higher than German and Greek hazard rates. For example, at age 18 the hazard rates from the Spanish data are 7.1%, while in Germany and in Greece about 2% of the respondents without previous cannabis experience reported to have used cannabis at age 18 (Figure 21). Moreover, up to age 18 the Greece hazard rates are lower than the German ones while in the age range 21-30, both Spanish and German hazard curves are rather similar but lower compared to the Greek curves. The wide gap between the Spanish and the other hazard rates in the age range 14 to 20 and the comparatively small differences at higher ages lead to much higher prevalence rates in Spain compared to Germany and Greece at all ages (Figure 20). At the end of the 90s, one in five respondents in Germany or Greece reported lifetime use of cannabis, while in Spain the lifetime prevalence rates add up to 30%.

Apart from smaller oscillations, the Spanish and German hazard rates of the birth cohort 1973-77 (Figure 23) show a common shape with a peak at age 18, whereas the Greek curve is bimodal with peaks at ages 20 and 23. Up to age 18, the Greek hazard rates are lowest, followed by the German hazard rates and the Spanish ones. The big differences in hazard rates – especially between Greece and both Germany and Spain – result in big differences concerning the lifetime prevalence rate at age 18: The lifetime prevalence rate of the Greek respondents (9.8%) is about one half of that of the German respondents (18.4%) and about one third of the respondents of the 1997 Spanish survey (27.4%). The corresponding lifetime prevalence rate in the 1999 Spanish survey is 23%. For ages 21

and older (i.e. end of the 80s/beginning of the 90s), the Spanish hazard rates are lower than those in Greece and Germany.

Figure 18. Empirical distribution function for age of first cannabis use: Germany, Greece, Spain, 1943-47

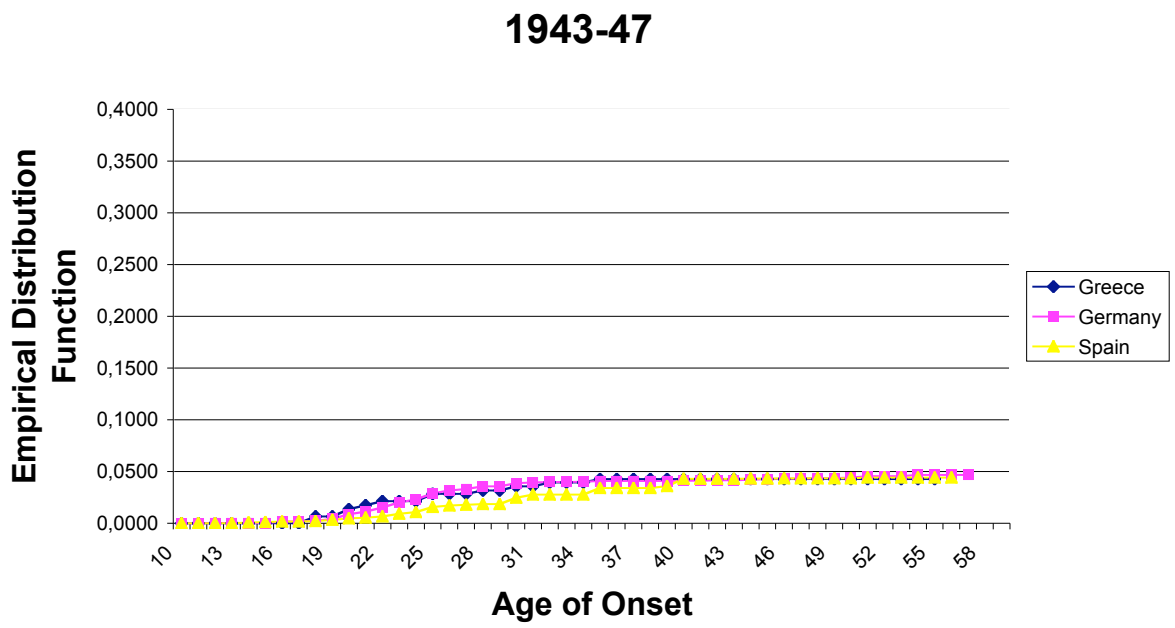


Figure 19. Hazard rates for age of first cannabis use: Germany, Greece, Spain, 1943-47

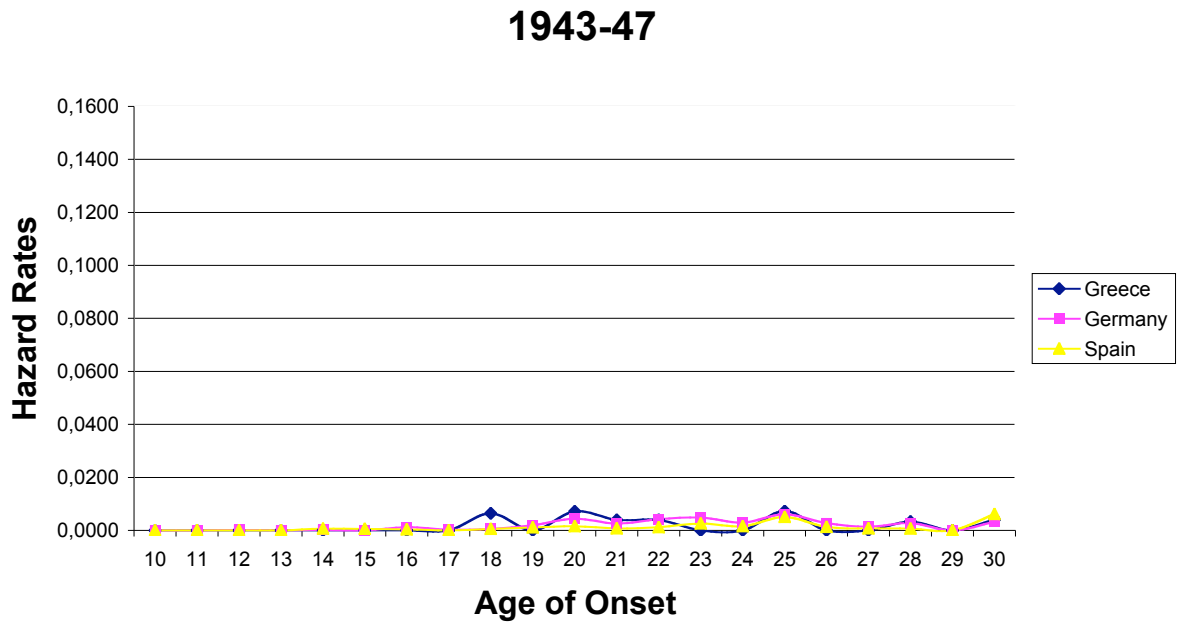


Figure 20. Empirical distribution function for age of first cannabis use: Germany, Greece, Spain, 1963-67

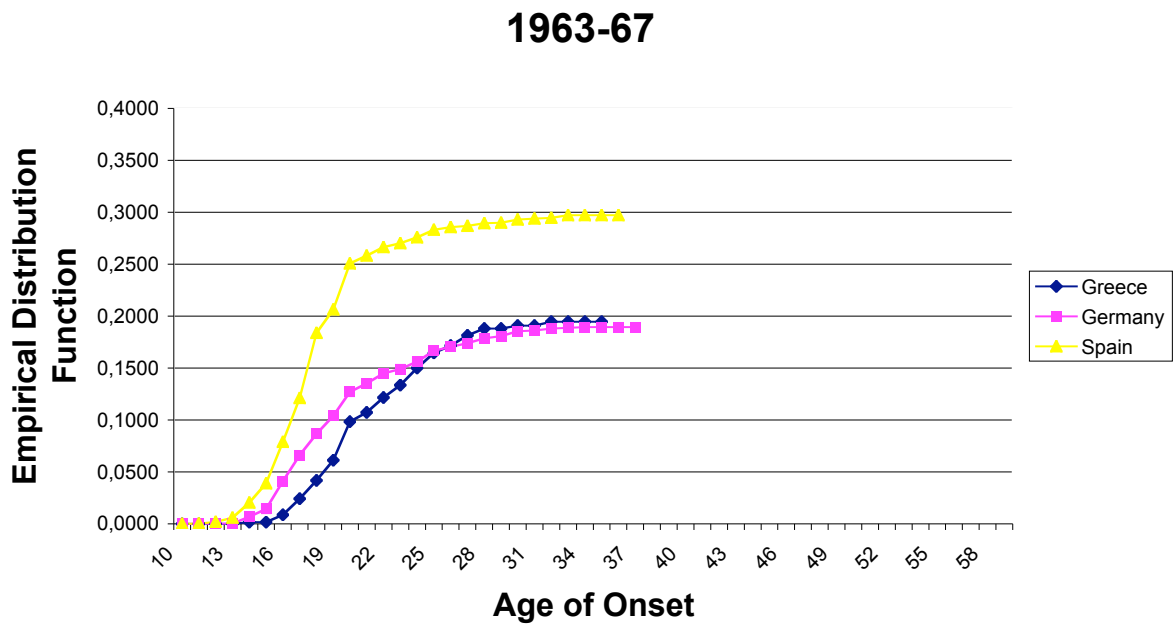


Figure 21. Hazard rates for age of first cannabis use: Germany, Greece, Spain, 1963-67

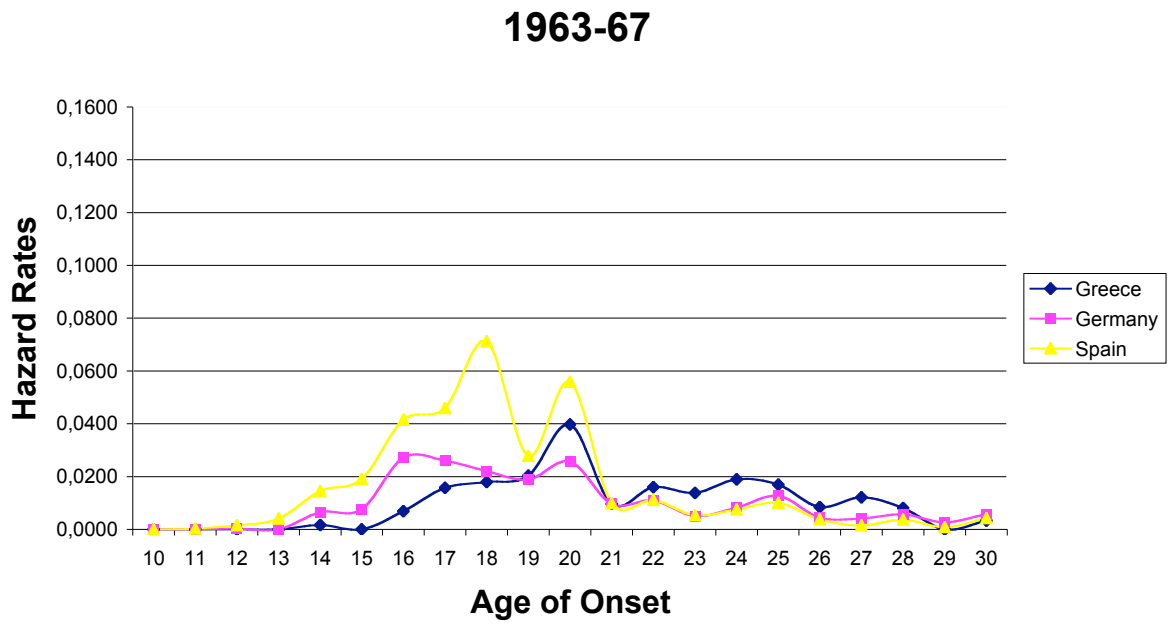


Figure 22. Empirical distribution function for age of first cannabis use: Germany, Greece, Spain, 1973-77

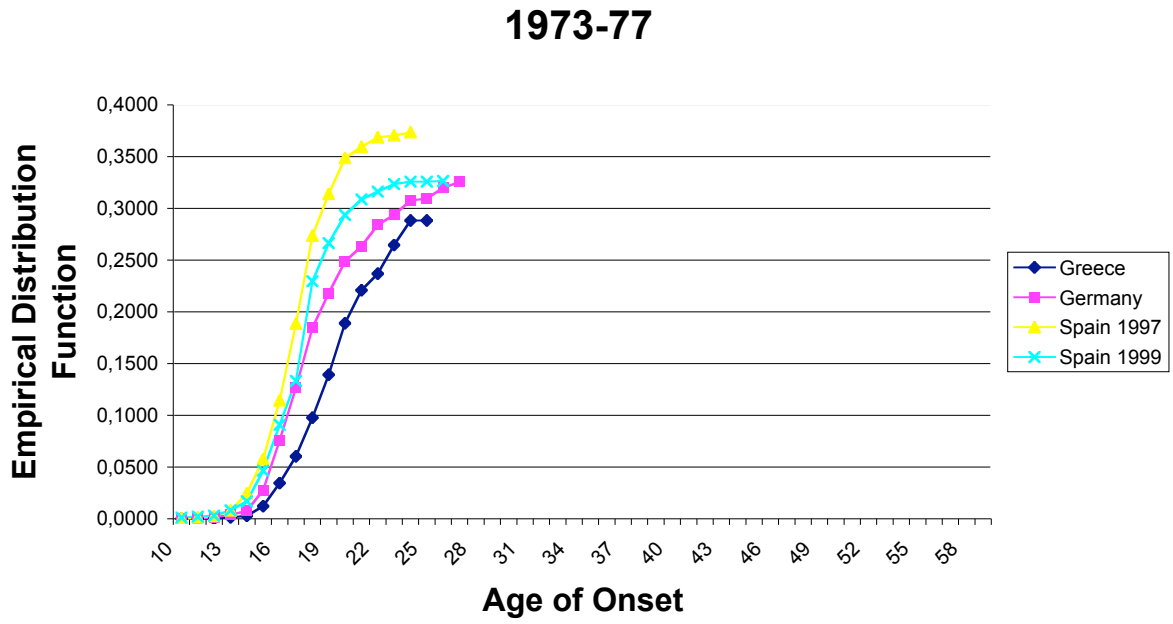
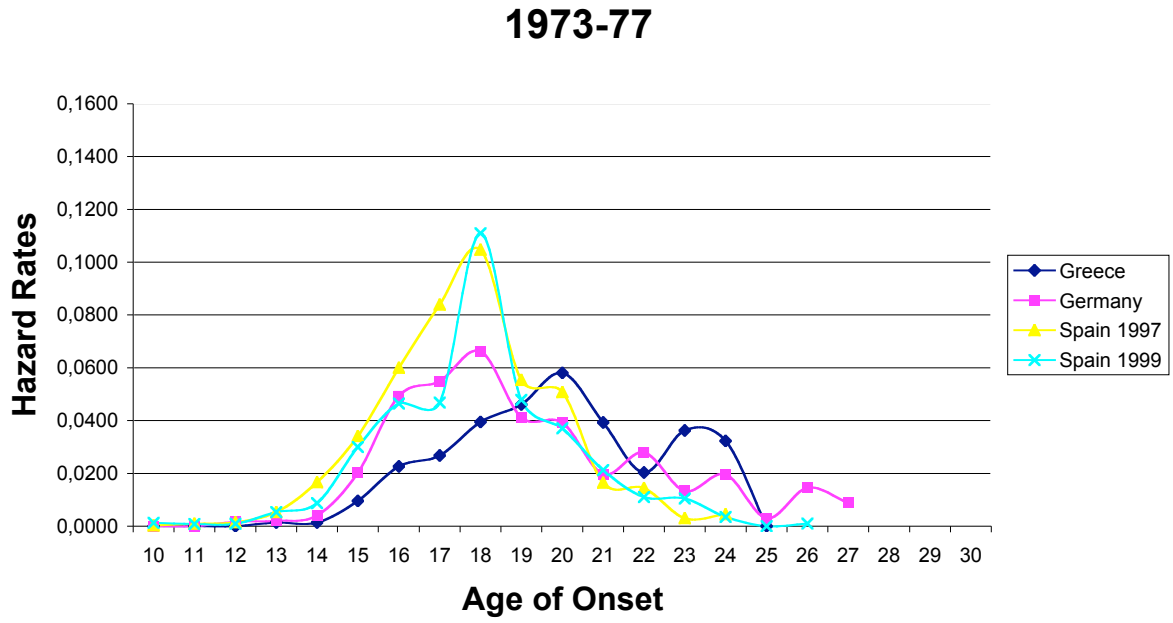


Figure 23. Hazard rates for age of first cannabis use: Germany, Greece, Spain, 1973-77



Discussion

Analyses of the cumulative incidence of age of first cannabis use in Germany, Greece and Spain show that cannabis lifetime-prevalence in each country increased from older to younger cohorts. In addition, in Spain a sharp increase in cannabis experience could be observed in the late 70s while in Germany this pattern could be observed in the early 90s. Finally, the maximum risk of first cannabis experience in Germany and Spain was found at age of 18 years and in Greece at age of 20 years.

