E.M.C.D.D.A.

EMCDDA SCIENTIFIC REPORT

European Network to Develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions

Final report: Part 4 – Time trends and incidence of problem drug use

EMCDDA / 2002

EMCDDA SCIENTIFIC REPORT

European Network to Develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions

Final Report: Part Part 4 – Time trends and incidence of problem drug use

EMCDDA / P1/ TSER Network/ 2002

This report (Part 4 – Time trends and incidence) was prepared by:

Carla Rossi (Work Group coordinator)

Work Group members Time trends and incidence:
Erik van Ameijden , Anna Maria Bargagli, Massimiliano Bultrini, Marcel Buster , Maria Grazia Calvani, Catherine Comiskey , Simon Heisterkamp , Matthew Hickman , Alessandra Nardi, Lucilla Ravà, Jorge Ribeiro, Francis Sartor, Katalin Veress, Denise Walckiers .

Full Network Details:

Project Partners (project and work group coordinators):
Lucas Wiessing, EMCDDA (project coordinator), Gordon Hay, Univ. Glasgow, Carla Rossi, Lucilla Ravà, Univ. Rome 'Tor Vergata', Martin Frischer, Heath Heatlie, Univ. Keele, Hans Jager, Wien Limburg, RIVM, Christine Godfrey, Univ. York, Chloé Carpentier, Monika Blum, Kajsa Mickelsson, Richard Hartnoll, EMCDDA

The important input of all network participants and invited experts is fully acknowledged. For a list of network participants per working group and email contacts see Final Report Part 1, Annex A.

Other Network Participants and Invited Experts:
Fernando Antoñanzas, Rita Augustin, Marcel Buster, Maria Fe Caces, John Carnavale, Gloria Crispino O'Connell, Antonia Domingo, Ken Field, Gerald Foster, Maria Gannon, David Goldberg, Peter Hanisch, Toon van der Heijden, Matthew Hickman, Neil Hunt, Claude Jeanrenaud, Pierre Kopp, Petra Kümmler, Mirjam Kretzschmar, Marita van de Laar, Nacer Lalam, Fabio Mariani, Linda Nicholls, Alojz Nociar, Deborah Olzewski, Laetitia Paoli, Paivi Partanen, Paulo Penna, Harold Pollack, Maarten Postma, Thierry Poynard, Jorge Ribeiro, Janusz Sieroslawski, Ronald Simeone, Filip Smit, Juan Tecco, Jaap Toet, Gernot Tragler, Giovanni Trovato, Alfred Uhl, Julián Vicente, Katalin Veress, Denise Walckiers, Robert Welte, Ardine de Wit, John Wong, Tomas Zabransky, Terry Zobeck, Brigitta Zuiderma-van Gerwen.

Project funded by the European Commission, DG Research, Targeted Socio-Economic Resarch (TSER). Project n°: ERB 4141 PL 980030, Contract n°: SOE2-CT98-3075
Starting date: 1st December 1998
Duration: 36 months
Date of issue of this report: 31st January 2002

European Monitoring Centre for Drugs and Drug Addiction
Rua da Cruz de Santa Apolónia 23-25
PT - 1149-045 Lisboa
Portugal
Tel: + 351 21 811 30 00
Fax: + 351 21 813 17 11
e-mail: info@emcdda.org
http://www.emcdda.org

Contents

Final Report Part 1: General Overview

For more detail see the full final reports of the six working groups:

1 Executive Summary

In the 1st 12-month period Kaplan Meyer and Cox Regression models were used to estimate the latency time between first drug use and first treatment. The group applied a Back-Calculation Model to estimate the incidence of problem drug use from the observed incidence in treatment, using the latency time. Latency time was remarkably similar in Amsterdam, London and Rome (on average between 6 and 8 years). However, this depended strongly on age at first drug use, latency time being longer in drug users who started at a younger age. This has important consequences for treatment centres, which might not be reaching young drug users sufficiently. Incidence curves were also estimated in the three cities on the basis of treatment data, using an updated version of the Empirical Bayesian Back Calculation procedure, showing important differences in the dynamic of the drug use epidemic. Data for Lisbon were analysed as well and showed similar results, despite being not from drugs treatment but from other services and therefore difficult to compare.
In the 2nd 12-month period Kaplan Meyer and Cox Regression models were used to estimate the latency time between first drug use and first treatment introducing epidemiological information about the phase of the sub-epidemics concerning the sub-groups defined by discrete covariates (ethnicity, gender, route…) in order to correct possible biases due to the different starting point of the sub-epidemics. The group generalised the Empirical Bayesian a Back-Calculation Model to estimate the incidence of problem drug use from the observed incidence in treatment, using the latency time as incubation distribution. User friendly interfaces were produced for the BC procedure written in S+ language. Latency time was remarkably similar in Amsterdam, London and Rome (on average between 6 and 8 years). However, this depended strongly on age at first drug use, latency time being longer in drug users who started at a younger age. Incidence curves were estimated for London, Amsterdam and Italy, showing important differences in the dynamic of the drug use epidemic. For Italy it was also possible to estimate some incidence curves at local (regional level). Data for Lisbon were analysed as well and showed similar results, despite being not from drugs treatment but from other services and therefore more difficult to compare. The incidence curve for Lisbon was estimated using a special snapshot method, namely the onset delay adjustment method (ODAM). Recent work from Dublin indicates that the situation might be very different, latency time on average about 2 years and not related to age at first use.
This might however be related to the stage of the drugs epidemic, which appears to be still in its early phase in Dublin, while already endemic in most other European cities. Unfortunately data from Dublin presently comprises only one year of treatment, thus it is not possible to check for the phase of the epidemic. Data from the French Community of Belgium were also analysed estimating the latency period with results comparable to the other sites (except Dublin). The incidence curve was estimated by the ODAM model. Similarly data from Budapest were also analysed producing results similar to those of Dublin. In this latter case, however, the phase of the epidemic could be considered due to the completeness of the data-set, which comprises more years of treatment. During the period 4 papers were written to be submitted to International Journals (2 of them are in press on UN Bulletin on Narcotics). Other papers related to the local analyses, results and policy implications were also produced. Preprints related to 3 of the 4 International papers can be presently downloaded from the website: http://mat.uniroma2.it/biometria/. A system dynamic model was developed and used to make some preliminary scenario analyses.

A structural equation model was also developed to evaluate prevention interventions, some very preliminary analyses were performed. The results were also presented at national and International meetings, such as, for example, the Harm Reduction Conference held in Jersey and the Workshop on "Dynamic drug policy: Understanding and controlling drug epidemics", held in Vienna.

In the 3rd 12th month period the work was concentrated mostly on methodological developments and in summarizing all the results obtained in form of guidelines. Further local analyses were performed in order to obtain better estimates of the latency time distribution. Unfortunately new high quality local data are available for Italy only from the VE.de.TTE (Evaluation of Efficacy of Treatments for Heroin Addiction) multi-regional study carried out between 1998 and 1999 from ASP Lazio in different Italian regions, no further data were provided by the other partners. The group used again the Empirical Bayesian a Back-Calculation Model to estimate the incidence of problem drug use from the observed incidence in treatment, using the latency time as incubation distribution. Incidence curves were estimated for various areas in Italy (regional level). Further analyses considered regional aggregates obtained by spatial analysis and geographical aggregation. The incidence curves obtained for the aggregates show higher regularity and lower uncertainties due to increased sample sizes. A paper presenting the methods and analyses has been submitted.

The system dynamic model already developed was studied in depth both from a qualitative point of view and from a numerical point of view by means of a simulation procedure developed using the language S-plus. Several scenarios were obtained. A paper presenting both analyses has been submitted. A structural equation model was also developed to evaluate prevention interventions, some analyses were performed on the basis of a large data set coming from 35,000 interviews to military conscripts in Italy. Such analyses permitted to identify 4 different causal models for 4 different primary use substances. Work is in progress to write a paper summarising the results. On the basis of the preliminary analysis a new survey among military conscript in Italy has been designed in order to analyse time trends. The study will be developed in the next three years. Some of the various results obtained were presented at national and International meetings, such as, for example, the Harm Reduction Conference held in New Delhi. Further methodological developments allow to obtain correction of biases due to truncation in the estimates of the latency period. The experimental work performed within the project and the results obtained allow to set up a preliminary draft of guidelines.

2 Scientific overview

2.1 General overview

The development of problem drug use over time in Europe is only known through indirect indicators like treatment presentations, drug seizures, or drug related deaths, and from a limited number of cities (Pompidou Group 1995). Estimates of prevalence and incidence at the country or city level generally are not available.
In the U.S., Hunt and Chambers derived estimates of the incidence or diffusion of heroin epidemics. They examined firstly the delay between "onset" of use and entry into treatment; and secondly the likelihood of heroin use transferring from a person to another. These insights presage new techniques used for understanding the epidemiology of infectious diseases and utilising surveillance data to estimate incidence and prevalence (e.g. Back-Calculation methods and system dynamic models).

On the basis of such considerations, the specific objectives to be attained in the first 12-months period were:

1. To form a working group of people interested in the analysis of the latency period, who could provide data and experience and to organise a first meeting of the working group to discuss and finalise the workplan,
2. To investigate some suitable methods, in particular survival analysis methods for estimating the latency period distribution.
3. To estimate the latency period distribution for some pilot sites.
4. To set up proper Back-Calculation and system dynamic models to estimate incidence of problem drug use from treatment data.

The working group was established and the first meeting was organised in Bilthoven (June 28th 1999).
The work was carried out in various countries in Europe, with a prime focus on estimating and analysing the latency period distribution in these sites: London (UK), Lazio region, Milan, Pisa, Trieste, Pescara (Italy), Amsterdam (The Netherlands), Dublin (Ireland) and, with some limitations, the site of Casal Ventoso in Lisbon (Portugal).
The latency period distribution was estimated from treatment data.
From the results obtained, the latency period appears to be remarkably similarly distributed over different metropolitan areas, with the exception of Dublin, with a median of between four and six years and an average of between five and seven years. The parametric models which seem to be more suitable to represent the distribution of the latency period belong to the gamma and Weibull families.
The parameters of such distributions allow to obtain the following values of the mean of the latency period for the major capital cities: Amsterdam 7 years, Rome 6.82 years, London 6.64 years, whereas the mean for Dublin is remarkably shorter (3.5). This time-lapse, however, appears to be much longer than this in young drug users and inner-city drug users. Differences relating to ethnicity also were observed, whenever this covariate could be included in the analysis, in particular for the Amsterdam, London and Lisbon data sets.
The estimated latency period distributions and the treatment data were then utilised to back calculate historical trends in incidence of problem drug use from first reports of people in treatment, by using the Empirical Bayesian Back Calculation procedure, that was specifically set up and applied for Italy, The Netherlands and United Kingdom. The epidemics in the three sites appear quite different in shape. In The Netherlands the problem drug use incidence curve peaked around 1984 with a fast decrease towards a low endemic phase in the following years. In Italy the epidemic seems to be much more important, still increasing at the end of the '80s with two peaks at 85 and 91 (higher) and an endemic phase at high level afterwards. Finally in the UK the incidence peaked in 1990, with new cases presenting to

treatment peaking in 1994/95 and a general flattening of the rise in the treatment population towards the end of the 1990s.
In order to evaluate the performance of the EB-BC in investigating the extent and the dynamics of problem drug use, a sensitivity analysis, particularly regarding the choice of the latency period distribution and the inclusion of the covariate “age” in the model, was performed.
The incidence curves provided by the BC estimation procedure are strongly dependent on the latency period model chosen, but the location of the peaks of the epidemic seems to be a robust parameter. Also the qualitative trends seem to be robustly estimated
In the following, first the latency period study and then the generalities and some applications of the models developed to estimate incidence curve of problem drug use are reported. First the Back-Calculation model, then the general features of the system dynamic model and some ideas about possibilities for identification and use of such model to make scenario analyses are outlined.

The specific objectives to be attained in the second 12-month period were:

1. Finalise 3 papers on latency analysis and Back-Calculation model on the basis of the results already obtained.
2. Generalise the BC model and the related procedure to incorporate the epidemic level at start among the parameters to be estimated.
3. Perform an extensive simulation study to assess robustness of estimates and the dependency upon the phase of the epidemic (latency period).
4. Preliminary qualitative study of the dynamic model already set up.
5. Simulation procedure for the dynamic model and some scenario analyses for the evaluation of possible control interventions.
6. Outline of a Structural Equation Model (SEM) to study the impact of possible risk factors on the epidemic of problem drug use.

The related tasks were:

1. Latency period estimation for other countries and comparisons. Analysis of the aspects useful to make policy monitoring and evaluation through synthetic indicators.
2. Incidence curve estimation for other countries and again for the previous ones using the updated procedure and other possible models.
3. Generation of simulated data and use of these to assess the qualitative behaviour of the estimation procedures. Estimates of latency period distributions as functions of external co-variates such as age, gender, year...
4. Development of software for back-calculation.

The main contribution from all the partners in this phase has been first to discuss the results obtained during the first year of work at the second meeting organised in Munich in order to suggest new approaches and interpretations and then to provide new or updated data-base, both for the latency period analysis and for the Back-calculation, and to discuss further results in separate small meetings aimed at finalising local reports and joint papers.
In the following, first the latency period study and then the incidence estimation and dynamic model are presented in order to show the updated results.

The specific objectives to be attained in the third12-month period were:

1. Finalise 2 papers on latency analysis and Back-Calculation model on the basis of the results already obtained.
2. Qualitative study of the dynamic model already set up and evaluation of interventions.
3. Further scenario analyses.

4. Finalise 1 paper on qualitative study of the dynamic model already set up and scenario analyses.
5. Finalise 1 paper on methodological aspect of the BC model with applications.
6. Structural Equation Model (SEM) to study the impact of possible risk factors on the epidemic of problem drug use and preliminary application.
7. Preliminary draft guidelines for incidence estimation.

The related tasks were:

1. Latency period estimation for other data sets. Analysis of the aspects useful to make policy monitoring and evaluation through synthetic indicators.
2. Incidence curve estimation for the new latency period estimates, possibly at regional level, wherever data is available.
3. Evaluation of intervention through the dynamic model and the SEM model.
4. Providing a preliminary document presenting guidelines for incidence estimation.

In the following, the preliminary document presenting guidelines is reported as an overall summary of the three years work. In the 4 appendices, further documents presenting recent results are reported.

2.2 The Latency period analysis

The pilot sites were chosen (some Italian sites, Amsterdam, London, Dublin and Casal Ventoso in Lisbon).
Some analyses using survival models that allow to estimate the latency period distributions as function of the possible covariates and assess the influence of the various external variables available were conducted for the available data sets provided by some sites in the format described below. Suitable statistical models can be used to study the latency period distribution. Exploratory analysis can be conducted by means of the **Kaplan-Meyer** method for the global sample and for various stratifications to identify important covariates. Using the most important covariates, the **Cox regression model** can then be applied to estimate the parameters and evaluate the different impact (prognostic analysis) of the covariates on the latency period distribution. Both techniques are very popular methods in survival analysis (Collet, 1994; Marubini & Valsecchi, 1995)[1]. Finally, the best parametric models of the latency period distribution can be estimated by means of the P-Plot method.

Data needed for the latency period analysis.
For the latency period analysis data from health care services has to be provided according to the following specification:

Raw (as opposed to aggregated) data, classified according to, at least, the following variables:

- Age at first problem drug use
- Age at first registration in some health care service
- Gender

[1] Collet D. "Modelling survival data in medical research", Chapman and Hall, London, 1994.
Marubini E., Valsecchi M.G., "Analysing survival data from clinical trials and observational studies", Wiley, NY, 1995.

and to any other variable that could be used as covariate in the latency period analysis, such as:

- Educational level
- Ethnicity
- Residence
- Health care type
- Route of administration

To study and apply univariate and multivariate survival models the statistical package SPSS was used.

- Some results are shown in Table 1.1.

Table 1.1. Kaplan-Meyer summaries for the latency period analysis and Cox multivariate analysis results.

Site	Group (size)	Mean	Quartiles		
Amsterdam	Total (1058)	7.1	2	5	11
	Males (870)	7.3	2	6	11
	Females (188)	6.3	2	4	10
	Turkish/Moroc (60)	4.6	2	3	6
	Antillians (35)	8	3	6	13
	Netherl (756)	7.1	2	5	10
	Surinam (196)	8	3	6	12
	Treat. started. 85-88 (377)	6.3	2	5	10
	Treat. started 89-92 (388)	7.1	2	5	10
	Treat. started 93-95 (293)	8.2	3	7	13
	Age at first use <20 (403)	9.6	5	9	14
	Age at first use 20-25 (334)	6.3	2	5	9
	Age at first use >25 (321)	4.7	1	3	7
London	Total (8817)	6.7	2	5	10
Analysis for year starting treatment 95-97, for reliability reasons	Males (6588)	7	3	5	11
	Females (2229)	5.8	2	4	9
	White (5408)	6.9	2	5	11
	Black (480)	7	3	6	11
	Indian, Pakistan…(292)	3.5	1	2	4
	Chinese and asians (123)	8	3	7	12
	Others and missing (2874)	6.8	2	5	10
	Injectors (4750)	7.8	3	7	12
	Smokers (2639)	5.3	2	4	7
	Age at first use <16 (828)	10.3	5	10	15
	Age at first use 16-19 (2984)	8	3	7	12
	Age at first use 20-24 (2713)	6	2	4	9
	Age at first use >24 (2292)	4.6	2	3	6
Rome (inner)	Total (4658)	6.5	3	5	9
Rome (prov.)	Total (1696)	6	2	5	8
Rome (inner)	Age at first use <16 (555)	9.2	6	8	13
	Age at first use 16-21 (2675)	7	3	6	10
1996 data	Age at first use >21 (1426)	4.7	1	3	7
	Sex is not significant				

Cox multivariate model: in the last column is reported the ratio between the expected latency of the group indicate with respect to the expected latency of the baseline group when the other covariates are fixed.

Site	Group (size)	Baseline group	Ratio: mean latency of the group/mean latency of the baseline group
Amsterdam	Total (1058)		
	Males (870)	Males	1
	Females (188)		0.8
	Turkish/Moroc (60)		0.43
	Antillians (35)		1
	Netherl (756)		0.74
	Surinam (196)	Suriname	1
	Treat. started. 85-88 (377)	85-88	1
	Treat. started 89-92 (388)		1.25
	Treat. started 93-95 (293)		1.64
	Age at first use <20 (403)	<20	1
	Age at first use 20-25 (334)		0.6
	Age at first use >25 (321)		0.38
London	Total (8817)		
Analysis for year starting treatment 95-97, for reliability reasons	Males (6588)	Males	1
	Females (2229)		0.81
	White (5408)		2.04
	Black (480)		2.04
	Indian, Pakistan…(292)	Indian, Pakistan	1
	Chinese and asians (123)		2.2
	Others and missing (2874)		2.04
	Injectors (4750)	Injectors	1
	Smokers (2639)		0.68
	Age at first use <16 (828)	<16	1
	Age at first use 16-19 (2984)		0.74
	Age at first use 20-24 (2713)		0.53
	Age at first use >24 (2292)		0.4
Rome (inner)	Total (4658)	Rome inner	1
Rome (prov.)	Total (1696)		0.79
Rome (inner) 1996 data	Age at first use <16 (555)	<16	1
	Age at first use 16-21 (2675)		0.71
	Age at first use >21 (1426)		0.45

On the basis of these preliminary results it was observed an impressive agreement of the estimated means and medians for the different data sets and in particular for the three major capital cities analysed (Rome, Amsterdam, London) and of the influence of the covariates (Table 1.2).

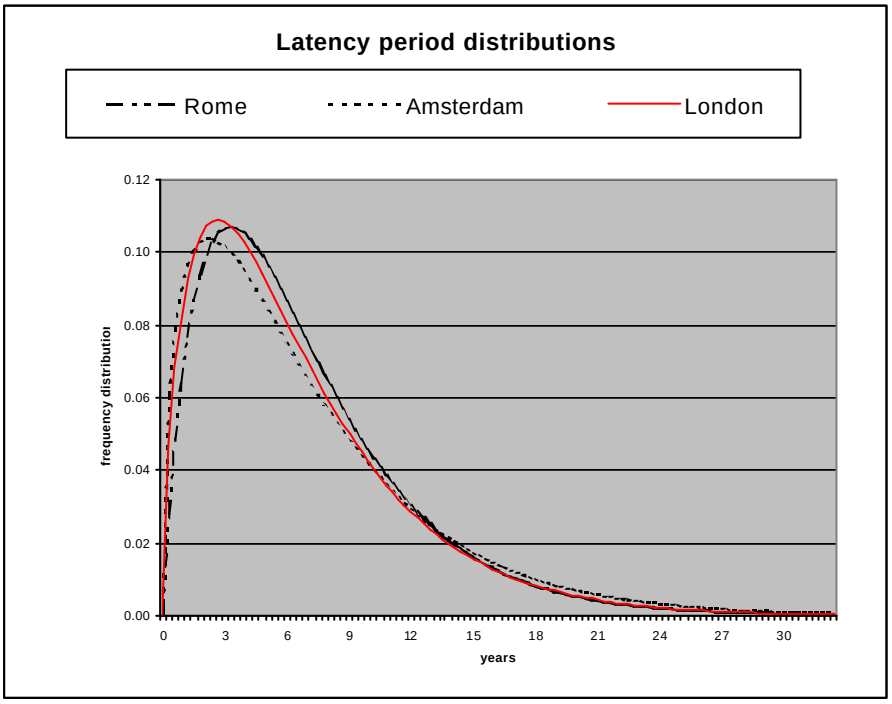
Table 1.2. Summary statistics for Kaplan-Meyer analysis for the Capital Cities

Country–Capital City	*Sample size*	*Mean*	*Stand.dev.*	*1° quart.*	*Median*	*3° quart.*
ITA – Rome metr.	4656	6.5	0.1	3	5	9
NET – Amsterdam	1058	7.1	0.2	2	5	11
UK – London	8817	6.7	0.1	2	5	10

From the parametric analysis by p-plot it resulted that the best models for the latency period distributions in the Capital cities is the gamma model (Figure 1.1). If these results will be confirmed by further comparisons of the survival functions, then there will be some evidence that the latency period relates much more to the "natural history" of drug addiction than to external aspects such as the availability of services, the waiting list and so on.
The parameters of the gamma distributions allow to obtain the following values of the mean of the latency period:

- Amsterdam: mean=7 years,
- Rome: mean=6.82 years,
- London: mean=6.64 years.

Figure 1.1. Gamma functions related to the latency period distributions of the capital cities.



The new analyses for the estimation of the latency period were performed on the basis of the updated data from UK, Italy (Lazio region, Trieste and Pisa), Amsterdam, Lisbon, Dublin and the French Community of Belgium. Recently a data set has been made available from Budapest (Dr. Katalyn Veress) and interesting results have been obtained from the analyses (updated Country report reported as appendix 1). Wherever possible, the Kaplan-Meyer method has been applied both for onset cohort (latency period) and for therapy cohort (backward latency period) comparing the results (This was not possible for Dublin only). The new analyses using Cox model differ from the previous ones mainly because the covariate

"age at first use" has been considered continuous in order to obtain comparability of the corresponding odd-ratios between the various countries.
In the following tables some summary results are reported both for the Kaplan-Meyer and for the Cox analyses.

Kaplan-Meyer analysis					
Site	**Group (size)**	**Median**	**Confidence interval 95%**		
Amsterdam	Total (1099)	5	4.5-5.5		
	Males (906)	6	5.5-6.5		
	Females (193)	4	3.1-4.9		
	Turkish/Moroc (73)	4	2.8-5.2		
	Antillians (37)	6	3-9		
	Netherl (788)	5	4.4-5.6		
	Surinam (201)	6	4.6-7.4		
	Treat. started. 92 (entry cohort)	4	2.8-5.2		
	Drug use started 86 (onset cohort)	5	2.4-7.6		
	Route of administration is not available				
	Best parametric model	***Gamma***	a ***=1.44***	l ***=0.20***	***Mean*** $=a/l=$ ***7.2***
London	Total (8817)	5	4.9-5.1		
Analysis for year starting treatment 95-97, for reliability reasons	Males (6588)	5	4.8-5.2		
	Females (2229)	4	3.8-4.2		
	White (5408)	5	4.8-5.2		
	Black (480)	6	5.3-6.7		
	Indian, Pakistan…(292)	2	1.7-2.3		
	Chinese and asians (123)	7	5.4-8.6		
	Others and missing (2874)	5	4.7-5.3		
	Injectors (4750)	7	6.7-7.3		
	Smokers (2639)	4	3.9-4.1		
	The analysis for year of onset versus year of entry is graphically reported below (the series F1 and F3 are latency curves and the series F2 is the backward latency which is statistically non-significant)				
	Best parametric model	***Gamma***	a ***=1.66***	l ***=0.25***	***Mean*** $=a/l=$ ***6.6***
	Group (size)	**Median**	**Confidence interval 95%**		
Rome (inner)	Total (4658)	6	5.8-6.2		
	Sex is not significant				
	Ethnicity and route of administration are not available				
	Best parametric model	***Gamma***	a ***=1.84***	l ***=0.27***	***Mean*** $=a/l=$ ***6.8***
Belgium	**Group (size)**	**Median**	**Confidence interval 95%**		
	Total (2608)	6	5.77-6.23		
	Males (2019)	7	6.75-7.25		
	Females (578)	5	4.56-5.44		
	Treat. started. 94 (entry cohort)	5	4.62-5.38		
	Drug use started 86 (onset cohort)	7	6.63-7.37		
	Route of administration is not significant				
	Ethnicity is not available				
	Best parametric model	***Gamma***	a ***=2.45***	l ***=0.34***	***Mean*** $=a/l=$ ***7.2***

Cox multivariate model: *in the last column is reported the ratio between the expected latency of the group of interest with respect to the expected latency of the baseline group when the other covariates are fixed.*			
Site	**Group (size)**	**Baseline group**	**Odd-ratio of risk functions (CI)**
Amsterdam	Total (1099)		
	Males (906)	Males	1
	Females (193)		1.25 (1.06-1.46)
	Surinam (201) Antillians (37)	Suriname/Antillians	1
	Netherl (788)		1.27 (1.09-1.47)
	Turkish/Moroc (73)		2 (1.54-2.62)
	Age (baseline 11)		1
	Each year		1.05 (1.03-1.07)
London	Total (8817)		
	Males (6588)	Males	1
	Females (2229)		1.23 (1.17-1.30)
	Indian, Pakistan…(292)	Indian, Pakistan	1
	White (5408) Black (480)		0.58 (0.52-0.66)
	Chinese and asians (123)		0.61 (0.49-0.76)
	Injectors (4750)	Injectors	1
	Smokers (2639)		1.39 (1.33-1.46)
	Age (baseline 12)		1
	Each year		1.06 (1.05-1.07)
Rome (inner)	Total (4658)		
	Age (baseline=14)		1
	Each year		1.07 (1.05-1.8)
Belgium	Total (2608)		
	Males (2019)		1
	Females (578)		1.2 (1.1-1.3)
	Age (baseline 10)		1
	Each year		1.07 (1.05-1.09)

Onset delay can be analysed by entry cohort (ie. by year of first report) or by “onset” cohort (ie. by year of first use). In the first case we are estimating the latency period distribution, in the second one the backward latency period distribution. All being equal they produce the same figures. However, if incidence changes over time analyses by entry cohort may be biased – because it will look like onset delay is decreasing over time if incidence is increasing, or vice versa.

The idea is that, if the two survival curves are not significantly different, as those reported in Figure 1, then we can infer that the epidemic is fairly stable over the period taken into account and the possibility of biases in the estimate of the latency period distribution is low.

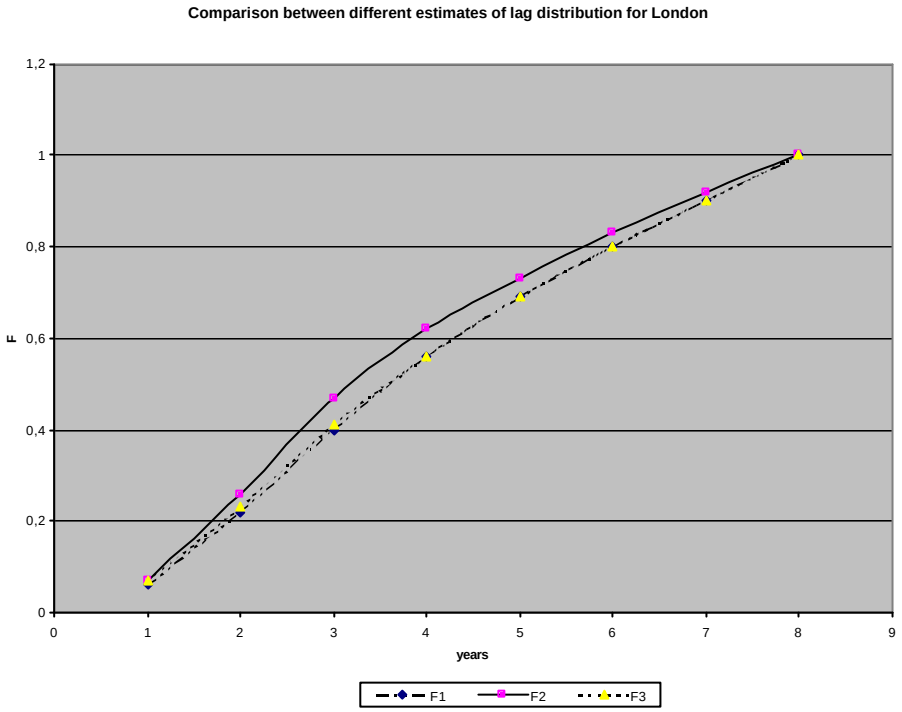
Similar results, with respect to those reported in Figure 1, were obtained for Amsterdam, Belgium and the various Italian sites, thus allowing to use the backward latency as an approximation of the latency distribution for most sites. Some results have been presented at the Harm Reduction Conference in Jersey (2000) and at the workshop on "Dynamic Drug Policy: Understanding and Controlling Drug Epidemics", May 22-24, 2000, Vienna International Centre (UNDCP) and will appear in the proceedings (UN Bulletin on Narcotics).

It must be noted that the odd-ratio corresponding to the covariate “age at first use” do not vary significantly among different sites.

A special analysis regards the study of the period between first use and first continuous use for the sites were this information is presently available (Trieste and Pisa). Preliminary analyses suggest that within the LP a non-problematic period of about 1 year exists, again with stable distribution, from first heroin use to first regular heroin use. In particular, data analysis, conducted with the same methodologies considered above, on such initial period of drug use by Christos Salamouras at the University of Rome Tor Vergata, provided important information about the natural history of heroin use. Main results can be summarized as follows. The period between first heroin use and first regular heroin use show the same distributional characteristics in the two sites analized with a mean of 2 years and a median of 1 year. The year of first treatment is not significant as well as sex. Age at first use is significant showing an adjusted odd ratio 1.03 for each year of age. Educational

level is significant as well, but further analyses are necessary to better investigate the effect of this covariate, due to incompleteness of data presently available.

Figure 1



New analyses were performed at local level for Italy (the only new data set presently available) by using the same methodology used previously. The results were compared with those obtained in the previous 12 month period and presented at the Harm Reduction Conference in New Delhi (2001). The results are also reported in a paper in preparation to be submitted to an epidemiological journal. Summary results are reported in the draft guidelines presented in the following.
New important methodological improvements were also obtained and are widely presented in the two papers reported as appendix 2 (accepted for publication in Computational Statistics and Data Analysis), and appendix 3 (preprint accepted for oral presentation at the Harm Reduction Conference in Lubjana 2002).

2.3 The Incidence estimation

The incidence of problem drug use was investigated, through the Back-Calculation (BC) methodology, by using treatment data and estimates of latency period distribution obtained as explained above.
The BC is a general class of deconvolution methods originally proposed as a tool for estimating the minimum number of HIV-infected people and making short-term projections of AIDS incidence (Brookmeyer and Gail, 1986). As the HIV/AIDS epidemic developed and knowledge of its elements increased, particularly of the incubation period distribution, more and more sophisticated BC methods were implemented and used to estimate the HIV-infection curve too (Brookmeyer and Gail, 1988, Rosemberg and Gail, 1991; Brookmeyer 1991).
The basic idea of each BC method is to reconstruct, through a deconvolution procedure, and

by using an estimate of the incubation period distribution, the numbers of individuals who must have been previously infected in order to yield the observed AIDS incidence cases. Then, by applying the assumed incubation distribution to the estimated HIV infection curve, and making some assumptions on future HIV infection rates, the AIDS incidence is projected forward.
Let *A(t)* be the expected cumulative number of AIDS cases diagnosed by calendar time *t*, *h(s)* the HIV infection rate at calendar time *s*, and *F(t)* the incubation period distribution, then the convolution equation

$$A(t) = \int_0^t h(s)F(t-s)ds \qquad (1)$$

is known as the *"fundamental Back-Calculation equation".* The equation (1) links, through the incubation period distribution, the HIV infection rate to the AIDS incidence. In fact an individual results diagnosed with AIDS at calendar time *t* only if he has been previously infected at a calendar time *s, s<t* and has an incubation period less then *t - s.* Therefore the basic idea of the BC is to use a realization of *A(t),* the AIDS incidence data, an estimate of *F(t)*, usually external, and to use the equation (1) in order to gain information about the past infection rates *h(s), s<t.*
The *fundamental Back-Calculation equation* (1) has various explicit formulations, corresponding to specific assumption on the expected AIDS incidence, the shape of the HIV infection curve, the incubation period distribution, and the estimation procedure. Each different combination of the above assumptions results in a different BC method.