Although an increase in prevalence of cannabis experience over a period of approximately 30 years can clearly be interpreted as cohort effect, hazard functions show higher rates for younger cohorts but age distribution across cohorts did not shift its maximum towards younger ages. This rather constant age distribution of the hazard rates points to a proportional increase in cannabis use initiation. Differences in the age distribution between countries may be the result of social or cultural influences, facilitating or preventing early initiation of drug use. Results further indicate, that risk for cannabis initiation begins to decline after age 20-21. After age 25 initiation happens rather rarely. This points at age as a protective factor, i.e. if someone has not used cannabis until age of 25 years he will most likely not start using cannabis thereafter.

Analyses of several cross-sectional samples demonstrate that today more young people are involved in cannabis use than in the 60s, 70s or 80s. However, as could be shown for Western Germany not the age of onset has changed, but the number of people getting involved with drugs increased at a proportional rate in all age groups (Kraus et al., 1998). Hazard rates observed for different cohorts in Spain and Greece do also not indicate a shift of first cannabis experience towards younger ages. Since hazard rates for the youngest cohorts are censored around the critical age these results should not be taken as evidence for possible changes.

Analyses of responses on age of first cannabis use from different surveys combined responses from different samples of the same cohort that differed only with respect to respondent's age at interview. The majority of responses of the same cohort from different surveys differed not significantly when statistically tested. This indicates that responses on age of first cannabis use may be considered independent of age at interview. Reliability studies found that responses on age of first cannabis use were highly consistent with earlier reported ages. Most interestingly, reports on age of cannabis use were more reliable compared to responses on age of first alcohol, tobacco or other illicit drug use (Johnson & Mott, 2001). In general it is found that respondents commonly forward telescope their answers, producing older estimated ages of substance use initiation upon reinterview (Engels et al, 1997; Johnson & Mott, 2001; Kandel & Yamaguchi, 1985). While information on substance use initiation is reported at good, but not excellent level of reliability, Johnson & Mott (2001) conclude that self-reports on age of first substance use as collected via survey questionnaires are of sufficient reliability for most current epidemiological applications.

References

Bachmann, J.G., Wadsworth, K. N., O`Malley, P. M., Johnston, L. D. & Schulenberg, J. E. (1997). *Smoking, drinking, and drug use in young adulthood.* New Jersey: L. Erlbaum Associates.

Bundeszentrale für gesundheitliche Aufklärung (BZgA) (2001). *Die Drogenaffinität Jugendlicher in der Bundesrepublik Deutschland*. *Wiederholungsbefragung 2000/01*. Köln: Bundeszentrale für gesundheitliche Aufklärung.

Engels, R.C.M.E., Knibbe, R.A. & Drop, M.J. (1997). Inconsistencies in adolescents' self-reports of initiation of alcohol and tobacco use*. Addictive Behaviors, 22*, 613-623.

European Monitoring Centre for Drugs and Drug Addiction (2001). *Annual report on the state of the drug problem in the European Union*. Lisbon: European Communities.

Hibell B, Andersson B, Ahlström S, Balakireva O, Bjarnason T, Kokkevi A and Morgan M (2000). *The 1999 ESPAD report. Alcohol and other drug use among students in 30 European Countries.* Stockholm: CAN.

Höfler, M., Lieb, R., Perkonigg, A., Schuster, P., Sonntag, H. & Wittchen, H.-U. (1999). Covariates of cannabis use progression in a representative population sample of adolescents: a prospective examination of vulnerability and risk factors. *Addiction, 94, 1679-1694.*

Johnson, T.P. & Mott, J.A. (2001). The reliability of self-reported age of tobacco, alcohol and illicit drug use. *Addiction, 96,* 1187-1198.

Kandel, D. B. & Logan, J. A. (1984). Patterns of drug use from adolescence to young adulthood: I. Periods of risk for initiation, continued use, and discontinuation. *American Journal of Public Health, 74*, 660-666.

Kandel, D. B., Yamaguchi, K. & Chen, K. (1992). Stages of progression in drug involvement from adolescence to adulthood: Further evidence for the Gateway Theory. *Journal of Studies on Alcohol, 53*, 447-457.

Kandel, D. B. & Yamaguchi, K. (1985). Developmental patterns of the use of legal, illegal, and medically prescribed psychotropic drugs from adolescence to young adulthood. In C. L. Jones. & R. J. Battejes (Eds.), *Etiology of drug abuse: Implications for prevention* (NIDA Research Monograph No. 56, DHHS Publication No. (ADM) 85-1335, pp. 193-235). Washington: Government Printing Office.

Kandel. D. B. & Faust, R. (1975). Sequence and stages in patterns of adolescent drug use. *Archives of General Psychiatry, 32*, 923-932.

Kirschner, R. (1997). *Jugend und illegale Drogen in Ostdeutschland. Eine bevölkerungsrepräsentative Längsschnittuntersuchung*. München: Profil.

Kraus, L., Bauernfeind, R. & Herbst, K. (1998). Hat sich das Alter des Erstkonsums illegaler Drogen verschoben? Survivalanalyse retrospektiver Querschnittsdaten 1980-1995.(Has age at onset of drug use changed over the past 25 years? Survival analysis of cross-sectional survey data 1980-1995). *Zeitschrift für Klinische Psychologie, 27 (1), 20-29.*

Lawless, J. F. (1982). *Statistical models and methods for lifetime data.* New York: John Wiley & Sons.

Müller, S. & Gmel, G. (2002). Veränderungen des Einstiegsalters in den Cannabiskonsum: Ergebnisse der zweiten Schweizer Gesundheitsbefragung. *Sozial- und Präventivmedizin, 47,* 14-23.

Perkonigg, A., Lieb, R., Höfler, M., Schuster, P., Sonntag, H. & Wittchen, H.-U. (1999). Patterns of cannabis use, abuse and dependence over time: incidence, progression and stability in a sample of 1228 adolescents. *Addiction, 94,* 1663-1687.

Robins, L. N. & Przybeck, T. R. (1985). Age of onset of drug use as a factor in drug and other disorders. In C. L. Jones & R. J. Battejes (Eds.)*, Etiology of drug abuse: Implications for prevention* (NIDA Research Monograph No. 56, DHHS Publication No. (ADM) 85-1335, pp. 178-192). Washington: Government Printing Office.

Schumann, J., Augustin, R., Duwe, A., Küfner, H. (2000). Welchen Einfluss haben erste Erfahrungen und Begleitumstände zu Beginn des Drogenkonsums auf den weiteren Verlauf? *Zeitschrift für Gesundheitspsychologie, 8,* 69-82. Göttingen: Hogrefe.

Schumann, J., Duwe, A., Küfner, H., Bühringer, G. (1999). Risk and Protective Factors for illegal drug use and the development of substance related diagnoses. In: Louis S. Harris (Ed.). NIDA Research Monograph 180. Rockville: National Institute on Drug Abuse.

Shah, B., Barnwell, B. and Bieler, G. SUDAAN User's Manual, Release 7.5, Research Triangle Park, NC: Research Triangle Institute, 1997.

Silbereisen, R.K. (1997). Missbrauch und Gebrauch von Alkohol und Drogen im Jugendalter. In R. Weinkraut, J. Haisch & M. Kessler (Hrsg.), *Public Health und Gesundheitspsychologie* (S. 170-178). Göttingen: Huber.

Singer, J. D. & Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin*, *110*, 268-290.

Willett, J. B. & Singer, J. D. (1993) Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurence, *Journal of Consulting Clinical Psychology*, 61, 952-965.

Yamaguchi, K. & Kandel, D. B. (1984). Patterns of drug use from adolescence to young adulthood: III. Predictors of progression. *American Journal of Public Health*, *74*, 673-681.

2.2. DRUG USE AND THE NARROWING GENDER GAP

DIRK KORF, ANNEMIEKE BENSCHOP
Bonger Institute of Criminology, University of Amsterdam, Netherlands

Summary

This paper analyses trends in gender differences in illicit drug use, as reflected in general population data from four EU countries (England & Wales, Germany, Spain and Greece) from 1993 to 2000. The introduction and spread of cannabis use in the past four decades showed temporal differences between European countries. National statistics thus seem to reflect different stages in the cannabis 'epidemic', with England & Wales as 'trendsetters' and Greece as a 'late developer'. Data from successive surveys reveals a steady increase in experimentation with cannabis in all four countries starting in the late 1960s (the 1948-1952 birth cohort). The diffusion of cannabis use also appears to have accelerated recently as a result of delayed drug initiation during the 1990s, when more people were taking their first cannabis at older ages.

The 'gender gap' is the differential in drug use prevalence between males and females. The absolute gender gap in the lifetime prevalences of cannabis use did not show a clear trend in life time use of cannabis. The relative gender gap did appear to have narrowed in all four countries, although both the pace of change and the current size of the gender gap varied between countries. Convergence between males and females was clearest in Greece (Athens) and Spain, where the gender gap seems likely to narrow further in the years to come.

Introduction

Drug use appears to have increased in the European Union during the past decade. Recent general population surveys in several EU member states show a particularly sharp rise in the lifetime prevalence of cannabis use (EMCDDA, 2001). Since lifetime prevalence rates are cumulative, however, a considerable degree of that rise may be attributable to a generation effect (Korf, 2002). The use of cannabis typically began spreading in Europe in the late 1960s and early 1970s, and successive generations of youth have been trying the drug ever since. All such generations are reflected in the lifetime figures, including people who have ever tried the drug but are not current users. Yet recent surveys of 15- and 16-year-old school pupils in many European countries have also revealed higher lifetime prevalences of illicit drug use in that age group for 1999 as compared to 1995 (Hibell et al., 2000). The last-year prevalence of cannabis use in the general population has risen too, albeit less sharply than lifetime use (EMCDDA, 2001). On balance, then, the rising lifetime prevalence rates indicate more than just a generation effect. The youth of today appear to be experimenting more widely with drugs than previous generations ever did.

From a socio-epidemiologic perspective, one might ask whether this quantitative trend in drug use has stayed confined to the same types of user populations as previously, or whether qualitative changes have occurred – for instance, in the form of differential diffusion to new groups of users, whereby the drug user profiles become altered over time. Traditionally, age and gender have been strong determinants of drug use. In terms of age, experimenting with drugs has typically been a phenomenon of adolescence and early adulthood. If the average age of first experimentation should decline, that might cause a temporary rise in lifetime prevalence within that age cohort, but a limit would probably be reached in the course of time. As it is, however, the average age of first cannabis use has proven relatively stable. Although the percentages of young adolescents first trying cannabis have increased,

the same is true of older adolescents and young adults, thus keeping the average age of first use virtually unchanged (Kraus et al, 2002).

With regard to gender as a determinant of substance use, males have consistently been found over the years to experiment with drugs more than females do. However, much has changed in the social status of women since the beginning of the drugs epidemic in the late 1960s and early 1970s. Women now have higher levels of both education and employment. Spurred by these social changes, alcohol researchers already set out in the 1970s to empirically test the so-called convergence hypothesis. The broader theoretical context of this hypothesis derives in particular from the work of feminist-inspired scholars such as Adler (1975), who predicted that as women made their way into a man's world, they would start behaving more like men. Briefly, then, when applied in alcohol research, the convergence hypothesis predicts that the gender gap in alcohol consumption will narrow, because behavioural norms and expectations for women will have changed along with their changing roles (Bloomfield et al., 2001). An increasing resemblance between male and female drinking behaviour will be seen, especially in groups where women's roles and positions most closely resemble men's (Neve et al., 1996). The general hypothesis of the narrowing gender gap in alcohol use is empirically tested by comparing gender differences in drinking behaviour over time, using such indicators as the prevalence of current drinking, the frequency of drinking, average consumption, and the prevalence of heavy or problem drinking.

The narrowing gender gap has been confirmed in the United States by data from annual, nationally representative surveys conducted between 1975 and 1995 (Johnston, O'Malley & Bachman, 1996), as well as by survey data on alcohol dependence (Nelson et al., 1998). In Europe, Sælan, Møller & Køster (1992) were among the first to test the convergence hypothesis in a cohort study of Danes, whom they followed from 1976 to 1987. They found that women increased their alcohol consumption and that a convergence in drinking frequency between the genders began to occur. Neve et al. (1996), by contrast, concluded on the basis of empirical data on alcohol use in the Dutch general population between the late 1970s and early 1990s that gender differences had largely remained stable. Some convergence in average consumption per drinker was observed for women in their 40s in the second half of the 1980s, with average consumption increasing for women while remaining stable for men. An analysis by Bloomfield et al. (2001) of data on alcohol use in four European countries from the early 1980s to the early 1990s found gender convergence in Finland in terms of current drinking and mean alcohol consumption. Per capita alcohol consumption rose across the board in Finland after a relaxation of national alcohol policies (a phenomenon not occurring in the other three countries, Germany, the Netherlands and Switzerland). The researchers suggest that convergence is more likely to occur at times when overall consumption is on the rise than when it is stable or declining.

Ahlström et al. (2001) analysed surveys from nine European countries in the late 1980s and early 1990s. Overall, men were slightly more likely than women to be current drinkers. In most countries, the monthly frequency of drinking among men was 1.5 to 2 times higher, mean monthly alcohol consumption was 2 to 3 times higher, and heavy drinking was 3 to 4 times higher than for women. Kraus et al. (2000) concluded from German surveys conducted in 1994 and 1996 that men on the whole were still more likely to drink alcohol, to be heavier drinkers and to experience more alcohol problems than women. Nevertheless, the German findings still pointed to a narrowing gender gap, by virtue of the increasing prevalence of regular alcohol use by females across cohorts. Regular drinking among younger females increased in recent decades, while male drinking remained constant. The most marked differences in drinking prevalence were between the oldest and the youngest female cohorts. The authors believe these findings reflect gender-specific transitions in German post-war society in the 1950s and 1960s. In an analysis of survey data on Swiss youth aged 10-17, Steinhausen and Winkler Metzke (1998) found an initial gender gap in alcohol use that tended to close with increasing age. Although the popularity of alcohol was greater among boys than girls, the gender gap in preadolescence narrowed in the course of adolescence, and had closed almost completely by the end of that stage of development.

One finding common to all these studies is that the gender gap in drinking behaviour is narrowing, though the observed convergence is more pronounced in some studies than others. Bloomfield et al. (2001) spoke of mixed results, but suggested that the failure to show clear overall convergence could have lain in the relatively brief time interval of their study. Neve et al. (1996) argued that their results were largely in line with most other studies in showing some convergence, while Kraus et al. (2000) concluded that their findings provided additional evidence for the convergence in drinking behaviour as shown in other studies. Such differences in interpretation seem to derive in part from the importance the researchers attached to alcohol-related problems. Findings with regard to the latter

issue were ambiguous, possibly due to the questionable validity of self-reports when it comes to alcohol problems (Knibbe & Bloomfield, 2001). The observed gender convergence chiefly involved women drinking greater amounts and more frequently, while men's alcohol use remained stable. This would seem to confirm that women's drinking behaviour has indeed moved towards that of men, though gender differences still remain.

The studies cited above also revealed differences between countries with regard to the point in time at which the gender convergence in alcohol use began and the pace at which it proceeded. Bloomfield et al. (2001) argued that convergence in Europe generally began among women who came of age in the 1960s, 'a period in which the contemporary feminist movement was born and new opportunities for women began to appear.' Though the finding by Neve et al. (1996) that convergence cannot be demonstrated until the 1980s might seem to contradict this, this convergence involved women in their 40s, who came of age in the 1960s. A possible explanation for the fact that convergence in Germany was already apparent in the 1950s and 1960s is that women in that country had swiftly taken on 'male' roles after so many German men lost their lives in the Second World War. The relatively late start of gender convergence in Finland could be explained by the specific temperance tradition in that country. With the decline of that tradition in the 1970s, attitudes and opinions about alcohol began to liberalise, and alcohol became more accessible to women. The drinking behaviour of Finnish women began more to resemble that of their other European counterparts (Bloomfield et al., 2001).

Yet there is still no evidence of full gender convergence in alcohol use. The assumption that women's higher labour participation rate will generally lead to convergence is also questionable. Neve et al. (1996) reported that employment did not account for any convergence, and they found no support for the thesis that women suffer more alcohol-related problems once they enter roles and positions formerly dominated by men. A possible explanation is that many women who have jobs still retain primary responsibility for childrearing. Ahlström et al. (2001) recently found that parenthood was profoundly and consistently negatively associated across societies with women's monthly alcohol consumption and their prevalence of heavy drinking. One qualification of this finding was that women with higher education, after adjustment for age, tended to consume more alcohol than those with less education, whereas no such pattern was found for men. 'This might be a reflection of the freedom higher education provides for women both economically and socially.'