2.3.1 The Empirical Bayesian Back-Calculation.

The BC methods which are based on step functions or splines for *h(s)* are the most flexible and those providing the best compromise between bias and variability of estimates.
Heisterkamp et coll. (1995, 1999) proposed a BC method based on an Empirical Bayesian approach where the HIV infection curve is represented by a step function, and a Poisson process is postulated for the occurrence of the infections in a single time interval. Thus, the AIDS incidence in each interval is assumed to be independently Poisson distributed. Without any constraint the estimated HIV incidences might be highly variable, therefore a smoothness restriction is adopted by placing a prior distribution for the infection parameters to be jointly estimated. The advantage of this BC method is that it provides the simultaneous estimation of the infection curve parameter and of the degree of smoothing by using the EM algorithm (Tanner, 1996). The penalty parameter of the penalized likelihood, which is directly linked to the degree of smoothing, determines the relative weights given to the data and to the prior distribution: large values for the penalty parameter give more weight to the prior information than to the data. To implement the model the following quantities can be defined:

$I_T(s)$: the incidence of drug users who eventually present to treatment at time *s, s = 1,…,S*. These individuals pass through a period of hidden drug use before they become visible by having their first contact with some health care service;
$F_T(v-s)$: the cumulative distribution of the period between the time *s* of the first problem use of drug, and the time *v* of the first presentation for treatment. This distribution is the "Latency period distribution" defined and estimated above.
The goal of the present application is to estimate the incidence of DUs eventually seeking for treatment $I_T(s)$ using data on the incidence of DUs in treatment, $I_{treat}(v)$ using the estimated latency period distribution $F_T(v-s)$ on the basis of data collected by the Focal Point and data provided by the health care services offering treatment (of any kind) to the drug users. The incidence $I_T(s)$ can be estimated through the EB-BC method, by deconvolving the following equation:

$$I_{treat}(v) = \int_0^v I_T(s)d(F_T(v-s))$$

This is analogous to the earlier equation (1).
A proper (new) EB-BC model, analogous to that developed by Heisterkamp et al. to estimate HIV incidence curve from AIDS incidence data, was developed and applied, in order to study the extent of problem drug use in Italy, The Netherlands and UK, through a software programme specifically written in S-plus language. In order to use the programme suitable data are required.

2.3.2. The data needed for the EB-BC analysis

The version of EB-BC software used in the present project has been developed on the basis of the structure of Italian data, provided by the national focal point which is presented in detail in the following of this section.
The data-file needed for the EB-BC could be derived either from the data used for the latency period analysis, yet if in a different format, or from different data sources, depending on the local availability of data.
For Italy the EB-BC was performed on the basis of national data, other than those used for the latency period analysis (local data). On the contrary for both the United Kingdom and The Netherlands, the EB-BC was applied to data-files which have been obtained by transforming, mirroring the Italian file-structure, the data used for the latency period analysis performed for each country.
For each country, data must be provided in just one file, whose format must be fixed or tab-delimited ASCII.
The EB-BC uses biannual incidence data of "new" individuals entering treatment in some health care services.
The Italian data file contains multiple records, each one with 47 fields (for a total of 200 bytes) as follows:

- Field 1: type of the health care service
- Field 2: geographic area
- Field 3: date of first registration to the health care service
- Field 4: total incidence (number) of new DUs under treatment: male
- Field 5: total incidence (number) of new DUs under treatment: female
- Field 6 - 21: incidence (number) of new DUs under treatment by age categories: male/female
- Field 22 - 35: incidence (number) of new DUs under treatment by occupation: male/female
- Field 36 - 47: incidence (number) of new DUs under treatment by education level: male/female

See the legend, reported in Table 2.1., for a detailed description of records.
Note that for the present application just the information contained in the first 21 fields were used, since this version of EB-BC include just the covariate "age at first treatment". Nevertheless it was decided to keep the other two variables, occupation and education level as well, since they could be included in the EB-BC in the future.
The data provided by UK required some cleaning and work to be properly used, as they presented in the wrong format, both for the latency analysis and for the BC procedure, this latter, in particular, had to be specifically adapted in order to perform the incidence estimation for UK. This problem should not be present anymore in the future when the data to be processed should be provided in the format described above.

Table 2.1. Legend: Description of the structure of treatment data file used the EB-BC

Variables	Field length	Notes	
Health care service type	3 bytes	A21 = Public health care services	
		B21 = Private health care services	
Geographic area	15 bytes		
Date (aammgg)	6 bytes	**0630 **1231	
Total incidence (number) of new DUs in treatment, male	4 bytes		
Total incidence (number) of new DUs in treatment, female	4 bytes		
Incidence (number) of new DUs in treatment by age category		*Age categories*	
		Until 31-12-1990	Since 1-1-1991
Male - age cat. 1	4 bytes	<15	<15
Female – age cat. 1	4 bytes	<15	<15
Male - age cat. 2	4 bytes	16-18	16-19
Female – age cat. 2	4 bytes	16-18	16-19
Male - age cat. 3	4 bytes	19-22	20-24
Female – age cat. 3	4 bytes	19-22	20-24
Male - age cat. 4	4 bytes	23-25	25-29
Female – age cat. 4	4 bytes	23-25	25-29
Male - age cat. 5	4 bytes	26-30	30-34
Female – age cat. 5	4 bytes	26-30	30-34
Male - age cat. 6	4 bytes	31-40	35-39
Female – age cat. 6	4 bytes	31-40	35-39
Male - age cat. 7	4 bytes	>40	>39
Female – age cat. 7	4 bytes	>40	>39
Male - age cat. 8	4 bytes	NA	NA
Female – age cat. 8	4 bytes	NA	NA
Incidence of new DUs in treatment by occupation		*Occupation categories*	
Male	4 bytes	No occupation	
Female	4 bytes	No occupation	
Male	4 bytes	Looking for the first occupation	
Female	4 bytes	Looking for the first occupation	
Male	4 bytes	Unoccupied	
Female	4 bytes	Unoccupied	
Male	4 bytes	Under-occupied	
Female	4 bytes	Under-occupied	
Male	4 bytes	With stable occupation	
Female	4 bytes	With stable occupation	
Male	4 bytes	Student	
Female	4 bytes	Student	
Male	4 bytes	NA	
Female	4 bytes	NA	
Incidence of new DUs in treatment by education level		*Education levels*	
Male	4 bytes	None	
Female	4 bytes	None	
Male	4 bytes	Junior degree	
Female	4 bytes	Junior degree	
Male	4 bytes	"Low medium degree"	
Female	4 bytes	"Low medium degree"	
Male	4 bytes	High school	
Female	4 bytes	High school	
Male	4 bytes	University degree	
Female	4 bytes	University degree	
Male	4 bytes	NA	
Female	4 bytes	NA	

2.3.3 The models presently allowed by the procedure as latency period distributions.

The EB-BC method has been modified, with respect to the previous version set up to study the HIV/AIDS epidemic, in order to allow the use of the various models for the latency period distributions, according to the results of the latency period analysis. In particular the following models have been considered:

1. Markov model (allowing for forward and backward jumping to stages)
2. Approximate Markov model when in fact a Gamma distribution is fitted (by equating the fisrt two moments of the Gamma to a sum of k independent exponentials with rate $?_i$ of which the first k-1 have an equal rate ?, k is chosen heuristically)
3. Gamma
4. Weibull
5. Log-Normal

When performing the EB-BC with age-covariate, once one of the five models above has been specified, it is possible to use just one set of parameter values for every age-category or a different set for each category. Clearly, the latter option should be chosen in case the latency period analysis would provide a different estimate of the latency period distribution for each age-category in parametric form.

2.3.4 Output provided by the EB-BC.

The output provided by the EB-BC consists of different incidence and prevalence figures and of the corresponding confidence intervals.
For each age category and for the total population, if the "age at first treatment" covariate is included in the EB-BC model, and just for the total population, if the covariate is not included, the following figures are provided:

1. observed yearly incidence of DUs in treatment
2. estimated yearly incidence of DUs in treatment (with confidence intervals)
3. estimated yearly cumulative incidence of DUs in treatment (with confidence intervals)
4. estimated yearly cumulative incidence of DUs (at first problem use of drug)
5. estimated yearly incidence of DUs (at first problem use of drug) (with confidence intervals)

Some other figures, such us the prevalences, could be easily provided as well, but they should be based on some hypothesis requiring information not available yet.

2.3.5 Some preliminary results

The EB-BC was applied separately for Italy, UK and The Netherland, by using the treatment data, which were provided by the health care services of each Country, and various estimates of the latency period distribution. In particular for each country, the EB-BC was performed both with and without the inclusion of the age-covariate in the model, and by modelling the latency period either with a Gamma or with a Weibull density. The other possible models were excluded on the basis of the latency period analysis performed in each site. In this section an overview of such results is given through the Figures 2.1 and 2.2. The two figures show, for each country, respectively the curves of yearly incidence of problem drug use, and of yearly incidence of DUs presenting to treatment, as obtained with and without the age-covariate and for the various latency period distribution models used. It is important to stress that such results must be considered just as preliminar. Nevertheless, they can provide some qualitative comparisons about the extent and the dynamic of problem drug use in the three countries. Clearly, the incidence curves represented in the figures do not allow us to compare the magnitude of the problem among the countries, since they do

not take into account the total number of inhabitants of each country or city. On the contrary the figures show the differences in the location and number of peaks, and in the stage of the problem drug epidemic in the three countries. In particular, some qualitative conclusions can be drawn. As a matter of fact, the epidemics in the three sites appear quite different in shape

The Netherlands: the shape is peaked around 1984 with a fast decrease towards a low endemic phase in the following years, the epidemic seems to have reached a low level endemic phase.

Italy: the epidemic seems to be much more important, still increasing at the end of the '80s with two peaks at 85 and 91 (higher) and an endemic phase at high level afterwards. Thus the epidemic seems to have reached a high level endemic phase.

UK: the incidence peaked in 1990, with new cases presenting to treatment peaking in 1994/95 and a general flattening of the rise in the treatment population towards the end of the 1990s.

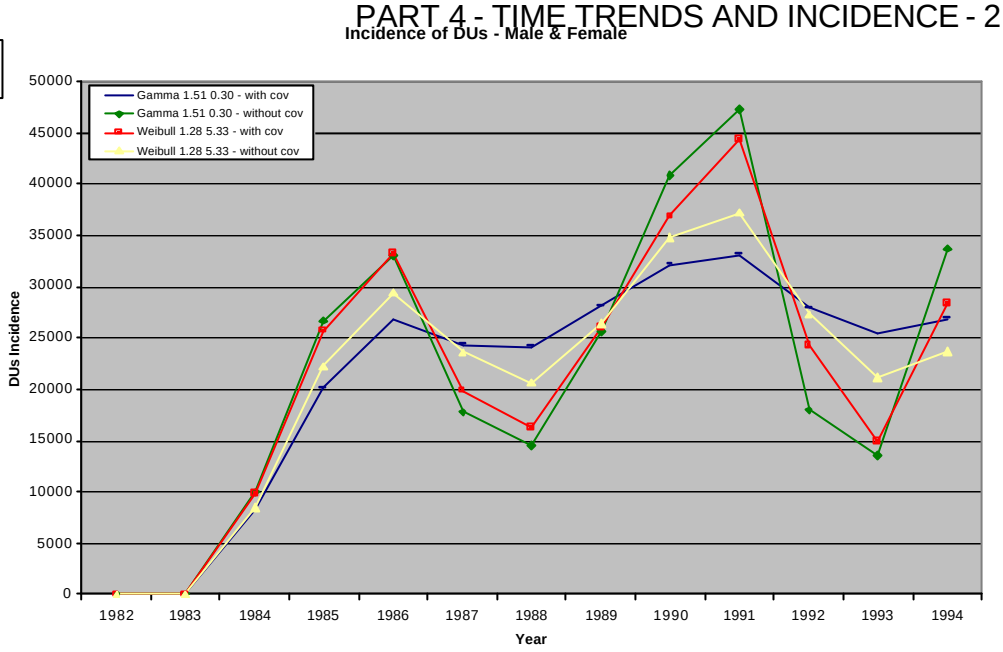
From the analysis of the results it appears that the therapy data available in Italy and UK captures the very beginning of the epidemic (therapy incidence data are well interpolated by the therapy projected data since the beginning), whereas the therapy data available for Amsterdam does not allow to estimate the very beginning of the incidence curve of problem drug use (observed therapy data are all above the predicted data at the beginning). This left truncation is reflected in the estimate of the incidence curve of problem drug use that should actually have a level higher that zero at the beginning. This problem will be addressed in the next phase of the present project. As a matter of fact, presently a new version of the procedure is under development (S. Heisterkamp and L. Ravà) to incorporate the initial level of the incidence curve in the set of parameters to be estimated.

Figure 2.1: Empirical Bayesian Back-Calculation analysis: incidence curves of problem drug use as estimated with and without the age-covariate and for the various latency period distribution estimates for Italy, The Netherlands and UK.

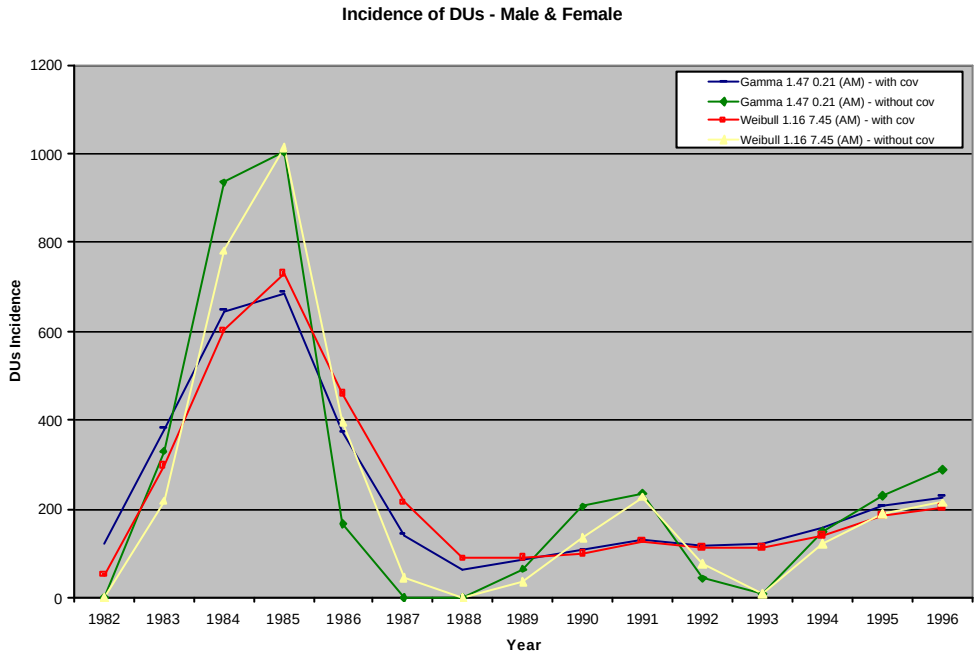
Figure 2.2: Empirical Bayesian Back-Calculation analysis: incidence curves of DUs presenting to treatment, as estimated with and without the age-covariate and for the various latency period distribution estimates used for Italy, The Netherlands and UK.

Figure 2.1

ITA



NET



UK

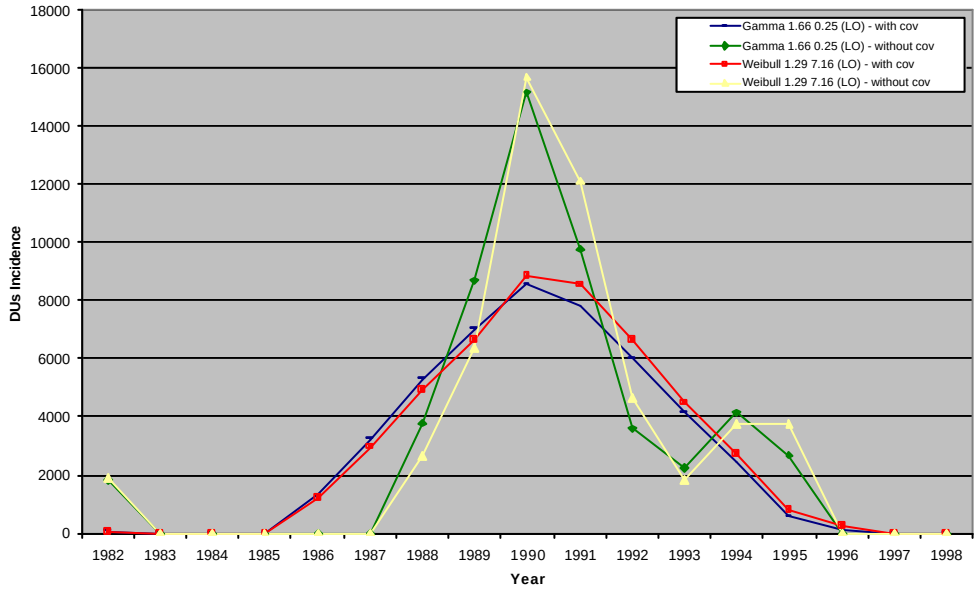
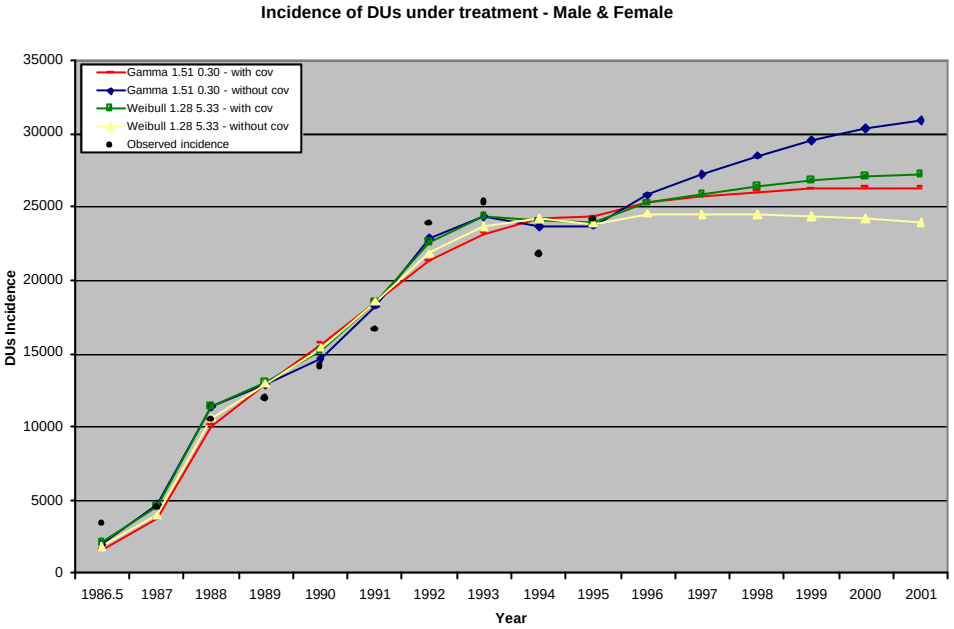
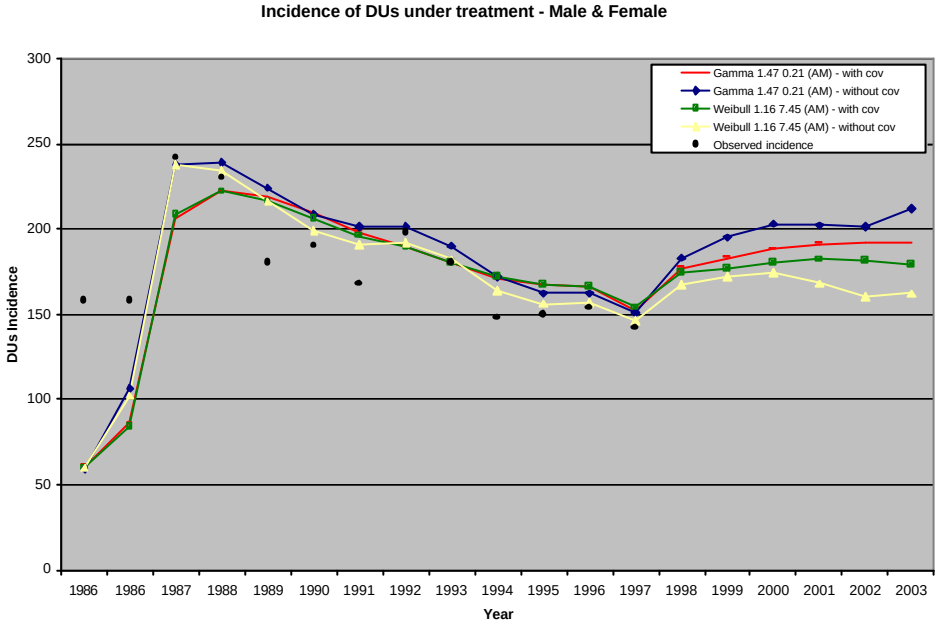


Figure 2.2

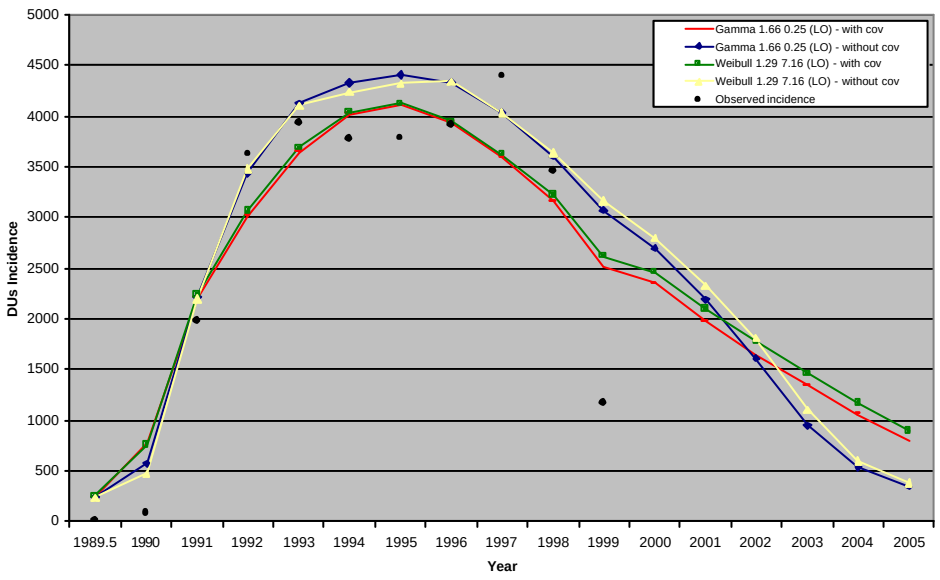
ITA



NET



UK



2.4 Sensitivity analysis of EB-BC method

In order to evaluate the performance of the EB-BC in investigating the extent and the dynamics of problem drug use, a sensitivity analysis, particularly regarding the inclusion of the age-covariate in the model and the choice of the latency period distribution, was performed using Italian data.
The Italian data about DUs in treatment in the public institutions, without any correction for the proportion of double counting and for the proportion of drug users in treatment who are not heroin users, were used; 8 estimates of latency period distribution (4 Gamma and 4 Weibull models), as provided by the latency period analysis performed for 4 Italian cities (Rome - RM, Milan - MI, Frosinone - FR, Latina - LT) were considered. The EB-BC was performed both with and without the age-covariates.
Figures 2.3 and 2.4 report the curves of the incidence of problem drug use and of the observed and expected incidence of DUs in treatment (therapy data-incidence) for the total population, as provided by the EB-BC performed with and without the age-covariate, and corresponding to the 8 latency period distributions. The problem drug use incidence curves showed in Figure 2.3 are all bi-modal and all included in a quite narrow range. Just 2 curves, both obtained by performing the EB-BC without the age-covariate and by using respectively the Weibull models estimated through the Rome and Milan data, are quite different from the others. At a first glance, it appears that the application EB-BC model with the age-at first treatment covariate provides curves that are smoother than those obtained by the model without the age covariate. Moreover Figure 2.3 provides evidence that the method is more sensitive to the values assumed by the parameters of the latency model rather than to the parametric form of these models. For example it can be seen that the curves obtained by using the gamma and the Weibull densities estimated for the city of Latina, are closer to each other than the curves based on the gamma densities estimated for Rome and Latina. On the other hand it should be stressed that, in order to obtain a good picture of the problem drug use at national level, it is important to use an estimate of the latency period distribution based on a data set representative of the whole country under study, rather than just of a "peculiar" city like Rome as in the present estimation (Rome is the Capital of Italy, the biggest, and the most densely populated city of the Country).
From Figure 2.4. it can be observed that the fitting of the expected to the observed treatment data is quite satisfactory, and similar to each other, for any one of the latency models. The same figure provides evidence that, for each latency period distribution, the right tails of the expected incidence of DUs (the predicted incidence) in treatment are heavily determined by the corresponding estimated curves of incidence of problem drug use reported in Figure 2.3.

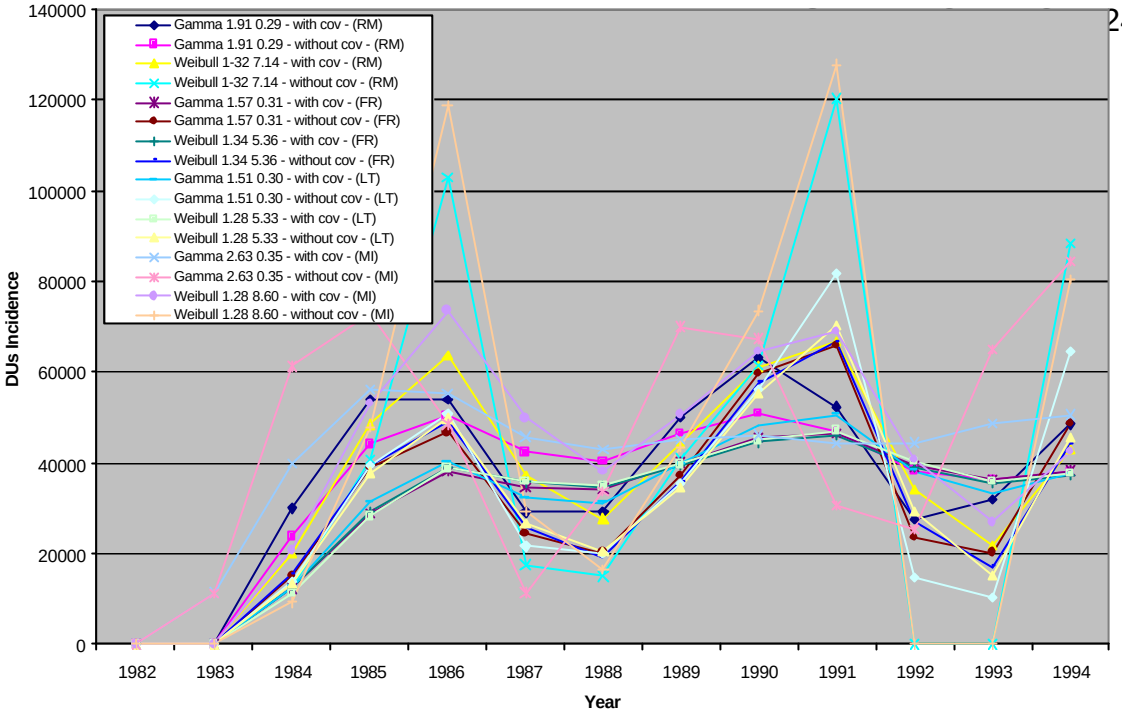


Figure 2.3. Sensitivity analysis for latency period distribution: incidence curves for the total population estimated by the Empirical Bayesian Back-Calculation procedure with different models of the latency period distribution.

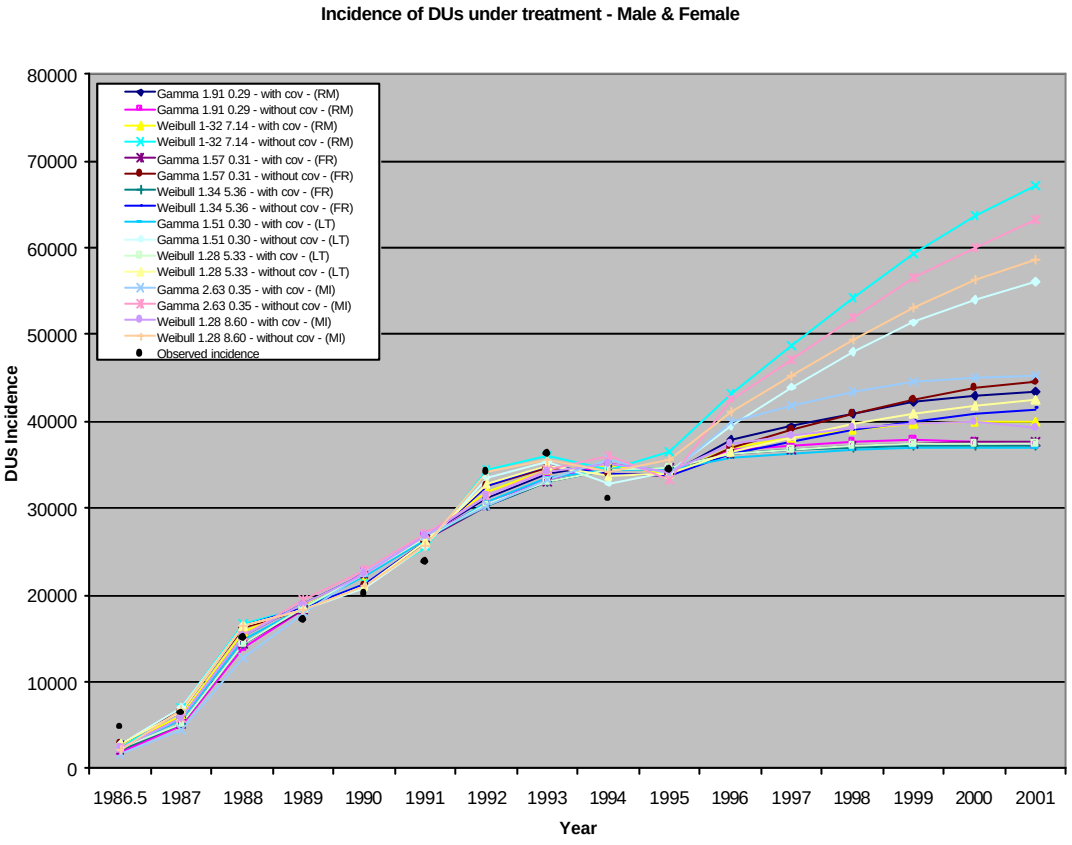


Figure 2.4. Sensitivity analysis for latency period distribution: Incidence curves for the presentation to therapy estimated by the Empirical Bayesian Back-Calculation procedure with different models of the latency period distribution and observed data.

8. Conclusive remarks and further developments.

From the results of the analysis of the different data sets provided by the participants, both for the estimation of the latency period distribution and of the incidence of problem drug use, we can conclude that:

1. The latency period appears to be remarkably similarly distributed over different sites, with a median of between four and six years and an average of between five and seven years. This time-lapse, however, appears to be much longer than this in young drug users and inner-city drug users. Differences relating to ethnicity also where observed, whenever this covariate could be included in the analysis. The parametric models which seem to be more suitable to represent the distribution of the latency period belong to the gamma and Weibull families.

2. The incidence curves provided by the EB-BC estimation procedure are strongly dependent on the latency period model chosen, but the location of the peaks of the epidemic seems to be a robust parameter. Also the qualitative trends seem to be robustly estimated. The cumulative incidence curves, which provide an overall size of the epidemic, also show a low sensitivity to the model chosen for the latency period distribution.

It must be noted that yet if the performance of the EB-BC applied to the HIV/AIDS field are known to be good, this is just the first time such methodology is used for studying the problem drug use, therefore the software used is presently a beta test version.
Other problems are related to the various possible biases which affect the data available for both the latency and the EB-BC analysis:

1) Latency period:

- there might be some bias because there is no standardised way to ask the age at first heroin use at the treatment centres. When, at a treatment centre, the question “how long are you using drugs?” is raised, the client can interpret this question as the period of uninterrupted drug use before treatment demand. When the period of drug use is interrupted, the latency period seems to be shorter, and the age of starting drug use will be higher than real figures.
- Age of first use is less reliable than age. This affects both ends of the distribution, in particular ages under 12 and those over 30 of “age at first use”. Short latency periods observed for older drug users are less reliable than short periods among younger users.
- Estimates of the latency period using treatment/surveillance data may under-estimate the true value because the data are right truncated. This bias is higher for recent epidemics whereas it will be minimal for older (stabilised) epidemics. Possibly the case of Dublin can be explained this way.
- The latency period can be analysed by entry cohort (i.e. by year of first report) or by “onset” cohort (i.e. by year of first use). All being equal they produce the same estimates. However, if incidence changes over time, analyses by entry cohort may be biased as they tend to produce decreasing observed periods when incidence is increasing (possibly the case of Dublin) and viceversa.
- There might be some bias due to local peculiarities of the therapy services. For example, in Amsterdam large scale methadone programmes started at 1980 and opiate users couldn’t apply for treatment during the ‘70s. Therefore, during the first years, the latency period will be prolonged. This bias could be corrected including in the study only opiate users who demanded for treatment for the first time after 1985.
- There seem to be differences in latency period between drug users originating from different countries. These differences could reflect differences in the onset of the heroin epidemic among different subgroups. In this case the heroin epidemic among those originating would be the oldest, followed by the epidemics among the other groups. When the epidemic grows the latency period will increase, especially when the incidence is decreasing. The same effect may affect the stratified analysis with respect to other variables, such as “route of administration” or “sex”.

2) Back-Calculation:

- the main problem related to the data used to apply the BC procedure is represented by the double counting, present in particular in the national data provided by the focal points, such as for example for Italy, which causes the actual incidence of DUs presenting to treatment be lower than the observed one, and, as a consequence a bias in the EB-BC estimates. An attempt to overcome such problem could be done by inflating the observed incidence data, on the basis of some information about the amount of double counting, if available. The possible biases introduced by the double counting in the estimated incidence curves of DUs and DUs presenting to treatment were investigated through the sensitivity analysis described below. The analysis was performed by applying the EB-BC to the complete Italian data-set and to the same data-set but with a 30% inflation, with and without the age-covariate and by using the Gamma and the Weibull models for the latency period distribution. Figures 2.5 and 2.6 show the incidence curves of DUs as estimated by the EB-BC, respectively with and without the inclusion of age-covariate, applied to the observed treatment data with (Total Population) and without (Inflated Population) double counting. It can be noted that, yet if the double counting does not appear to have any effect on the location and on the number of peaks of the estimated DUs incidence curves, the curves are different not just in level but also in shape. Moreover the effect of double counting strongly depends on the latency period distribution used, and on the age-covariate being included or not in the EB-BC model. In particular, the differences in the shape of the incidence curves are more evident if the Weibull rather than Gamma distribution was used. Similar, and even stronger, are the effects of double counting on the estimated incidence of DUs presenting to treatment, whose curves, corresponding to the total and inflated population, and to the various latency period distribution are reported in Figures 2.7 and 2.8, respectively for the EB-BC performed with and without the age-covariate. From Figure 2.7, i.e. EB-BC with age covariate, it is evident that the model fitting is poorer when the total data rather than the inflated data were used. Finally, it should be noted that, when applying the EB-BC without age-covariate (Figure 2.8), the double counting produces a rather dramatic overestimation of future incidence of DUs presenting to treatment.
- Since the DUs epidemic starting year needs to be inputed as an external parameter in the EB-BC model, a further sensitivity analysis was carried out, on the basis of the Italian data, in order to evaluate the effect of different starting point on the results. The analysis was performed by applying the EB-BC to the observed incident cases of DUs in treatment in Italy, ("inflated" as much as 30%, for taking into account the double counting and proportion of DUs non-heroin users), up to the end of 1991, and by using the Gamma latency period distribution. Four different starting points were considered (1978, 1980, 1981 and 1982). In order to assess the "best starting year", the Pearson Chi-square index was then used for comparing, for the years 1992-1994, the incidence of DUs in treatment projected by the EB-BC and the observed incidence data. Figure 2.9 and 2.10 report respectively the incidence curves of problem drug use and of DUs presenting to treatment and the Chi-square values.

The methods should be applied to new data sets from other sites and countries for better analysing the biases and the potentialities as tools to provide information to policy makers on the problem drug use epidemic in different situations.

Several developments can be proposed to improve the results obtained, in particular to correct the possible biases discussed above:

- Future analyses need to estimate latency period by “onset” cohort, in order to test further whether it does vary by the cofactors evidentiated in the present work.
- Introducing specific questions on type and age of “first treatment” demand into routine surveillance would improve the above analyses and help the interpretation of trends of problem drug users seeking treatment. Introducing questions on age at first use would also be of value.
- The bias introduced in the estimates of the latency period distributions, due to the right truncation of the data available for the analysis, might be corrected using other external information and standard models and methods.
- Another interesting analysis relates to further sensitivity studies on the results of the Back-calculation procedure, possibly using boostrap methods.

Figure 2.5. Sensitivity analysis for double counting: Incidence of DUs estimated through the EB-BC with age covariate.

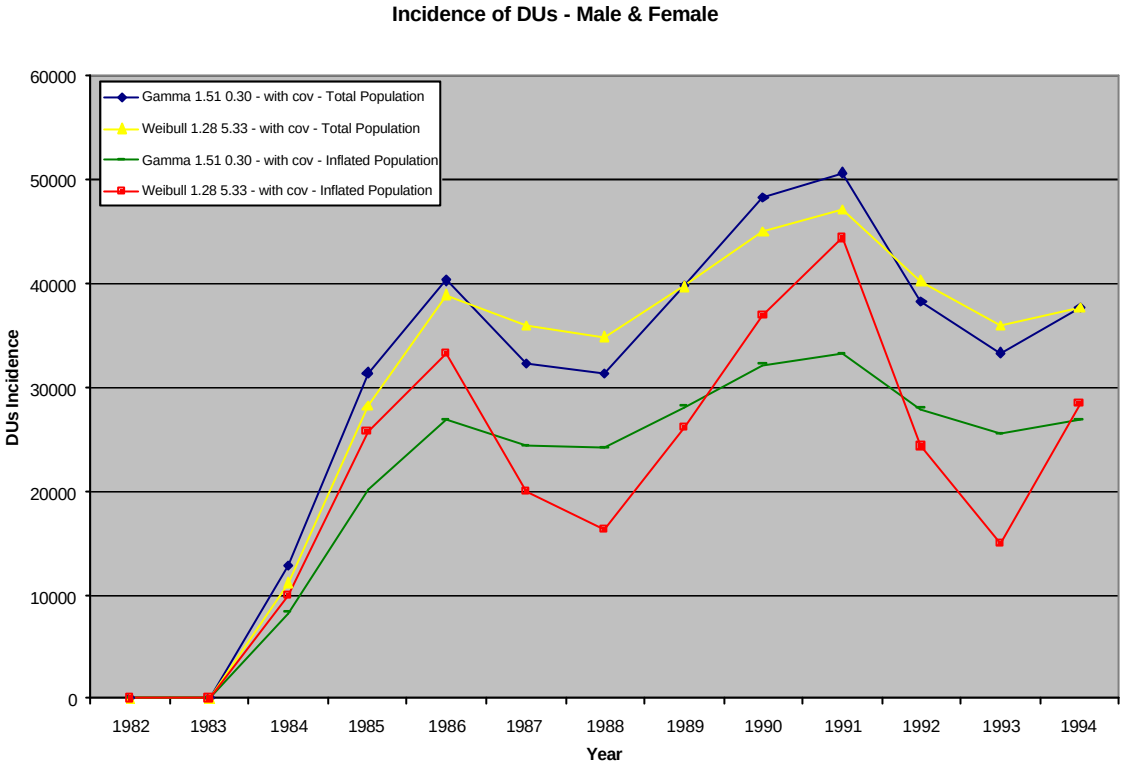


Figure 2.6. Sensitivity analysis for double counting: Incidence of DUs estimated through the EB-BC without age covariate.

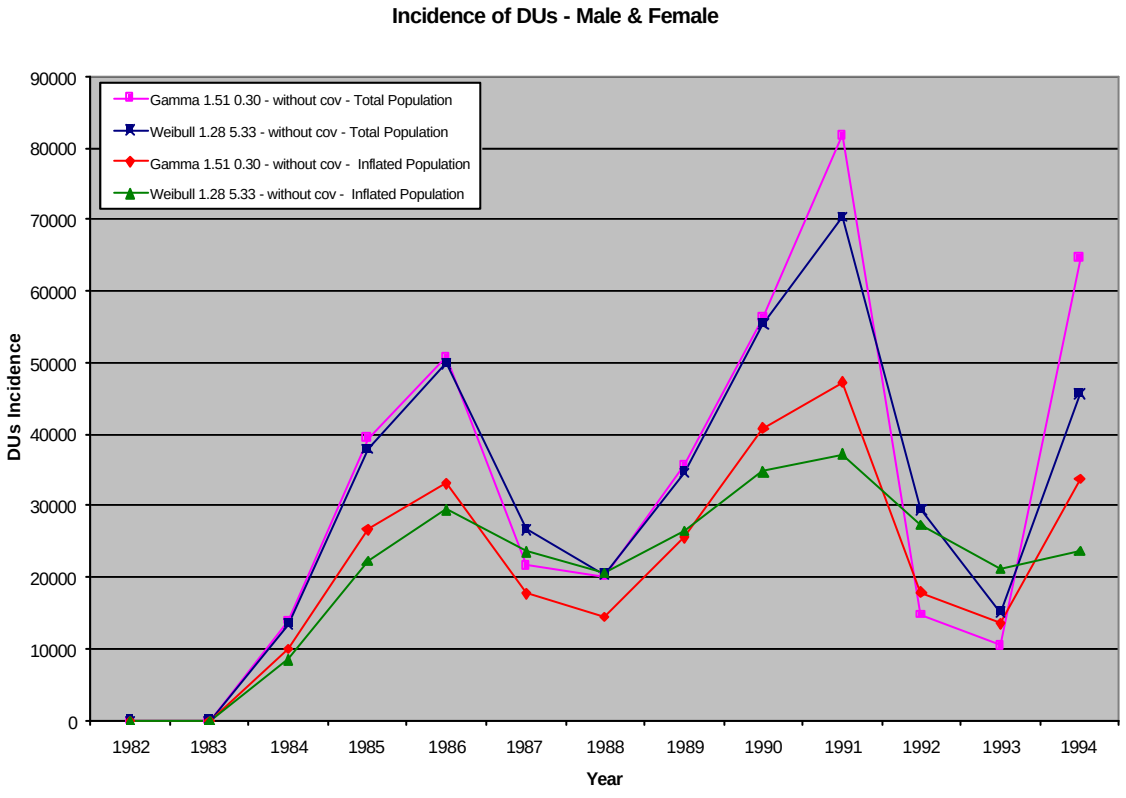


Figure 2.7. Sensitivity analysis for double counting: Incidence of DUs presenting to treatment estimated through the EB-BC with age covariate.

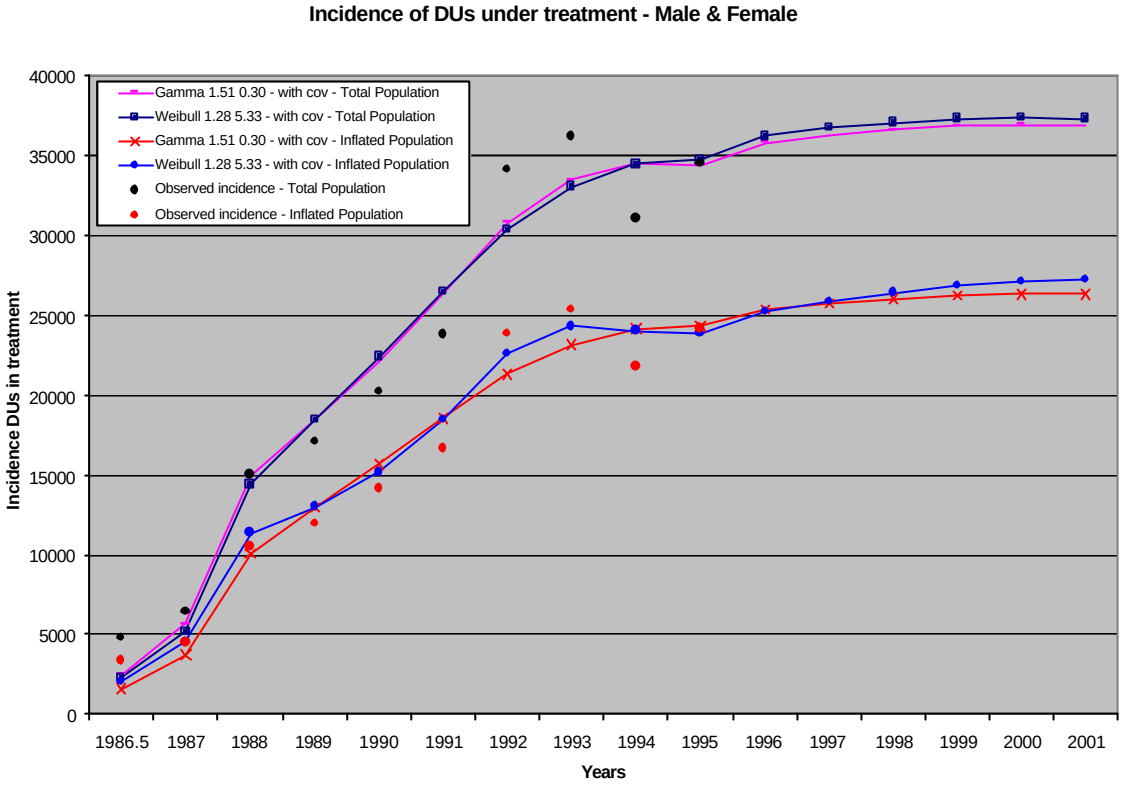


Figure 2.8. Sensitivity analysis for double counting: Incidence of DUs presenting to treatment estimated through the EB-BC without age covariate.

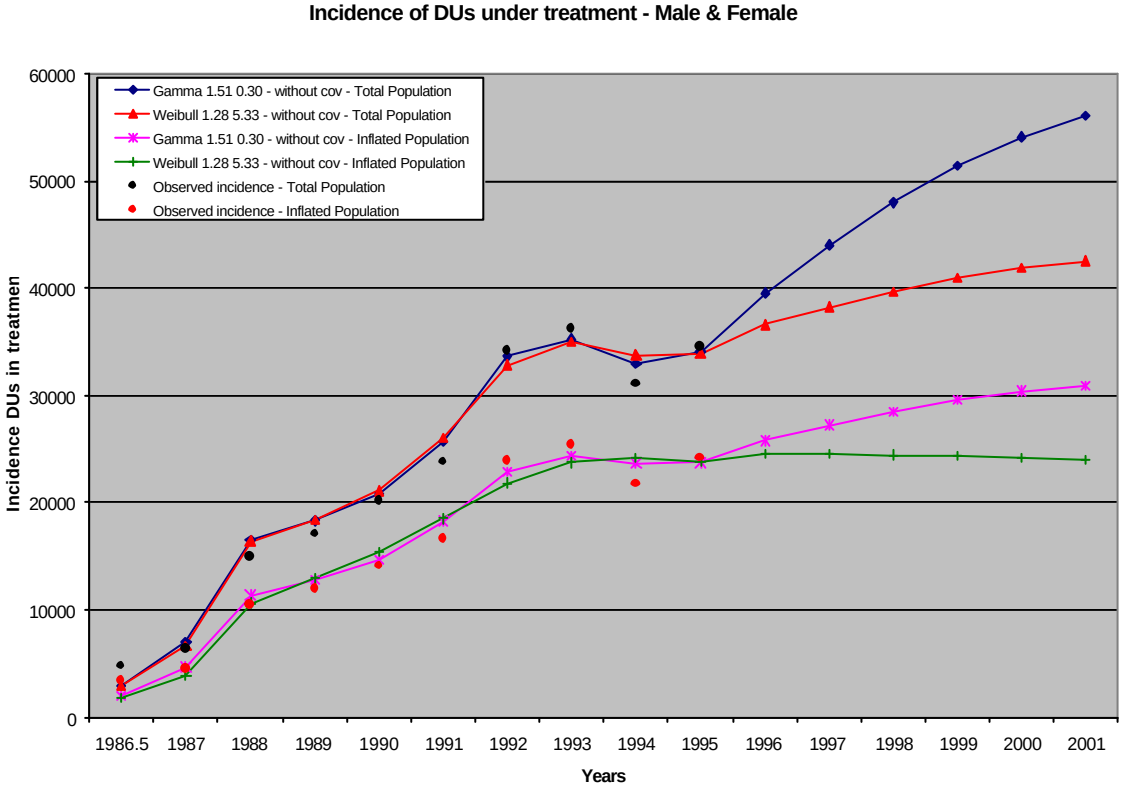


Figure 2.9. Sensitivity analysis for epidemic starting year: Incidence of DUs estimated through the EB-BC with age covariate.

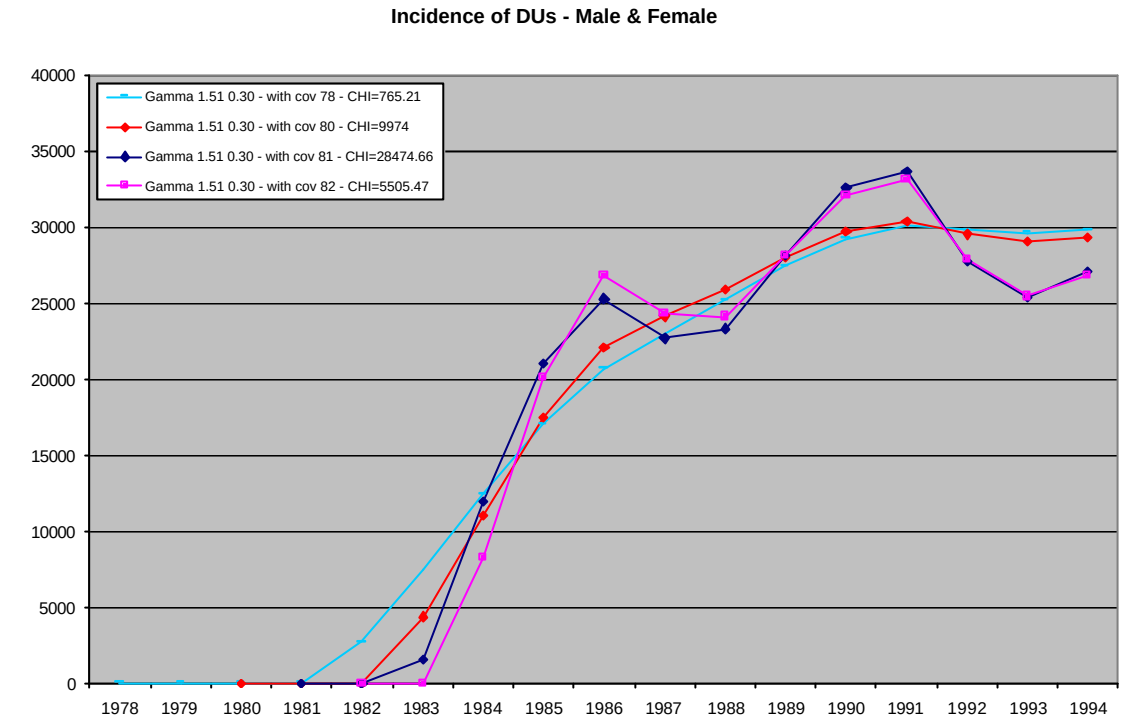
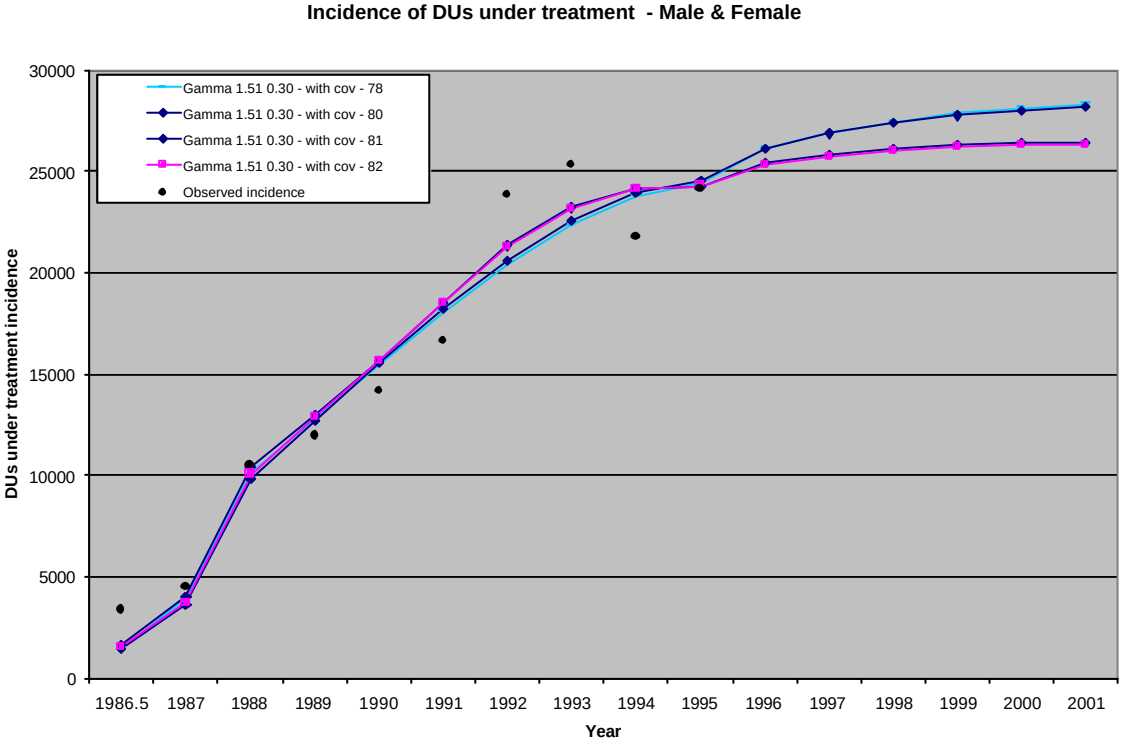


Figure 2.10. Sensitivity analysis for epidemic starting year: Incidence of DUs presenting to treatment estimated through the EB-BC with age covariate.



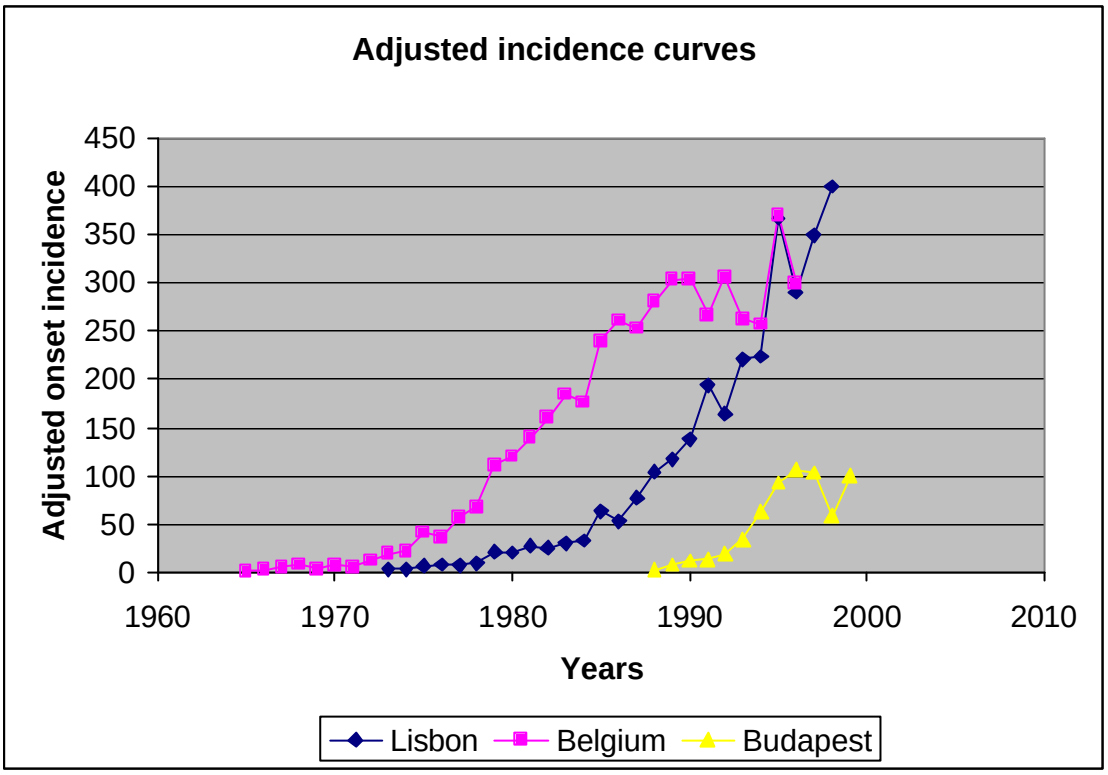
The Back-calculation procedure has been generalised by Siem Hesterkamp to incorporate the level of the epidemic at start (left truncation of therapy data) and it is available with new user friendly interfaces. The new procedure has been tested by Lucilla Ravà and then applied to therapy incidence data provided by the participants from Amsterdam, Italy and London. Sensitivity analyses have been performed and most results are included in a paper on incidence estimation. Some results have been presented at the Harm Reduction Conference in Jersey (2000) and at the workshop on "Dynamic Drug Policy: Understanding and Controlling Drug Epidemics", May 22-24, 2000, Vienna International Centre (UNDCP) and will appear in the proceedings (UN Bulletin on Narcotics, 2001).
Various other methods have been taken into consideration for the incidence estimation. A short overview of the main features and applicability of the various methods is reported in the following table.
The possibility of applying the various methods has been widely discussed with the participants during small ad hoc meetings. In particular, the local group in UK applied the RDA method obtaining some results presented in a published paper. Recently some preliminary results (Figure 2) obtained by using the RDA method applied to data from Lisbon, Belgium and Budapest allow to make comparisons of interest from an epidemiological point of view. In particular the different initial point for the heroin epidemic can be appreciated and the late starting point for Budapest is a possible explanation of the shorter latency periods observed for that site with respect to the others.

Methods for incidence estimation (Overview).

Characteristics	Back-Calculation (BC)	Mover-Stayer models (MS)	Reporting Delay Adjustment (RDA) (*)	Snapshot Models (SM)
Class of the method	Non-parametric deconvolution	Parametric Semi-Markov model (Markov approximation)	Statistic calibration (specific)	Statistic calibration (general)
Main uses	Incidence curve estimation (inaccurate for recent times)	Incidence curve estimation, prevalence estimation, scenario analyses	Estimation of recent incidence (possibly biased depending on the truncation of the data used to estimate the conditional lag distribution)	Estimation of recent incidence by using various sources of data. The specific model is adapted to data available
Data needed	Latency distribution esternally estimated + therapy incidence curve	This is a class of models adapted to the various situation	Latency distribution esternally estimated + onset incidence curve	This is a class of models adapted to the various situation
Availability of a general procedure	Yes	Yes	Easy to get	The procedure depends on the specific model
Possibility to be used by policy makers with routine data	Yes	Yes	No, because, usually, onset incidence data is not available to policy makers	The possibility depends on the specific model
Procedure user friendly	yes	yes	Not yet	Not yet

(*) This method is a particular snapshot model.

Figure 2



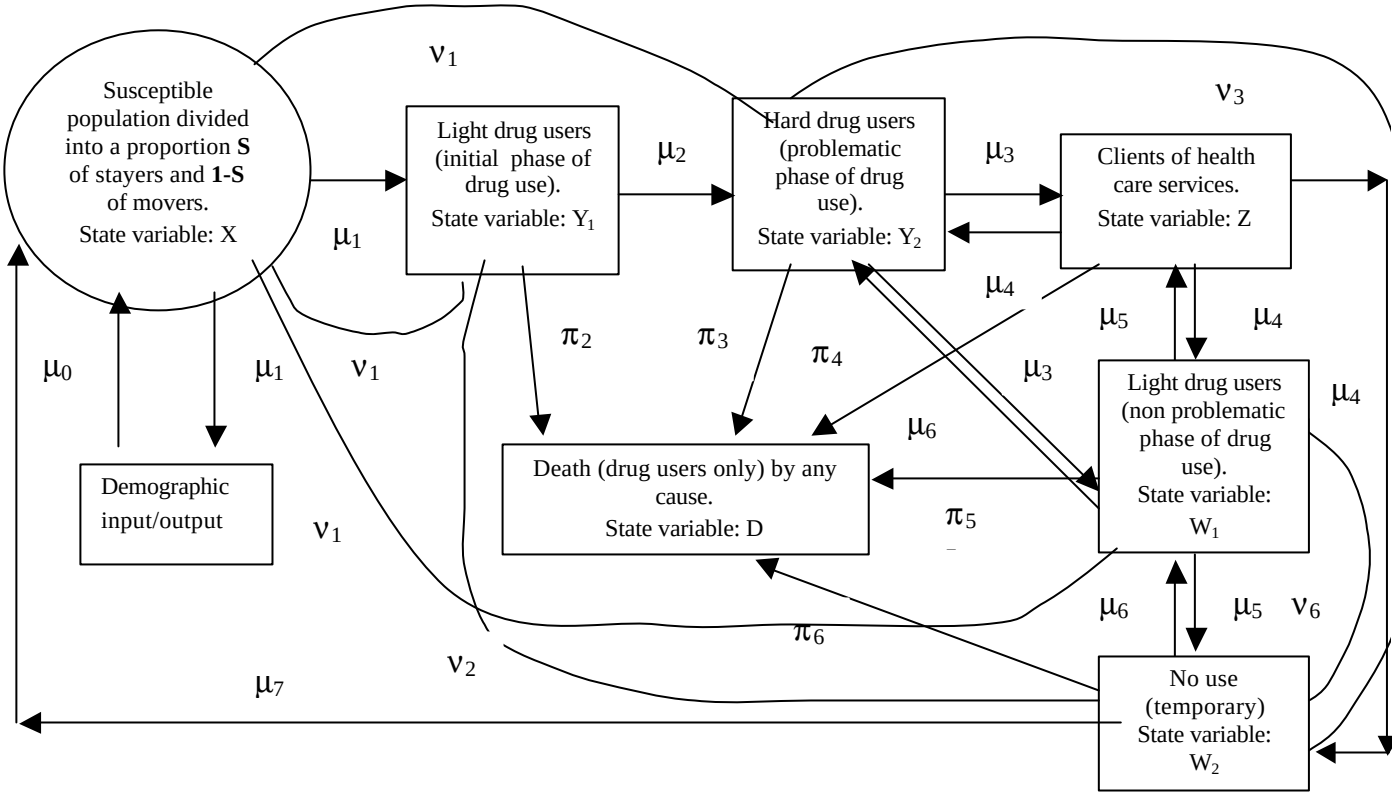
The BC procedure was applied at local level obtaining interesting information on the time-space trends of heroin epidemic. The results were compared with those obtained in the previous 12 month period and presented at the Harm Reduction Conference in New Delhi (2001). The results are also reported in a paper already submitted and in a paper in preparation to be submitted to an epidemiological journal. Summary results are reported in the draft guidelines presented in the following. Also RDA method is presented in the guidelines.

2.5 The System Dynamic Model

Dynamical modelling of epidemic processes occurs through the usual representation of the dynamics involved by a compartmental system, which finds its mathematical formulation in a system of stochastic or deterministic differential equations; this is the case for both the operational and the transmission models. The main difference between the two kind of models is in the fact that transmission models take into account the dynamic processes at micro level, modelling the interactions between individuals belongings to the different sub-groups involved in the epidemic, whereas operational models work on macro-variables or indicators suitable to be used to estimate the size of the phenomena or monitoring the impact of various interventions. Many models of the two types have been developed to study the HIV/AIDS epidemic and are suitable to be used, with some changes, to model the problem drug use epidemic as well.