What implications to these findings have for the question of gender convergence in the use of illicit drugs? First, one could argue that any narrowing of the gender gap in drug use would be more pronounced than for alcohol. In relation to alcohol, we saw that such a process occurred most distinctly in Finland at a time when the overall consumption of alcohol was rising. Whereas alcohol drinking has a very long tradition and is broadly integrated into society, the phenomenon of drug use is still relatively new. The demand side of the drugs market (theoretically at least) is still a long way from reaching the saturation point. If we assume that gender convergence is most probable in a growth market, then the gender gap is likely to narrow more rapidly for drugs than for alcohol. A second implication is that drug use, unlike alcohol use, is often a transitory activity. Although people generally first try both drugs and alcohol during adolescence, drug use usually begins somewhat later than alcohol use; and in contrast to alcohol, most people stop taking drugs once they reach adulthood. Current users of drugs are predominantly adolescents and young adults, who thereby constitute a more distinctive age category than current drinkers. This means that the norms and statuses inherent in that age group predominate in the use of drugs, and that parenthood plays little or no role. Education is a factor central to the lives of many adolescents and young adults, and usually the higher the educational track, the longer the educational career. It therefore seems plausible that education would be a crucial factor in explaining any narrowing gender gap in drug use. For one thing, young people are exposed to drugs for a longer time span while in a student role, and people with higher levels of education are also more likely to occupy social positions in which male and female roles are less polarised.

Methods

Population

Our analysis is based on eleven general population surveys conducted in four countries between 1993 and 2000: England & Wales (in 1994, 1996 and 1998), Germany (1995, 1997 and 2000), Greece (1993 and 1998) and Spain (1995, 1997 and 1999). Because the 1993 Greek survey was confined to Athens, we have limited our analysis of Greece to that region. In Germany, lifetime cannabis prevalence in respondents aged 35 and older was consistently far lower in the 'new' federal states (Eastern Germany) than in the 'old' Western states. A rapid diffusion of cannabis use in Eastern Germany appears to have occurred in the course of the 1990s – by the year 2000, lifetime prevalence rates for cannabis among 20-29-year-olds were very similar in both parts of Germany. Because drug use in the former East Germany did not manifest itself until the past decade, we have restricted our analysis here to the data from the former West Germany. Our entire analysis is based on weighted data. For an overview of the surveys involved, see the paper by Kraus et al. (2002).

Age range and birth cohorts

The lowest age categories in the surveys varied from below 15 (Greece) to 15-19 years (the other three countries). For the latter, minimum ages varied from 15 to 18. As a consequence, the lowest age category suitable for cross-national comparison was 20-24 years. The highest age categories varied from 55-59 (in England & Wales, Germany and Greece) to 65+ (Spain). Thus, the highest age category for cross-national comparison was 55-59. This means that to incorporate all four countries into a single analysis, we were restricted to an age range of 20-59. Since the survey years did not correspond in all countries, we defined birth cohorts (survey year minus age). Analysing birth cohorts instead of age categories made it possible to merge data from successive surveys within each country. Using data from uniform birth cohorts also enabled us to compare gender differences in drug use for the various countries. Following Kraus et al. (2002), we used 5-year birth cohorts.

Gender gap

We constructed two indicators for the gender gap. The first is the 'absolute gender gap', or the prevalence rate for males minus that for females. Expressed as a formula for lifetime prevalence (LTP) of cannabis, this is:

$$(\text{Absolute Gender Gap}_{\text{LTP cannabis}} = \text{LTP cannabis}_{\text{males}} - \text{LTP cannabis}_{\text{females}}).$$

The closer the absolute gender gap came to 0, the greater the gender convergence.

Our second indicator was the 'gender rate ratio', the prevalence rate for females relative to that for males. As a formula for lifetime prevalence of cannabis, this was $(\text{Gender Rate Ratio}_{\text{LTP cannabis}} = \text{LTP cannabis}_{\text{females}} / \text{LTP cannabis}_{\text{males}})$.

If this 'relative gender gap' came to 0.5, then the lifetime prevalence of cannabis was twice as high for males as for females. If the relative gender gap was 1.0, then lifetime prevalence for males would equal that for females, indicating full gender convergence.

Results

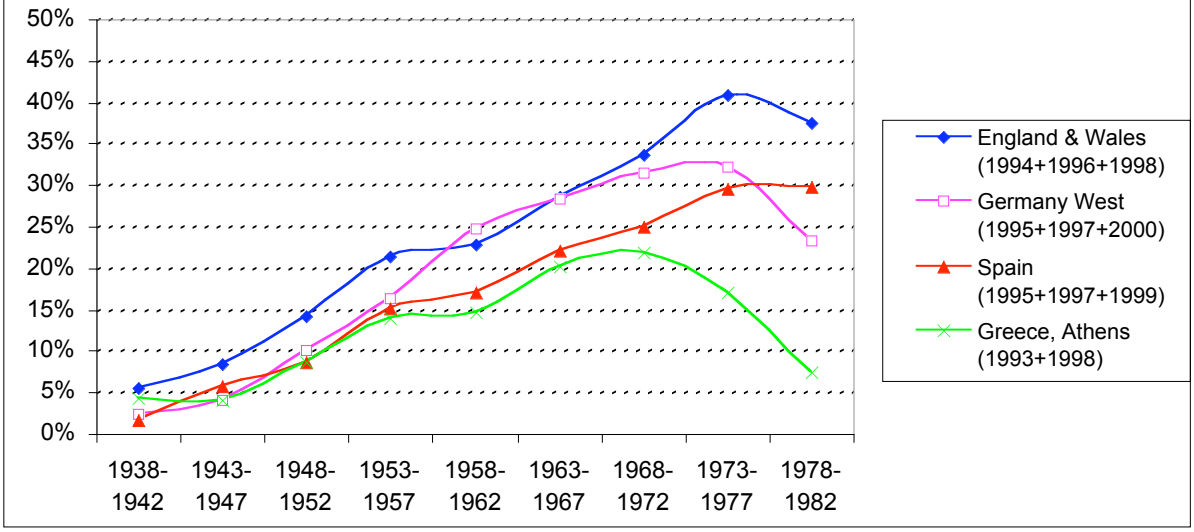
Consistent trends?

Exploration of the English & Welsh data revealed a steady increase in lifetime cannabis use between 1994 and 1998 for all age categories, and especially for 20- to 24-year-olds. The West German data showed little change between 1995 and 1997, but higher overall figures for 2000. Comparison of the 1995 and 2000 German data suggests that greater numbers of people had recently been taking their first cannabis at older ages ('catching up'). The data from Spain showed comparable patterns for 1997 and 1999, but lower lifetime cannabis use in 1995, especially for those aged 40-44. This indicates that diffusion accelerated within that particular age group, probably at two different points in time: through a growing popularity of cannabis during the late adolescence of its members (the early 1970s) as well as through a catch-up effect at a later age. The 1993 Greek data for Athens indicated a steady increase in lifetime cannabis use with decreasing age, but the 1998 survey data suggested more of a wave-like trend, with a first peak in the 40-44 age category (aged 20 around 1980). In sum, the data from successive surveys consistently indicated a **steady increase** in experimentation with cannabis from the late 1960s onwards in all four countries studied. It seems plausible that **diffusion accelerated through delayed drug initiation** in the 1990s, when more people were taking their first cannabis at older ages (even though the average age of first use may not have changed). However, it is also conceivable that the differences between successive years were largely coincidental or arose from measurement errors. For this reason it seemed wise to merge the data for the different survey years per country. As explained above, the best way to do this was to generate birth cohorts.

General trends in cannabis prevalence

In all four countries, lifetime prevalence of cannabis use was low in the older birth cohorts (figure 1), but the figures show rather steady increases in lifetime prevalence rates in all countries, with younger cohorts having higher prevalences than older ones. Although the very youngest cohort seems to show a drop in prevalence, its lower rate is artificial, because the cohort is likely to still contain many young people who have not yet used cannabis but will do so later. For the older birth cohorts, lifetime prevalence can be seen to be higher in England & Wales than in the other three countries, with the rates for Germany approaching them from the 1958-1962 cohort onwards, and those for Spain and Greece (Athens) lagging at least a few cohorts behind. This may indicate that the spread of cannabis began later in Southern Europe. For the 1973-1977 cohort, however, cannabis prevalence in Spain comes close to the West German figures, while Greek prevalence remains lower. The spread of cannabis thus seems to have accelerated earlier in Spain than in Greece.

FIGURE 1 LIFETIME CANNABIS PREVALENCE

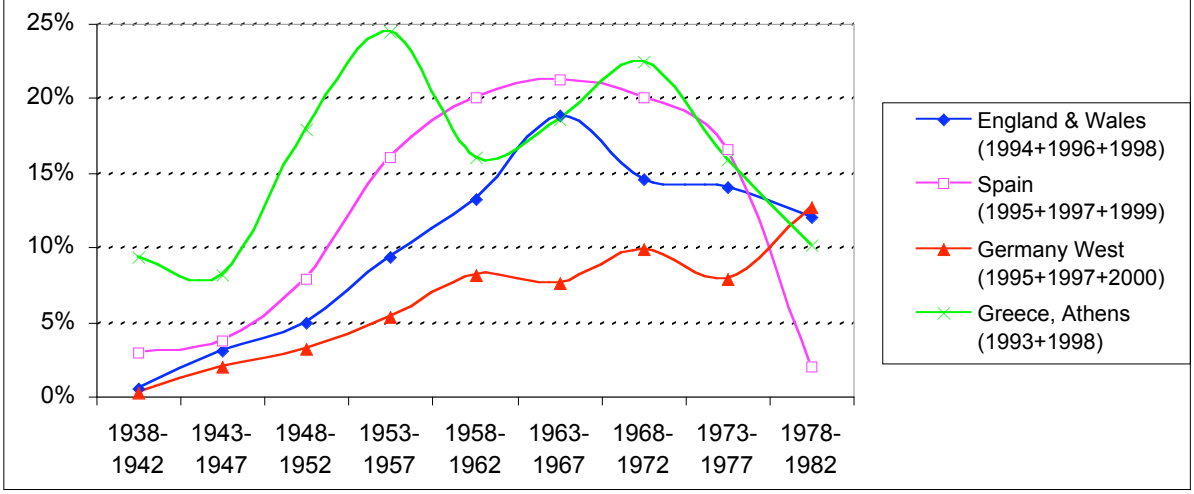


To summarise, **temporal differences are evident between European countries** with regard to the introduction and diffusion of cannabis. The national statistics discussed here may therefore reflect different stages in the cannabis 'epidemic', with England & Wales as 'trendsetters' and Greece as a 'late developer'.

The absolute gender gap in lifetime cannabis prevalence

Although the absolute gender gap appears to have narrowed sharply in the youngest birth cohort (1978-1982, figure 2), the lower figure is probably artificial, as a result of the artificially low lifetime prevalences in this cohort (as noted above, many non-users may later become users). No consistent narrowing of the absolute gender gap can be observed in the remaining cohorts. In fact, the gap generally appears to have widened.

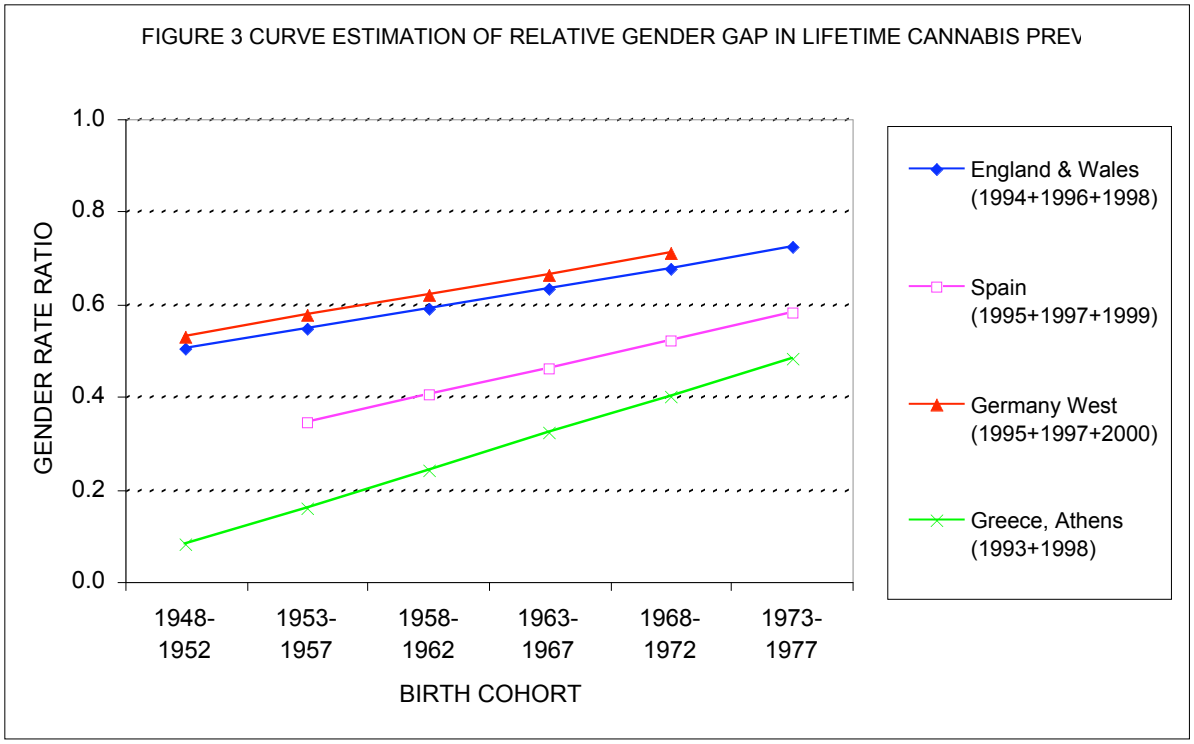
FIGURE 2 ABSOLUTE GENDER GAP IN LIFETIME CANNABIS PREVALENCI



The relative gender gap in lifetime cannabis prevalence

In calculating the relative gender gap, a first complication we encountered was that male and female lifetime prevalence rates in the oldest two birth cohorts fluctuated erratically for the different survey years. This is mainly attributable to the small numbers of reported users. An artificially small absolute gender gap may generate an artificially high, and perhaps misleading, gender rate ratio. On the other hand, in some of the older birth cohorts in the Greek surveys, lifetime cannabis prevalence for females was 0%, resulting in a gender rate ratio of 0. A second difficulty was that the gender rate ratios for the youngest birth cohort were markedly different in some countries from the ratios for the next oldest cohort. The discrepancy can probably again be explained by the fact that many potential users in the youngest cohort had not yet reached their age of first use. Such complications led us to exclude the two oldest as well as the youngest birth cohort from the further analysis, thus restricting our comparative analysis to the birth years 1948-1977.

We used curve estimation to assess the relationship between birth cohort and gender rate ratio (figure 3). The graphs show linear relationships, but some relationships can probably be better understood as loglinear or exponential.



The curve estimation procedure exposed two problems: the linear relationships were not significant for Spain and for West Germany. Closer inspection revealed that this was due to outliers: the 1948-1952 cohort in Spain and the 1973-1977 cohort in Germany. Once these outliers were excluded from the procedure, the curve reached significance. The curves for England & Wales and for West Germany ran parallel and lay very close together. Gender rate ratios increased there from about 0.5 to 0.7. The Spanish curve showed a stronger increase, from about 0.35 for the 1953-1957 cohort to almost 0.6 for the 1973-1977 cohort. The Greek (Athenian) curve was even steeper, with the gender ratio increasing from 0.1 for the 1948-1952 cohort to 0.5 for the 1973-1977 cohort.

To conclude, the relative gender gap appears to have narrowed in all four countries, although both the increases and the current levels of the gender rate ratios varied between countries. Convergence between males and females was the most pronounced in Greece (Athens) and in Spain, and the gender gap may be expected to narrow further in those countries in the years to come. The gender rate ratios for England & Wales and West Germany were a good deal higher than those for Spain and Greece, and one might be tempted to interpret the former figures as evidence that the gender gap will soon stabilise at its upper limit. However, the ongoing slow growth of the ratio suggests that the gap may continue to narrow for the time being.

The relative gender gap: other indicators of illicit drug use

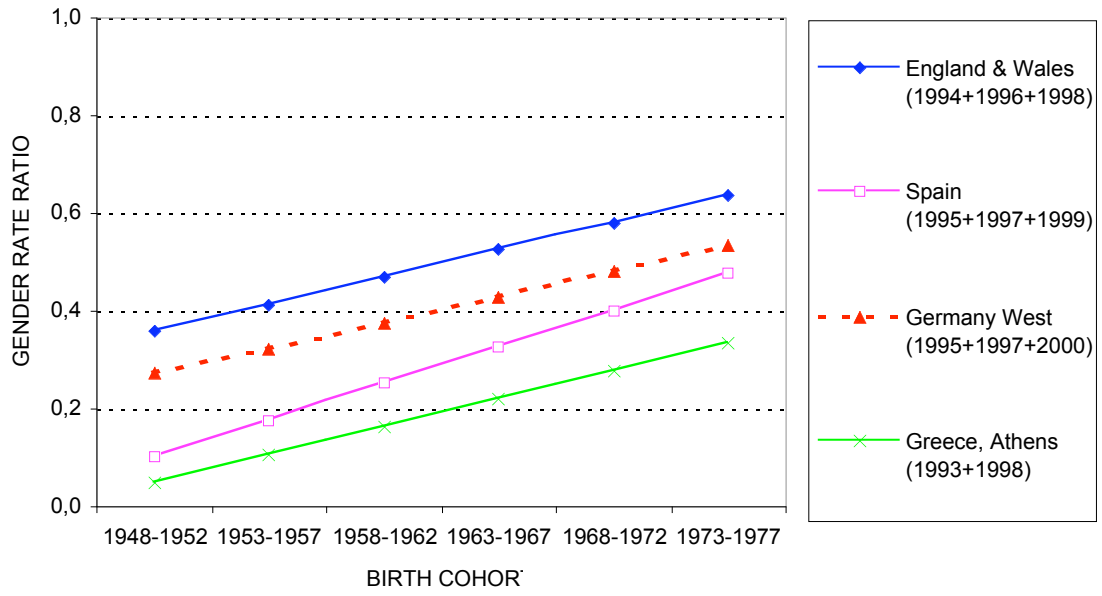
Using the same procedure (figure 4), we calculated gender rate ratios for last-year prevalence (LYP) of cannabis use

(Gender Rate Ratio $_{\text{LYP cannabis}}$ = LYP cannabis $_{\text{females}}$ / LYP cannabis $_{\text{males}}$).