2.5.1 The operational model for problem drug use epidemic.

The graph describes the main features of the proposed model. The model is a Mover-Stayer type model and allows for non randomness among the susceptibles. Such kind of models considers the susceptible population as subdivided into two groups: the group of stayers that is the group of individuals who are considered not at risk of „infection" (these model are suitable to make scenario analyses to study the impact of various proportions of vaccinated or immune persons to estinguish a given epidemic) and the group of movers who are at risk. Due to the interactions between infected individuals (for our problem we can imagine these are the problem drug users who are also drug dealers) and the susceptibles and to the pressure of the black market, some of these may pass to the drug user compartment and begin a „drug user career". Similarly to the model proposed by Behrens et al., the present model comprises two different stages of drug use. The first (light use) stage is the initial stage of drug use, then light drug users can either stop using drug or pass to hard drug use (or death). The other arrows in the graph completely describe the other possible steps of a drug user career.



To write the corresponding equations (either deterministic or stochastic) some hypotheses have to be set up and the known and unknown parameters have to be chosen. A first approximation may be through a Markov model, possibly a marked Markov process. In such a case the length of stay in each compartement is assumed to be exponentially distributed and the results are useful to get a first qualitative insight in the epidemic process. A more realistic approximation is by using semi-markov processes. In such a case the length of stay in each compartment may be assumed to be distributed differently. The study is more complex but suitable mathematical techniques can be used. For what concern the parameters and distributions of length of stay, some are already available from the study of the latency period. Some can be derived by therapy data already available in some sites. The length distribution of the initial phase can be estimated by using the information from the data set available for some sites. The demographic parameters regulating the dynamics

of the susceptible population are supposed to be known and are country specific. The development of the model and the simulation and scenario analyses are planned for the second period.

The simulation procedure related to the system dynamic model has been completed and is available with user friendly interfaces. Some results have been recently presented at the workshop on "Dynamic Drug Policy: Understanding and Controlling Drug Epidemics", May 22-24, 2000, Vienna International Centre (UNDCP) and will appear in the proceedings (UN Bulletin on Narcotics, 2001). Such paper is reported as appendix 4.

The qualitative analysis of the system dynamic model was completed and permitted to obtain information about the optimal scheduling of control interventions.
In particular, it could be shown that the impact of a primary prevention intervention, with efficacy parameter Δ is bilinear with respect to such parameter and to the proportion of movers among susceptibles. This implies, due to the results of the qualitative analysis showing that the proportion of movers is monotonically decreasing during the epidemic phase, that the effect of a primary prevention intervention is higher at the beginning of the epidemic. It also implies that the effect of the observation of the adverse consequences of drug abuse cannot be by itself highly effective as primary prevention, due to the long latency time. This fact, unfortunately, prevents from observing such consequences for several years since the beginning of the epidemic. When starting observing them, most movers will already be drug users.
Similarly, the effect of law enforcement interventions can be evaluated. It can be immediately derived that the impact of such interventions is more effective during a mature phase of the epidemic when the level of Stayers is high and the sizes of the three drug use compartments is high as well.
Several new scenario analyses (what if) were also performed using the simulation procedure written in S-plus. The results are reported in a paper already submitted. The results are completely general and can be applied to any country.

2.6 The Structural Equation Model

Drug policy implementation needs to be evaluated in order to set up efficient strategies to control the various drug related phenomena that have a large impact at population level. One of the main issues in assessing the costs and evaluate the benefits of various kinds of interventions in the illicit use of drugs field is to attempt to conceptualize and assess the efforts required for effective prevention in comparison to the cost borne by society in their absence. Due to the intrinsic nature of the subject matter it is difficult to evaluate preventive efforts. Data regarding the cost of programmes with specific illicit use prevention objectives can be collected. However, preventive efforts cannot be seen in isolation from other social policy fields. It has long been established that no prevention campaign or activity per se can be effective in the absence of a conductive environment and/or policy regarding the family, education, employment situation, positive role models and positive alternatives. In this sense an income generating project could be seen as a prevention tool. As a prevention tool is any programme aimed to increasing population education level in terms of a higher proportion of highly educated people and/or a lower proportion of ineducated ones.
Path analysis and Structural Equation Models (SEM) have been applied to study several complex situations in different fields. A Structural Equation Model has been developed to study the links between socio-economic indicators and problem drug use indicators in order to evaluate the impact of modifications of the first ones on the latter ones (Appendix 5). A preliminary application of the model to Italian regional data allows to evaluate the effect of a reduction, for example, of the school drop out rate or of the unemployment rate on the drug

related phenomena. The results obtained applying the model in different years (from 1991 to 1996) show the existence of specific correlation and “causality” among the indicators of interest which can be measured and remains stable over time.
The methods should be applied to new data sets from other sites and countries for better analysing the biases and the potentialities as tools to provide information to policy makers in different situations. In particular the model should be applied using data with a lower level of aggregation with respect to the regional data used in the pilot application.

In order to better analyse the potentiality of the method, a new structural equation model (SEM) was developed to evaluate prevention interventions and some analyses were performed on the basis of a large data set coming from 35,000 interviews to military conscripts in Italy. Such analyses permitted to identify 4 different causal models for 4 different primary use substances. Work is in progress to write a paper summarising the results. On the basis of the preliminary analysis a new survey among military conscript in Italy has been designed in order to analyse time trends. The study will be developed in the next three years. The methods should be applied to new data sets from other sites and countries for better analysing the biases and the potentialities as tools to provide information to policy makers in different situations.

2.7 Preliminary draft guidelines for incidence estimation

The development of the various methodologies and models for incidence estimation within the present project has provided very interesting and useful results valuable for policy makers. In order to disseminate the results and extend the experimentation to other sites a preliminary document presenting guidelines has been written and is reported in the following.

EMCDDA Recommended Draft Technical Tools and Guidelines

Key Epidemiological Indicator 'Prevalence and Patterns of problem drug use':

Draft Guidelines for Estimating Incidence

EMCDDA/ 2001

The following have contributed to the development of these draft guidelines:

EMCDDA

Lucas Wiessing, Richard Hartnoll

University of Tor Vergata, Rome

Massimiliano Bultrini, Maria Grazia Calvani, Lucilla Ravà, Laura Re, Carla Rossi

Participants in current and previous EMCDDA projects on trends/incidence of problem drug use

Belgium: Francis Sartor, Denise Walckiers

Hungary: Katalin Veress

Ireland: Catherine Comiskey

Italy: Anna Maria Bargagli, Katiuscia Berretta, Marina Davoli, Giovanna Jona-Lasinio, Fabio Mariani, Alessandra Nardi

The Netherlands: Erik van Ameijden, Marcel Buster, Simon Heisterkamp

Portugal: Jorge Ribeiro

United Kingdom: Daniela De Angelis, Matthew Hickman, Seamus Seaman

Status of these Guidelines

These guidelines are based on a series of seminars and projects on incidence estimation, organised by the EMCDDA and co-ordinated by the University of Rome Tor Vergata, Department of Mathematics. Two suitable estimation methods emerged that have subsequently been tested in the participating countries. First results have been reported in the 1999 and 2000 EMCDDA Annual Reports. Additional, supporting documents and publications on the estimation of incidence, patterns and careers of problem drug use are listed below. These draft guidelines form part of the EMCDDA key indicator 'Prevalence and Patterns of Problem Drug Use', and are additional to, and partly based on, previously developed guidelines on prevalence estimation (EMCDDA 2000 b).

Supporting documents

- European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Study on incidence of problem drug use and latency time to treatment in the European Union. (CT. 99.EP.05) Lisbon: EMCDDA, December 2000 a.
- European Monitoring Centre for Drugs and Drug Addiction. Pilot Project to Estimate Time Trends and Incidence of Problem Drug Use in The European Union. (CT.98.EP.07) Lisbon: EMCDDA, March 1999.
- Heisterkamp S.H., Downs A.M., van Houweling J.C., Empirical Bayesian Estimators for Reconstruction of HIV Incidence and Prevalence and Forecasting of AIDS. I. Method of Estimation. In *Quantitative Analysis of HIV/AIDS: Development of Methods to Support Policy Making for Infectious Disease Control.* Ph.D thesis: University of Leiden, 65-98, 1995.
- Heisterkamp S.H., van Houwelingen J.C., Downs A.M., "Empirical Bayesian Estimators for a Poisson Process propagated in time", *Biometrical Journal*, 41-4, 358-400, 1999.
- Hickman M., Seaman S., De Angelis D., Estimating the relative incidence of heroin use: application of a method to adjust observed reports of presentations at specialist treatment agencies, *American Journal of Epidemiology*, in press.
- Pasqualucci C., Ravà L., Rossi C. and Schinaia G. (1998), "Estimating the size of the HIV/AIDS epidemic: complementary use of empirical bayesian back calculation and the mover-stayer model for gathering the largest amount of information", *SIMULATION*, 71-4, 213-227;

- Ravà L., Calvani M.G., Heisterkamp S., Wiessing L., Rossi C. "Incidence indicators for policy making: models, estimation and implications", UN Bulletin on Narcotics, 2001, in press.
- Rossi C., Monitoring drug control strategies: hidden phenomena, observable events, observable times, *International Journal of Drug Policy*, 10-1, 1999, 131-144.
- Rossi C. "A Mover-Stayer type model for epidemics of problematic drug use", UN Bulletin on Narcotics, 2001, in press.
- Wiessing LG, Hartnoll R, Rossi C. Epidemiology of drug use at macro level: indicators, models and policy-making. in "Modelling drug use: methods to quantify and understand hidden processes", EMCDDA Scientific Monograph Series, No 6, 2001, 19-32.
- European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). EMCDDA Recommended Draft Technical Tools and Guidelines - Key Epidemiological Indicator: Prevalence of problem drug use. Lisbon: EMCDDA, December 2000 b.

European Monitoring Centre for Drugs and Drug Addiction
Rua da Cruz de Santa Apolónia 23–25
PT-1149-045 Lisboa
Portugal
Tel: +351 21 811 30 00
Fax: +351 21 813 17 11
e-mail: info@emcdda.org
http://www.emcdda.org

Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
BC	Back Calculation method
DUs	**Drug users**
EB-BC	Empirical Bayesian Back Calculation method
EMCDDA	European Monitoring Centre for Drugs and Drug Addiction, Lisbon, Portugal
EU	European Union
HIV	Human immunodeficiency virus
IDUs	**Injecting drug users**
LP	Latency period
RDA	Reporting Delay Adjustment
UK	United Kingdom

Summary and Overview

Incidence is an important indicator to follow trends over time in the rate of new cases, or new recruitment, of problem drug use. Although total incidence cannot be estimated, a lower bound of incidence ('relative incidence') can be derived from observed registry data (e.g. drug treatment), which can be very useful for policy making.

Two statistical methods are being proposed to estimate relative incidence from drug treatment data: 1) The back-calculation method (which uses an independent estimate of the latency period from first use to first treatment but cannot always be applied) and 2) the reporting delay adjustment method (which directly corrects for LP but has heavier data demands and some other drawbacks). In addition, 3) it is suggested to obtain a qualitative impression of incidence by using the age distribution or age trends among treated cases (and/or other registries), the proportion or trends in first treatment demands, or trends in other indicators of drug problems.

The definition of the target group is closely linked to the availability of data. With data bases identifying:

a) the main drug consumed (e.g., polydrug use including opiates, monodrug use of opiates, monodrug use of cocaine),
b) Age at first use,
c) Age at first treatment,
d) Gender,
e) Any other variable to include in the LP model, such as route of administration.

The validity of the estimate is dependent on

a) Definition of target group and quality of data sources

An operational definition of the target group of interest is

a) Injecting drug use (IDU) or long duration/regular use of opiates, cocaine or amphetamines, eventually presenting for treatment.

Back calculation method

a) This method has a clearly defined target group: incidence of problem drug use for those users who will eventually be recorded by some Agency.
b) An appropriate computer programme is available.
c) The method is not advisable for countries with a short time series for therapy incidence (aggregated data) or other similar incidence, nor for small areas.

Brookmeyer and Liao RDA method

a) This method has a clearly defined target group: incidence of problem drug use for those users who will eventually be recorded by some Agency.
b) An appropriate computer programme is needed.
c) The method is more data demanding and sensitive to bias from missing data on year (age) of first use.
d) The method is not advisable for large areas.

The two methods can be used as alternative depending on the data availability and quality. The first one is more suitable for regional or national incidence estimation, whereas the second is more suitable for local (small areas) incidence estimation. Both require observed incidence (either therapy incidence (BC), or onset incidence (RDA)) for several points in time.

Introduction

Modelling the incidence of first drug use is important for understanding the diffusion of drug use in space and time. Incidence data can be used to evaluate current and future needs for, and effects of, services and interventions. In particular, incidence figures may provide an indication of whether the number of problem drug users is growing (epidemic phase), falling or stable (endemic phase).

Incidence is defined as the number of new cases occurring within a given time period, usually a year. Compared to prevalence, which is usually defined as the total number of existing cases in a given year, incidence gives more direct information on the recruitment of new cases. It is therefore a more sensitive indicator of prevention efforts than prevalence. Incidence is directly related to prevalence: adding up the cumulative incidence over a number of years, and subtracting the number of cases that died or ceased to be cases for other reasons, gives prevalence. A general and simple formula is that: prevalence is incidence times duration of disease. Incidence data can thus also serve to estimate prevalence.

It is acknowledged that accurate information on the incidence of drug use and especially that of heroin and other opiate use is difficult to obtain. Evidence from national surveys and other sources indicate that the prevalence of heroin use in the general population is relatively low and has shown a decreasing trend in most countries, although some new increases are reported. Despite this low prevalence, most of the widespread drug-related health and social problems in EU countries are caused by the use of heroin and other opiates. The substances causing health and social problems as well as the route of administration of these substances vary across Europe. In Sweden, for example, the drugs causing most of the problems are amphetamines, although like in Norway and Finland a new heroin wave is challenging the drug help system. On the other hand, the most common practice of using heroin in the UK is smoking ("chasing the dragon") whereas in Germany, France and Italy injecting drug use is the common way of administration.

These differences in substances and routes of administration make a common definition of the target group rather difficult, since different substances and practices of use are related to different health problems. Nevertheless, comparisons across countries with regard to the trend of drug use call for a common definition of the target group, the use of equivalent data sources, and the application of the same methodology. In an attempt to find a common definition in spite of the differences between the EU countries we use the term *problem drug use*, which includes all different forms of problems due to the use of opiates, cocaine, and amphetamines irrespective of the route of administration.

These guidelines describe two relatively sophisticated statistical methods to estimate incidence from observed cases. Until now these methods have been used to estimate incidence of first heroin use from drug treatment data only, however other sources could as well be used if data quality were sufficient. One major drawback of the methods described, is that they cannot estimate total incidence of heroin use from the observed cases in treatment. The methods can only estimate the trends in ‘relative incidence’, i.e. the incidence of cases that will eventually (up to many years later) show up in the registry (e.g. drug treatment). This means that they underestimate, or give a lower bound for, total incidence. However, they can indicate whether the number of new cases is rising, stable or falling between years. Under the assumption that relative incidence correctly indicates the trends in total incidence (‘the form of the curve is the same’), relative incidence is a very important indicator of the rate of new cases.

In addition to the two sophisticated statistical methods described, for which a certain level of statistical expertise and data quality/availability are needed, one might at least obtain some global impression of incidence (mainly is it increasing or not), by looking at very simple

parameters of observed cases e.g. in treatment. These include: a) comparing the age distribution of cases on a certain moment in time with the age distribution of non-cases, or following trends in the age distribution of cases over time, b) comparing the proportion of first registrations (e.g. first treatment demands) among all registrations, or following trends of first registrations over time. Knowing that the average duration of problem drug use is at least more than five and possibly around 8 or even more years, one can easily deduct a rising epidemic of heroin use when these parameters would show very marked results. E.g. if treatment data for example contains 50% first treatment demands, or when half of the treatment cases are under age 25. When results are not very marked however it is much better to correctly estimate the curve of relative incidence.

Target Group

The definition of the population targeted at by any incidence estimation in the drug field is one of the most difficult tasks. The identification of substance users can only be derived from the known population, i.e., only when an individual comes into contact with the legal, medical or social system do we know that he or she is a user. Any definition of problematic drug use should therefore consider these three perspectives (legal, medical or social) with their different interests, norms and values. In practice, however, the simultaneous consideration of these three perspectives is often not possible, may be e.g. due to the non-availability of data bases or lacking links between data bases. Thus, the researcher is very often left with a pragmatic definition of the target group.

Demands on a Definition

Definitions of target group may combine a certain time period (e.g., a certain year), a specific substance group (e.g., opioids, amphetamines), the route of administration (e.g., injecting, smoking), frequency of use (e.g., experimental, occasional, habitual, regular, long duration), legal status (illicit, licit), and clinical diagnoses (dependence, abuse). As the utilised databases, e.g. treatment monitoring systems, usually report data for a calendar year it is natural to use this time frame also for the incidence estimation of problem drug use. Even when referring to the broadest possible target group, the „drug users“, any definition should include

- a time period
- an age group
- and a definition of substances.

Pragmatic Definition

Not only does substance use vary between countries, birth cohorts, and even gender, also route of administration differs greatly between substances, countries and cohorts. Even if route of administration and frequency of use could clearly be related to a more or less hazardous consumption pattern, this information may not be directly available. Furthermore, substances are used in quite mixed, often chaotic patterns. Only very few opiate users do not use other drugs as well. On the other hand opiates, especially heroin, are the drugs which cause most of the problems. If the pattern of drug use has to be labelled and categorised in a simple way, it can be done on the basis of the drug which causes the highest risk. Complexity can be further reduced by omitting the notion of primary and secondary drugs, and not accounting for polydrug use.

A pragmatic definition for coding according to this concept can be summarised as follows:

- If a person uses heroin or other opiates he or she is always classified as opiate user regardless whether other drugs are taken as well.
- If no opiates are used then the person is a non-opiate user. He or she can then be classified as cocaine user (disregarding other drugs) or, if no cocaine is used, as amphetamine user.

Although not all groups of problem drug users are covered by this definition (e.g. problematic users of cannabis), they are not included in the target group as the estimation methods described in these guidelines are not appropriate for these groups.
For example, therapy data used as a basis for BC mostly relates to opiate users.
This logic of categorising patterns of drug use is summarised in table 1 below.

Table 1: **Groups of problem drug users**

	Groups	**Opiates**	**Cocaine**	**Ampheta-mines**
Opiate users	Problem Opiate user	Yes	Yes/No	Yes/No
Non-opiate users	Problem Cocaine user	No	Yes	Yes/No
	Problem Amphetamine user	No	No	Yes

In summary, we arrived at the following:

An operational definition of the target group of interest would be

- Injecting drug use (IDU) or long duration/regular use of opiates, cocaine or amphetamines, eventually requiring treatment.

Data Sources

Two possible data sources can be considered, in theory, as basis for incidence estimates. Even though the applications are, at present, only based on data coming from the first, both are briefly described in the following of this Section.

Clinical, Medical, and Social System

The clinical, medical and social systems collect the most detailed information on drug users. A drug user can come into contact with:

- Drug treatment agencies (inpatient versus outpatient, specialised on drug care versus general treatment agencies; e.g., drug counselling centres, general counselling centres, psychiatric hospitals, and specialised hospitals),
- Low threshold agencies (e.g., needle and syringe-exchange schemes, drop-in centres),
- Substitution services (dependent on the regulation in each country these may be general practitioners, substitution ambulances, hospitals, treatment agencies),
- General practitioners (medical reasons),
- Emergency ambulances (mobile or stationary),
- HIV/hepatitis related services,
- Clinical psychologists,
- Psychiatrists.

Problems associated with these data sources are:

- In general, treatment monitoring systems do not cover all treatment facilities of a country. Besides, the treatment facilities represented in the treatment monitoring system may differ from for one country to another or within the same country in different areas.
- Agencies usually collect their data for themselves, and in rather few countries a general treatment monitoring system covering most of the treatment centres has been established. Furthermore, double counting cannot be excluded, as many drug users will come into contact with a variety of treatment facilities. Due to privacy laws utilising unique personal identifiers to prevent double counting is impossible in some European countries.
- Drug users in urban areas have more options for receiving treatment in comparison to rural areas. On the other hand, the 'latency period' (time from first opiate use to first treatment) seems to be shorter in less urbanised areas which might be related to stronger social cohesion in rural areas. Therefore, the probability of coming into contact with treatment facilities may be nationwide not constant.
- Some treatment agencies are interrelated and send patients to certain other centres.
- The capacity of facilities is limited, drug users may be set on waiting lists. They might break off contact again, before any data could be collected, or they may leave incomplete, unreliable or even wrong data. There exists a population of drug users not covered by the drug-related social and medical system.
- It is very likely, that every drug user comes into contact with a general practitioner, but this rather seldom utilised data base has manifold problems. It is not clear, how many drug users will be recognised while visiting a doctor for medical reasons not obviously related to drug-use. Furthermore, in case of a non-fatal accident under influence of psychotropic substances it is almost impossible to distinguish between regular, occasional or experimental users.

Legal System

Data of drug users can be found in registers

- on convictions because of offences against laws on consumption, possession or supply of illegal drugs,
- on convictions because of secondary crimes, i.e. offences associated with the obtaining of drugs (theft, shoplifting, prostitution, forgery),

- on convictions because of offences under the influence of psychotropic substances (e.g. driving, violence),
- on detainees in connection with the above mentioned categories,
- on mortality (e.g., all-cause deaths of registered drug users, drug-related deaths).

Problems associated with these data sources are:

- It is difficult to distinguish between consumers and dealers of drugs. Frequency of drug use cannot easily be obtained, in order to separate regular from experimental users. Police recordings do usually not distinguish between minor experimentation with drugs, severe drug problems, mere trade without consumption, and long or regular users.
- Secondary crimes may not be registered at all, and dependent on the drug policy of a country no special attention may be paid on detecting them as connected to drug use.
- Age at first use is not always available.

Data requirements for single methods are considered in the specific chapters.

Methods

Qualitative analysis of age, first treatment, other indicators

Statistical Background
Important qualitative information on drug policy implementation useful to forecast future service needs can be obtained by analysing time trends of suitable epidemiological indicators based on secondary data. Secondary data can be defined as existing statistical and documentary information that is routinely collected and readily available, such as treatment presentations, drug seizures, infectious diseases indicators, or drug related deaths. Standard statistical summaries (statistical distributions, time series graphs and tables…) can then be utilized to properly present the available information on time trends. If the population of addicts asking for treatment (or addicts inprisoned) is ageing this may be assumed as an evidence of a decreasing incidence of problem drug use. An example of this kind of analysis based on observable data can better clarify the approach.

Example: Description of time trends of opiate users in treatment and police stations in Amsterdam.
Data of the methadone register show a decreasing number of clients and an ageing treatment population. These time-trends are not limited to the treatment population. Data of methadone prescription to arrested opiate users at police stations show similar trends.
Figure 1 shows that the annual number of treatment participants of the methadone programmes in Amsterdam is decreasing from ± 4900 at 1985 to ± 3100 at 1997. The number of opiate users who are actually in treatment does not change. Weekly, ± 2000 opiate users participate in methadone treatment during. This figure is stable during the whole period.

Figure 1: *number of opiate users in methadone treatment.*

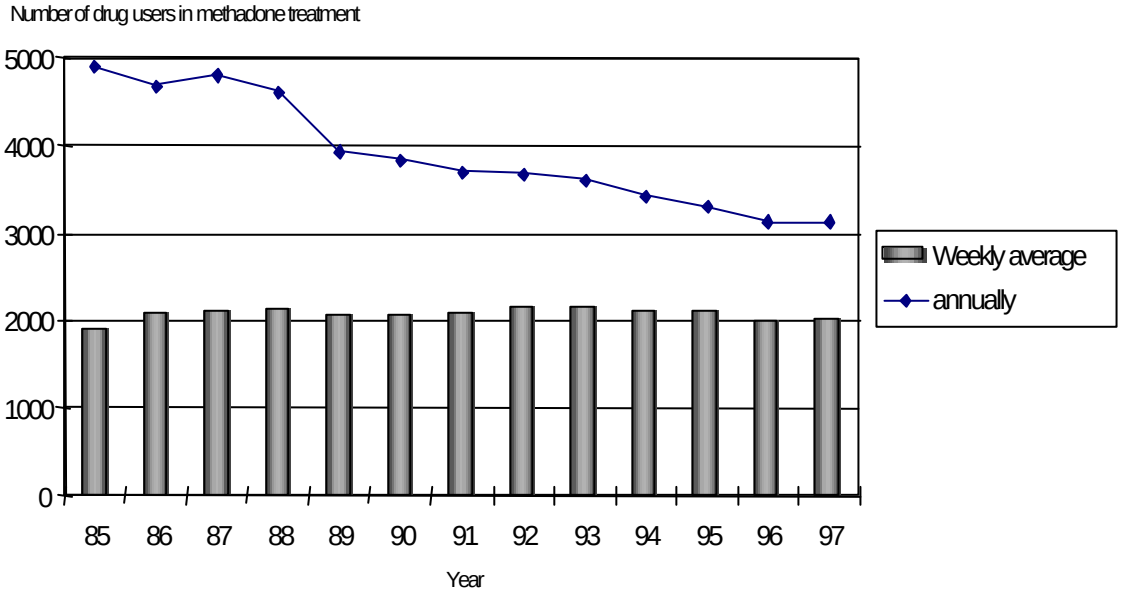
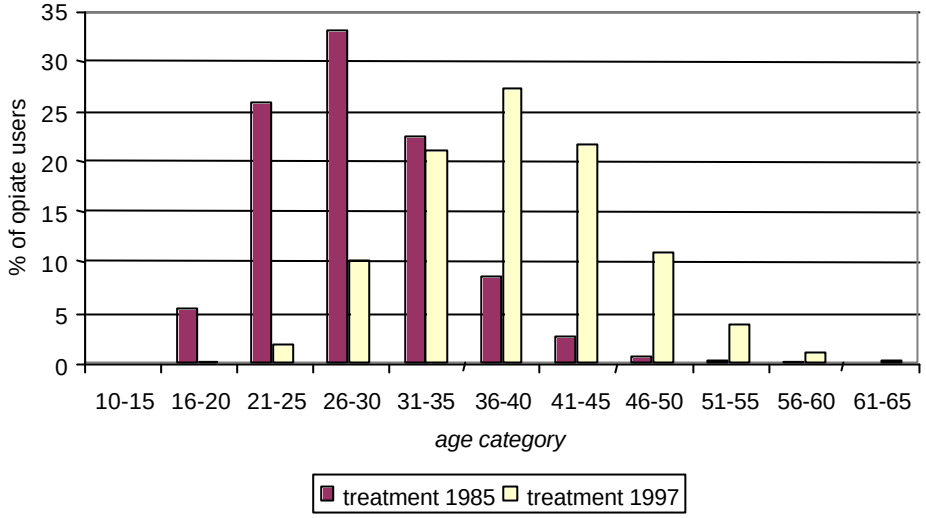


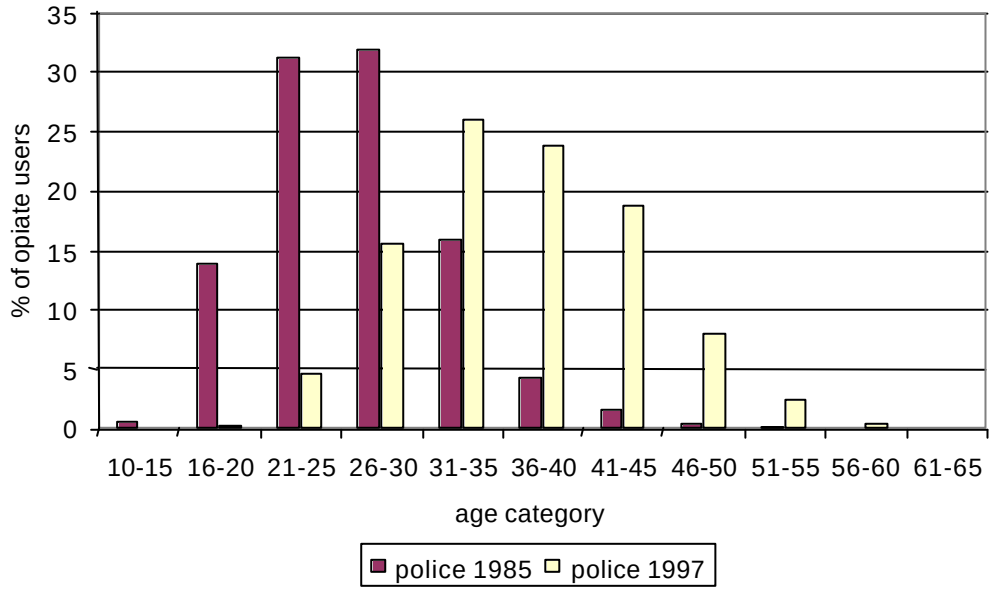
Figure 2 shows the age distribution of the treatment population at 1985 and 1997. It shows a clear ageing population. At 1985 the majority (65%) was thirty years or younger, at 1997 only 12.5% belongs to this younger age categories.

***Figure 2**: age distribution of participants of methadone treatment*



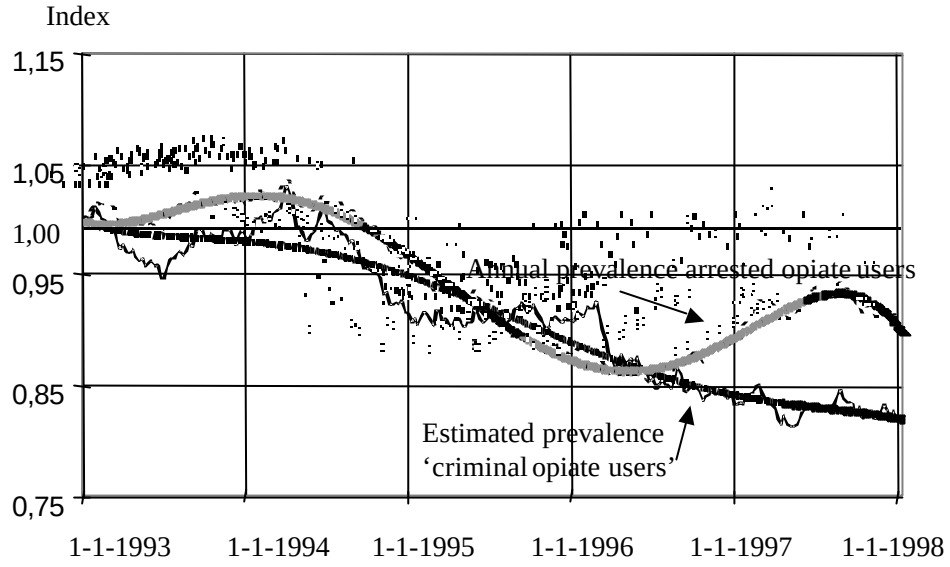
The decreasing number of opiate users in treatment and the ageing population could indicate that the incidence of opiate use is decreasing but could also indicate that younger drug users avoid methadone programmes. Therefore data of methadone prescriptions at police stations are studied. Methadone is prescribed at the police stations to prevent an opiate abstinence syndrome. These data are registered at the Central Methadone Register as well. Both at 1985 and 1997 the arrested opiate users are younger than those participating in treatment. However, a similar trend of ageing can be observed.

***Figure 3:** Age distribution of arrested opiate users*



Estimating the prevalence of criminal drug users is more difficult because the police activities vary over time. Figure 4 shows the index of the annual prevalence of arrested opiate users over time. A gradual decrease of the prevalence of 'criminal' opiate users is observed when these figures are adjusted for fluctuations for the chance of being arrested. The variation of the number of arrests per arrested drug user per year used as an indicator of chance to be arrested.

Figure 4: *development of the number of arrested opiate users and 'criminal' opiate users*



Back Calculation Method

Statistical Background

The back calculation method (BC) was developed to describe the dynamics of the AIDS epidemic. Based on AIDS incidence data and on medical knowledge as well as on statistical assumptions the HIV prevalence and incidence in the past years is estimated. The results allow for projections in the (near) future.

The number of new AIDS infection cases in a given year is the sum of those infected the year before with an incubation period of one year and those infected two years before with an incubation period of two years and those infected three years before with an incubation period of three years and so on. As AIDS cases are registered by the health authorities the number of new AIDS infection cases in a given year is known. Combining medical knowledge and appropriate statistical assumptions leads to an estimate of the incubation period distribution which in turn allows the back-calculation of the number of HIV infected in the preceding years. The estimated number of HIV infected in the preceding years and the estimate of the incubation period distribution are then utilised to estimate the future AIDS cases. The different back calculation methods differ in the numerical procedure of back calculation, in the assumed shape of the HIV infection curve as well as in the assumptions on the incubation period distribution.

The particular BC method adopted in the context of incidence estimation of problematic drug use is the Empirical Bayesian Back-Calculation (EB-BC), which was proposed by Heisterkamp et coll. (1995, 1999). The curve of the incidence of problem drug use is represented by a step function, a Poisson process is postulated for the occurrence of the new cases in a single time interval, and the incidence of DUs in treatment in each interval is assumed to be independently Poisson distributed.

Application (data from Clinical, Medical and Social System)

In order to apply the EB-BC the following quantities must be defined:

- *I(s)*: the onset incidence of DUs (who present to treatment at least once) at time *s, s = 1,…,S*. These individuals pass through a period of hidden drug use before they become visible by having their first contact with some health care service;
- *F(v-s)*: the cumulative distribution of the period between the time *s* of the first use of drug, and the time *v* of the first presentation for treatment, the "Latency period (LP) distribution";
- $I_{treat}(v)$*:* the incidence of DUs enrolled in treatment at time ν.

The goal of the method is, therefore, to estimate the incidence of DUs eventually seeking for treatment *I(s,)* using data on the incidence of DUs in treatment $I_{treat}(v)$ and an estimate of the latency period distribution *F(v-s),* externally obtained by using standard methods and models for the analysis of duration or survival data (Collet, 1994; Marubini & Valsecchi, 1995). The incidence *I(s)* can be estimated through the EB-BC, by deconvolving the following equation:

$$I_{treat}(v) = \int_0^v I(s)\, d(F(v-s))$$

The therapy incidence data $I_{treat}(v)$ and the data needed to estimate *F(v-s)* should be provided by the health care services offering treatment (of any kind) to the drug users. In particular the present version of the EB-BC uses aggregated biannual incidence data of "new" individuals under treatment in some health care services (at their first treatment), and classified according to: date of first registration to the health care service, and any other variables that could be used as covariate or stratification variables such as gender, age, geographic area, educational level and type of the health care service (EMCDDA,1999). Moreover the method, allows the use of various models for the latency period distributions: Markov model (allowing for forward and backward jumping to stages), Gamma, Weibull, Log-Normal.
The software for the EB-BC is written in S-plus language.

In summary the application follows the 2 steps:

Step 1: Estimate the latency period distribution *F(n-s)* and find a suitable parametric model to model it, possibly depending on covarites such as age, sex…by standard methods used in Survival Analysis.

Step 2: Apply the EB-BC procedure to aggregated therapy incidence $I_{treat}(v)$ using the LP distribution estimated in Step 1.

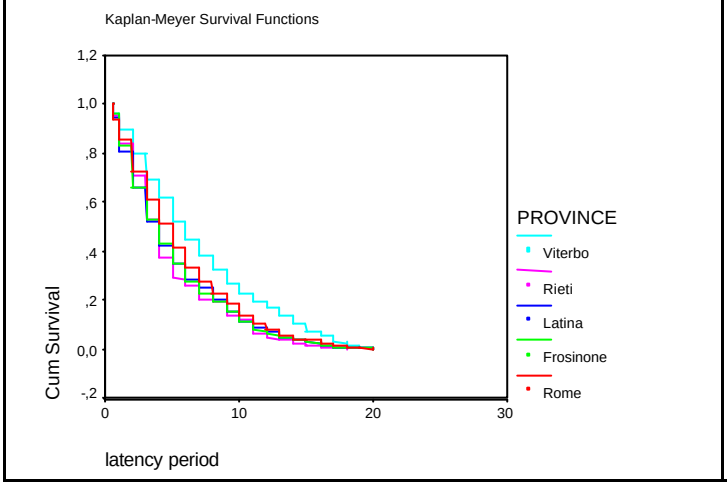
Example: Estimating incidence of heroin use in Italy
Heroin by injecting caused, during the '80s and '90s, the majority of consequences for both health and criminal justice departments in Italy (EMCDDA, 2000). The number of people between the ages of 15 and 54 who have used heroin at some time in their life is estimated to be not less than 300,000 people (Ravà and Rossi, 1999). Thus, in the following, the application of the EB-BC method is focussed on estimating the onset incidence curve of just heroin use.

Step 1. Latency period analysis
The latency period was estimated, by using various local datasets provided by the different regional health authorities containing individual data on drug users in treatment. From the results obtained, such period appears to be remarkably similarly distributed over the different sites, with a median of between five and six years and an average of between five and seven years. Figure 1 a) reports the survival curves, estimated by Kaplan Meier method (Rossi, 1999a; EMCDDA, 1999), only for the data related to the 5 provinces of the Lazio region. This time-lapse, however, appears to be much longer than this in young drug users suggesting that age at first use is an important covariate to be included in any model (EMCDDA, 1999).

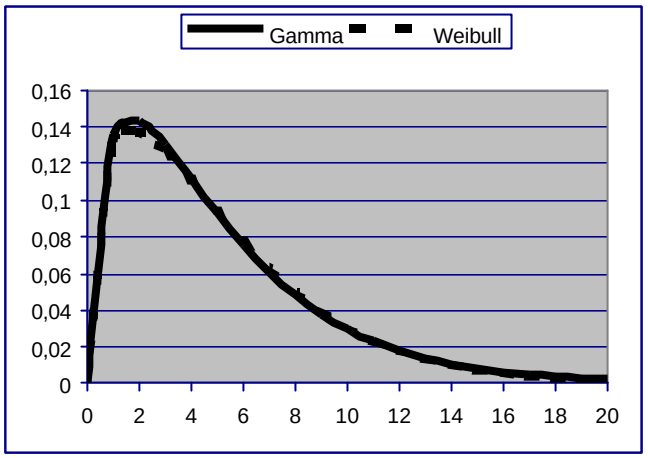
For each dataset, the best parametric estimate of the latency period distribution was obtained by using the P-plot method (quantile best fitting) which is available in the SPSS package used for the present application. The best model resulted, for each site considered, either a Gamma or a Weibull density (Figure 1, c and d) with slightly different parameters (EMCDDA, 1999). On the basis of a sensitivity analysis (EMCDDA, 2000), based also on further secondary data, the most representative estimates at national level were chosen as a Gamma density with parameters 1.51 (shape) and 0.30 (scale) and a Weibull density with parameter 1,28 (shape) and 5.33 (scale) (Figure 1 b). The Gamma, being less dispersed than the Weibull, is always associated with smoother incidence curves as results of the estimation procedure based on EB-BC. The similarities between the different Italian sites and over time suggest that LP is not strongly related to external factors.

Figure 1: *Some results of the estimation of the latency period.*

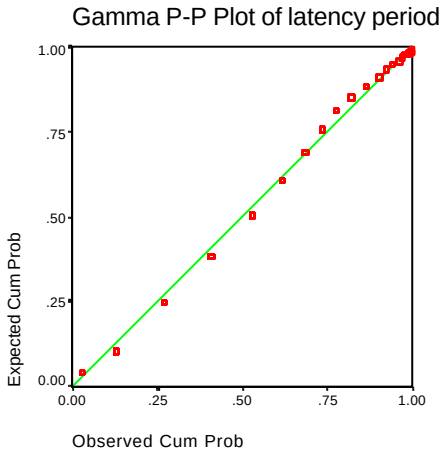
a)



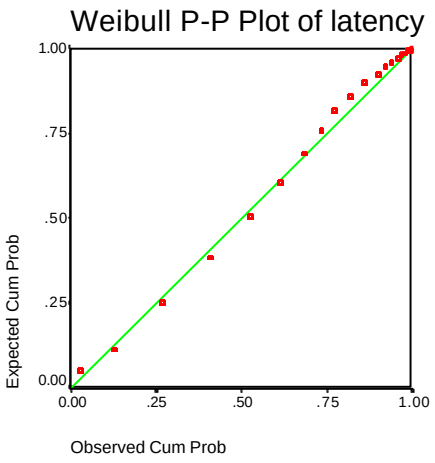
b)



c)



d)

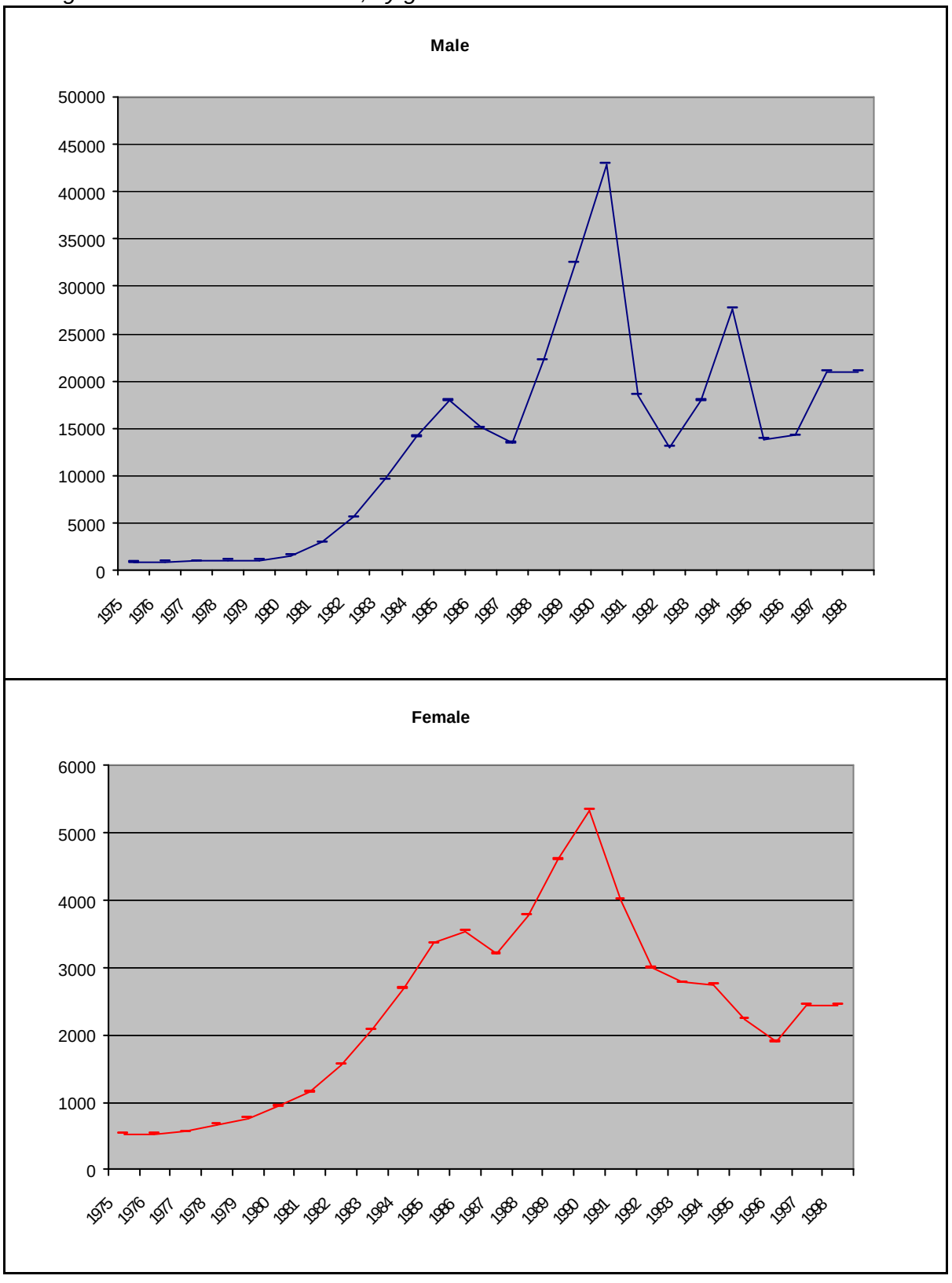


Step 2. Estimating the onset incidence curve from therapy incidence data through the EB-BC.

The EB-BC was first applied on national data provided, for the period 1986-1998, by the Ministry of Health, who routinely collects data on clients in public services. Data on clients in private services were excluded as these are mostly referred from public centers and double counts would have been high. Data include only clients who enrolled in treatment for the first time and do not include those who only contacted services but did not receive treatment. Treatment refers to any therapeutic and rehabilitation procedure - either pharmacological or not - offered by the service, even outside the premises (prisons, therapeutic communities, hospitals).

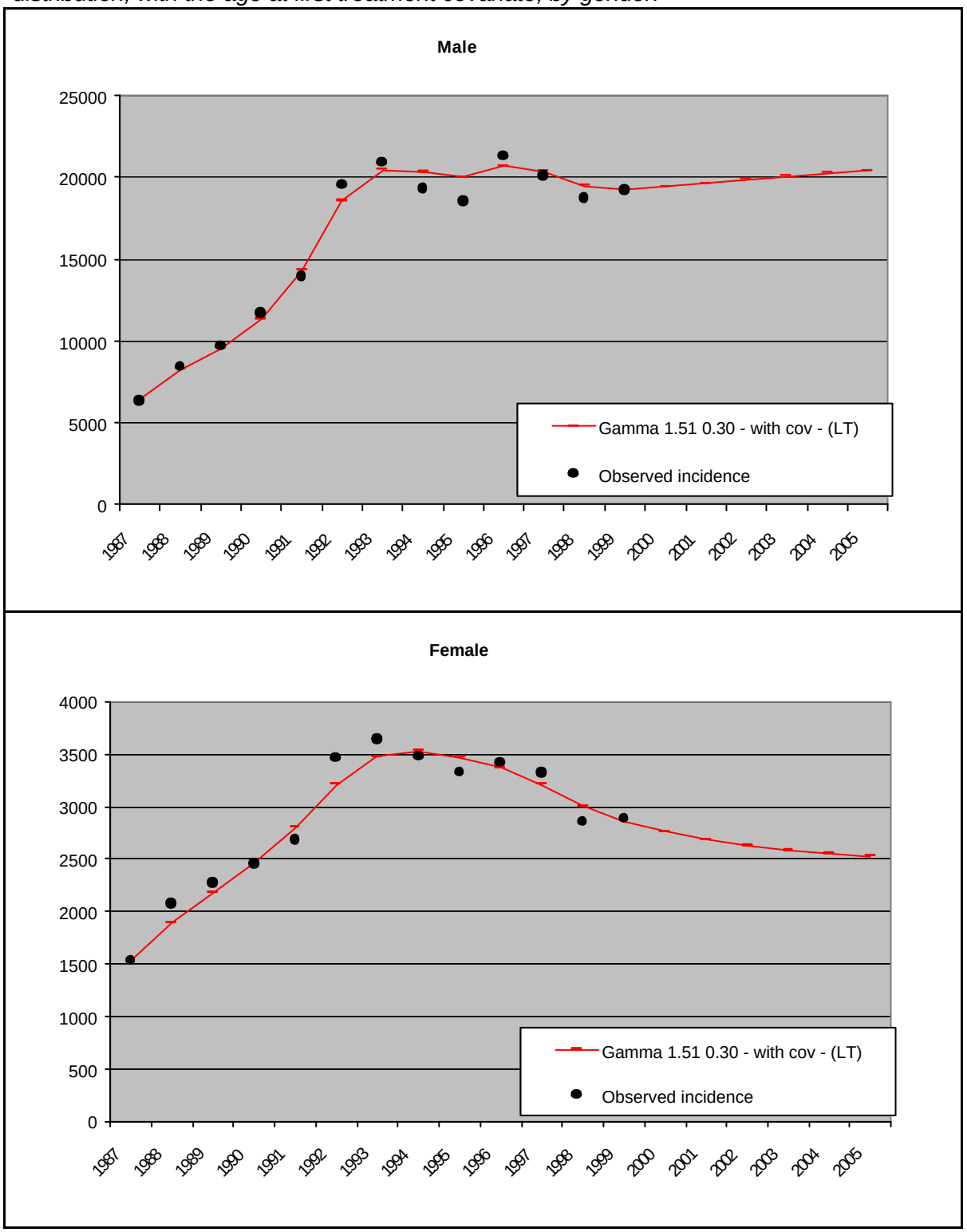
In order to apply the EB-BC, the original therapy incidence data were classified according to age-classes, gender and region, and multiplied by 0.70 to take into account the estimated proportion of double counting and the proportion of drug users in treatment who are not heroin users (about 10%).

Figure 2: *Incidence of DUs estimated through the Gamma latency period distribution, with the age at first treatment covariate, by gender.*



The incidence curve for the total population estimated through the EB-BC shows two peaks respectively in 1985 (the lowest one) and 1991 (the highest one). Nevertheless, this is not necessarily an evidence of two different epidemics in time, since the two peaks could be determined by the combination of different local sub-epidemics developing differently in time. The same trend, in particular the location of the peaks, can be observed also for the incidence curve estimated for males and females separately (Figure 2). The comparison of observed and estimated therapy incidence (Figure 3) shows a satisfactory agreement for both sexes.

Figure 3: *Incidence of Dus under treatment estimated through the Gamma latency period distribution, with the age at first treatment covariate, by gender.*



From further analyses reported below (Figures 4-6), it appears that the application of the method at regional level provides the picture of several sub-epidemics with peaks differently located in time and with evident space-time trends. Figure 4 reports the onset incidence estimates for six of the biggest regions in Italy. As can be seen, the principal peak of each curve is differently located. In particular, the northern regions (Lombardia and Emilia-Romagna) peaked before the others, more or less simultaneously the peak of Puglia can be observed (this region can be considered a border region with respect to the Balkan traffic route) then the peaks of central and central-southern regions (Lazio and Campania) can be observed and, finally, the peak in Sicilia, where, due to the wider uncertainties corresponding to recent years, it is also possible that the curve is still increasing. Statistics on deaths from drug abuse allow to confirm this last observation for Sicilia. Figure 5 reports the forecasts of therapy incidence in the same regions which can be used to estimate the health care needs at a short-medium term. The observed therapy incidence data are in good agreement with the curves shown in the graph (not shown in the present paper). Figure 6 reports the estimated cumulative incidences which measure the overall impact of the epidemic since the beginning in each region.

***Figure 4**: Incidence of DUs estimated through the Gamma latency period distribution, with the age at first treatment covariate for six Italian Region - Total population.*

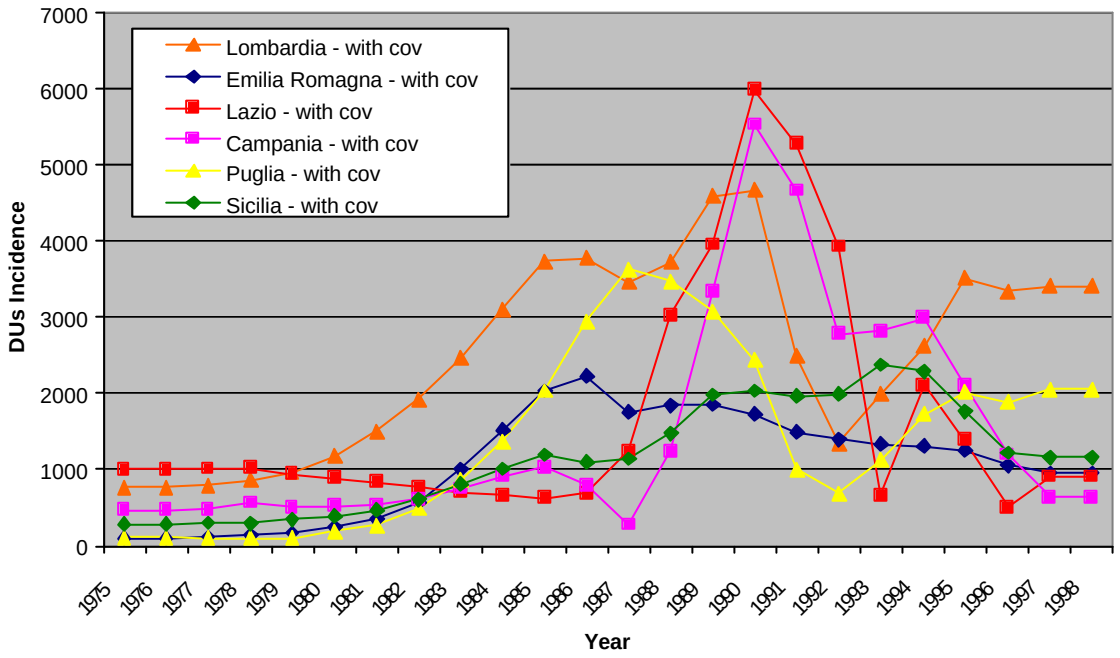


***Figure 5**: Incidence of DUs under treatment estimated through the Gamma latency period distribution, with the age at first treatment covariate for six Italian Region - Total population.*

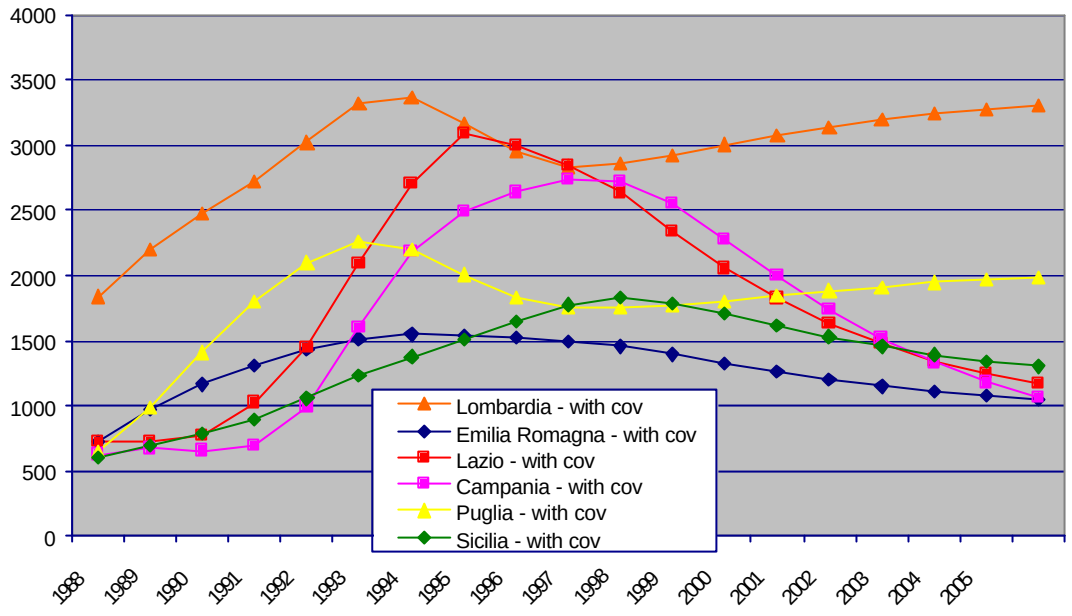
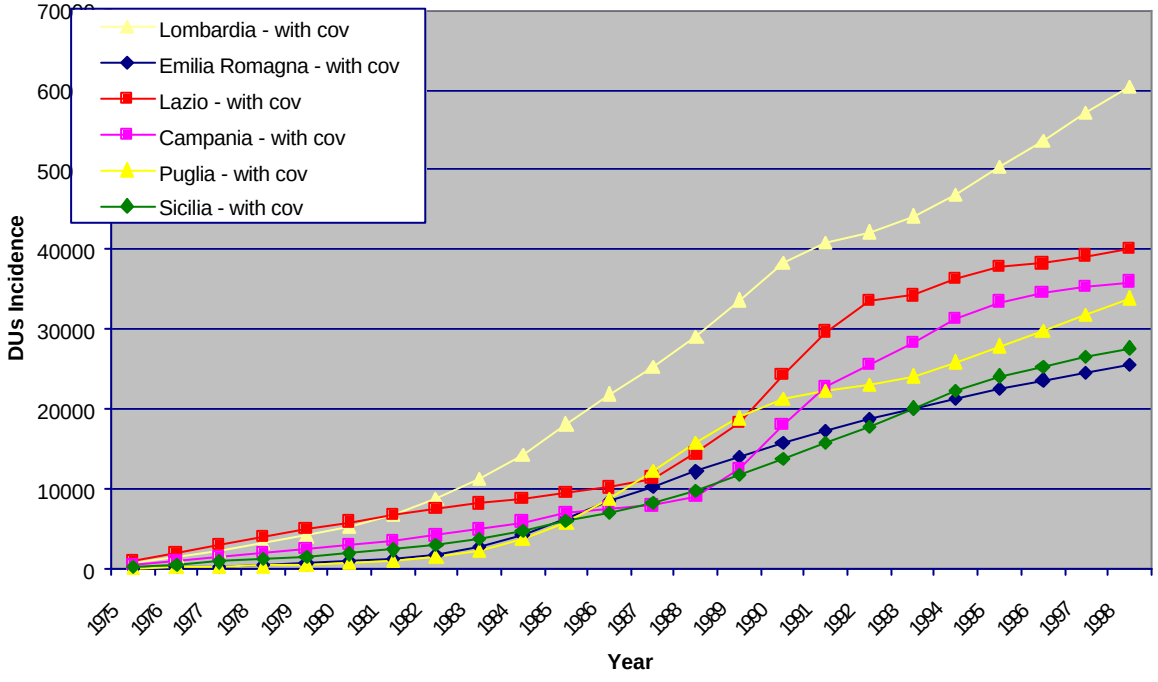


***Figure 6**: Cumulative incidence of DUs estimated through the Gamma latency period distribution, with the age at first treatment covariate for six Italian Region - Total population.*



Data Demands

Step 1:

For the latency period analysis data from health care services has to be provided according to the following specification:

Raw (individual as opposed to aggregated) data, reporting, at least, the following variables:

- Age at first problem drug use
- Age at first registration in some health care service
- Gender

and any other variable that could be used as covariate in the latency period analysis, such as:

- Educational level
- Ethnicity
- Residence
- Health care type
- Route of administration

Step 2:

The EB-BC procedure uses periodic incidence data of "new" individuals under treatment in some health care services classified, at least, by age and gender.

The data-file needed for the EB-BC could be derived either from the data used for the latency period analysis, yet if in a different format, or from different data sources, depending on the local availability of data.

For Italy the EB-BC was performed on the basis of national and local data, other than those used for the latency period analysis (local data).

Limitations

3) Latency period:

- There might be some bias because there is no standardised way to ask the age at first heroin use at the treatment centres. When, at a treatment centre, the question “how long are you using drugs?” is raised, the client can interpret this question as the period of uninterrupted drug use before treatment demand. When the period of drug use is interrupted, the latency period seems to be shorter, and the age of starting drug use will be higher than real figures.
- Age of first use is less reliable than age. This affects both ends of the distribution, in particular ages under 12 and those over 30 of “age at first use”. Short latency periods observed for older drug users are less reliable than short periods among younger users.
- Estimates of the latency period using treatment/surveillance data may under-estimate the true value because the data are right truncated. This bias is higher for recent epidemics whereas it will be minimal for older (stabilised) epidemics.
- The latency period can be analysed by entry cohort (i.e. by year of first report) or by “onset” cohort (i.e. by year of first use). All being equal they produce the same estimates. However, if incidence changes over time, analyses by entry cohort may be biased as they tend to produce decreasing observed periods when incidence is increasing and viceversa.
- There might be some bias due to local peculiarities of the therapy services. For example, in Amsterdam large scale methadone programmes started at 1980 and opiate users couldn’t apply for treatment during the ‘70s. Therefore, during the first years, the latency period will be prolonged. This bias could be corrected including in the study only opiate users who demanded for treatment for the first time after 1985.
- There might be differences in latency period between drug users originating from different countries. These differences could reflect differences in the onset of the heroin epidemic among different subgroups. In this case the heroin epidemic among those originating would be the oldest, followed by the epidemics among the other groups. When

the epidemic grows the latency period will increase, especially when the incidence is decreasing. The same effect may affect the stratified analysis with respect to other variables, such as “route of administration” or “sex”.

4) Back-Calculation:

- The main problem related to the data used to apply the BC procedure is represented by the double counting which causes the actual incidence of DUs presenting to treatment be lower than the observed one, and, as a consequence a bias in the EB-BC estimates. An attempt to overcome such problem could be done by inflating the observed incidence data, on the basis of some information about the amount of double counting, if available.
- Due to the rather long latency period therapy data provide little information on recent onset. Therefore, estimation of recent incidence of drug use is more imprecise than estimation of long ago incidence. For the same reason BC cannot be applied when the time series of aggregated therapy incidence data is too short (less than twice the mean LP).

- Back-calculation does involve several important assumptions and parameters that need to be estimated for drug users (e.g. the shape of the "incubation" or «latency» distribution, the influence of covariates such as age, sex, education level…..). These are based on external information coming both from observational secondary data and from survey. Both the incubation period and the subsequent estimates are highly sensitive to the quality and completeness of data.

Literature

1. Brookmeyer R., Gail H.G., Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States, Lancet, 2, 1320-1322, 1986.
2. Brookmeyer R., Gail H.G., A Method for Obtaining Short-term Projections and Lower Bounds on the Size of the AIDS Epidemic*. J. of the American Statistical Association* 83, 301-308, 1988.
3. Brookmeyer R., Reconstruction and Future Trends of the AIDS Epidemic in the United States. *Science Articles*, 37, 37-253, 1991.
4. Collet D. “Modelling survival data in medical research”, Chapman and Hall, London, 1994.
5. EMCDDA, “Pilot project to estimate time trends and incidence of problem drug use in the European Union”, Final report, Lisbon, 1999a.
6. EMCDDA, "1999 Annual Report on the State of the Drugs Problem in the European Union", Lisbon, 1999b.
7. EMCDDA, “Study on incidence of problem drug use and latency time to treatment in the European Union”, Lisbon, 2000.
8. Heisterkamp S.H., Downs A.M., van Houweling J.C., Empirical Bayesian Estimators for Reconstruction of HIV Incidence and Prevalence and Forecasting of AIDS. I. Method of Estimation. In Quantitative Analysis of HIV/AIDS: Development of Methods to Support Policy Making for Infectious Disease Control. Ph.D thesis: University of Leiden, 65-98, 1995.
9. Heisterkamp S.H., van Houwelingen J.C., Downs A.M., “Empirical Bayesian Estimators for a Poisson Process propagated in time”, Biometrical Journal, 41-4, 358-400, 1999.
10. Hunt L.G., Chambers C.D., The Heroin Epidemics, SPECTRUM PUBLICATIONS INC, NY, 1976.
11. Marubini E., Valsecchi M.G., “Analysing survival data from clinical trials and observational studies”, Wiley, NY, 1995.
12. Ravà L., Rossi C., “Estimating the size of a hidden population involved in the HIV/AIDS epidemic: a method based on Back-Calculation and dynamical models”, in

Simulation in the Medical Sciences, Anderson & Katzper eds., The Society for Computer Simulation, San Diego, California, 57-62, 1999.
13. Ravà L., Calvani M.G., Heisterkamp S., Wiessing L., Rossi C. "Incidence indicators for policy making: models, estimation and implications", UN Bulletin on Narcotics, 2001, in press.
14. Rosemberg P., Gail M.H., Backcalculation of Flexible Linear Models of the Human Immunodeficiency Virus Infection Curve. *Applied Statistics* 40, 269-282, 1991.
15. Rossi C., Monitoring drug control strategies: hidden phenomena, observable events, observable times, International Journal of Drug Policy, 10-1, 131-144, 1999.

The Brookmeyer and Liao Method.

Statistical Background
A possible alternative method for onset incidence estimation can be considered: the Brookmeyer and Liao Reporting Delay Adjustment (RDA) method. The RDA method was first developed by Brookmeyer and Liao (1990) in the AIDS field and the re-modulation for application to the estimation of the incidence of problem drug use is due to Seaman, Hickman and De Angelis (2000).
In the AIDS context, the problem of adjusting for reporting delay arises when complete information on AIDS incidence is required. Typically, the number of AIDS cases reported to surveillance centres seriously underestimate the number of recent AIDS diagnoses, because of substantial delay in reporting. Therefore, an estimate of AIDS incidence is obtained by adjusting reported data for reporting delay (Brookmeir and Damiano, 1989, Brookmeyer and Liao, 1990, Zeger and Lai-Chu, 1989). The interval of time between "onset of drug use" and "presentation to treatment" can be seen as analogous to the time between "AIDS diagnosis" and "AIDS report", making the problem of estimating drug use incidence similar to that of estimating AIDS incidence. The methods developed for adjusting AIDS reports can thus be adapted to the drug use context, to estimate the lag between onset of heroin use and treatment presentation and, hence, the historical trends in heroin incidence.

Application (data from Clinical, Medical and Social System)
In order to apply the RDA method the following quantities must be defined:

- *I(t)*: the onset incidence of DUs (who present to treatment at least once) at time t*;*
- *F(s-t)*: the cumulative distribution of the period between the time *t* of the first use of drug, and the time *s* of the first presentation for treatment, the "Latency period (LP) distribution";
- *O(t)*: the observed onset incidence of DUs at time *t*.