The initial overall impression was that the gender gap had likewise narrowed in terms of the recent use of cannabis, though the gap was still wider than for lifetime use. However, only the Athens data produced a clear linear curve without outliers. Once again, analysis of the Spanish data identified the 1948-1952 birth cohort as an outlier (with an extraordinarily high gender rate ratio of 0.62). For England & Wales, the outlier was found in the 1958-1962 cohort (with an excessively high ratio of 0.60). Both curves were significant nonetheless. Analysis for West Germany did not result in a

significant curve, even though no outliers were present. Excluding the older and younger cohorts from the analysis did not solve the problem.

FIGURE 4 CURVE ESTIMATION OF RELATIVE GENDER GAP IN LAST-YEAR CAN PREVALENCE



Curve estimations of gender ratios for recent continuation rates of cannabis use (last-year cannabis prevalence as a percentage of lifetime prevalence) yielded no significant outcomes for any of the four countries. Nor did curve estimations in terms of the lifetime prevalence of stimulant use (ecstasy, cocaine, amphetamines) show any significant outcomes.

Discussion

The prevalence of illicit drug use remained higher for men than for women in the countries we studied[1], and the gender gap was greater for recent drug use than for lifetime use. Nevertheless, the *relative* differential in drug use prevalence appears to have diminished, at least in the case of cannabis. There were temporal differences between countries both in prevalence rates and in the narrowing of the gender gap.

Temporal differences in the diffusion of cannabis use may be observed *within* countries as well as between them. The clearest example is Germany, where the spread of cannabis in the new federal states occurred decades later than in the older states (beginning after the fall of the Berlin Wall). In Greece, the lifetime prevalence of cannabis use was higher in the Athens metropolitan area than nationwide, especially among younger respondents. This probably reflects the more general phenomenon of delayed diffusion into rural areas.

The analyses conducted so far have made no allowances for dissimilarities in national drug policies, but the available literature suggests that such policies have little influence on the prevalence of cannabis use. Recent general population surveys in several EU member states have reported similar figures for lifetime prevalence of cannabis, roughly 20% to 25% (EMCDDA, 2001). Another factor that has not yet been taken into account is the variation in background characteristics between countries, and within countries, over longer periods of time, such as changes in the ethnic composition of the population. Most important, we have not yet incorporated education levels and labour participation into

[1] In most ESPAD countries (European School Survey Project on Alcohol and Other Drugs), the use of illicit drugs is more prevalent among young males than among young females. However, in nearly all countries that show increasing rates of drug use, an upward trend is visible for both males and females (Hibell et al., 2000).

the analysis, even though education clearly came forward in the studies cited in the introduction as an key explanatory factor for the narrowing gender gap in alcohol use.

References

Adler, F. (1975) *Sisters in crime.* Prospect Heights: Waveland.

Ahlström, S., Bloomfield, K. and Knibbe, R. (2001) Gender differences in drinking patterns in nine European countries: Descriptive findings. *Substance Abuse*, 22(1):69-85.

Bloomfield, K., Gmel, G., Neve. R., and Mustonen, H. (2001) Investigating gender convergence in alcohol consumption in Finland, Germany, the Netherlands, and Switzerland: A repeated survey analysis. *Substance Abuse*, 22(1):39-53.

EMCDDA (2001) *Annual report on the state of the drugs problem in the European Union.* Lisbon: European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).

Hibell, B., Andersson, B., Ahlström, S., Balakivera, O., Bjarnasson, T., Kokkevi, A., and Morgan, M. (2000) *The 1999 ESPAD report. Alcohol and other drug use among students in 30 European countries*. Stockholm: The Swedish Council for Information on Alcohol and Other Drugs (CAN).

Knibbe, R.A. and Bloomfield, K. (2001) Alcohol consumption estimates in surveys in Europe: Comparability and sensitivity for gender differences. *Substance Abuse*, 22(1):23-38.

Korf, D.J. (2002) Dutch coffee shops and trends in cannabis use. *Addictive Behaviors,* 27 (6): 851-866.

Kraus, L., Bloomfield, K., Augustin, R., and Reese, A. (2000) Prevalence of alcohol use and the association between onset of use and alcohol-related problems in a general population sample in Germany. *Addiction*, 95(9):1389-1401.

Kraus. L. et al. (2002) *Analysis of age of first cannabis use in Germany, Greece and Spain.* Report to European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).

Neve, R.J.M., Drop, M.J., Lemmens, P.H., and Swinkels, H. (1996) Gender differences in drinking behaviour in the Netherlands. *Addiction*, 91(3):357-373.

Johnston, L.D., O'Malley, P.M. and Bachman, J.G. (1996) *National survey results on drug use from the monitoring the future study, 1975-1995.* Rockville, MD: National Institute on Drug Abuse.

Sælan, H., Møller, L. and Køster, A. (1992) Alcohol consumption in a Danish cohort during 11 years.

Scandinavian Journal of Social Medicine, 20:87-93.

Steinhausen, H.C. and Winkler Metzke, C. (1998) Frequency and correlates of substance use among pre-adolescents in a Swiss epidemiological study. J Child Psychol. Psychiat., 39(3):387-379.

3. RE-ASSESSMENT OF CORE ITEMS AND DATA COLLECTION METHODOLOGY

New core items

Based on the experience of the harmonisation process and the joint analysis exercises two proposals have been formulated for addition and amendment of the EMQ. The proposals below have been discussed and in principle adopted by the Expert Group and they have also been taken into account in the harmonisation process of the deposited survey data.

Other potential additions and amendments to the EMQ have also been discussed in the Expert Group, but these discussion have not yet resulted in concrete recommendations. With regard to the discussed item of "availability of drugs" EMCDDA has in the meantime launched a separate project to assess the options and pro's and con's to include this item in general population surveys.

Age of first use

Proposal

Include the variable **AGE_…** "Age of first use of <name of drug>" for all drug types covered by the EMQ.
The corresponding model on would then as: "At what age did you take <name of drug> for the first time ?

At present the EMQ only recommends to ask for age of first use of cannabis. The main reason for this limitation was the assumption that the key interest would be the age of initiation to any illicit drugs, which in the vast majority of cases is cannabis, and the aim to restrict the number of questions of the EMQ. In retrospect this seems not justified by the growing policy interest in incidence and initiation with regard to different types of illicit drugs. At the same time it has already become common practice in many prevalence surveys to ask for age of first use for all drugs covered in the survey. Inclusion of the variable "age of first use" for all drug types allows the assessment of "drug careers" and it is a crucial variable in analysing survival rates and differential developments of drug use by cohorts as demonstrated in the joint analysis studies.

Last month frequency of use

Proposal

Assign the values of the variables LMF_… "Last month frequency of use of < name of drug> " for all drugs covered by the EMQ as either:

Number of days having taken <drug> in the last 30 days
OR
20 times or more
10-19 times
4-9 times
1-3 times

The corresponding model question would then read as: "During the last 30 days, on how many days did you take <name of drug>" ?

At present this item is included with the following model categories: daily or almost daily, several times a week, at least once a week, less than once a week.

The argument for this amendment is more of a practical nature. The original categories with colloquial names were chosen to accentuate that we are only interested in differentiation along ordinal categories varying from high to low and we did not assume that respondents can recall current drug use in a book-keeping manner, where for example the answer “on 17 days” can really be interpreted as la lower frequency than the answer “on 18 days”. In retrospect we acknowledge however that most surveys ask for exact number of days or distinct categories and that the categories in the current EMQ suppose a following order of presenting the answers in order to choose the first that applies, which would not be applicable in all survey modes.

Proposing this amendment however does not imply that we have changed our original arguments. In other words, the responses indicating exact number of days or distinct exact categories can only be interpreted as ordinal variables, but such “exact” categories will make harmonisation efforts more easy and avoid disputes about overlaps between answer categories.

Data collection methodology

In project CT.97.EP.08 about the harmonisation of prevalence surveys the methodological discussions were largely focussed on modes of interviewing and sampling procedures had only been covered in a summary manner. Although mode differences should be taken into account when comparing different surveys, differences between modes often relate to in sampling frames and sampling procedures which are inherent to the interview modes. In the context of the re-assessment of data collection methodology the Expert Group has expressed a demand to include more explicit recommendations about sampling procedures into the standards for general population surveys. Following this demand we have we have elaborated the suggestions for good practise in sampling. This resulted in the revised text below, which in the meantime also has been incorporated in the final report of project CT.99.EP.08 B.

Sampling Design

As in other domains of social science general population surveys on drug use are always executed among a sample of the entire target population, since it is neither practicable nor cost- and time-effective to interview every single individual in the population. The purpose remains nevertheless to make conclusions about attributes or behaviours of the whole target population. Sampling is a critical factor in any survey design that determines to what extent the survey results allow reliable inferences within acceptable margins of error to the population.

A sample design should deal with both the selection of individuals to be included in the sample and the process of estimation of population values from the sample values. Selection and estimation are interlinked as selection rules affect the methods of estimating population values and the precision required for population estimates influence the selection rules. The precision needed depends on the general survey aims and selection depends on possibilities or feasibilities to identify and approach the members of the universe of the target population, which in turn depend on survey mode and in particular survey budget. In principle therefore, survey design and sampling design should go hand in hand.

Sampling

The assessment of population estimates from sample data requires that the sample is « representative » of the total population. Careful selection can make a sample more or less representative. This is best achieved by *probabilistic sampling,* whereby each individual of the population has a known non-zero probability of being selected allowing inferences to population values by means of statistics computed from the sample data without having to make assumptions about the distribution of the survey variables in the population. In prevalence studies, as in social studies in general, we usually cannot make such assumptions and as a consequence probability sampling should be considered almost mandatory.

The basic selection method in probability sampling would be simple random sampling in the universe of the target population. However this might not always be possible or practical. For examples: the sampling process might not be perfect because we do not have correct information about the target population; operational aspects of actual survey execution can distort a theoretical good sample; budget constraints might enforce alternative strategies; survey aims may call for various levels of precision of estimates for different segments of the population; etc.

In many cases therefore simple random selection will be or has been replaced by other methods or combinations of methods. Common methods are :

- simple random sampling in which each individual of the population has an equal probability of being selected ;
- varying probability sampling, in which the probability of being selected varies according to the magnitude of another variable (e.g. household size, city size)
- stratification : a priori selection of subpopulations from which samples are drawn
- multi-stage sampling, in which first groups of individuals are sampled (e.g. people in a certain area, city blocks, households) and individuals are selected in the final stage within a group
- multi-phase sampling in which a final sample is taken from a previous sample that provides information to improve the final selection.

The methods applied dictate the computation of statistics to estimate the population values and the statistical errors or precision of the estimates. These various methods are explained in any textbook on survey sampling and the techniques involved do not require further discussion here.
Although we may be able in the context of improving European comparability of prevalence surveys to harmonise survey aims and to set criteria for precision of population estimates, it will be difficult to create uniform conditions for sampling applicable to all countries. This implies that for the time being we cannot identify a particular method as a European standard of sampling for general population surveys on drug prevalence, other than the requirement that probability sampling should be applied.
Nevertheless, we can present some general considerations about three aspects of sampling that could help future surveyors in elaborating their sampling designs : sample size, sampling frames, implementation of sampling rules.

Sample size

The size of the sample is a critical factor with regard to the precision of population estimates resulting from survey data. It is also a critical factor in the costs of surveys.

Required sample sizes should be determined before starting any survey. In probabilistic samples that are small compared to the target population, they can be calculated from the following general formula:

$$1{,}96*SE(p) = \sqrt{\frac{P*Q}{n'}}$$

Where SE(p) is the error margin (in percentages) of the population estimate, the factor 1,96 is taken from a table of the normal distribution at the usual 95% confidence interval, *P* is the expected population percentage (e.g. prevalence measure), *Q=(1-P)* and *n'* is the estimated sample size. For a large target population, normally the case in national surveys, n' equals n, the real sample size. The sample size n can then be calculated if we decide on an acceptable margin of error and have some notion about expected prevalence rates. In table 1 we calculated n for different levels of precision and different expected prevalence rates.

It should be noted that the formula above supposes simple random sampling and needs modifications in other sample designs; calculation formulas can be found in standard textbooks on survey sampling.

Acceptable levels of precision depend on survey aims and on expected prevalence levels. If we expect actual prevalence levels of say 1% or less, we might find a margin of error of 2% (i.e. a population estimate of 1% ± 2%) not acceptable, whereas if we expect a rate of 40% even a margin of 5% might be acceptable to some. In general we would not accept that the ratio of the error margin to the population estimate is more than 0,5 or even less. So 40% ± 5% is acceptable (ratio = 0,125), but 1% ± 2% with a ratio = 2,0 is not. In table 1 sample sizes where the margin of error relative to the

expected prevalence rate equals 0,5 or more have been shaded and these sample sizes will not result in prevalence estimates with an acceptable margin of error; the minimum sample sizes at different levels are printed in red.

Drug use is still a relatively rare phenomenon, in particular if we talk about last 30 days prevalence (LMP) and for most drugs expected LMP-rates will be very low. Tracing such low rates within acceptable margins requires large sample sizes, particular if we also consider survey aims that call for prevalence rates for subgroups.

From the perspective of EMCDDA a general survey aim should be to obtain population estimates corresponding to the report format of the key indicator on prevalence rates from general population surveys. In other words estimates for the drugs included in the indicator and for males and females of each 10-year group between 15-64. Theoretically this means that we a minimum sample size should apply for each of the ten age-gender groups. According to table 1 a sample size of over 1500 is needed to assess a population rate of 1% ± 0,5%, which might be considered just acceptable. If this should apply to each age-gender group the minimum sample size raises to at least 15.000. In a simple random sample this figure will be higher as we have to ensure a priori that we get a minimum of 1500 for the smallest age-gender group in the population. This further increase of sample size can be avoided by stratification, but this is only possible if we have a sample frame that allows stratification by age and gender (see below).

Large sample sizes also increase survey costs. In practice therefore we have to compromise on the above precision requirements, in particular in case of low prevalence rates and eventually also for subgroups. It should also be noted that figures which as a one-time prevalence estimate will not result in acceptable margins, can still be used for trend and multivariate analysis. For example, rates of 0,8% ± 2% in one year increasing to 1,1% ± 2% in later years might in each individual year not constitute an acceptable population estimate, but they might nevertheless reveal a statistical significant trend.

Table 4. Sample sizes for different levels of expected population prevalence rates and accepted margins of error; both in percentage points. Shaded cells indicate sample sizes where the ratio of expected prevalence and error margin is > 0.5 (50%). Red figures indicate minimum sample size at given combination of expected prevalence rate and margin of error.

Margin of error (in %)	**Expected prevalence rates (%)**					
	50	**25**	**10**	**5**	**1**	**0.5**
0.5	38416	28812	13830	7299	**1521**	764
1	9604	7203	3457	1825	380	191
2	2401	1801	864	**456**	95	48
3	1067	800	384	203	42	21
4	600	450	216	114	24	12
5	384	288	**138**	73	15	8
10	96	**72**	35	18	4	2
15	43	32	15	8	2	1
20	24	18	9	5	1	0
25	**15**	12	6	3	1	0
30	11	8	4	2	0	0

Another factor to consider in defining the required sample size is the expected non-response. At the end of the day population estimates are calculated for survey variables, which values can only be assessed for the response. Levels of non-response vary between countries and survey modes and what we call minimum sample sizes should actually be read as minimum sizes of the response. In this context we should be aware of a possible confusion of terminology as studies often report net response as the sample size.

In reality sample sizes are often decided upon a mixture of more or less explicit arguments and budget considerations usually play a major role. However, if we take EMCDDA's report format and a precision of at least 50% for prevalence rates of 5% (i.e. 5% ± 2,5%) as a minimum requirement, the minimum

response size will be at least 10 * 456 = 4560 and accounting for optimistic response levels of 60-80% this minimum sample size will be between 5700 and 7600. Even with such substantial samples we have to take for granted that we don't get population estimates with an acceptable margin of error when they fall below 5% (cf. Table 1), which applies for most drugs, in particular with regard to recent or current use.