The method, based on one year of therapy data, say time s, simply adjusts by calibration the observed onset incidence (we define onset as a person's first use of heroin) at time t (t<s) by dividing it by the probability that an individual starting heroin use at time t asks for therapy for the first time not later than s, i.e. by the probability that his/her latency period is less than s-t, that is *F*(s-t), the cumulative distribution function at (s-t):

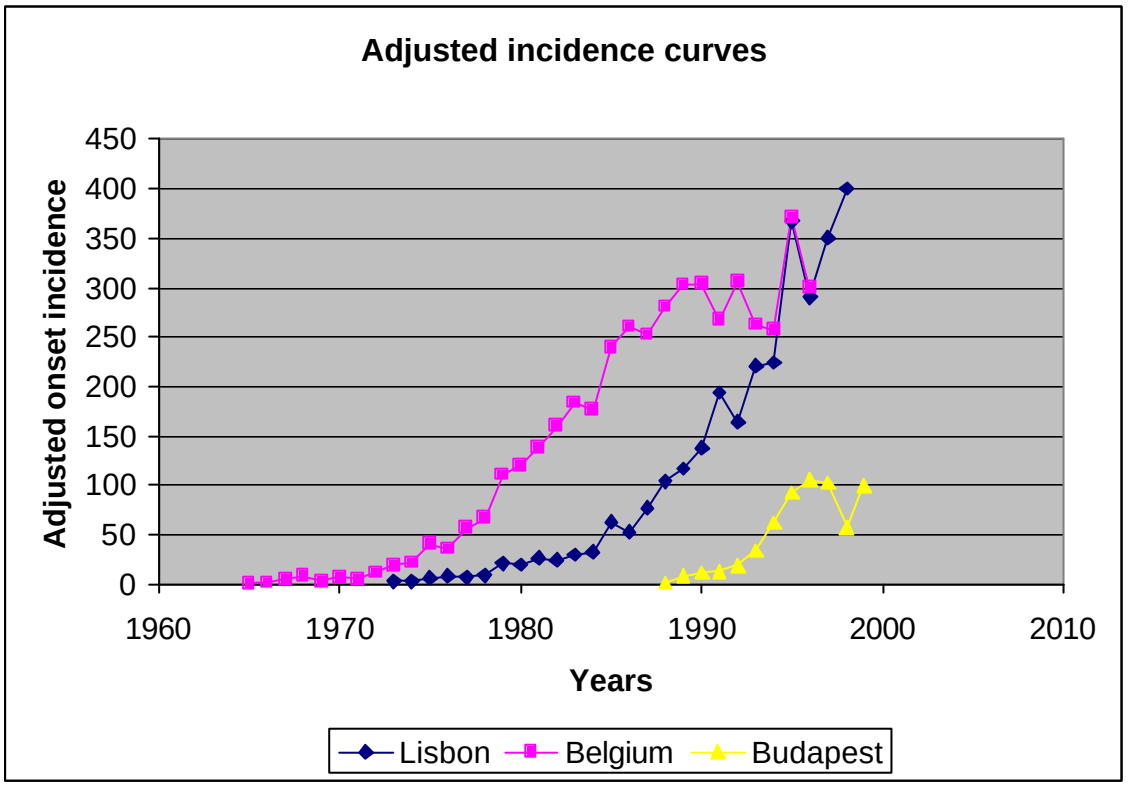
Thus, for the present method too, the latency period (LP) analysis is a prerequisite.
In summary the application follows the 2 steps:

Step 1: Estimate the latency period distribution *F(s-t),* possibly depending on covariates such as age, sex… by standard methods used in Survival Analysis.
Step 2: Apply the calibration procedure to the observed onset incidence of DUs at time *t:* O(t), using the LP distribution estimated in Step 1.

Example: Estimating incidence of heroin use in Lisbon, Budapest and the French Community of Belgium

RDA method was applied to various data-sets. The results obtained for Lisbon, Budapest and the French Community of Belgium are summarized in the following Figure 1. The different phases of the epidemics in the various sites can be easily appreciated. No covariate was included.

Figure 1: *Incidence curves estimated by the Brookmeyer and Liao method for Belgium, Lisbon and Budapest.*



Data Demands

Step 1:

For the latency period analysis data from health care services has to be provided according to the following specification:

Raw (individual as opposed to aggregated) data, reporting, at least, the following variables:

- Age at first problem drug use
- Age at first registration in some health care service
- Gender

and any other variable that could be used as covariate in the latency period analysis, such as:

- Educational level
- Ethnicity
- Residence
- Health care type
- Route of administration

Step 2:

It must be stressed that LP method used to estimate the incidence curves for Lisbon, Belgium and Budapest requires different data with respect to the Back-Calculation. For the

RDA method, data are required with individual records containing year of first use and all the covariates included in the LP model, thus, essentially, same data as those used for LP analysis.

The RDA method also necessarily uses a biased estimate of the latency period as it by definition cannot include latency periods that exceed the length of the observed time series of treatment data. This results in an incidence curve that may not only be much lower than the BC one, but also not parallel to it. If the real incidence curve is very marked there should be no problem and the RDA method should correctly indicate the direction of the incidence (rising or falling). However in a less marked situation the RDA curve may wrongly indicate rising or falling incidence.

Limitations

1) Latency period:

- same as for the BC method

2) RDA method:

- Inaccuracy in the data may introduce biases. Whilst dates of birth are generally accurate, "age at first use" may be less reliable. Overstatement or understatement of age at first use could lead to under-estimation or inflation of trends in incidence. Most heroin users start their drug use between the ages of 17 and 20, but there are a significant number that start drug use earlier or later, and the true distribution of age at first heroin use is not known.

- For the RDA method, data are required with individual records containing year of first use, often in treatment centres this item is not asked for many (e.g. half) of the individuals. The BC method can reconstruct incidence using all individuals because latency time is estimated separately, while the RDA method will in this case result in an incidence estimate based only on those individuals reporting year of first use and may therefore underestimate incidence much more severely than the BC method.

- As the RDA method requires individual data records, while the BC method can use aggregated treatment data, the RDA method can only reasonably be used in local or multi-local studies. The BC method can easily be used to cover a large country as well as local data.

Literature

16. Brookmeyer R and Damiano A. Statistical methods for short-term projections of AIDS incidence. Statistics in Medicine 1989; 8: 23--34.
17. Brookmeyer R., Liao J., “The analysis of delays in disease reporting: methods and results for the acquired immunodeficiency syndrome”, American Journal of Epidemiology, Vol. 132, No. 2, 355-365, 1990.
18. Collet D. “Modelling survival data in medical research”, Chapman and Hall, London, 1994.
19. EMCDDA, “Study on incidence of problem drug use and latency time to treatment in the European Union”, Lisbon, 2000.
20. Hickman M., Seaman S., De Angelis D., “Estimating the relative incidence of heroin use: application of a method to adjust observed reports of presentations at specialist treatment agencies”, *American Journal of Epidemiology*, 2000.
21. Marubini E., Valsecchi M.G., “Analysing survival data from clinical trials and observational studies”, Wiley, NY, 1995
22. Zeger SL and Lai-Chu S. Statistical methods for monitoring the AIDS epidemic. Statistics in Medicine 1989; 8: 3--21.

2. Management and co-ordination aspects

The main results outlined in this report are summarised and presented in the mentioned papers. Further results will be presented in the papers presently in preparation.

Addresses of participants (representatives of different countries).

Organization	Contact Person	Address	Telephone	Fax	e-mail
University of Rome Tor Vergata Dipartimento Di Matematica	Rossi, Carla	Via Ricerca Scientifica 00133 Rome Italy	+39 06 7259 4676	++39 06 7259 4699	c.rossi@agora.stm.it
National University of Ireland, Mathematics Department	Comiskey, Catherine	Maynooth Co. Kildare, Ireland	+353 1 70839-14	++353 1 70839-13	Cc@maths.may.ie
Centre for Research on Drugs and Health Behaviour - Social Science and Medicine - Imperial College	Hickman Matthew	200 Seagrave Road London SW6 1RQ United Kingdom	0181 846 6567/ 6565	0181 846 6555	m.hickman@ic.ac.uk
Head Epidemiology and Information Municipal Health Service	Van Ameijden Erik	P. O. Box 2423 NL-3500 GK Utrecht The Netherlands	+31-30-286.3519	+31-30-286.3344	e.van.ameijden@utrecht.nl
Gabinete de Planeamento e de Coordenação do Combate à Droga (GPCCD)	Ribeiro Jorge	Rua de Alcolena 1 1400 Lisboa Portugal	+351-1-3015953	+351-1-3010988	jsribeiro30@hotmail.com
Scientific Institute of Public Health - Louis Pasteur	Walckiers Denise	Wytsmanstreet, 14 B-1050 Brussels, Belgium	+32/2/642.50.35	+32/2/642.54.10	denise.walckiers@ihe.be

Appendix 1

Country Report : Hungary

Katalin Veress*

* Institute of Hygiene and Epidemiology, Semmelweis University, Medical School, Budapest

Introduction

Hungary is located in Central Europe surrounded by the Ukraine and Romania from the East and South-East, by the countries of the former Jugoslavia from the South, South-West and by Austria from the West. The geographical location provides a certain "border-crossing" situation between Western and Eastern Europe. The international airport, Ferihegy with heavy traffic, - in close proximity (1-2 hours flight) with any European capital, as well as with its worldwide connections - facilitates this role. Just like the river Danube that crosses the country, and has connections to big European ports. The Jugoslavian war of the !990's forced the Balcanic route of drug smuggling to divert and drive across Budapest. International trafficking with drugs has only widened since than, Budapest or any of the Western border-crossing points of Hungary become frequent meeting points for dealers from Africa and Western Europe, intermediated sometimes by Hungarian carriers.
The country covers a total area of 93.000 km^2. With 10 million inhabitants, Hungary has a population density around 110 inhabitants per km^2. One fifth of the total population of Hungary are residents of Budapest, the capital city .
The age group 15-44 years, relevant from the view point of drug use, represents 43% of the population (see **Table I**). Because of the low birth rate the population has been on decrease for the recent couple of years.

Table 1 City population by age and gender 1998

age groups	Male	Female	Total (M+F)
< 15	137496	130927	**268423**
15-19	63848	62504	**126352**
20-24	82460	85035	**167495**
25-29	67007	71363	**138370**
30-34	55572	59510	**115082**
35-39	48157	54133	**102290**
40-44	69791	80690	**150481**
45-49	64309	77827	**142136**
50-54	62387	75603	**137990**
55-59	49327	61899	**111226**
60-64	37909	53482	**91391**
65+	109944	200203	**310147**
Total	**848207**	**1013176**	**1861383**

Ref.: Multi-City Network on Drug Misuse Trends
Co-operation Group to combat drug Abuse and
Illicit Trafficking in Drugs (Pompidou Group)
P-PG/Epid (98) Budapest City Report by Katalin Veress
Strasbourg 1998

In economics prevailed certain negative characteristics. The effect of the austerity package called "Bokros-csomag" in Hungary, in fact "took its toll" in 1997: people's real income has reduced significantly on an economic background when in the first half of the 1990s the GDP has declined each year . The number of available work places lessened, people became threatened by the earlier almost “unknown” - unemployment.

Drug use has been spreading among the circumstances of developing market economy. Threatening economical crisis, confusion created by the so called “health-care reform”, and the hardly developed protective grass-route cooperation culminated in a flimsy social net, in almost full lack of institutions and possibilities for rehabilitation and social reintegration of addicts. The country has been characterized by high premature mortality in general.

Nevertheless, a boom of tourism and entertaining industry with quick proliferation of discos, bars and night-clubs have got characteristic. Spread of a new type of youth culture became noticable. Discos, night clubs (often based on the investments of offenders) are key-spots of drug distribution.

Drug-tourism of the 80s has imported drugs in smaller amounts from abroad, mainly the Netherland. This has been changed by smuggling large quantities of illicit drugs through the borders, especially, since the time of the Jugoslavian war.

Supplies of heroin are under the control of Kosovo-Albanien drug traders, and as another option: a country of origin is quite often again the Netherland.

Opiate use started in Hungary at the end of the 70’s when beside prescription drugs (frequently with codeine) youngsters started to use home made preparations of poppy seed and poppy straw. Opiates became the primary drugs of abuse by the mid 80s, poppy milk has frequently been used intravenously as well. Heroin – smuggled from the Golden Triangle through Turkey - appeared in street markets around railway stations, etc. It’s around the mid 90’s when the first heroin users presented in treatment sites. As for trafficking Hungary counted as a transit country between the Middle East and Western Europe until the start of the civil war in Jugoslavia. That’s when the South-Balkan Route diverted, crossing Hungary and the transformation into a consumer country has started. Numbers of opiate addicts and intravenous drug users increased steadily each year in the period 1994-98 (see **Tables VI and VII** and **Figure I**).

While there is a lack of information on the level of drug use among the general population, a few selected subpopulations, different level risk-groups have been surveyed several times. Hungary has been participating in the European School Survey from the very start.

In the ESPAD study there were questionned 17 year old students of the secondary schools in Hungary. The main advantage of the study is using standardized study set-up through all the surveys that took place annually for the last three years. Life time prevalence of illegal drugs and inhalants among the surveyed school population almost has doubled, grown from 10.0% to 19.1% in the country between 1995-1999. **Table II** demonstrates life time prevalences (%) of different substances used among the 2nd grade students in Budapest, while in **Table III** there are the results given as 99% confidence intervals to the values produced in the surveys during the period 1992 – 2000. There is an explicit rise in the prevalence of each and every substance. The life prevalence of heroin use ab. has doubled between 1995 and 1999, and is far beyond the one of marijuana/hashish or the life-prevalence of amphetamines/ ecstacy. The heroin life time prevalence - as all estimates - has seemed to level off in 2000.

Table II-1: Life time prevalence of different substances in school population 1995 / 1998

DRUGS experienced	1995 BUDAPEST	1998 BUDAPEST
Marijuana or hashish	5,8	16,9
Amphetamines	0,8	4,5
Ecstacy	1,2	5,4
LSD/ other hallucinog.	2,0	7,4
Heroin	0,4	1,3
cocain	0,2	1,5
Crack	0,0	1,6
any drug by injection	0,8	1,5
Tranquilizers	9,0	10,5

Table II-2: Life time prevalence of different substances in school population

DRUGS experienced	1999 BUDAPEST
Marijuana or hashish	14.1
Amphetamines	3.5
Ecstacy	4.2
LSD/ other hallucinogenes	4.3
Opiates	3.2
cocain	1.1
Crack	1.0
any drug by injection	0.6
Tranquilizers	10.6
Hypnotics	4.7

Table III: Life time prevalence of drug use among second grade (17 year old) secondary school students in Budapest

Substances	1992	1995	1998	1999	2000
Opiates	2.4-3.8	0.8-2.0		3.4-5.8	1.3-3.9
Marijuana/Hashish	5.4-7.2	6.0-8.6	12.9-20.9	20.9-28.1	21.1-28.3
Inhalants	2.7-4.1	3.5-5.5		2.3-5.5	1.6-4.4
Amphetamines	2.4-3.8	0.5-1.5	3.3-6.7	5.5-10.1	3.0-6.6
LSD	1.0-2.0	2.0-3.6	4.6-10.2	5.9-10.5	3.5-7.3
Crack		0.2-0.8	0.3-2.9	0.5-2.5	0.2-1.8
Cocaine	0.3-0.9	0.2-1.0	0.2-2.8	0.9-3.3	0.2-1.8
Heroin		0.2-0.8	0.1-2.5	0.3-2.3	0.2-2.0
Ectasy		0.8-2.0	3.0-7.8	4.3-8.3	3.5-7.3
Injected drugs		0.2-0.8	0.2-2.8	0.0-1.1	0.3-2.1
Illegal drugs and Inhalants	10.4-12.8	10.5-13.7		25.0-32.6	23.4-30.8
N	4518	2762	597	932	946

Ref.: Elekes, Zs. and Paksi, B.:Drogok és Fiatalok (Drugs and Youth), ISM, Hungary 2000

Treatment services and data collection system

Services available today were established partly using and transforming existing psychiatric health care institutions (inpatient psychiatric wards and outpatient psychiatric or mental health clinics). At the same time there have been developed new institutions like the specialised drug outpatient clinics. Partly, outpatient services, developed on the previously existing base, built on the logistics and professional experience of institutions caring earlier for alcoholics. These clinics provide care for patients with any kind of addiction nowadays. Budapest is the city best supplied with services, although low threshold facilities, outreach settings aiming at harm reduction, and rehabilitation care has been poorly developed all over the country. A targeted, wide scale developmental process has been planned in the recently accepted national strategy on the fight against drugs.

Data on treatment demand have been collected since 1994.The most reliable data source in the country - is the still in force National Statistical Data-Collection Program (OSAP).
In the moment the system collects an aggregated, minimal core dataset (type of substance(s) used; mode of administration; frequency of use) on patients, rerequesting or starting treatment in anyone of the above institutions during a given year; and age and gender distributions as for a December, 31 census annually. The Programme for Data-Collection introduced more than five years ago, works with nearly full coverage of the due institutions by now. At the same time, the system collects only aggregated data pulling out those from different clinical documents once a year (the standard format recommended in the Definitive Protocol by the Pompidou Group for this purpose has not yet been introduced officially in Hungary).

Study population

It was exactly the lack of routine registration of the appropriate data that made the especially good collaboration with clinicians in treatment sites a necessary precondition for data collection for the particular study. It meant several communications to doctors, asking initially, and periodically reminding them to obtain answers to the key questions when interrogating the individual patient. It was necessary to make them understand that without having the anamnestic information, - in particular the year when a patient started his/her drug carrier and the year of first treatment demand - it would not have been possible to estimate the latency time, and that finally could make this patient to drop out from the study even if to have abundant other information related to him/her.
Data on heroin addicts has been collected in a central psychiatric ward and its integrated outpatient clinic, both specialized in care for drug addicts, providing the whole range of therapies in demand (from detoxification to methadon substitution and maintenance). This was the only treatmant institution providing this type of care continuously without any limitations regarding treated users' numbers since the first appearance of patients using opiates.

Table IV-1
Treatment demand by age and gender
(Budapest)

Year: 1998

age group	First treatment Male	Female	Total (M+F)	All treatment of 31 Dec census Male	Female	Total (M+F)
< 25				1783	679	**2462**
25-34				1022	386	**1408**
35 +				362	337	**699**
Total	**1924**	**1058**	**2982**	**3167**	**1402**	**4569**
mean age						
% < 25 yrs				**56,3%**	**48,4%**	**53,9%**
% females			**35,5 %**			**30,7%**

Table IV-2:
First treatment demand in Budapest in 1994 - 1998

		1991	1992	1993	1994	1995	1996	1997	1998
FIRST	male				418	679	561	1793	**1924**
TREATMENT	female				129	303	277	883	**1058**
	total (M+F)				547	982	838	2676	**2982**
	% females								**35,5%**
ALL	male								**3167**
TREATMENT	female								**1402**
	total (M+F)				1541	1965	2457	3920	**4569**
	mean age								
	% < 25 yrs								**53,9%**
	% females								**30,7%**

Ref.: Multi-City Network on Drug Misuse Trends
Co-operation Group to combat drug Abuse and Illicit Trafficking in Drugs (Pompidou Group)
P-PG/Epid (98) Budapest City Report by Katalin Veress
Strasbourg 1998

The age structure of heroin users can be derived from **Table V.**
As data from the 1999 census shows most of the addicts demanding treatment are of the age 15 – 29 years.
86% of heroin users belonged to this age group, while 48% of them were within 20-24, which does not make them different from the rest of clients using other substances.

Table V.: Distribution of Treatment Demand by age groups and types of substances, Budapest, as of 31st December, 1999

Illegal Drugs Types	Males								Females								M+
	< 13	13-14	15-19	20-24	25-29	30-34	35-X	Total	< 13	13-14	15-19	20-24	25-29	30-34	35-X	Total	
Opiate types	**1**	**9**	**427**	**982**	**553**	**221**	**99**	**2292**		**2**	**157**	**341**	**206**	**46**	**27**	**779**	**30**
Heroin	1	1	124	487	252	102	36	1003		2	78	164	79	23	12	358	13
Cocaine types			**12**	**34**	**19**	**15**	**23**	**103**			**2**	**11**	**7**	**2**	**2**	**24**	**12**
Cannabis types		**2**	**76**	**158**	**75**	**18**	**26**	**355**		**1**	**18**	**33**	**19**	**10**	**10**	**91**	**44**
Hallucinogenes		**1**	**29**	**61**	**23**	**13**	**12**	**139**			**10**	**22**	**8**	**2**	**8**	**50**	**18**
Amphetamine type		**3**	**91**	**183**	**97**	**30**	**35**	**439**		**2**	**47**	**62**	**21**	**6**	**9**	**147**	**58**
Sedatives/Hypnotics		**1**	**25**	**29**	**43**	**34**	**108**	**240**	**1**	**9**	**24**	**40**	**47**	**44**	**172**	**337**	**57**
Inhalants	**3**	**9**	**34**	**53**	**19**	**15**	**19**	**152**	**2**	**2**	**16**	**11**	**11**	**5**	**1**	**48**	**20**
Others			**7**	**6**	**8**	**10**	**26**	**57**			**3**	**5**	**24**	**36**	**97**	**165**	**22**

Ref.: Jelentes a kabitoszerfogyasztokrol es kezelesukrol [OSAP], 1999. Information Technology Centre of the Institute of Psychiatry and Neurology, Budapest 2000

Just like in all substance types - opiate, including heroin users' Treatment Demand has been on the rise for all the years throughout the period 1996-99. As compared to all the other ones, it is heroin users amongst whom there is the highest proportion of intravenous use. It is and remains well above 80% throughout the period (see **Table VI**)

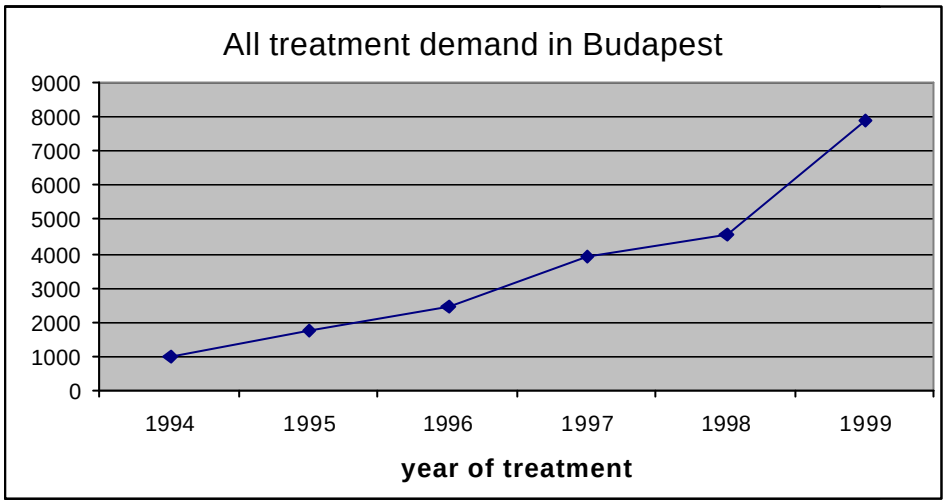
Table VI. Heroin Users' Treatment Demand and Percentage of Injecting Users, Budapest 1996 – 1999

	1996			1997			1998			1999		
	(Tr. Dem.) N	Injecting N	%	(Tr. Dem.) N	Injecting N	%	(Tr. Dem.) N	Injecting N	%	(Tr. Dem.) N	Injecting N	%
Opiate type	977	633	64.8	1986	1245	62.7	2257	1596	70.7	3005	2229	74.2
Heroin	446	372	83.4	1442	1171	81.2	1158	925	80.0	1358	1149	84.6
Cocaine type	29	9	31.0	97	15	15.5	120	27	22.5	138	13	9.4
Hallucinogens	54	6	11.1	158	14	8.9	112	7	6.3	229	17	7.4
Amphetamine type	195	22	11.3	483	60	12.4	526	88	16.7	713	74	10.4
Sedatives	434	13	3.0	518	12	2.3	408	34	8.3	836	39	4.7

Table VII: All treatment demand Budapest

1994	982
1995	1771
1996	2457
1997	3920
1998	4569
1999	7889

Figure I: All Treatment Demand in Budapest, Hungary, 1994-1999



Ref.: Jelentes a kabitoszerfogyasztokrol es kezelesukrol [OSAP], 1997-2000
Information Technology Centre of the Institute of Psychiatry and Neurology, Budapest 2000

Results of the descriptive statistical analysis of the study population

Frequencies

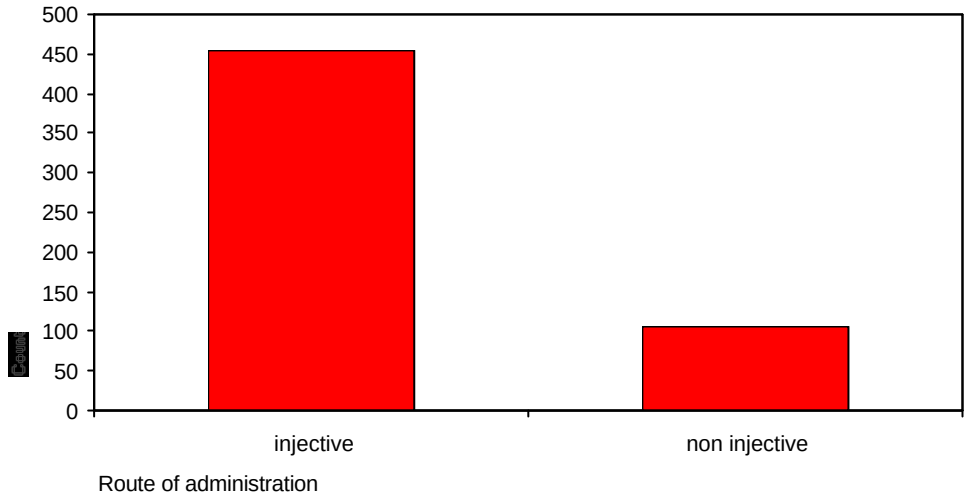
As for different characteristics of the study population – from the view point of the route of administration (**Table 1.1**) - significant majority (2/3) of heroin addicts presenting in clinics inject the substance (see **Figure 1.1**) while the others inhale it from prehated foil, sniff or infrequently though consume it per os. Usually addicts start with some other route of administration, and later on in their carrier they turn to injecting.

Table 1.1

Route of administration

	Frequency	Percent	Cumulative Percent
injective	455	81,11	81,11
non injective	106	18,89	100,00
Total	561	100,00	

Figure 1.1



As it has been shown in **Table 1.2** - about 80% of heroin addicts in treatment are males.

Table 1.2

Gender

	Frequency	Percent	Cumulative Percent
Male	451	80,39	80,39
Female	110	19,61	100,00
Total	561	100,00	

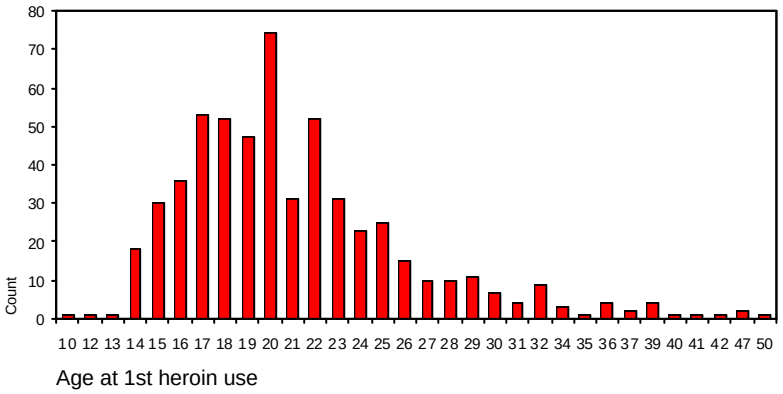
Young people starting heroin at an age between 15 and 20 years constitute about 50% of the treated population (**Table 1.3**). Moreover, more than 90% of the population will be represented by a somewhat wider age interval at start – between 11 and 28 yrs as it has been shown by the bar chart (**Figure 1.3**) below .

Table 1.3

Age at 1st heroin use

Age	Frequency	Percent	Cumulative Percent
10	1	0,18	0,18
12	1	0,18	0,36
13	1	0,18	0,53
14	18	3,21	3,74
15	30	5,35	9,09
16	36	6,42	15,51
17	53	9,45	24,96
18	52	9,27	34,22
19	47	8,38	42,60
20	74	13,19	55,79
21	31	5,53	61,32
22	52	9,27	70,59
23	31	5,53	76,11
24	23	4,10	80,21
25	25	4,46	84,67
26	15	2,67	87,34
27	10	1,78	89,13
28	10	1,78	90,91
29	11	1,96	92,87
30	7	1,25	94,12
31	4	0,71	94,83
32	9	1,60	96,43
34	3	0,53	96,97
35	1	0,18	97,15
36	4	0,71	97,86
37	2	0,36	98,22
39	4	0,71	98,93
40	1	0,18	99,11
41	1	0,18	99,29
42	1	0,18	99,47
47	2	0,36	99,82
50	1	0,18	100,00
Total	561	100,00	

Figure 1.3



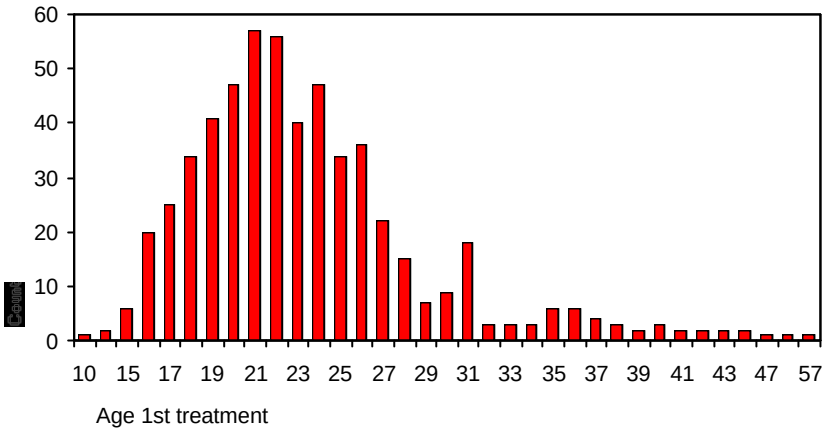
On the other hand about 50% of the study population were between 17 and 23 years when first requesting treatment. See **Table 1.4** and **Figure 1.4** below.

Table 1.4

Age at 1st treatment

Age	Frequency	Percent	Cumulative Percent
10	1	0,18	0,18
14	2	0,36	0,53
15	6	1,07	1,60
16	20	3,57	5,17
17	25	4,46	9,63
18	34	6,06	15,69
19	41	7,31	22,99
20	47	8,38	31,37
21	57	10,16	41,53
22	56	9,98	51,52
23	40	7,13	58,65
24	47	8,38	67,02
25	34	6,06	73,08
26	36	6,42	79,50
27	22	3,92	83,42
28	15	2,67	86,10
29	7	1,25	87,34
30	9	1,60	88,95
31	18	3,21	92,16
32	3	0,53	92,69
33	3	0,53	93,23
34	3	0,53	93,76
35	6	1,07	94,83
36	6	1,07	95,90
37	4	0,71	96,61
38	3	0,53	97,15
39	2	0,36	97,50
40	3	0,53	98,04
41	2	0,36	98,40
42	2	0,36	98,75
43	2	0,36	99,11
44	2	0,36	99,47
47	1	0,18	99,64
52	1	0,18	99,82
57	1	0,18	100,00
	561	100,00	

Figure 1.4



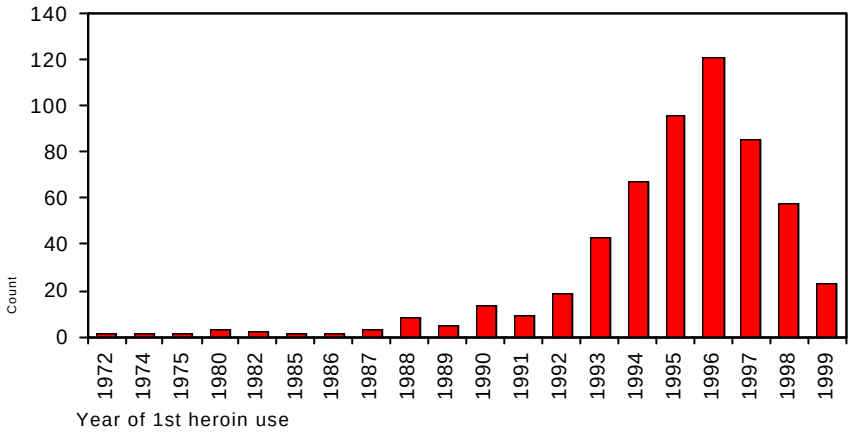
As it's been shown on **Table 1.5** if to take calendar years from 1991 to 1999 than we have seen ab. 90% of the whole population. As demonstrated on Figure 1.5 the most frequented calendar years are from 1994 – 1997, 70% of the treated heroin population could be placed in this interval.

Table 1.5

Year of 1st heroin use

Year	Frequency	Percent	Cumulative Percent
1972	1	0,18	0,18
1974	1	0,18	0,36
1975	1	0,18	0,53
1980	3	0,53	1,07
1982	2	0,36	1,43
1985	1	0,18	1,60
1986	1	0,18	1,78
1987	3	0,53	2,32
1988	8	1,43	3,74
1989	5	0,89	4,63
1990	13	2,32	6,95
1991	9	1,60	8,56
1992	19	3,39	11,94
1993	43	7,66	19,61
1994	67	11,94	31,55
1995	96	17,11	48,66
1996	121	21,57	70,23
1997	86	15,33	85,56
1998	58	10,34	95,90
1999	23	4,10	100,00
Total	561	100,00	

Figure 1.5



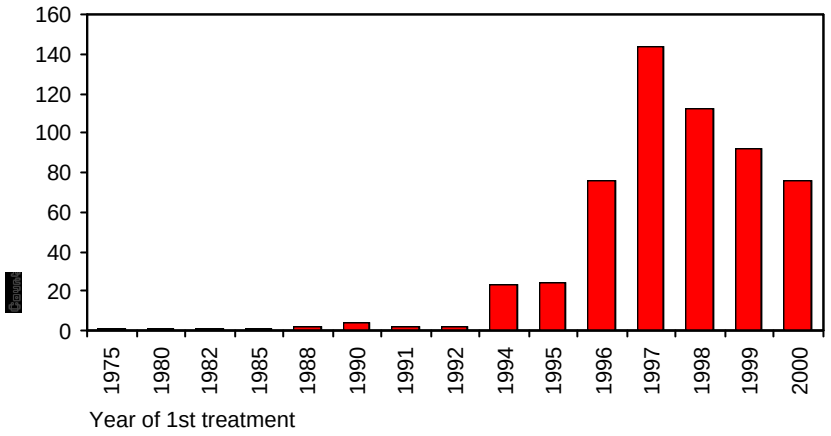
In 90% of the cases in our study the first treatment demand occurred in the period 1996-1999 (see **Table 1.6** and **Figure 1.6)**
That is why we believe that these were the years - meanly - to be considered in Kaplan-Meier as well as for the Cox analysis.

Table 1.6

Year of 1st treatment

Year	Frequency	Percent	Cumulative Percent
1975	1	0,18	0,18
1980	1	0,18	0,36
1982	1	0,18	0,53
1985	1	0,18	0,71
1988	2	0,36	1,07
1990	4	0,71	1,78
1991	2	0,36	2,14
1992	2	0,36	2,50
1993	7	1,25	3,74
1994	16	2,85	6,60
1995	24	4,28	10,87
1996	76	13,55	24,42
1997	144	25,67	50,09
1998	112	19,96	70,05
1999	92	16,40	86,45
2000	76	13,55	100,00
Total	561	100,00	

Figure 1.6



The latency time (or period) - according to definition - is the time, in years, between - when in the patient's carrier - the 1st heroin use, and further on - the 1st treatment demand have happened.

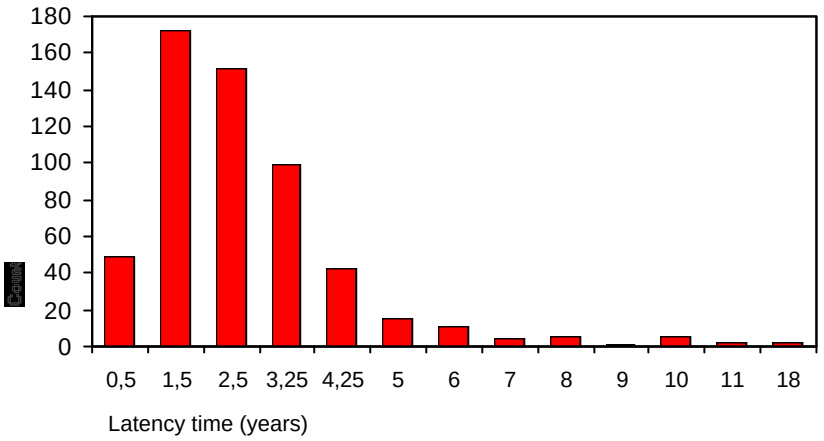
E.g. we observed that in 90% of the cases the latency period is less than or equals 5 years. The median latency time is at 2.5 years (see **Table 1.7** and **Figure 1.7**).

Table 1.7

Latency time (years)

	Frequency	Percent	Cumulative Percent
0,5	49	8,73	8,73
1,5	172	30,66	39,39
2,5	152	27,09	66,49
3,25	99	17,65	84,14
4,25	43	7,66	91,80
5	15	2,67	94,47
6	11	1,96	96,43
7	4	0,71	97,15
8	5	0,89	98,04
9	1	0,18	98,22
10	6	1,07	99,29
11	2	0,36	99,64
18	2	0,36	100,00
Total	561	100,00	

Figure 1.7



Incidence

If to examine distributions of the variables according to different levels of the exposure - as **Figures 2.1** and **2.2** show - distribution of the year of 1st treatment stratified by route of administration is the same for the two levels of the variable (for the injecting / non-injecting subgroups).

Figure 2.1

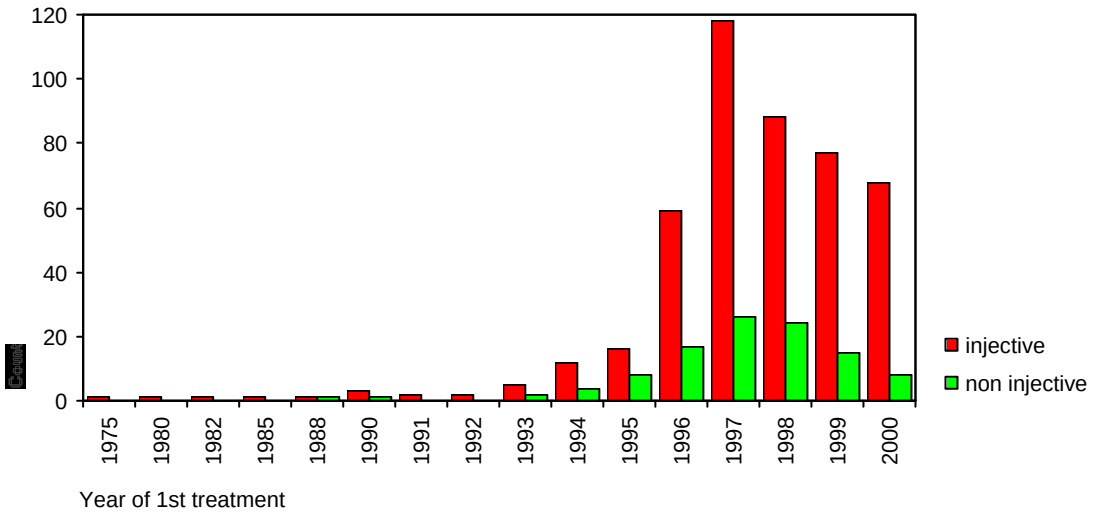
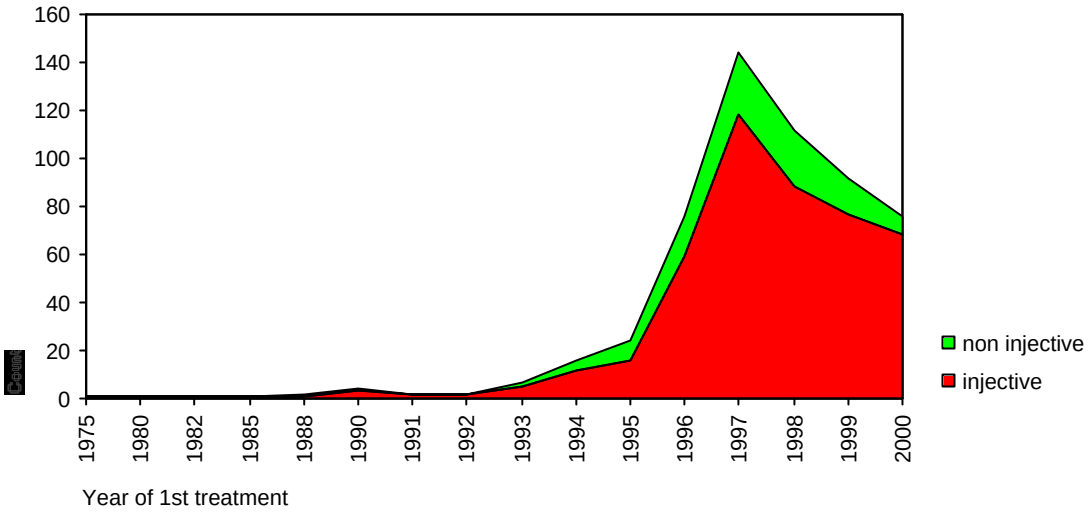


Figure 2.2



The situation is the same when stratifiing by gender. The year of 1st treatment shows the same distribution whether it's been shown for males or females **Figures 2.3** and **2.4.**

Figure 2.3

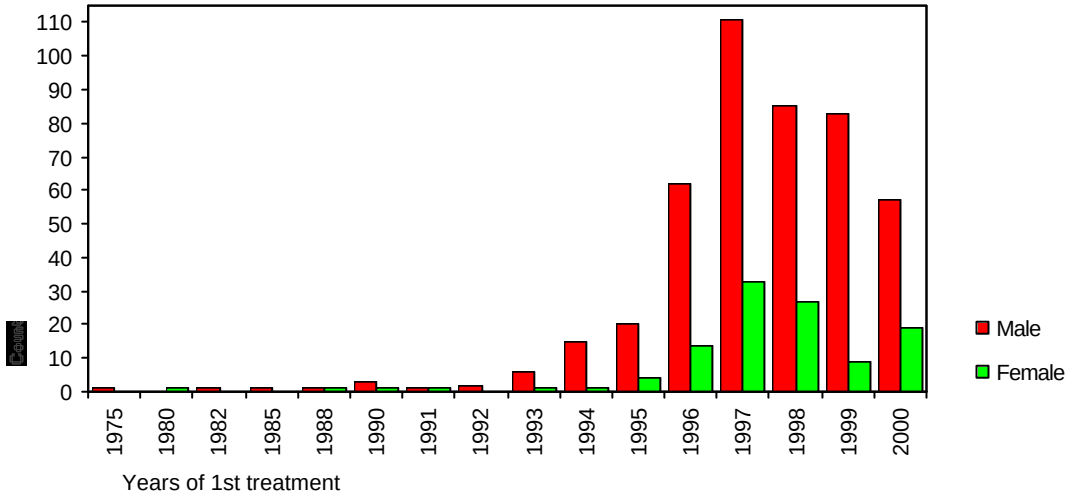


Figure 2.4

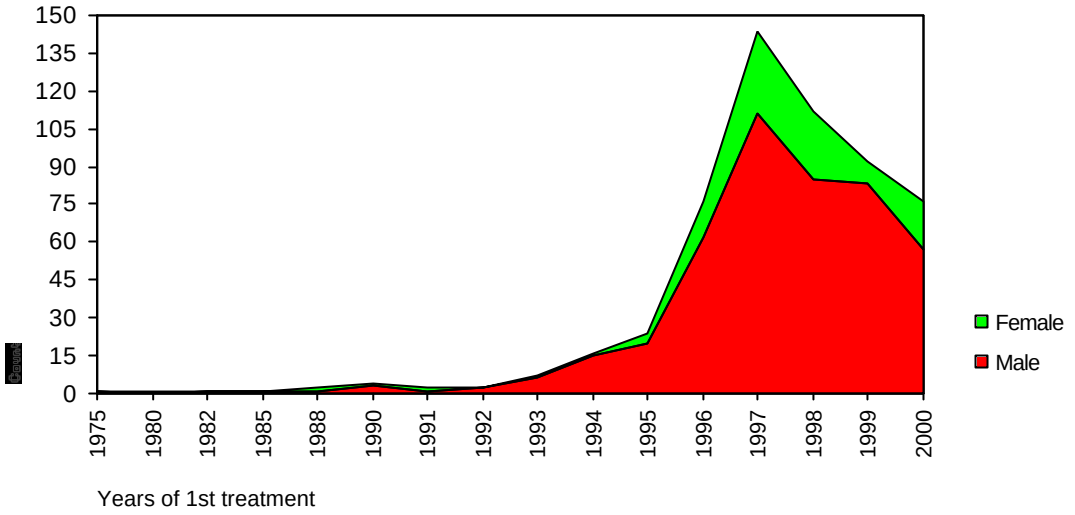
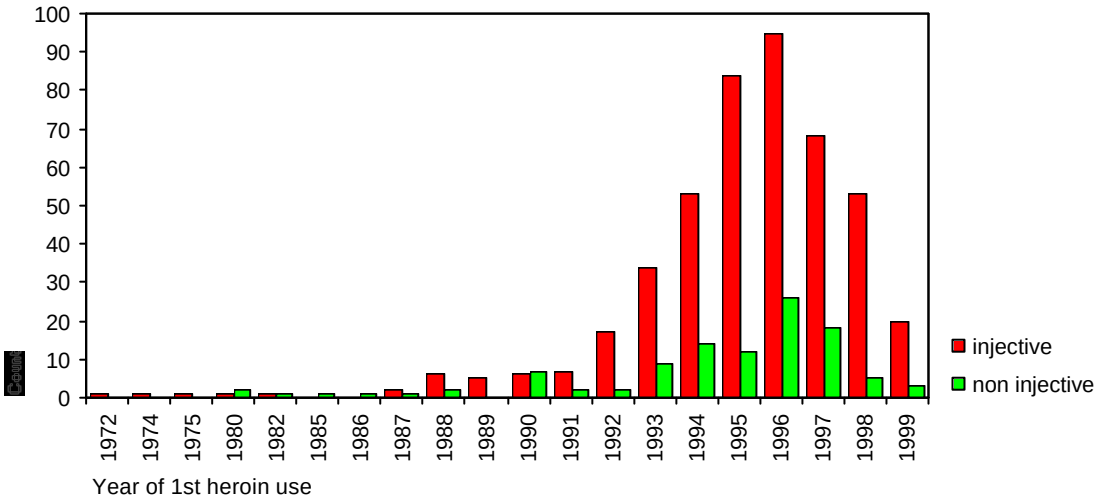
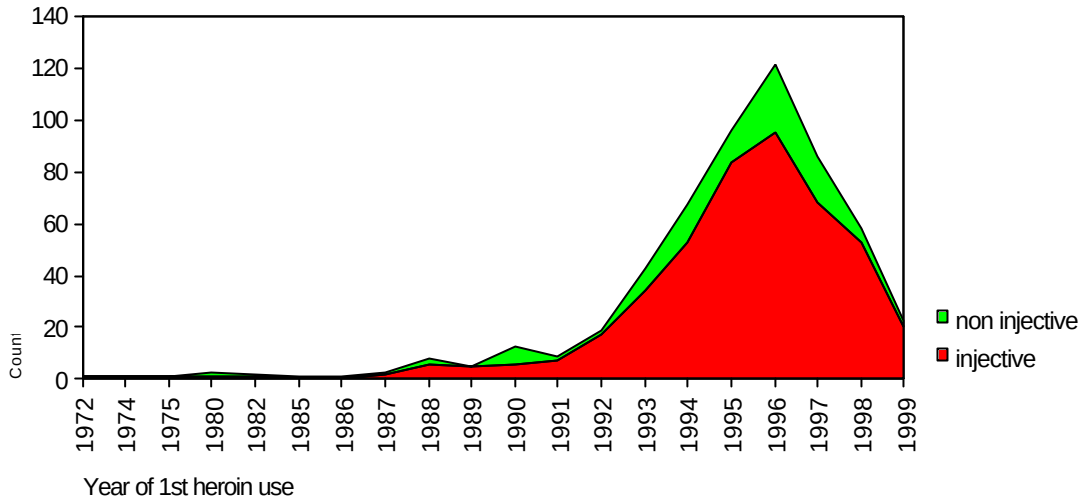


Figure 2.5



According to the area chart below (**Figure 2.5** and **2.6**) - the variable "Year of 1st treatment" shows "holes" for non-injectors, as if non-injectors started their heroin use later as compared to injectors.

Figure 2.6



However, there is no difference in the year of 1st heroin use for males vs females (**Figure 2.7**; **2.8**).

Figure 2.7

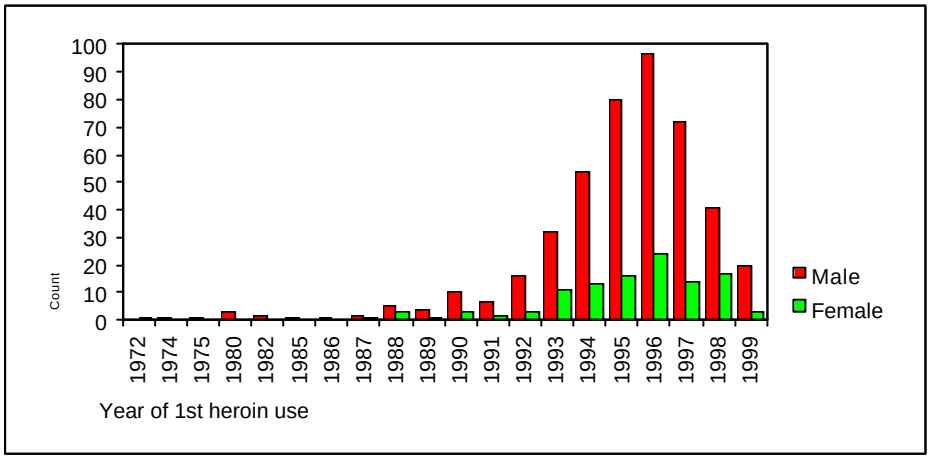
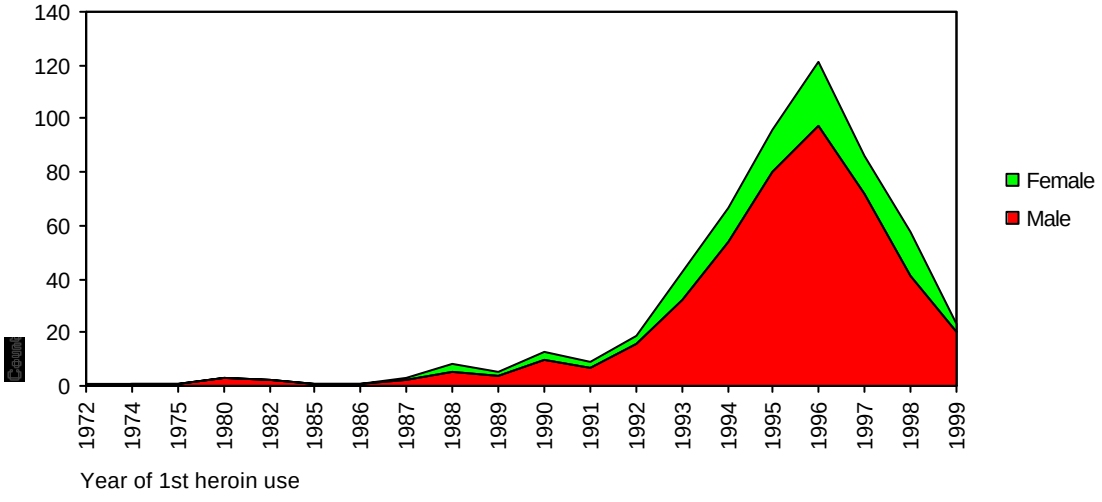


Figure 2.8



We estimated the latency period by Kaplan-Meier method, to see if there existed a difference, or there was no difference in the various subgroups. At the same time the Cox regression model was applied for the estimation of the latency period to assess the effects of different variables (covariates).

Kaplan-Meier And Cox Results

We considered as variables in the Kaplan-Meier and Cox analyses - gender, age at 1st heroin use, route of administration and year of 1st treatment. We did not consider other available information such as educational level because this kind of information has been referred to the year of 1st treatment, so it could be different for the moment of starting his/ her carrier.

Table 3.1

Kaplan-Meyer summaries for the latency period analysis

Group (size)	Mean	1° quart.	Median	3° quart.
Total (561)	2.35	1.00	2.00	3.00
Males (451)	2.35	1.00	2.00	3.00
Females (110)	2.39	1.00	2.00	3.00
Injectors (455)	2.25	1.00	2.00	3.00
Non injectors (106)	2.80	1.00	2.00	3.00
1993 (9)	2.94	3.00	3.00	3.00
1994 (17)	2.88	1.00	2.00	4.00
1995 (26)	2.17	1.00	2.00	3.00
1996 (85)	2.07	1.00	2.00	3.00
1997 (158)	2.24	1.00	2.00	3.00
1998 (119)	2.32	1.00	2.00	3.00
1999 (96)	2.54	1.25	2.00	3.00
2000 (90)	2.89	1.50	2.00	3.00

From the Kaplan-Meier analysis - the mean survival time, that is the time between the 1st heroin use and the 1st treatment demand, is 2.35 years, which is higher than the median survival time (2.00 years).
According to the means, males demand treatment earlier (0.04 years) than females (see **Table 3.1)** even if the difference is not quite marked.
As well as the non-injectors are to come in for treatment somewhat earlier (on average 0.55 years) than the injectors. This result could be explained by the fact that persons injecting heroin usually start by inhaling or sniffing it first.
As for the year of 1st treatment one can see a decreasing trend for the years 1993-1996 (2.94-2.07 years), while from 1997 there seems to be a rise in the mean survival time (2.24-2.89 years - see **Table 3.1**).
We showed survival function stratified by gender, route of administration and year of 1st treatment in **Figures 3.1-3.3**

Figure 3.1 Survival function of the variable “latency time” according to gender

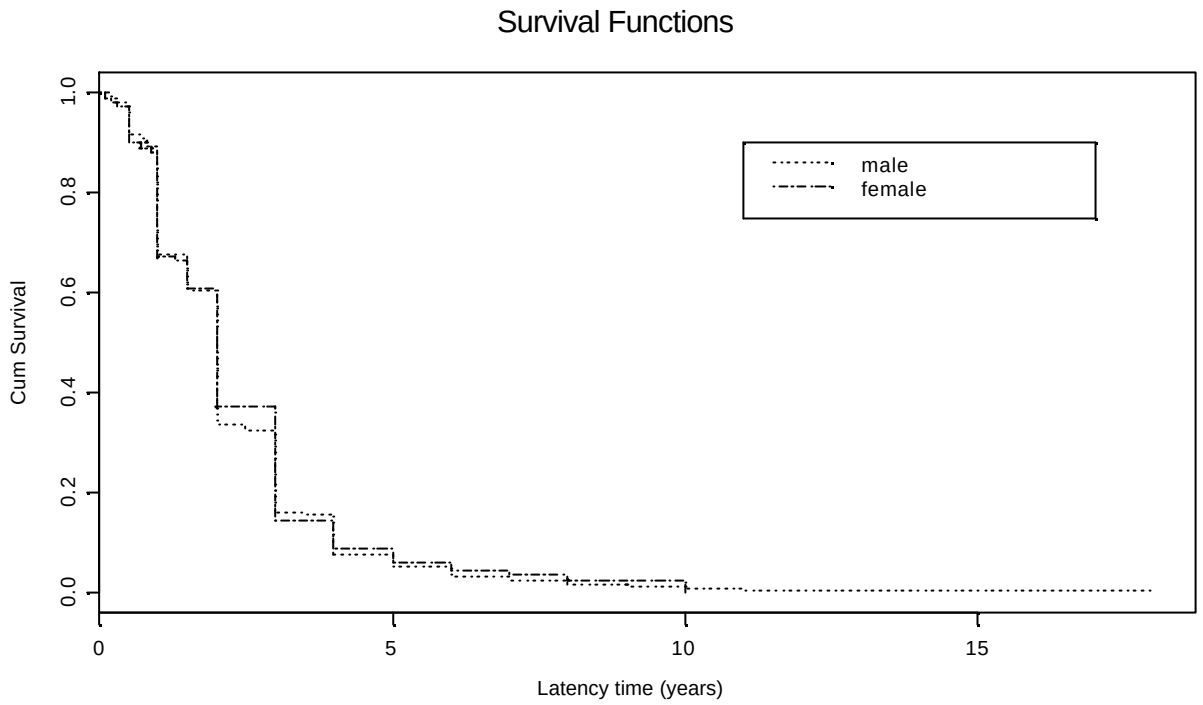


Figure 3.2 *Survival function of the variable “latency time” according to the route of administration*

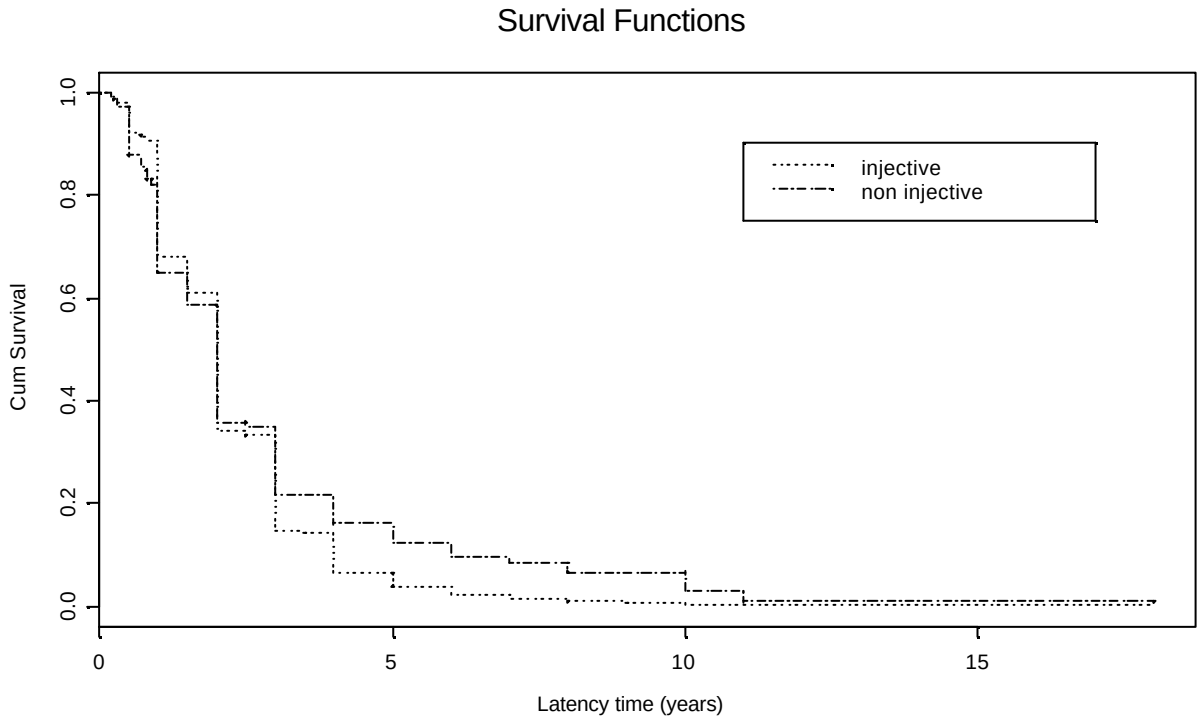
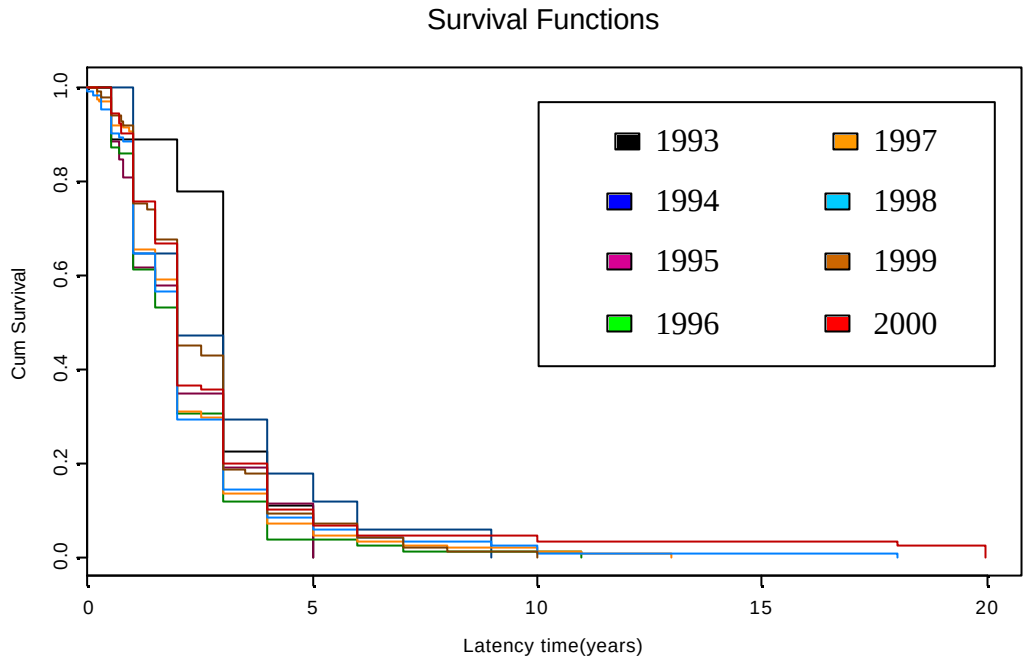
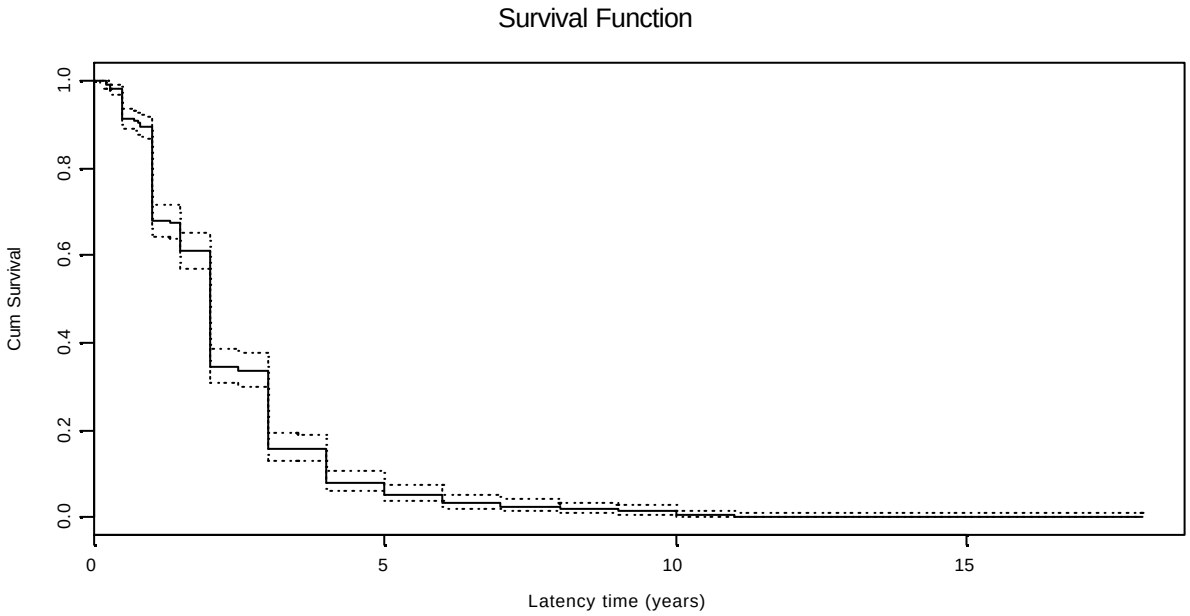


Figure 3.3 *Survival function of the variable "latency time" according to the year of 1st treatment*



The effects of all covariates (variables) on the latency time were estimated by Cox regression method. The only significant variable seems to be the route of administration (0.047 < 0.05). The estimate of the relative risk of 1st treatment for two patients in the study differing in this covariate (injector / non-injector) is exp($\hat{b}$) = 0.797. This result indicates that a non-injector has an estimated 0.797 times lower risk than the injector.
The estimate of survival function of the variable "latency time" given by the Cox model is shown in Figure 3.4.

Figure 3.4 *Cox Regression estimate of Survival function of the variable "latency time"*.