Sampling frames

Probability sampling requires that we have an enumeration of the target population from which we can draw a sample. Without such a sample frame we cannot select individuals at random.

It may be obvious that one should select a sample frame that provides the best possible coverage of the target population. The best possible frame might be different in each EU Member State. Moreover, the optimal choice may depend on survey mode and budget constraints.

In theory, for national population surveys, the optimal frame would be a list of all the residents of a country that belong to the target group. Not all countries have such lists however and if they do the lists might not be complete, accurate, up-to-date or accessible for survey purposes. Examples of such lists are population registers and election registers, though the latter will be limited to people eligible for voting, which excludes the lower age group of 15-18 year olds of the recommended target population of prevalence surveys.

A complicating factor often is that these registers are not centralised, but can only be addressed at the level of many administrative units, which for practical reasons may require two-stage sampling (first on administrative units, then on individuals).

Telephone registers are also widely used, in particular since CATI has developed into a fast an low-cost survey mode. Apart from complications due to the CATI mode itself and often unknown coverage of the population by phone lists accessible for researchers, using telephone registers always implies a two-stage sampling process. In the first stage random phone numbers are selected, but as they connect to households, a second stage is needed to select an individual within the household. This final selection however is left to the person who answers the phone (albeit with a randomisation instruction provided by the interviewer) and may create an uncontrollable bias towards a random sample.

If registers are not available or cannot be used, a general accepted alternative is to create a sample frame of its own. A common example is so-called random route sampling, which consists of a multi-stage combination of methods. For example : random selection of areal units, with or without proportionality to (population) size, followed by random selection of starting points and routes for random walks within units, systematic selection of dwellings along the routes and systematic selection of individuals among the inhabitants of the dwellings. There are many variations on this model and the selection processes and rules can become quite complicated and may involve the construction of auxiliary enumeration lists (e.g. dwellings within blocks, inhabitants of dwellings).

It should be remarked here that the common practice to select only one person in each household (as in the case of CATI surveys) provokes an under-representation of people in large households. In many cases this cannot be corrected in subsequent weighting procedures as information about household sizes in the population is often not available. A feasible solution however is to ask for household size in the survey and include inversely proportional weights in the assessment of population estimates.

The important issue here is that no sample frame is perfect and perfect probability sampling does not exist. Any sample frame will have imperfections which need to be addressed. Sometimes this can imply the introduction of remedies to overcome or reduce imperfections, for example by making additional samples based on other frames to select individuals which might not be represented in the original frame. Acknowledgement of imperfections of sample frames always implies the making of assumptions about the effects on population estimates.

In any case imperfections, remedies and their likely effects need to be accounted for in the sampling design and should be reported in the technical survey report. In this context it is useful to distinguish between frame bias and frame errors.

Frame bias relates to technical or theoretical imperfections with regard to the accurate coverage of the target population. For instance, the fact that some people by definition are not included in the frame or by experience or general assumption are missing or underrepresented in the frame. Examples might

be the exclusion of foreigners from electoral registers or homeless, prisoners and soldiers from population registers (they may registered, but might not be found at their registered addresses).

Frame errors refer to imperfections encountered in the field, for example, non-existing addresses, people not residing (anymore) at their registered address, etc.

Implementation of sampling rules

The proper choice of sampling frame and sampling methods does not necessarily result in a correct probability sample. The actual implementation of operational sampling procedures and rules in the field can play a major role as well.

Fieldwork is human work and prone to errors and mistakes. So a perfect sample on paper might become less perfect in the execution of the fieldwork. Sometimes practical circumstances can enforce deviations of sampling rules (e.g. deviation from pre-assigned random routes due to blocked roads), interviewers in the field can make mistakes (e.g. address not found, wrong address selected) or some stages in the sampling are left to the potential non-controllable bias of others. For the latter see the example in telephone surveys above, but the same can apply when the sampling rules include enumerations of dwellings in housing blocks taken from key informants.

An important and often neglected deviation of the theoretical sample design may be induced by the economics of fieldwork. Fieldwork contracts are usually based on the achievement of a set number of completed interviews ; interviewers are usually paid for completed interviews and for practical reasons they receive multiple sample addresses at the same time. As there will be many addresses where at first call nobody can be approached, the initial waves of the survey will result in responses from people who are more likely to be at home than others. The initial waves may already result in the contracted number of completed interviews. This result is biased towards the likelihood of people being often at home. Common remedies are to instruct interviewers to visit addresses at different times and to hand out only limited numbers of addresses at the same time. But at the end of the day most agencies will stop the fieldwork when they assume that the completed interviews are representative for the target population on the basis of some criterion variables (usually gender, age and/or locality). The net effect however is that quotas creep into a nice probability sample, leaving the bias of the « not-being-at-home » characteristics of the sampled population.

This type of deviation is difficult to avoid. Remedies can be costly and field agencies are not always willing or able to reveal the details of actual implementation of fieldwork. It is not in the interest of agencies or interviewers to state that they have approached mainly the « easy » addresses (i.e. those directly resulting in completed interviews) and the client is usually happy if the response is proportional to known distributions (age, gender, etc.) in the target population.

Asking for statistics about date and time of interview and realisation of interview at first, second or later attempts can indicate the possible amount of such bias and should therefore be standard practice.

Summarising, a perfect probability sample on paper is not always a perfect probability sample in real life. Even if we accept that imperfections due to implementation cannot be fully avoided, their possible effects need to be addressed in the sampling design and accounted for in the technical report.

4. IMPLEMENTATION OF THE KEY INDICATOR

Implementation of the key indicator by countries

Survey plans and work plans of National Focal Points have been continuously monitored in cooperation with EMCDDA. The latest overview based on the situation of early June 2001 is added to this report as Annex II. Although a lot of work still needs to be done, many countries already comply to a large extent with the EMQ most countries are making progress in this direction.

The key obstacles today are not so much the comparability of the survey content (questions, categories) but the comparability of survey methodology, in particular interview mode and sampling. For this reason the development of generally accepted standards on methodology should be a main issue in the further pursuing of harmonisation of national general population prevalence surveys.

Aggregated data

According to the project design the NPSD-EU will also store and make accessible aggregated data in the table formats of EMCDDA's key indicator on general population prevalence survey data. These tables are provided regularly as Excel files by the National Focal Points in the framework of their annual reports to EMCDDA.

At first we developed a uniform table format that could appear as pages to view and download from the website of EMCDDA. In the meantime however EMCDDA has developed a database structure that allows dedicated queries on aggregated data in MS Access and may generate tailor-made tables with more flexibility in presenting comparable information. Under the terms of contract the contractor has assisted EMCDDA in the data entry and testing of a data mask for aggregated general population survey results into EMCDDA's database of indicator data. The test has been successful and will allow direct update of the aggregated tables included in the NPSD-EU.

In theory the aggregated data can also be generated from the deposited (derived) survey datasets. This may reduce the workload of the National Focal Points and also ensure more consistency between the aggregated information of individual countries. The development of model routines for this aggregation process however was not included in the terms of contract of project CT.00.EP.14 .

Advise to national focal points

Comments and advises have been given to the research team responsible for the development of the first Italian national survey on drug prevalence. At present Italy is carrying out a pilot survey, which questionnaire complies with the European Model Questionnaire.

Meetings of the Expert Group on Population Surveys

During the execution of project CT.00.EP.14 two meeting of the Expert Group have been organised in collaboration with EMCDDA. Both meeting were held in Lisbon, the first in February 2001, the second in May 2002 .

ANNEX 1: MINUTES EMCDDA Annual Expert Meeting (23-24 May 2002)

Key Indicator

"Extent and patterns of drug use among general population" (Population Surveys)

ANNUAL EXPERT MEETING

May 23rd and 24th 2002. EMCDDA, Lisbon

MINUTES

EMCDDA (Julian Vicente)

Agenda: Draft agenda is enclosed in **Annex 1**.

List of participants: List of participants is enclosed in **Annex 2**.

1. Progress in the implementation of the Key Indicator

Julian Vicente and Richard Hartnoll (EMCDDA) gave and overview of work done in previous years in the development of the key indicator and informed participants of progress obtained during 2001 and first months of 2002.

It was recalled that the EMQ is primarily based on concepts more than on wording (although there is a "demonstration questionnaire" translated to several languages) and that the EMQ can be used to design broader national questionnaires, or to report and harmonise data from previously existing questionnaires (post-ex harmonisation). Also it was recalled that the EMQ is a minimum set of items, and national questionnaires can collect more information, according to national information needs and objectives of their surveys.

Participants were informed that the five Key Indicators were adopted by the EMCDDA Scientific Committee in December 2000 and by the Management Board in September 2001. In addition, a Council Resolution of December 2001 urged the Member States and the European Commission to support the implementation of the key Indicators in collaboration with the EMCDDA. In response to this resolution, the Commission and the EMCDDA are looking for ways to reinforce the legal framework (competences covering Community Statistics and Public Health), and alternatives for practical ways to collaborate with Eurostat and Sanco are being analysed.

2. Progress in the implementation of the Key Indicator by each Member States

Each Member State (and Norway and Poland) presented briefly the progress obtained in their country during the last year. The presentations were based in a "National Form" prepared by national experts beforehand, including relevant contact persons, achievements, problems and future perspectives. The National Forms from all countries were copied and distributed to all participants.

A summary of main national developments (up to May 2002) is presented in **Annex 3**.

It was stressed that significant progress is being obtained, and that many countries have already conducted (or plan to conduct shortly) high quality surveys or are establishing surveys series. It was acknowledged that in many countries National Focal Points are playing a key role on population surveys promotion (implementation, improvement of quality, and establishing consistent series).

3. Assessment of the current compliance on national questionnaires with the EMQ

Based on the National Forms filled in by countries, the compliance between national surveys and the EMCDDA set of common core items (EMQ) was assessed. The compliance regarding illicit drugs and the basic prevalence measures can be considered very high. Compliance regarding respondents attributes and alcohol and tobacco items is high, although a detailed analysis of variable's categories is needed. Compliance regarding pharmaceuticals can be considered as medium and also categories equivalences need closer analysis. Finally, compliance is considered lower regarding opinions and risk perceptions.

Ruud Bless explained his experience in harmonising several national surveys during the project to develop the EMCDDA Databank. He stressed that although in many national surveys the EMQ items are present, the exact correspondence of categories is difficult (in particular respondents attributes) and close attention has to be paid to this issue.

Annex 4 presents an assessment of compliance between national questionnaires and the EMQ on an item by item basis.

4. Information of developments from Eurostat (Health Interview Surveys) and ESPAD project (School Surveys).

Jaap van den Berg gave a broad overview of Eurostat work in the field of Health Interview Surveys. Eurostat has a Task Force on health and health-related survey data (TF His/Morb). This Task Force is analysing existing survey instruments in Member States (a database has been created), developing recommended instruments and collecting available data. The Task Force has identify 18 health topics, being drug use one of them. The European Model Questionnaire will be taken into account in the development of recommended instruments by Eurostat. EMCDDA prepared a document for the Eurostat Task Force meeting of April 2001 based on the Key Indicator guidelines (European Model Questionnaire).

Björn Hibell participated in the meeting as Swedish National Expert and as ESPAD coordinator. ESPAD is a European project on school surveys coordinated and supported by CAN (Sweden), and supported also by the Pompidou Group of the Council of Europe. Mr Hibell explained that the ESPAD school survey has been carried out already twice (1995 and 1999) at European level using the same questionnaires and methodology. The next survey will be conducted in 2003 with participation of more that 30 European countries, maintaining basically the same questionnaire.

5. Development of the EMCDDA Databank on population surveys

Ruud Bless (Quinx Research, project contractor) summarised briefly the progress in the EMCDDA Databank on population surveys.
Ten databases have been deposited by four countries; Germany (2), Greece (2), Spain (3) and U. Kingdom (3). Some countries had offered to deposit school surveys but these surveys are not covered by the present project. During the meeting, Finland informed that the National Institute of Statistics agreed with depositing their national surveys, after discussion

and agreement of its Ethics Committee. Also Ireland expressed its interest in depositing the next national survey.

Mr Bless recalled the difficulties experienced during the process. These may be in part because many drug researchers are not familiar with depositing data in their national archives for social sciences. The detailed Licence Agreements, which were aimed to give full guarantees to national data owners and were based on UK Data Archive long experience, seem to have cause concerns in some countries.

Mr Bless also recalled the detailed work that was necessary to harmonise the national databases to make them compatible with the EMQ (post-ex harmonisation). This work highlights the need to have reports with detailed documentation about all methodological aspects of population surveys, and common practices regarding technical aspects such as missing values, etc

Hilary Beedham (UK Data Archive, associated to project contractor) gave an overview of the rational of data archiving. Ms Beedham recounted that many data owners have concerns about protection of intellectual property rights, fear of criticism from third parties about data quality, and data protection issues. In contrast with these concerns, she stressed that Licence Agreements strongly protects the property rights of data owners, that data archives facilitate improvement of data quality, and that data protection can be ensured through technical methods and through the contractual commitments of archives and data users not to misuse the data. Some researchers may find it easier to share data informally, but data archives facilitate access to data by doing substantial preparatory work.

See **Annex 5** with a summary of Ms Beedham presentation.

Finally Julian Vicente and Richard Hartnoll informed participants that after the developmental phase, the Databank will be located at the EMCDDA offices in Lisbon (See also point 7). They thanked Mr Bless (Quinx Research) and Ms Beedham (UK Data Archive) for their essential contribution to the development of the Survey Databank, and expressed EMCDDA interest to continue information exchange and find ways for future collaboration.

6. Presentation of first results of joint analysis base on EU Databank

Ludwig Kraus and Dirk Korf presented the first results based on the joint analysis of national surveys deposited on the EMCDDA survey Databank. They used for their analysis the harmonised data file created with common procedures, following the EMQ items.

The results had been presented and discussed in a small working group (22 May) that included EMCDDA staff, researches, experts from countries that had deposit their data, and other experts, with the aim to act as a focused “advisory group” to help researchers in validation and interpretation of results.

Mr Kraus presented the analysis of age of first cannabis use in Germany, Greece and Spain (deposited surveys from U. Kingdom do not include age of first use).

Mr Korf presented and analysis on changes over time in gender differences in illicit drug use in Germany, Greece, Spain and U. Kingdom.

After the presentations a general discussion took place. It was stressed the need to conduct high quality national surveys to ensure quality of results of joint analysis based on these surveys. It was recalled that surveys deposited in the Databank are considered of high quality, conducted by experienced researchers, and that the harmonisation process

contributed to the quality control of data analysed. Also, it is expected that research conducted on Databank data will follow the standard scientific review procedures.

7. EMCDDA standard reporting process and the EMCDDA Integrated Key Indicator Database

Norbert Frost (EMCDDA) presented an overview of the first developments of the EMCDDA Epidemiological Database System. This system includes, in one hand, the information collected on a regular an systematic basis through the REITOX Standard Tables (reported every year by National Focal Points together with their National Reports) and, in the other hand, the ad hoc databases generated by specific projects that have developed and tested the Key Indicators.

The EMCDDA Databank on Population Surveys will be a specialised database within the EMCDDA Database System, maintaining the legal and technical safeguards developed during the projects that created the Databank.

The EMCDDA will maintain a reinforced collaboration with those countries that deposit their data for further developments and utilisation of the Databank.

8. Summary and conclusions about the KI implementation and progress

Julian Vicente presented a summary and conclusions of discussions of previous sessions. These conclusion are integrated in Point 11 (Summary and conclusions of the meeting).

9. Worksession 1 (Assessment of existing methodological guidelines of the Key Indicator)

This session had as objective to discuss the Section 4 (Good Practice in Survey Methodology) of the Handbook that includes the EMCDDA guidelines for the Key Indicator. This Section of the Handbook does not intend to prescribe a unique survey methodology, but rather describes alternatives for different issues, discussing advantages and limits of the different options. In some cases, the alternatives may be conditioned by the eventual different objectives of the survey.

The session was initiated with two “case studies” that presented innovative methodological approaches: the U. Kingdom (British Crime Survey) regarding sampling methods and The Netherlands (Licit and illicit drug use in The Netherlands) regarding methods of data collection.

> Tom Bucke and Rebecca Aust presented “Recent developments in measureing drug consumption using the British Crime Survey”. These included important increases in sample sizes and ad-hoc sampling strategies. From the 2001 sweep of the BCS, the sample size was increased to 40.000 people, including a boost sample of 4000 non-white people and a youth boost of 1.600 additional respondents (16-24 years olds). Interviewing will take place on a continuous basis. The aims of sample increases are to allow a greater detailed analysis, and the need to measure changes robustly to assess government’s drug strategy targets.
>
> Manja Abraham presented “A Multi Method Approach to Measuring the Prevalence of Drug Use in the Netherlands”. Previously used CAPI methodology presented the difficulties to find interviewers and the reduced willingness of people to participate (in

particular in big cities). With the MM protocol, sampled people can choose a preferred method of interview (self completion of paper questionnaire, on disk or through internet) with telephone or mail follow up of non respondents. The MM method was considered feasible although complex to manage by the fieldwork agency. In addition, the MM method produced higher prevalences than the CAPI method.

Following "case-studies" presentations and discussion, participants were asked by Mr Bless to comment and discuss Section 4 (methodological questions) of the draft Handbook, in particular regarding detection of concrete mistakes or identification of gaps to be developed in future editions of the handbook.

Some participants explained that they had not had the time to revise the Section or were not aware of the Worksession requirements. Therefore, participants were asked to send, if they wish, further contributions in written form within four weeks.

10. Worksession 2 (Perspectives for future developments of the Key Indicator)

Julian Vicente (EMCDDA) explained that future developments of the key indicator will be considered with caution, maintaining stability of existing guidelines and building on them. Developments should have a clear added value in terms of understanding drug use, or policy developments. Proposals may be elaborated, assessed and tested by subgroups of the EMCDDA expert group.

There is not yet a formal procedure to modify guidelines of the Key Indicator. The first step will imply the experts' consensus based on the relevance of the topic, scientific literature and empirical testing. Some developments may become part of the core guidelines and others may remain as recommendations.

Small modifications of two existing items of the EMQ were proposed and adopted. The rational was to adapt the items to their practical formulation in most existing national surveys. The items concerned were the "Age of first use" (to be asked in all illegal drugs) and the "Last month frequency of use" (wording of categories revised, although with close correspondence with previous categories).

Annex 6 presents the modified variables.

Other proposals were discussed, regarding eventual recommendations of standard geographical breakdowns in national surveys designs, and regarding recommendation to oversample subgroups of the population where illegal drug use concentrates (e.g. young people). It was considered that more elaboration was necessary before taking decisions.

Chloe Carpentier (EMCDDA) presented the results of a first expert meeting (21-22 May, Lisbon) to develop a module on drug availability within survey questionnaires. The meeting involved a subgroup of the group of national experts. After an inventory of existing questions, literature searches and, when possible, secondary analysis of existing data, a draft proposal for a module will be developed. This proposal will be tested and presented to the complete EMCDDA expert group for discussion and eventual adoption.

The report of this expert meeting was distributed (19th July 2002) among all members of the EMCDDA expert group on population surveys, with an invitation to voluntary participation to further analysis of national data.

11. Summary and conclusions of the meeting

Julian Vicente and Richard Hartnoll summarised the discussions of the previous sessions regarding progress in Key Indicator implementation, compliance of national surveys with EMQ, development of the EMCDDA Databank, and future developments of the key indicator.

The *basic guidelines* of the Indicator are already developed after five years of work. The guidelines have been developed through an expert consensus method, by means of iterative discussions. Some aspects of the guidelines were field tested (pre-test of "demonstration questionnaire"). A parallel project carried out empirical research on methodological aspects (effects of data collection mode).

The guidelines were endorsed by the EMCDDA Scientific Committee and Management Board, and they have a wide acceptance among national experts.

These guidelines can be improved on a step-by-step basis, after identifying the gaps. Developments should be made cautiously, and justified in terms of improved knowledge or policy relevant needs. Subgroups of the whole expert group may be involved in formulation of new developments. The EMCDDA Databank may be useful to obtain empirical evidence in some developments.

In several countries, substantial progress has been obtained in recent years, such as conduction of the first national surveys, improvement in quality –including increases in sampling sizes-, or establishment of series of national surveys.

However, in some countries substantial work is still necessary. In addition, participants and EMCDDDA considered that political and institutional changes may influence future progress (decrease funding, interruption of series of surveys, etc), and therefore it is necessary to consolidate progress obtained.

Compliance of national surveys with the EMQ is good in items on illegal drugs, and more limited in other areas (legal substances and respondent attributes). In addition, compliance should be analysed also at a detailed level (categories within variables).

Information obtained from population surveys has a broad potential for understanding and monitoring drug use, although in many cases this potential is not fully used. It is not unusual that resources for in-depth analysis of collected data are limited, and researches and funding institutions have to limit themselves only to basic descriptive analysis.

Therefore, in order to improve quality and usefulness of survey data, two interdependent processes should be carried out simultaneously:

> promotion of analysis and use of survey information at national and EU level, by addressing relevant policy and research questions.
>
> continued improvement of survey quality and comparability, developing common quality criteria and improving reporting of survey methodology.

Eurostat has substantial experience in harmonisation and use of population surveys from European countries in different domains (health, employment, living conditions, etc). EMCDDA and the European expert group should profit of this experience.

The EMCDDA Databank on Population surveys is a voluntary initiative based on previous experiences (UK Data Archive, or Eurostat). It implies a trust building process in a sensitive area such as drug use. The process of depositing national surveys took longer than expected, but progress is been achieved. The Databank development has highlighted that surveys reports should provide detailed methodological information.

It was stressed that even when survey data are not deposited in the EMCDDA Databank, deposit in national archives of social sciences should be encouraged.

Surveys already deposit in the EMCDDA Databank are of good methodological quality, and deposited databases have been subjected to a detailed work of harmonisation following EMQ items. Incorporation of national surveys to the Databank should include a process of quality assessment.

The joint analysis of deposited surveys is producing promising results. As in any other field of research, scientific accepted practices should be followed. Data depositors can contribute to interpretation and validation of results. National surveys methodologies should be carefully accounted for when conducting joint analysis.

(Minutes) ANNEX 1

Indicator prevalence of drug use among general population

Annual Expert meeting

May 23rd and 24th. EMCDDA, Lisbon

Draft agenda

Thursday 23 May 2002

Chair: J Vicente

09.00-09.15: Overview of the meeting *(J Vicente)*

09.15-09.45: Implementation of Key Indicator: Progress by EMCDDA *(J Vicente and R. Hartnoll)*

09.45-10.45: Implementation of the key indicator: Progress by countries

10.45-11.15: Coffee break

11.15-12.00: Implementation of the key indicator: Progress by countries

12.00-12.30: Questions and discussion

12.30-13.00: Assessment of current compliance with EMQ *(J Vicente and R Bless)*

13.00-14.30: Lunch at the EMCDDA

Chair: R Hartnoll

14.30-15.00: Information of developments from Eurostat (Health Interview surveys) and ESPAD project (School surveys) *(J van den Berg and B Hibell)*

15.00-16.00: Developments of EU Databank on population surveys (*R Bless and H Beedham)*

16.00-16.30: Coffee break

16.30-17.30: Presentation of first results of joint analysis based on EU Databank *(L Kraus and D Korf)*

Friday 24 May 2002

09.00-09.45 EMCDDA standard reporting process and EMCDDA Integrated Key Indicator Database *(N Frost and J Vicente)*

9.45-10.15 Summary and conclusions about KI implementation progress *(J Vicente and R Hartnoll)*

WORKSESSION 1 (Chair R Bless)

10.15-10.45 Assessment of existing methodological guidelines of the Key Indicator *(R Bless)*

Case studies *(National experts – The Netherlands, U. Kingdom)*

10.45-11.15 Coffee break

11.15-12.15 Worksession in plenary

12.15-12.30 Summary and conclusions *(R Bless and J Vicente)*

12.30-13.30 Lunch at the EMCDDA

WORKSESSION 2 (Chair J Vicente)

13.30-14.00 Perspectives for future developments of the Key Indicator *(J Vicente, R Bless and C Carpentier)*

Presentation of workshop

14.00-15.00 Workshop in parallel groups
Group 1 *(Chair and rapporteur –national experts-)*
Group 2 *(Chair and rapporteur –national experts-)*

15.00-15.30 Plenary: Presentation of results of groups

15.30-16.00 Summary of and conclusions of the meeting *(J Vicente and R Hartnoll)*

16.00 End of the meeting

(Minutes) ANNEX 2

Annual meeting of the EMCDDA Expert Group on the key indicator "Population Surveys"
23-24 May 2002

LIST OF PARTICIPANTS

Austria

Martin Busch
Österreichisches Bundesinstitut für Gesundheitswesen (ÖBIG)
Austrian Health Institute
Stubenring 6
A-1010 Wien
T: +43 1 515 61-124
F: +43 1 513 84 72
busch@oebig.at

Belgium

Francis Sartor
Scientific Institute of Public Health Louis-Pasteur
Rue Juliette Wytsman, 14
B-1050 Brussels
Francis.sartor@iph.fgov.be

Denmark

Niels Kr. Rasmussen
Deputy Director
National Institute of Public Health
25 Svanemollevej
D-2100 Copenhagen
T: +45 39 20 7777
F: +45 39 20 8010
nkr@dike.dk
nkr@si-folkesundhed.dk

Finland

Pekka Hakkarainen
National Research and Development Centre for Welfare and Health
P. O. Box 220
SF-00531 Helsinki
T: +358 9 3967 2161
F: +358 9 3967 2170
pekka.hakkarainen@stakes.fi

France

Francois Beck
Pôle Enquêtes en population générale
OFDT (Observatoire Français des Drogues et Toxicomanies)
(French Monitoring Centre for Drug and Drug Addiction)
105, rue Lafayette - 3ème étage
F-75010 PARIS
T: +33 1 53 20 16 08
F: +33 1 53 20 16 00
frbec@ofdt.fr

Stephane Legleye
stleg@ofdt.fr

Germany

Ludwig Kraus
Institut für Therapieforschung (IFT)
Parzivalstr. 25
D - 80804 München
T: + 49 89/36080430
F: + 49 89/36080469
kraus@ift.de

Greece

Manina Terzidou
(UMHRI) University of Mental ealth Research Institute
12 Eginitou Str.
GR-115 28 Athens
T: +30 (1) 6536 902
F: +30 (1) 6537 273
ektepn@hol.gr

Ireland

Hamish Sinclair
73 Lower Baggot St
IE - Dublin 2
T: +353 1 6761176 Ext 160
F: +353 1 6618567
hsinclair@hrb.ie

Italy

Stefano Salvatori
CNR - Istituto di Fisiologia Clinica
Sezione di Epidemiologia
Via Moruzzi
I - 56124 Pisa
T: +39 050 315 21 06
F: +39 050 315 20 95
stefsa@ifc.pi.cnr.it

Luxembourg

Pascale Straus
CRP-Santé
18, rue Dicks
Luxembourg
T:+352 26846131
F: +352 45 32 19
pascale.straus@crp-sante.lu

Netherlands

Peter Cohen
Centre for Drug Research
University of Amsterdam
Faculty of Social and Behavioural Sciences
P.O.Box 94208
NL-1090 GE Amsterdam
T: + 31 20 5254278
F: + 31 20 5254317
pcohen@bariaeus.ic.uva.nl

Manja Abraham
m.d.abraham@frw.uva.nl

Norway

Sturla Nordlund
Norwegian Institute for Alcohol and Drug Research (SIRUS)
P.O. Box 565 Sentrum
NO-0105 Oslo
T: +47 22 34 04 12
F:
sn@sirus.no

Poland

Boguslawa Bukowska
Deputy Director of National Bureau for Drug Prevention
Warsaw
T: +48 22 6411 501
F: +48 22 6411 565
bdsnark@medianet.pl

Portugal

Fernanda Feijão
IPDT
Av.ª João Crisóstomo, 14
PT-1000-179 Lisboa
T: +351 213 104 127
F: +351 213 104 192
ffeijao@ipdt.pt

Casimiro Balsa
CEOS – Investigações Sociológicas
Faculdade de Ciências Sociais e Humanas
Universidade Nova de Lisboa
Av. de Berna, 26C
PT-1069-061 Lisboa
T/F: +351 21 796 1021
cm.balsa@fcsh.unl.pt

Spain

Jacinto Rodriguez-Osuna
Mateo Lopez, n° 7
E-28007 Madrid
T: +34 91 5743544
Maherrerias@cyii.es

Sweden

Björn Hibell
CAN
Box 70412
S - 10725 Stockholm
T: +46 8 412 4607
F: +46 8 10 4641
bjorn.hibell@can.se

UK

Tom Bucke
Drugs and Alcohol Research Unit
Home Office
50 Queen Anne's Gate
UK-London SW1H 9AT
T: +0207 273 2243
F: +0207 273 4491
Tom.Bucke@homeoffice.gsi.gov.uk

Rebbecca Aust
Rebbecca.Aust@homeoffice.gsi.gov.uk

ESPAD

Björn Hibell
CAN
Box 70412
S - 10725 Stockholm
T: +46 8 412 4607
F: +46 8 10 4641
bjorn.hibell@can.se

Eurostat

Jaap van den Berg
Bâtiment Joseph Bech
5 Rue Alphonse Weicker
L-2721 Luxembourg
T: +352 4301 32693
F: +352 4301 35399
Jaap.VAN-DEN-BERG@cec.eu.int

Project Team

Ruud Bless
Quinx Research
In de Poldermolen 3
NL-1115 GR Duivendrecht
T: +31 20 600 87 60
F: +31 20 600 65 21
bless@quinx.nl

Dirk Korf
University of Amsterdam
Faculty of Law
'Bonger' Institute of Criminology
PO Box 1030
NL - 1000 BA Amsterdam
T: +31 20 5253930
F: +31 20 5253495
korf@jur.uva.nl

Hilary Beedham
UK Data Archive
University of Essex
Wivenhoe Park
Colchester
UK - Essex CO4 3SQ
T: +44 1206 872570
F: +44 1206 872003
beedh@essex.ac.uk

EMCDDA

Richard Hartnoll
Rua da Cruz de Santa Apolónia, 23-25 A
1149-045 Lisboa
Portugal
T: +351 21 811 30 06
F: +351 21 813 79 43
Richard.hartnoll@emcdda.org

Julian Vicente
Rua da Cruz de Santa Apolónia, 23-25 A
1149-045 Lisboa
Portugal
T: +351 21 811 30 23
F: +351 21 813 79 43
Julian.vicente@emcdda.org

Norbert Frost
Rua da Cruz de Santa Apolónia, 23-25 A
1149-045 Lisboa
Portugal
T: +351 21 811 30 53
F: +351 21 813 79 43
Norbert.frost@emcdda.org

Chloe Carpentier
Rua da Cruz de Santa Apolónia, 23-25 A
PT-1149-045 Lisboa
T: +351 21 811 30 65
F: +351 21 813 79 43
chloe.carpienter@emcdda.org

(Minutes) ANNEX 3

Key Indicator "Drug use among the general population (population surveys)"
Annual expert meeting; May 2002

Relevant recent developments in Member States

Based on National Forms filled in by countries in May 2002

	Latest progress	Series of surveys
Austria	First national survey 2002 Face to face 15-64 y.o. High compatibility EMQ Sample: 4000	Planned
Belgium	Previous survey 2001 First time a module on illegal drugs (although quite limited) in national HIS	Not for illegal drugs
Denmark	Previous survey 2000 (HIS) Next HIS survey in 2004/2005	Planned in HIS
Finland	Previous surveys 2002 repetition of 1998 survey 2000 repetition of 1992 survey	Planned for alcohol and drugs from 2004 onwards
France	Previous survey 1999 Next survey 2003	Yes
Germany	Previous survey 2000 Next survey 2003	Yes
Greece	Previous survey 1998 Next survey 2003	Planned series
Ireland	Previous survey 2000 Next survey 2002-03 Face to face 15-64 y.o. High compatibility EMQ Sample 5000	Not decided
Italy	First national survey 2002 Postal 15-44 y.o. Sample 6241	No
Luxembourg	Proposed but difficult	-
The Netherlands	Previous survey 2001 Next survey 2004 Methodological work on data collection	Yes
Norway	Previous survey 1999 Next survey 2004-05	Undecided
Portuqal	First national survey 2001 Face to face Complete compatibility EMQ Sample 14184	-
Spain	Previous survey 2001 Next survey 2003?	Yes
Sweden	2002 merge of traditional alcohol and drugs survey with youth survey	Yes
U. Kingdom	Previous survey 2001 Next survey 2002 Annual continuous survey Sample size 40000 Methodological on sampling methods	Yes

HIS: Health Interview survey

(Minutes) ANNEX 4

Key Indicator "Drug use among the general population (population surveys)"
Annual expert meeting; May 2002

Compliance of national surveys with EMCDDA set of common core items (EMQ)

Based on National Forms filled in by countries in May 2002

	Item not available	Differences in categories
1. Tobacco		
(1) SMOKING	nl uk	no sp
2. Alcohol		
(2) LYP_ALC	fi it	
(3) DRINKING	it nl	bl fr no sp
(4) BINGING	fr it sp	bl no
(5) LMP_ALC	bl fi	fr gr no
(6) LMF_ALC	bl fi no	fr nl sp
3. Pharmaceuticals		
(7) LYP_MED	bl sp	it
(8) MEDHABIT	bl fi nl sp uk	fr it no
(9) LMP_MED	bl no sp	it
(10) LMF_MED	bl fi nl no sp uk	fr it
(11) LASTMED	bl fi no sp uk	gr nl
4. Illicit drugs		
4.1 Cannabis		
(12) KNO_CAN	bl dk fr nl no sp uk	fi gr
(13) LTP CAN		
(14) AGE_CAN	bl dk	
(15) LYP_CAN	bl	
(16) LMP_CAN		
(17) LMF_CAN	bl dk no	fi fr nl sp uk
4.2 Ecstasy		
(18) KNO_XTC	bl dk fr ge nl no sp uk	fi
(19) LTP_XTC		
(20) LYP_XTC	bl	
(21) LMP_XTC	bl	
(22) LMF_XTC	dk no	fi fr nl sp uk
4.3 Amphetamines		
(23) KNO_AMP	dk fr ge nl no sp uk	fi gr
(24) LTP_AMP	gr	
(25) LYP_AMP		
(26) LMP_AMP		
(27) LMF-AMP	dk no	fi fr nl sp uk

4.4 Heroin		
(28) KNO_HER	bl dk fr ge nl no sp uk	fi gr
(29) LTP_HER	bl	
(30) LYP_HER	bl	
(31) LMP_HER	bl	
(32) LMF HER	bl dk no	fi fr nl sp uk
4.5 Cocaine		
(33) KNO_COC	bl dk fr ge nl no sp uk	fi gr
(34) LTP_COC	bl	
(35) LYP_COC	bl	
(36) LMP_COC	bl	
(37) LMF_COC	bl dk no	fi fr nl sp uk
4.6 Relevin *(not mandatory items)*		
(38) KNO_REL	bl fr ge nl no uk	
(39) LTP_REL	bl nl no	
(40) LYP_REL	bl nl no	
(41) LMP-REL	bl nl no	
(42) LMF_REL	bl nl no uk	fr
4.7 LSD		
(43) KNO_LSD	bl dk fr ge nl no sp uk	
(44) LTP_LSD	bl gr no	
(45) LYP_LSD	bl no	
(46) LMP_LSD	bl no	
(47) LMF_LSD	bl dk no	fi fr nl sp uk
5. Opinions		
Opinions about drug addicts		
Q1 *(not mandatory item)*	bl fr nl no uk	
Opinions about drug policies		
Q2 *(not mandatory item)*	bl nl uk	no
Q3 *(not mandatory item)*	bl nl no uk	
Opinions about behaviours		
Q4	at bl fr ge nl no sp uk	
Q5	at bl fr ge nl no sp uk	
Q6	at bl fr ge nl no sp uk	
Q7	at bl fr ge it nl no sp uk	
Q8	at bl fr ge nl no sp uk	
Perceptions of risks		
Q9	bl fr ge it nl no uk	sp
Q10	bl fr ge it nl no uk	sp
Q11	bl fr ge it nl no uk	sp
Q12	bl fr ge nl no uk	sp
Q13	bl ge nl no uk	sp
6. Respondent attributes		
(48) SEX		
(49) AGE		
(50) HOUSEHOLD		nl no sp
(51) ACTIVITY		nl no sp
(52) EDUCAT		Sp
(53) URBANISATION		it nl no sp uk

NOTES
Austria: refers to survey to be carried out during 2002
Belgium: module on illegal drugs within HIS (only Cannabis and XTC/Amphetamines; LTP and LMP)
France: frequency last months asked as number of times
Greece: compliance 73% (without relevin and political questions 87%)
Ireland: refers to survey to be carried out late 2002 or early 2003
Italy: refers to survey which data collection will be carried out during 2002. Pharmaceuticals without a medical prescription
Luxembourg: Survey foreseen but not carried out.
Portugal: a very detailed comparative analysis of EMQ and Portuguese questionnaire was provided in separate document
Spain: questions with categories slightly different from EMQ can provide information relatively compatible with EMQ (frequency?)
UK: age of first use (16-24 yr olds only). Frequency last month: some compliance (different wording)

(Minutes) ANNEX 5

Brief discussion of issues that deter data producers from depositing data (particularly individual level data) in archives or databanks for secondary analysis.

Hilary Beedham UK Data Archive, University of Essex.

Summary of presentation to Annual Expert Meeting, EMCDDA, Lisbon. May 23rd. 2002.

There are a number of frequently cited reasons for not making individual level data available for secondary analysis. These include the following:

Protection of intellectual property rights
Concern that secondary users may be critical of data quality
Data protection, ethics and statistical disclosure
Data are available elsewhere or by informal exchange agreements

The UK Data Archive has experience of dealing with all of these concerns and is confident that in the majority of cases, they need not present a reason for withholding data if the appropriate measures are taken.

Protection of intellectual property rights. Data held in archives are normally held under licence so that the archives do not become owners of the data but are agents for the depositors. Licences are similar across borders and ensure the protection of intellectual property. By allowing dissemination through archives, depositors can ensure that their copyright is acknowledged whenever data are re-used, including the use of data in web-based systems. As part of the licence agreement there is usually a clause requiring that secondary users are made aware of their obligations before being given access to data, this usually takes the form of agreement to a document which has legal status.

Quality control. Experienced agencies such as data archives implement basic procedures to ensure levels of data quality. Internal processing (in the archive) can isolate some problems, allowing depositors to correct them before data are disseminated. In addition, archives provide a link between users and suppliers of data and some hold regular meetings for key datasets at which depositors and users openly and constructively discuss data quality issues. A number of archives are also working on cross national projects with national statistical institutes to develop more structured and ultimately automated methods for the collection of metadata that are essential for informed and accurate use and understanding of data.

Data protection, ethics and statistical disclosure. There are a number of ways in which the responsibilities of data producers can be met without preventing data being made available for secondary analysis. Data entering archives should be in anonymised form. On occasions this may not be the case but routine internal processing will identify such instances before data are made available to external users. Even were this not the case, the licence agreement usually includes the requirement that users agree not to misuse data or attempt to disclose information about individuals, as part of their legal commitment in using data. In addition, archives are increasingly aware of and involved in the development of statistical disclosure software that enables data to be checked for disclosive information before they are made available on the web. Software has also been developed for the web that allows data owners and publishers to control who can access what datasets or parts of datasets, and for what purposes. Finally, some of the problems in this area can be overcome by including a question at the data collection stage which allows respondents to give their informed consent to the use of the material for secondary research.

Data are already widely available. Whilst it is always good to have data available, difficulties can arise when they are made available informally either through collection organisations directly or via a network of researchers. In particular, not all researchers (especially internationally) may be able to determine that the resource is in fact available because they don't know who or where to ask. Archives will always incorporate studies into their catalogues, which are searchable and available to all researchers via the web. This system provides the same information to all researchers about where to go to get access to particular data. Placing data in a single repository also reduces the administrative burden of managing access control and allows easier monitoring of use. Increasingly archives are at the forefront of work to provide automated systems to manage these aspects of their work.

(Minutes) ANNEX 6

Proposals for new or modified core items

(New) item

AGE_... Age of first use of any drug for which prevalence measures are assessed

At present the item is only included for Cannabis: AGE_CAN

Arguments for inclusion

- Allows assessment of initiation to drug use (incidence). Early detection of drug trends, value for prevention formulation
- Inclusion is already common practice in most surveys

Modified item

LMF_… Last month frequency of use of any drug for which prevalence measures are assessed (including alcohol and sedatives/tranquillisers)

New categories: Number of days having taken <drug> in the last 30 days
OR
1. 20 days or more
2. 10-19 days
3. 4-9 days
4. 1-3 days

At present this item is included with the following model categories:

1. daily or almost daily aprox. correspondence to new (1)
2. several times a week aprox. correspondence to new (2)
3. at least once a week aprox. correspondence to new (3)
4. less than once a week aprox. correspondence to new (4)

Arguments for modification

- Modification corresponds with common practice in most current surveys
- Improving comparability
- Facilitating harmonization of survey data

www.emcdda.eu.int

ANNEX 2: CORRESPONDENCE BETWEEN MODEL DATA HARMONISATION AND THE EMQ

The table below indicates the correspondence between the European Model Questionnaire and the variables and categories, which we have tried to construct in the harmonisation process. The listing of variables in the model harmonised derived datasets does not imply that real values for all these variables could be assigned for each derived dataset that actually has been created. For the correspondence between the model harmonised dataset and the derived datasets that have been created in the NPSD-EU, see Annex 2

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
TOBACCO				
SMOKING	Smoking	1 Current smoker 2 Quitter 3 Never smoked 4 Quitter or never smoked	**SMOKING**	1 Current smoker 2 Quitter 3 Never smoked
ALCOHOL				
LTP_ALC **LYP_ALC** **LMP_ALC**	Life time prevalence of alcohol Last year prevalence of alcohol Last month prevalence of alcohol	1 Yes 2 No	 **LYP_ALC** **LMP_ALC**	
LMF_ALC	Last month frequency of alcohol	1 Very high = 20-30 times 2 High = 10-19 times 3 Low = 5-9 times 4 Very low = 1-4 times 5 Not once = 0 times	**LMF_ALC**	1 Daily or almost daily 2 Several times a week 3 At least once a week 4 Less than once a week

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
TOBACCO				
SMOKING	Smoking	1 Current smoker 2 Quitter 3 Never smoked 4 Quitter or never smoked	**SMOKING**	1 Current smoker 2 Quitter 3 Never smoked
ALCOHOL				
LTP_ALC **LYP_ALC** **LMP_ALC**	Life time prevalence of alcohol Last year prevalence of alcohol Last month prevalence of alcohol	1 Yes 2 No	 **LYP_ALC** **LMP_ALC**	
DRINKING **BINGING**	General frequency of drinking General frequency of binge drinking	1 High = 20-30 times 2 Medium = 10-19 times 3 Low = 2-9 times 4 (Almost) never = 0-1 times *If refers to last year or if period unspecified* *If refers to the last 30 days*	**DRINKING**	1 4 times a week or more often 2 2 to 3 times a week 3 2 to 4 times a month 4 once a month or more seldom
			BINGING	1 Daily or almost daily 2 Every week 3 Every month 4 More seldom than once a month 5 Never

MODEL DATA HARMONISATION				EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**		**Variable**	**Categories and remarks**
PHARMACEUTICALS					
HHO_SED **HHO_TRA**	Having heard of sedatives Having heard of tranquillisers	1 Yes 2 No			
LTP_SED **LTP_TRA** **LYP_SED** **LYP_TRA** **LMP_SED** **LMP_TRA**	Life time prevalence of sedatives Life time prevalence of tranquillisers Last year prevalence of sedatives Last year prevalence of tranquillisers Last month prevalence of sedatives Last month prevalence of tranquillisers	1 Yes 2 No		**(LYP_MED)** **(LMP_MED)**	MED refers to “sedatives and/or tranquillisers”. Existing surveys do not include this combination of substances.
LTF_SED **LTF_TRA**	Life time frequency of sedatives Life time frequency of tranquillisers	1 High 2 Low 3 Not once	= 20 + times = < 20 times = never		
LMF_SED **LMF_TRA**	Last month frequency of sedatives Last month frequency of tranquillisers	1 Very high 2 High 3 Low 4 Very low 5 Not once	= 20-30 times = 10-19 times = 5-9 times = 1-4 times = 0 times	**(LMF_MED)**	MED refers to “sedatives and/or tranquillisers”. Existing surveys do not include this combination of substances.
AGE_SED **AGE_TRA**	Age of first use of amphetamines Age of first use of amphetamines	nn <age>			
				(MEDHABIT)	Not included in existing surveys
				(LASTMED)	Not included in existing surveys

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
ILLICIT DRUGS				
HHO_AMP **HHO_CAN** **HHO_COC** **HHO_CRA** **HHO_DUM** **HHO_HAL** **HHO_HER** **HHO_LSD** **HHO_MET** **HHO_MUS** **HHO_OPI** **HHO_XTC**	Having heard of amphetamines Having heard of cannabis Having heard of cocaine Having heard of crack Having heard of dummy drug Having heard of hallucinogens Having heard of heroin Having heard of LSD Having heard of methadone Having heard of magic mushrooms Having heard of (any) other opioïd Having heard of ecstasy	1 Yes 2 No	**(KNO_AMP)** **(KNO_CAN)** **(KNO_COC)** **(KNO_REL)** **(KNO_HER)** **(KNO_LSD)** **(KNO_XTC)**	The EMQ includes a proxy question “Do you personally know people who take <name of drug> ?” In the EMQ the dummy drug is called Relevin (REL)
LTP_AMP **LTP_ANY** **LTP_CAN** **LTP_COC** **LTP_CRA** **LTP_DUM** **LTP_HAL** **LTP_HER** **LTP_LSD** **LTP_MET** **LTP_MUS** **LTP_OPI** **LTP_XTC**	Life time prevalence of amphetamines Life time prevalence any (illicit) drug Life time prevalence of cannabis Life time prevalence of cocaine Life time prevalence of crack Life time prevalence of dummy drug Life time prevalence of hallucinogens Life time prevalence of heroin Life time prevalence of LSD Life time prevalence of methadone Life time prevalence of magic mushrooms Life time prevalence of other opioids Life time prevalence of ecstasy	1 Yes 2 No	**LTP_AMP** **LTP_CAN** **LTP_COC** **LTP_REL** **LTP_HER** **LTP_LSD** **LTP_XTC**	The EMQ does not ask for crack, methadone and magic mushrooms separately, and does not include the combined drug classes hallucinogens, other opioids or any (illicit) drug. In the EMQ the dummy drug is called Relevin (REL).

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
ILLICIT DRUGS (continued)				
LYP_AMP **LYP_ANY** **LYP_CAN** **LYP_COC** **LYP_CRA** **LYP_DUM** **LYP_HAL** **LYP_HER** **LYP_LSD** **LYP_MET** **LYP_MUS** **LYP_OPI** **LYP_XTC**	Last year prevalence of amphetamines Last year prevalence any (illicit) drug Last year prevalence of cannabis Last year prevalence of cocaine Last year prevalence of crack Last year prevalence of dummy drug Last year prevalence of hallucinogens Last year prevalence of heroin Last year prevalence of LSD Last year prevalence of methadone Last year prevalence of magic mushrooms Last year prevalence of other opioids Last year prevalence of ecstasy	1 Yes 2 No	**LYP_AMP** **LYP_CAN** **LYP_COC** **LYP_REL** **LYP_HER** **LYP_LSD** **LYP_XTC**	The EMQ does not ask for crack, methadone and magic mushrooms separately, and does not include the combined drug classes hallucinogens, other opioids or any (illicit) drug. In the EMQ the dummy drug is called Relevin (REL).
LMP_AMP **LMP_ANY** **LMP_CAN** **LMP_COC** **LMP_CRA** **LMP_DUM** **LMP_HAL** **LMP_HER** **LMP_LSD** **LMP_MET** **LMP_MUS** **LMP_OPI** **LMP_XTC**	Last month prevalence of amphetamines Last month prevalence any drug Last month prevalence of cannabis Last month prevalence of cocaine Last month prevalence of crack Last month prevalence of dummy drug Last month prevalence of hallucinogens Last month prevalence of heroin Last month prevalence of LSD Last month prevalence of methadone Last month prevalence of magic mushrooms Last month prevalence of other opioids Last month prevalence of ecstasy		**LMP_AMP** **LMP_CAN** **LMP_COC** **LMP_REL** **LMP_HER** **LMP_LSD** **LMP_XTC**	The EMQ does not ask for crack, methadone and magic mushrooms separately, and does not include the combined drug classes hallucinogens, other opioids or any (illicit) drug. In the EMQ the dummy drug is called Relevin (REL).

MODEL DATA HARMONISATION				EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**		**Variable**	**Categories and remarks**
LTF_AMP **LTF_CAN** **LTF_COC** **LTF_CRA** **LTF_HAL** **LTF_HER** **LTF_LSD** **LTF_MET** **LTF_MUS** **LTF_OPI** **LTF_XTC**	Life time frequency of amphetamines Life time frequency of cannabis Life time frequency of cocaine Life time frequency of crack Life time frequency of hallucinogens Life time frequency of heroin Life time frequency of LSD Life time frequency of methadone Life time frequency of magic mushrooms Life time frequency of other opioids Life time frequency of ecstasy	1 High 2 Low 3 Not once	= 20 + times = < 20 times = never		Not included in the EMQ
LMF_AMP **LMF_CAN** **LMF_COC** **LMF_CRA** **LMF_HAL** **LMF_HER** **LMF_LSD** **LMF_MET** **LMF_MUS** **LMF_OPI** **LMF_XTC**	Last month frequency of amphetamines Last month frequency of cannabis Last month frequency of cocaine Last month frequency of crack Last month frequency of hallucinogens Last month frequency of heroin Last month frequency of LSD Last month frequency of methadone Last month frequency of magic mushrooms Last month frequency of other opioids Last month frequency of ecstasy	1 Very high 2 High 3 Low 4 Very low 5 Not once	= 20-30 times = 10-19 times = 5-9 times = 1-4 times = 0 times	**LMF_AMP** **LMF_CAN** **LMF_COC** **(LMF_REL)** **LMF_HER** **LMF_LSD** **LMF_XTC**	In the original EMQ: 1 daily or almost daily 2 several times a week 3 at least once a week 4 less than once a week In the revised version (see Ch.3): nn <number of days in last 30 days> OR 1 20 times or more 2 10-19 times 3 4-9 times 4 1-3 times The EMQ also includes Last month frequency of the dummy drug (“Relevin”)

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
ILLICIT DRUGS (continued)				
AGE_AMP **AGE_ANY** **AGE_CAN** **AGE_COC** **AGE_CRA** **AGE_HAL** **AGE_HER** **AGE_LSD** **AGE_MET** **AGE_MUS** **AGE_OPI** **AGE_XTC**	Age of first use of amphetamines Age of first use of any (illicit) drug Age of first use of cannabis Age of first use of cocaine Age of first use of crack Age of first use of hallucinogens Age of first use of heroin Age of first use of LSD Age of first use of methadone Age of first use of magic mushrooms Age of first use of other opioids Age of first use of ecstasy	nn <age>	**AGE_CAN**	In the original EMQ Age of first use is only asked for Cannabis (CAN). In the revised version it should be asked for every drug covered in the survey questionnaire.

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
OPINIONS				
CRIMOPAT	Criminal or patient?	1 More criminal 2 More patient 3 Neither 4 Both 5 Cannot decide		Variable names have not been specified; answer categories are corresponding.
DISP_ALC **DISP_CAN** **DISP_HER** **DISP_SMO** **DISP_XTC**	Disapprove drinking several times a week? Disapprove smoking cannabis occasionally? Disapprove trying heroin once or twice? Disapprove smoking > 10 cigarettes a day? Disapprove trying ecstasy once or twice?	1 Do not disapprove 2 Disapprove 3 Strongly disapprove 4 Don't know		Variable names have not been specified; answer categories are corresponding.
LEGA_CAN **LEGA_HER**	Cannabis should be legal Heroin should be legal	1 Fully agree 2 Largely agree 3 Agree nor disagree 4 Largely disagree 5 Fully disagree		Variable names have not been specified; answer categories are corresponding.
RISK_ALC **RISK_CAN** **RISK_COC** **RISK_HER** **RISK_SMO** **RISK_XTC**	Risk perception of drinking > 5 drinks each weekend? Risk perception of smoking cannabis regularly? Risk perception of trying cocaine once or twice Risk perception of heroin once or twice? Risk perception of smoking > 1 pack of cigarettes a day? Risk perception of ecstasy once or twice?	1 No risk 2 Small risk 3 Moderate risk 4 Great risk		Variable names have not been specified; answer categories are corresponding.

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
RESPONDENT ATTRIBUTES				
AGE	Age	nn <age>	**AGE**	
AGEGR	Age group	1 < 15 2 15-19 3 20-24 4 25-29 5 30-34 6 35-39 7 40-44 8 45-49 9 50-54 10 55-59 11 60-64 12 65 +		Not included in the EMQ, but can be calculated from AGE.
COUNTRY	Country	276 Germany 300 Greece 724 Spain 826 United kingdom		Not specified in the EMQ.
EDUCAT	Level of highest completed education	1 Low 2 Medium 3 High	**EDUCAT**	1 Primary education or less 2 Lower secondary education 3 Higher secondary education 4 Higher education 5 Cannot be classified
EMPLOY	Main activity	1 Employed 2 Student 3 Unemployed 4 Disabled, retired, full-time housekeeping	**ACTIVITY**	1 Employed or self-employed 2 Full-time student 3 Unemployed 4 Other
HOUSHLD	Household composition	1 One person 2 More than one person	**HOUSHOLD**	1 one person living alone 2 two partners without children at home 3 two partners with children at home 4 one adult with children at home 5 other situation

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
RESPONDENT ATTRIBUTES (continued)				
INC_HH	Income level of household	1 Bottom 25% (+/- 10%) of national distribution 2 Middle 50% (+/- 20%) of national distribution 3 Top 25% (+/- 10%) of national distribution		Not included in the EMQ
INC_RES	Income level of respondent	1 Bottom 25% (+/- 10%) of national distribution 2 Middle 50% (+/- 20%) of national distribution 3 Top 25% (+/- 10%) of national distribution		Not included in the EMQ
MARITAL	Marital status	1 Married 2 Cohabiting 3 Single 4 Widowed 5 Divorced 6 Separated 7 Other		Not included in the EMQ
SEX	Gender	1 Male 2 Female	**SEX**	
URBAN	Level of urbanisation	1 Metropolitan 2 Urban 3 Rural	**URBANISATION**	The EMQ gives code 4 for “cannot be classified”, which in the harmonisation is replaced by code 9997 (see Table 3)

MODEL DATA HARMONISATION			EUROPEAN MODEL QUESTIONNAIRE	
Variable	**Label**	**Categories**	**Variable**	**Categories and remarks**
SURVEY IDENTIFIERS				
YEAR	Year of survey	nn <year>		Not specified in the EMQ
SURVEY	Country + year.	1 England and Wales 1994 2 England and Wales 1996 3 England and Wales 1998 4 Spain 1995 5 Spain 1997 6 Spain 1999 7 Greece (Athens) 1993 8 Greece 1998 9 Germany 1995 10 Germany 1997 11 Germany 2000		Not specified in the EMQ
COUNTRY SPECIFIC VARIABLES				
REGIO_DE	East-West split for Germany	1 East-Germany 2 West-Germany		Not specified in the EMQ
WEIGHT FACTORS				
WTCCYY	Survey specific weight factor, whereby CC = two digit country code YY = year			Not specified in the EMQ

www.emcdda.eu.int

ANNEX 3: CONTENT OF DERIVED DATASETS IN THE NPSD-EU

The table below indicates if a variables constructed in the data harmonisation have valid values in derived datasets of the countries that have survey data deposited in the NPSD-EU. Cells in the columns pertaining to the derived datasets are shaded blue, when a one-to-one correspondence of variables and categories could be achieved by manipulations on the original survey data. Cells are shaded rose when the correspondence match is not perfect or should be considered dubious; the discrepancies are explained in the notes below the table.

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
TOBACCO												
SMOKING	Smoking		1					2	2			
ALCOHOL												
LTP_ALC	Life time prevalence of alcohol											
LYP_ALC	Last year prevalence of alcohol	3										
LMP_ALC	Last month prevalence of alcohol										4	4

[1] No distinction between “quitter” and “never smoked”.

[2] Quitter = self-declared or not smoked in last 30 days.

[3] Refers to drinking “nowadays”.

[4] Very high = > 4 days a week; High = 1-3 days a week; Low = 1-3 days a month; Very low = 1-2 days a year

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
LMF_ALC	Last month frequency of alcohol				1	5	5					
DRINKING	General frequency of drinking	3 6	3 2	3 6	5	5	5					
BINGING	General frequency of binge drinking				7	7	7	3	7		7	7
PHARMACEUTICALS												
HHO_SED	Having heard of sedatives											
HHO_TRA	Having heard of tranquillisers											
LTP_SED	Life time prevalence of sedatives							4	1			
LTP_TRA	Life time prevalence of tranquillisers	1	1	1				1	1			
LYP_SED	Last year prevalence of sedatives							1	1			
LYP_TRA	Last year prevalence of tranquillisers	1	1	1				1	1			
LMP_SED	Last month prevalence of sedatives							1	1			
LMP_TRA	Last month prevalence of tranquillisers	1	1	1				1	1			
LTF_SED	Life time frequency of sedatives							1	1			
LTF_TRA	Life time frequency of tranquillisers							1	1			
LMF_SED	Last month frequency of sedatives							1	1			
LMF_TRA	Last month frequency of tranquillisers							1	1			
AGE_SED	Age of first use of amphetamines							1	1			

[1] Assessed as maximum of either beer, wine or spirits.

[2] (1994) High = drink heavily/quite a lot; Medium = drink moderate amount; Low = drink little/hardly at all; (1996,1998) High = > 5 days a week; Medium = 1-4 days a week; Low = once a year till 2-3 days a month; (Almost) never = less than once a year

[3] Refers to 5 glasses or more on same occasion during last 30 days.

[4] Refers to substance not prescribed by a doctor.

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
AGE_TRA	Age of first use of amphetamines							[1]	[1]			
ILLICIT DRUGS												
HHO_AMP	Having heard of amphetamines											
HHO_CAN	Having heard of cannabis											
HHO_COC	Having heard of cocaine											
HHO_CRA	Having heard of crack											
HHO_DUM	Having heard of dummy drug											
HHO_HAL	Having heard of hallucinogens											
HHO_HER	Having heard of heroin											
HHO_LSD	Having heard of LSD											
HHO_MET	Having heard of methadone											
HHO_MUS	Having heard of magic mushrooms											
HHO_OPI	Having heard of (any) other opioïd											
HHO_XTC	Having heard of ecstasy											
LTP_AMP	Life time prevalence of amphetamines							[1]	[1]			
LTP_ANY	Life time prevalence any (illicit) drug											
LTP_CAN	Life time prevalence of cannabis											
LTP_COC	Life time prevalence of cocaine							[2]	[2]			
LTP_CRA	Life time prevalence of crack											
LTP_DUM	Life time prevalence of dummy drug											
LTP_HAL	Life time prevalence of hallucinogens											
LTP_HER	Life time prevalence of heroin											

[1] Refers to substance not prescribed by a doctor.

[2] Explicitly includes crack cocaine.

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
LTP_LSD	Life time prevalence of LSD											
LTP_MET	Life time prevalence of methadone	1	1	1								
LTP_MUS	Life time prevalence of (psycho-active) mushrooms											
LTP_OPI	Life time prevalence of other opioids									1	3	3
LTP_XTC	Life time prevalence of ecstasy											
LYP_AMP	Last year prevalence of amphetamines							1	1			
LYP_ANY	Last year prevalence any (illicit) drug											
LYP_CAN	Last year prevalence of cannabis											
LYP_COC	Last year prevalence of cocaine							2	1			
LYP_CRA	Last year prevalence of crack											
LYP_DUM	Last year prevalence of dummy drug											
LYP_HAL	Last year prevalence of hallucinogens											
LYP_HER	Last year prevalence of heroin											
LYP_LSD	Last year prevalence of LSD											
LYP_MET	Last year prevalence of methadone	3	2	2								
LYP_MUS	Last year prevalence of magic mushrooms											
LYP_OPI	Last year prevalence of other opioids									4	3	3
LYP_XTC	Last year prevalence of ecstasy											
LMP_AMP	Last month prevalence of amphetamines							2	2			

[1] As examples of “other opioids” only brand names of medical opiates are listed.

[2] Explicitly includes crack cocaine.

[3] Refers to substance not prescribed by a doctor.

[4] As examples of “other opioids” only brand names of medical opiates are listed.

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
LMP_ANY	Last month prevalence any drug											
LMP_CAN	Last month prevalence of cannabis											
LMP_COC	Last month prevalence of cocaine							1	1			
LMP_CRA	Last month prevalence of crack											
LMP_DUM	Last month prevalence of dummy drug											
LMP_HAL	Last month prevalence of hallucinogens											
LMP_HER	Last month prevalence of heroin											
LMP_LSD	Last month prevalence of LSD											
LMP_MET	Last month prevalence of methadone	1	1	1								
LMP_MUS	Last month prevalence of magic mushrooms											
LMP_OPI	Last month prevalence of other opioids										2	2
LMP_XTC	Last month prevalence of ecstasy											
LTF_AMP	Life time frequency of amphetamines							1	1			
LTF_CAN	Life time frequency of cannabis											
LTF_COC	Life time frequency of cocaine							3	3			
LTF_CRA	Life time frequency of crack											
LTF_HAL	Life time frequency of hallucinogens											
LTF_HER	Life time frequency of heroin											
LTF_LSD	Life time frequency of LSD											
LTF_MET	Life time frequency of methadone											
LTF_MUS	Life time frequency of magic mushrooms											

[1] Refers to substance not prescribed by a doctor.

[2] As examples of “other opioids” only brand names of medical opiates are listed.

[3] Explicitly includes crack cocaine.

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
LTF_OPI	Life time frequency of other opioids											
LTF_XTC	Life time frequency of ecstasy											
LMF_AMP	Last month frequency of amphetamines							[1]	[1]			
LMF_CAN	Last month frequency of cannabis											
LMF_COC	Last month frequency of cocaine							[3]	[3]			
LMF_CRA	Last month frequency of crack											
LMF_HAL	Last month frequency of hallucinogens											
LMF_HER	Last month frequency of heroin											
LMF_LSD	Last month frequency of LSD											
LMF_MET	Last month frequency of methadone											
LMF_MUS	Last month frequency of magic mushrooms											
LMF_OPI	Last month frequency of other opioids										[1]	[1]
LMF_XTC	Last month frequency of ecstasy											
AGE_AMP	Age of first use of amphetamines							[2]	[2]			
AGE_ANY	Age of first use of any (illicit) drug											
AGE_CAN	Age of first use of cannabis											
AGE_COC	Age of first use of cocaine							[3]	[3]			
AGE_CRA	Age of first use of crack											
AGE_HAL	Age of first use of hallucinogens											
AGE_HER	Age of first use of heroin											
AGE_LSD	Age of first use of LSD											

[1] As examples of “other opioids” only brand names of medical opiates are listed.

[2] Refers to substance not prescribed by a doctor.

[3] Explicitly includes crack cocaine.

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
AGE_MET	Age of first use of methadone											
AGE_MUS	Age of first use of magic mushrooms											
AGE_OPI	Age of first use of other opioids									[1]	[1]	[1]
AGE_XTC	Age of first use of ecstasy											
OPINIONS												
CRIMOPAT	Criminal or patient?											
DISP_ALC	Disapprove drinking several times a week?											
DISP_CAN	Disapprove smoking cannabis occasionally?											
DISP_HER	Disapprove trying heroin once or twice?											
DISP_SMO	Disapprove smoking > 10 cigarettes a day?											
DISP_XTC	Disapprove trying ecstasy once or twice?											
LEGA_CAN	Cannabis should be legal				[1]	[1]		[2]	[2]			
LEGA_HER	Heroin should be legal							[2]	[2]			
RISK_ALC	Risk perception of drinking > 5 drinks each weekend?											
RISK_CAN	Risk perception of smoking cannabis regularly?											
RISK_COC	Risk perception of trying cocaine once or twice											
RISK_HER	Risk perception of heroin once or twice?											
RISK_SMO	Risk perception of smoking > 1 pack of cigarettes a day?											
RISK_XTC	Risk perception of ecstasy once or twice?											
RESPONDENT ATTRIBUTES												
AGE	Age											
AGEGR	Age group											

[1] Question reads: Soft drugs should be free? Yes/no recoded as: largely agree / largely disagree.

[2] Yes/no recoded as: largely agree / largely disagree; in 1993 both lega_can and lega_her were combined in one question.

MODEL DATA HARMONISATION		ENGLAND AND WALES			GERMANY			GREECE		SPAIN		
Variable	**Label**	**1994**	**1996**	**1998**	**1995**	**1997**	**2000**	**1993**	**1998**	**1995**	**1997**	**1999**
COUNTRY	Country											
EDUCAT	Level of highest completed education [1]											
EMPLOY	Main activity [2]		3	5								
HOUSHLD	Household composition											
INC_HH	Income level of household	1	1	1	1	1	1	1	1			
INC_RES	Income level of respondent											
MARITAL	Marital status									6	6	6
SEX	Gender											
URBAN	Level of urbanisation				2	2		3	4	5	5	5
SURVEY IDENTIFIERS												
YEAR	Year of survey											
SURVEY	Country + year.											
COUNTRY SPECIFIC VARIABLES												
REGIO_DE	East-West split for Germany											
WEIGHT FACTORS												
WTCCYY	Survey specific weight factor, whereby CC=country code and YY year of survey											
CASCCYY	Survey specific case identifier, whereby CC=country code and YY year of survey											

[1] The EMQ categories, based on the ISCED codes have been assessed from the listed answer categories, but listed categories could not always be found in the official ISCED classification; moreover listed categories are often examples which leave room for misclassification of less common educational degrees.

[2] As all surveys use different classifications of responses; in some cases the responses on several questions had to be assessed. The resulting harmonised categories are not fully comparable between surveys and countries.

[3] From the source files only the main activity of the head of household can be assessed and the variable is therefore not included.

[1] Resulting Bottom-Middle-Top distributions in % of the weighted total response are:

England/Wales 1994	36 – 33 – 32	Germany 1995	28 – 52 – 20	Greece 1993	8 – 56 – 36
England/Wales 1996	32 – 31 – 38	Germany 1997	25 – 51 – 24	Greece 1998	20 – 51 – 29
England/Wales 1998	28 – 48 – 24	Germany 2000	25 – 46 – 29		

[2] Rural = < 20.000; Urban = 20-500.000; Metropolitan = > 500.000

[3] Only Athens surveyed; classified as Metropolitan.

[4] Athens and Thessaloniki classified as Metropolitan; “urban” as Urban and “rural” as Rural.

[5] Rural = < 10.000; Urban = 10-500.000; Metropolitan = > 500.000.

[6] The category “cohabitating” does not exist.