Appendix 2

CLASSIFICATION OF "CURED" INDIVIDUALS IN SURVIVAL ANALYSIS: THE MIXTURE APPROACH TO THE DIAGNOSTIC-PROGNOSTIC PROBLEM[2]

Marta Morbiducci[(1)], Alessandra Nardi[(2)], Carla Rossi[(1)]

[(1)]Department of Mathematics,
University of Rome "Tor Vergata"
Via Ricerca Scientifica 1, 00133 Rome, Italy

[(2)] Department of Systems Theory
University of Teramo
Viale Crucioli 122, I-64100 Teramo, Italy

Abstract

Various mixture survival models have been presented in the literature since 1982 (Farewell, 1982). Such models have been widely applied in various frameworks due to their flexibility in mirroring complex situations, even though a general theory has not yet been fully developed .
In the present contribution a general 'coherent' setting is outlined for survival models dealing with non homogeneous populations, with a special focus on cure-rate models and some exploratory and diagnostic tools are studied. The class of cure-rate models proposed here is a generalization of the previous one presented in Schinaia and Rossi (2000) to address the diagnostic-prognostic problem, now allowing for unknown prior probability of long survivorship.
The classification (diagnostic) problem is developed using the bayesian paradigm. The parameter estimation procedure is based on the development of a class of GEM algorithms, in the framework of the MLE approach. The model diagnostic is based on the asymptotic properties of the survival and risk functions and of the censoring process and on the generalization and application of the theory of log-odds residuals (Nardi and Schemper, 1999).
The results of an extended experimentation by means of simulated data is also reported in order to present the main features of the model and the methods and procedures proposed.

Keywords: survival analysis, latent variables, mixture models, cure-rate model, EM algorithm, log-odds residuals.

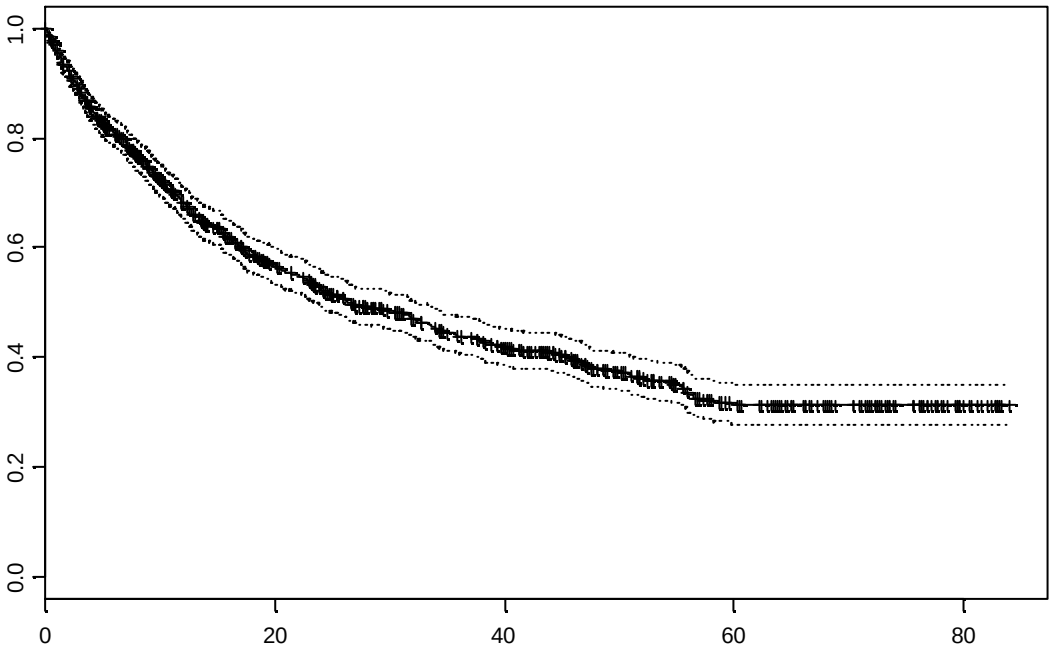
[2] The work contributes to the EU funded European Network to develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions (TSER/DGXII project ERB 4141 PL980030).

1. Introduction.

In some survival studies, specifically in clinical trials, a proportion of patients who respond to treatment appear subsequently to be free of signs of the disease for a long time being possibly cured. As a consequence, long term censored times usually appear in observed follow up data and it can be shown that, if the population of interest comprises cured individuals the proportion of censored times, higher than a specified time τ, tends to 1 for τ increasing (see appendix). With the purpose of estimating the proportion of cured patients and the survival function of uncured patients, possibly depending on observable covariates with distribution functions to be estimated as well, suitable mixture survival models can be used. They generally are referred to as cure-rate models. These models have been extensively studied and applied by various authors, including, among others, Farewell (1982, 1986), Sposto, Sather and Baker (1992), Cantor and Shuster, 1992, Brambilla et al. (1990), Brambilla, Rossi and Schinaia (1998), Ghitany, Maller and Zhou (1994), Maller and Zhou (1992, 1994, 1995, 1996) Peng and Dear, 2000, Schinaia and Rossi, 2000, Chen and Ibrahim (2001), Ibrahim, Chen and Sinha (2001). The possible fields of applications range from engineering and economics to bio-medicine and social sciences. The main characteristic of these models is the possibility to incorporate diagnostic information, aimed at classification of "cured" individuals, in a suitable multivariate survival model.

A typical application is in the treatment of tumours, where patients may or may not respond to therapeutic agents: response can, however, only be recognized after long follow-up periods. It is quite common in these cases, when the survival function is estimated, to obtain a plot with a qualitative behaviour such as that reported in Figure 1, where a plateau occurs at a level of approximately 35%, suggesting that this fraction of patients is cured.

Figure 1. Typical survival curve in a cure-rate model.



Medical expertise is usually employed in order to determine what "sufficient follow up" means in this framework and a qualitative assessment is used to incorporate information on observable covariates which can influence the expected cure-rate for a particular individual.
A suitable mixture model can improve this situation by a more quantitative approach and help in obtaining early classification of cured individuals on the basis both of observable covariates and of time elapsed "disease free" since the beginning of the follow up.
In the following we present a general mixture model. The diagnostic problem, i.e. the classification of individuals into diagnostic classes, is addressed within a bayesian framework, whereas the Maximum Likelihood Estimation approach is chosen to estimate the survival and probability distribution functions, because of the usually "large" sample size available with respect to the "vague" prior information.
The parameter estimation procedure is based on the development and implementation of a

The a priori and a posteriori diagnostic is based on the asymptotic properties of the survival and risk functions and of the censoring process and on the generalization and application of the theory of log-odds residuals (Nardi and Schemper, 1999).
class of GEM algorithms.
In Section 2 the general model for the survival and hazard function is presented and discussed in the case of interest. Section 3 describes the EM procedure providing estimates for the relevant multivariate functions (prior probabilities, survival function, pdf's of covariates).The diagnostic through log-odds residuals is presented in Section 4, numerical applications follow in Section 5 and the conclusions are in Section 6.

2. The general setting of a cure-rate model.

Let T be a nonnegative random variable denoting the failure time of interest and denote by $\tilde{X} = [X, Z]$ the set of covariates, which typically includes the patients classification Z into K classes of different survival experiences (diagnostic class information) and the set of observable covariates X. Then S(t/X=x,Z=i) and f(t/X=x,Z=i) represent respectively the survival function and the probability density (possibly generalized) function of T, conditional on the observed values of the vector covariate $\tilde{X} = [X, Z]$. If the diagnostic class cannot be observed, then Z is a latent random variable with a finite number of possible determinations $i = 1, \ldots, K$.

2.1. The general discrete mixture model for survival data.

In terms of survival function $S(t/X=x)$, the prognostic problem (Schinaia and Rossi, 2000) can be expressed as a linear convex combination (mixture) of survival functions $S(t \mid X = x, Z = i)$ $(i = 1, \ldots, K)$ for each of the K diagnostic classes, corresponding to the various determinations of Z:

$$S(t \mid X = x) = \sum_{i=1}^{K} S(t \mid X = x, Z = i) P(Z = i \mid X = x) = \\ = \frac{\sum_{i=1}^{K} S(t \mid X = x, Z = i)\, f_X(x \mid Z = i) P(Z = i)}{\sum_{i=1}^{K} f_X(x \mid Z = i) P(Z = i)} \qquad (1)$$

the corresponding hazard function being:

$$r(t \mid X=x)=\frac{\sum_{i=1}^{K} r(t \mid X=x, Z=i) S(t \mid X=x, Z=i) P(Z=i \mid X=x)}{\sum_{i=1}^{K} S(t \mid X=x, Z=i) P(Z=i \mid X=x)}= \tag{2}$$

$$=\sum_{i=1}^{K} r(t / X=x, Z=i) P(Z=i / X=x, T \geq t)$$

where $f_X(x \mid Z=i)$ is the generalized density function of X in the class $Z=i$.
The survival function $S(t/X=x)$, representing the "global" survival model for the population of interest, is thus expressed as a mixture of the survival functions of each diagnostic class $S(t \mid X=x, Z=i)$, using the *prior* distribution of Z as mixing distribution, whereas the "local" model, represented by the hazard function $r(t/X=x)$, is the combination of the class hazard functions $r(t \mid X=x, Z=i)$, using the *posterior* distribution of Z as mixing distribution.
From (1) and (2) it is seen that a correct estimation procedure of the survival and hazard function involves the estimation of a number of other elements, such as the pdf's of the covariates within each class identified by Z.

2.2. The general mixture cure-rate model.

The cure-rate model is obtained from the general formulation (1) and (2) by allowing one of the classes, say Z=1, to comprise only cured individuals who are not at risk of experimenting the event of interest, whereas no cured individuals belong to any other class. The failure times of cured individuals can be conveniently defined as infinite. This implies that:

$$S(t/X=x,Z=1)\equiv 1 \text{ and } r(t/X=x,Z=1)\equiv 0 \quad \forall x$$

whereas, $\forall i>1$, $\lim_{t\to\infty} S(t/X=x,Z=i)=0$. This also implies that the survival function $S_0(t/X=x)$, defined as the mixture of the survival functions of the classes Z=i, i=2,3,…,K satisfies the two general properties $S_0(0/X=x)=1$ and $\lim_{t\to\infty} S_0(t/X=x)=0 \quad \forall x$. This function can, thus, model any "usual" situation, either homogeneous (K=2) or not (K>2), whenever no cured individuals are present.
Thus, the study can be restricted, without loss of generality, to the model with only two diagnostic classes, identified by the latent variable:

$$Z=\begin{cases} 0 & \Leftrightarrow \text{ non respondent} \\ 1 & \Leftrightarrow \text{ respondent} \end{cases}$$

and the general formulas (1) and (2) take the form:

$$S(t \mid X=x)=S(t \mid X=x, Z=0)P(Z=0 \mid X=x)+P(Z=1 \mid X=x)=p_0(x)S_0(t/X=x)+p_1(x)$$

$$r(t \mid X=x)=r(t \mid X=x, Z=0)P(Z=0 \mid X=x, T>t)=r_0(t/X=x)p_0(x,t)$$

where:

$$p_1(x)=P(Z=1 \mid X=x)=\frac{f_X(x \mid Z=1)p_1}{f_X(x \mid Z=1)p_1+f_X(x \mid Z=0)p_0}, \quad p_0(x)=1-p_1(x);$$

$$p_0(x,t)=P(Z=0 \mid T>t, X=x)=\frac{S(t/X=x,Z=0)\, f_X(x \mid Z=0)p_0}{f_X(x \mid Z=1)p_1+S(t \mid X=x,Z=0) f_X(x \mid Z=0)p_0}$$

and $p_1=1-p_0=P(Z=1)$ is the prior probability of the class Z=1.
It follows immediately that S(t/X=x) is a monotone decreasing function such that:

$$S(t/X=x) \geq p_1(x), \lim_{t\to\infty} S(t/X=x) = p_1(x) \quad \text{and} \quad \lim_{t\to\infty} r(t/X=x) = 0 \ \forall x$$

Thus the level of the plateau in the estimated survival curve, if the period of follow up is long enough, can be used as a naïve or preliminary estimate of $p_1(x)$ and, if it is greater than zero, provides evidence of the presence of long survivors in the population (a priori diagnostic). Similarly, as an a priori diagnostic tool, it can be also used the fact that the expected proportion of censored times, higher than a specified time τ tends to 1 for τ increasing. It must be observed that $\lim_{t\to\infty} r(t/X=x) = 0$ implies that the risk function is either decreasing or eventually decreasing.
Inference on Z can be carried out using information provided by X and incorporating the additional information given by the event $E_t=\{T>t\}$, by considering data for early censored individuals as incomplete and using a suitable form of the EM algorithm to numerically obtain the Maximum Likelihood Estimates of the unknown parameters appearing in the various functions. It must be observed that the approach is completely general and allows using parametric or non parametric models, as it is shown in the previous papers where special cases are considered: in Brambilla et al. (1990), in fact, S(t) was estimated using the Life Table approach and no observable covariate was included in the model, whereas in Schinaia and Rossi (2000) a parametric model was used for S(t/X=x).

2.3. The bayesian classification approach: the diagnostic values of t and X.

Our aim is to separate cured individuals from the individuals at risk (see also Brambilla et al. 1990) and to estimate the various distributions appearing in the model.
Let us first consider the classification problem, which is addressed using a bayesian approach. Any individual, who has survived until time t, has a posterior probability:

$$P(Z=1 \mid T>t, X=x) = \frac{f_X(x \mid Z=1)p_1}{f_X(x \mid Z=1)p_1 + S(t \mid X=x, Z=0) f_X(x \mid Z=0)p_0} \qquad (3)$$

to be a respondent and the evaluation of (3) involves the estimation of $S(t \mid X=x, Z=0)$, $f_X(x \mid Z=0)$ and $f_X(x \mid Z=1)$. However, being Z not always observable, with no further information available, the classification of early censored individuals is not straightforward. In fact he can be classified either as a censored non respondent individual or as a respondent who is not at risk of failure. This fact does not allow to separately estimate the functions of interest. The problem of incompleteness concerning the estimation of such functions appearing in (3), besides the usual censoring, is thus due to the non observability of the diagnostic class Z, and the EM scheme can be properly applied to solve the problem, as it is shown in Section 3. It is very important to observe that a special diagnostic covariate (Brambilla, Rossi and Schinaia, 1998) is the value of the censored times. Generally, this covariate has a high impact in term of identification of cured individuals mirroring the actual decision process based on expertise, as mentioned in Section 1. It is easily seen from (3), in fact, that, for any value x of X, P(Z=1/T>t,X=x) is an increasing function of t converging to 1 as t goes to infinity, the convergence being modulated mostly by the distributions of the observable vector covariate X in the two diagnostic classes (Schinaia and Rossi, 2000). This result can be better analysed by studying the behaviour, with respect to t, of the posterior odd ratio:

$$O_{01}(t,x)=\frac{P\left(Z=0\middle|T\geq t,X=x\right)}{P\left(Z=1\middle|T\geq t,X=x\right)}=S(t/X=x,Z=0)\frac{f_X(x/Z=0)}{f_X(x/Z=1)}\frac{p_0}{p_1}$$

In the following section the estimation and diagnostic (a posteriori) problems are addressed and some applications are presented.

3. The EM estimation algorithm.

In the previous section, we have seen that the global survival function is a mixture of the survival function in the class Z=0 and of that in the class Z=1. We have also assumed that the respondents have a non-proper survival function $S(t|\ X=x,\ Z=1) \equiv 1$. We can now consider the complete log-likelihood for the mixture model proposed:

$$\cdot\left\{\ [S(t_i\mid X=x_i,Z=0)P(X=x_i\mid Z=0)\,p_0\,]^{1-z_i}\cdot[P(X=x_i\mid Z=1)p_1]^{z_i}\right\}^{1-d_i}\ )\quad(4)$$

where d_i is a censoring indicator that is equal to zero if t_i is censored and to one otherwise. It

$$\ln L=\ln(\prod_{i=1}^{n}[\,f(t_i\mid X=x_i,Z=0)P(X=x_i\mid Z=0)\,p_0\,]^{d_i}\cdot$$

must be observed that the contribution of a censored individual to the log-likelihood function takes into account the possibility for such individual of being a censored respondent or a censored non-respondent.
We can also observe that if $d_i = 1$, i.e. if the i-th individual has experimented the event of interest, then $z_i = 0$, i.e. the i-th individual is classified as non respondent.
Separating the components depending on the survival time, on the distribution of the covariates and on the prior probabilities p_0 and p_1, we obtain the following three expressions:

$$\ln L_X(j\mid x,z)=\sum_{i=1}^{n}(1-z_i)\ln P(X=x_i\mid Z=0,j)+\sum_{i=1}^{n}z_i\ln P(X=x_i\mid Z=1,j)$$

$$\ln L_T(q\mid t,x,Z=0)=\sum_{i=1}^{n}d_i\ln f(t_i\mid X=x_i,Z=0)+\sum_{i=1}^{n}(1-d_i)(1-z_i)\ln S(t_i\mid X=x_i,Z=0)=$$

$$=\sum_{i:d_i=1}\ln f(t_i\mid X=x_i,Z=0,q)+\sum_{i:d_i=0}(1-z_i)\ln S(t_i\mid X=x_i,Z=0,q)$$

$$\ln L_Z(y\mid z)=\sum_{i=1}^{n}(1-z_i)\ln p_0(y)+\sum_{i=1}^{n}z_i\ln p_1(y)$$

where φ and θ denote the parameters of the distribution functions of covariates and of survival times respectively, while ψ is an artificial parameter as the "real" parameter on which the third expression depends is p_0 itself.
On the basis of the log-likelihood function, we can define the steps of the EM algorithm[3].

[3] The convergence is implied by the fact that the loglikelihood function is limited, as can be easily verified.

The E-step calculates the expectation of the three expressions reported above. These expressions must then be maximized in the M-step and each of them depend on the posterior probability of being respondent, i.e. on $E(z_i)$:

$$E(z_i) = P(Z=1 \mid T_i > t_i, X = x_i) = \frac{f_X(x_i \mid Z=1)p_1}{f_X(x_i \mid Z=1)p_1 + S(t_i \mid X = x_i, Z=0)f_X(x_i \mid Z=0)p_0}$$

Therefore, the E-step simply calculates the current value of $E(z_i)$, whereas the M-step maximizes the first, the second and the third expression with respect to φ, to θ and to p_0 respectively.
The initial estimates for $S(\cdot)$, $f(\cdot)$, $f_X(\cdot|Z=0)$ and $f_X(\cdot|Z=1)$ can be obtained from observed data by initially classifying as respondent all censored individuals and as non respondent all those who have died. The EM procedure has been implemented using S-plus; an S-plus macro is available on request.

4. Log-odds residuals.

Log-odds and normal deviate residuals have been introduced in 1999 (Nardi and Schemper) for the purpose of screening for outliers. We will focus our attention on log-odds residuals which are defined as:

$$l = \log \frac{\hat{S}(t \mid X = x)}{1 - \hat{S}(t \mid X = x)}$$

where $\hat{S}(t \mid X = x)$ denotes the estimated survival function for a single individual evaluated at his observed failure time. They are 0 if the observed failure time coincides with the estimated median failure time, i.e. a median calibration is assumed. Increasing departures from the predicted median time are reflected by increasing absolute values of log-odds residuals that can be intuitively interpreted as a "distance" between the observed failure time and the median failure time estimated by the fitted model. Large negative residuals identify too long survival times while large positive values correspond to very early events. Assuming a correctly specified model, their distribution converges in probability to a standard logistic distribution. This results is based on the idea that, assuming the survival function as known, $U_i = S(T_i) \quad i = 1,...,n$ represent a set of n independent random variables, each having [0,1] uniform distribution. When the unknown survival function is replaced by its estimator, the convergence to the reference logistic distribution can be proved, provided that the estimator is consistent. In particular it holds true both for Cox's model and for the class of accelerated failure times models.

Comparison of the empirical distribution of log-odds residuals with the reference logistic density is of limited value in Cox's model where the non-parametric approach to the baseline hazard tends to guarantee a global goodness of fit and important cases of misspecification might be missed (Crowley and Storer, 1983). Conversely a graphical inspection of residuals, on the ground of their distributional properties, may be very informative for parametric models where the choice of a family of distributions for the baseline survival time should be verified. In particular such an investigation may lead to the conclusion that none of the standard parametric models fits satisfactorily the observed data, providing evidence in favour of a mixture model (Nardi and Schemper, forthcoming).

The definition of log-odds residuals for censored times deserves some comments. Censoring affects the empirical distribution of residuals which departs from the reference standard logistic density even under a correctly specified model (Nardi, 2000). The left tail of the distribution, which corresponds to long survival times, is cut and residuals are concentrated around central values. In order to remove the confounding effect of censoring we propose to randomly sample from the residuals' conditional distributions and to proceed in the spirit of Rubin's multiple imputation (Rubin, 1987, Little and Rubin, 1987). Assume that individual i is censored at c_i. Then, we randomly generate m log-odds residuals from the reference logistic density given that $T_i > c_i$, m being the number of imputations. Each of the imputed residuals for individual i is weighted 1/m. It is worth to remark that the imputation is done under the null hypothesis of a correctly specified model. This may lead to a conservative behaviour in assessing departures from model assumptions when the percentage of censoring is high. In this case an artificial agreement between the empirical and the reference distribution on the left tail simply means that sampling information is not sufficient the detect any departure from the fitted model for long survival times.

The extension of log-odds residuals to mixture models requires some adjustments. We will restrict our discussion to the model with only two diagnostic classes introduced in Section 2.2 . In this case the survival function can be written as:

$$S(t \mid X = x) = p_0(x) S_0(t \mid X = x) + p_1(x)$$

As $\lim_{t \to \infty} S(t / X = x) = p_1(x)$, U_i is no more uniformly distributed in [0, 1]. The distribution function of U_i is in the present situation:

$$F_U(u \mid x) = \begin{cases} 0 & u < p_1(x) \\ u & p_1(x) \le u < 1 \\ 1 & u \ge 1 \end{cases}$$

that is an improper uniform distribution in $[p_1(x), 1]$. When we move on the scale of log-odds residuals we obtain a logistic density with the left tail truncated in $l_0 = \log(p_1(x)/(1 - p_1(x)))$, where the support of this density depends on the vector of explanatory variables. Actually, when the probability of being cured depends on x, the distribution of log-odds residuals is a mixture of truncated logistic densities with mixing distribution given by the distribution of the vector of covariates X in the total sample:

$$f(l) = \int_{\mathfrak{X}} f(l / x) dH(x) \quad [l_0 \le l \le +\infty]$$

where:

$$f(l / x) = \frac{1}{[1 - p_1(x)]} \frac{\exp(l)}{[1 - \exp(l)]^2} \quad [l_0 \le l \le +\infty]$$

has been normalised.

Being a mixture of unimodal distributions having mode in 0, under a properly specified model we still expect a unimodal distribution.

5. Applications.

In order to show the main features of the model and of the algorithm, we have applied it to simulated data. The data were simulated to mirror the experimental situation of the Greater London HIV/AIDS application (Schinaia and Rossi, 2000). Thus we considered the various distributions estimated for that application as the basic statistical models to generate the data needed for the present experimentation. The covariates involved in that model were just two: a continuous one, the age at diagnosis with a Gamma distribution in both classes, and a discrete one, a categorical variable (six categories) representing the type of diagnosis. The survival function in the class Z=0 was a Weibull model.

We have performed various simulation studies, generating samples with 1000, 500 and 250 units and generating, for each of them, various samples by modifying the value of the prior probability of being respondent (p_1) from 5% to 50% (Table 1). We have used the Weibull distribution to generate survival times for non respondents, whereas, for long-survivors, we have generated uniformly distributed times in [0,96]. The values for the covariate "Age at diagnosis" were obtained generating from a Gamma distribution for each class. The values for the discrete variable were generated according to the probability distribution obtained in Rossi and Schinaia (2000) for each class (the parameters of the various distributions are given in the tables reported below). We have also run the EM algorithm with different values of the parameters in order to test the procedure under different conditions. The parameters have been modified one by one to better control the results.

With respect to parameter estimates, we have obtained almost the same results obtained by Schinaia and Rossi (2000), considering p_0 as a known constant. The new procedure provides better estimates of the Weibull parameters and better classification of the long-survivors, despite p_0 is unknown. The estimates obtained for p_1 are always very satisfactory (Table 1). In some cases the procedure by Schinaia and Rossi provides better estimates for the Gamma distribution in the class Z=1.

Table 1. Point estimates of p_1 obtained in the various simulation experiments.

p_1	**Sample size**		
	1000	*500*	*250*
0.05	*0.05*	*0.04*	*0.04*
0.10	*0.09*	*0.08*	*0.11*
0.15	*0.12*	*0.15*	*0.13*
0.20	*0.20*	*0.18*	*0.15*
0.25	*0.21*	*0.21*	*0.23*
0.30	*0.28*	*0.23*	*0.27*
0.35	*0.32*	*0.31*	*0.30*
0.40	*0.38*	*0.37*	*0.39*
0.45	*0.44*	*0.46*	*0.41*
0.50	*0.51*	*0.47*	*0.46*

The parameters of the Weibull distribution are well estimated by the algorithm for any sample size, as well as the probability of the various diagnosis in the class Z=0. The situation is slightly different for the estimates of such probabilities for long-survivors. A limited sample size, in fact, may lead to unstable estimate for the probability of rare diagnoses, such as Diagnosis 6 in Table 10.

In the following tables some summary information is reported. The worst results always correspond to sample size of 250 units, whereas the best to sample size of 1000 units.

Table 2. Summary results of the simulation runs concerning the parameter estimates for the Weibull distribution.

Parameters	*l*	b_0	b_1	b_2
True values	*2.1*	*-14.71*	*0.195*	*0.43*
Worst estimate	*2.20*	*-10.85*	*0.189*	*0.41*
Best estimate	*2.13*	*-14.84*	*0.201*	*0.428*

Table 3. Summary results of the simulation runs concerning the parameter estimates for the Gamma distribution in the class Z=1.

Parameters	*l*	*g*
True values	*43*	*1.3*
Worst estimate	*33.83*	*1.00*
Best estimate	*43.8*	*1.31*

Table 4. Summary results of the simulation runs concerning the parameter estimates for the Gamma distribution in the class Z=0.

Parameters	*l*	*g*
True values	*56*	*1.43*
Worst estimate	*42.98*	*1.10*
Best estimate	*56.28*	*1.45*

Table 5. Summary results of the simulation runs concerning the parameter estimates for the distribution of the variable "Diagnosis" in the class Z=1.

Category **Probability**	*Diagnosis =1*	*Diagnosis= 2*	*Diagnosis= 3*	*Diagnosis =4*	*Diagnosis =5*	*Diagnosis =6*
True values	*0.76*	*0.19*	*0.02*	*0.02*	*0.004*	*0.006*
Worst estimate	*0.63*	*0.29*	*0.02*	*0.03*	*0.01*	*0.008*
Best estimate	*0.77*	*0.188*	*0.02*	*0.024*	*0.006*	*0.005*

Table 6. Summary results of the simulation runs concerning the parameter estimates for the distribution of the variable "Diagnosis" in the class Z=0.

Category **Probability**	*Diagnosis =1*	*Diagnosis =2*	*Diagnosis= 3*	*Diagnosis= 4*	*Diagnosis= 5*	*Diagnosis =6*
True values	*0.70*	*0.15*	*0.02*	*0.05*	*0.03*	*0.05*
Worst estimate	*0.63*	*0.18*	*0.04*	*0.09*	*0.01*	*0.06*
Best estimate	*0.704*	*0.154*	*0.020*	*0.055*	*0.028*	*0.053*

The general considerations reported above can be better enlightened by means of an extended example based on the results of the simulation run corresponding to a sample of 1000 individuals, 707 uncured (with 139 censored non respondents) and 293 long-survivors. The value of p_0 is 0.70. The following tables report the expected, the initial and the final values of the various parameters. The well known good convergence properties of the EM algorithm are evident.

Table 7. Weibull distribution.

Z=0	*g*	b_0	b_1	b_2
Expected	*1*	*-5.1*	*0.195*	*0.43*
Initial	*1.16*	*-2.09*	*0.126*	*0.33*
Final	*1.12*	*-4.3*	*0.161*	*0.38*

Table 8. Gamma distribution in the class Z=1.

Z=1	*a*	*l*
Expected	*43*	*1.3*
Initial	*36.43*	*1*
Final	*38.65*	*1.17*

Table 9. Gamma distribution in the class Z=0.

Z=0	*a*	*l*
Expected	*56*	*1.43*
Initial	*39.16*	*1.12*
Final	*53.12*	*1.37*

Table 10. Distribution of the variable "Diagnosis" in the class Z=1.

Z=1	*Diagnosis= 1*	*Diagnosis= 2*	*Diagnosis= 3*	*Diagnosis= 4*	*Diagnosis= 5*	*Diagnosis= 6*
Expected	*0.76*	*0.19*	*0.02*	*0.02*	*0.004*	*0.006*
Initial	*0.67*	*0.185*	*0.023*	*0.055*	*0.021*	*0.037*
Final	*0.71*	*0.196*	*0.021*	*0.03*	*0.003*	*0.03*

Table 11. Distribution of the variable "Diagnosis" in the class Z=0.

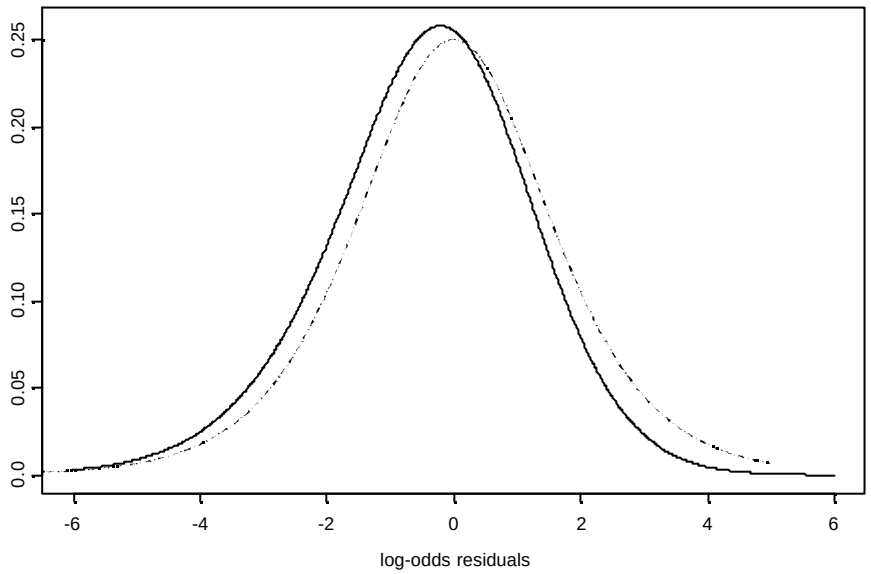
Z=0	*Diagnosis= 1*	*Diagnosis= 2*	*Diagnosis= 3*	*Diagnosis= 4*	*Diagnosis= 5*	*Diagnosis= 6*
Expected	*0.70*	*0.15*	*0.02*	*0.05*	*0.03*	*0.05*
Initial	*0.75*	*0.158*	*0.016*	*0.03*	*0.018*	*0.025*
Final	*0.72*	*0.160*	*0.017*	*0.040*	*0.022*	*0.042*

The initial and final values of p_0 are 0.568 and 0.68 respectively. The initial classification is obtained by considering all censored patients as long-survivors leading to 432 respondents at the first step. After the reconstruction, the estimated number of long-survivors is 276.

We can now show how the log-odds residuals have been used in the present framework on the basis of the example related to the simulation run reported above. At first we analysed their

behaviour when the statistical model is correctly specified and then we compared this behaviour to that obtained for two different cases of misspecification.
In Figure 2 the residuals' empirical distribution under a properly specified model is reported (smoothed density (continuous line) obtained by kernel method) together with the reference standard logistic distribution (dotted line). It appears that the empirical distribution, even if deformed as a mixture of truncated logistic distributions, still maintains the characteristic unimodality expected when the model is correctly specified.

Figure 2. Residuals' empirical distribution under a properly specified model: smoothed density (continuous line) obtained by kernel method and standard logistic distribution (dotted line).



This result can be compared with those obtained in two different cases of uncorrect model specification.
First we wrongly assumed an exponential distribution for Z=0 instead of a Weibull. The residuals' empirical distribution is shown in Figure 3. It appears that the observed distribution, is deformed due to the misspecification showing a clear bimodality.
A different case of misspecification is considered in Figure 4 where a simple Weibull model was fitted ignoring the fraction of cured individuals. It appears that the residuals' empirical distribution is deformed due to the misspecification showing again a clear bimodality.
In both cases the wrongly specified model cannot simultaneously describe short and long survivors and residuals keep the original bimodality of failure times.
According to these preliminary results log-odds residuals appear to be a useful diagnostic tool to detect various kind of misspecification related to cure-rate models. However, further analyses and experimentation are needed in order to better assess this possibility using both simulated and real data.

Figure 3. Residuals' empirical distribution under a misspecified model (exponential instead of Weibull in the class Z=0): smoothed density (continuous line) obtained by kernel method and standard logistic distribution (dotted line).

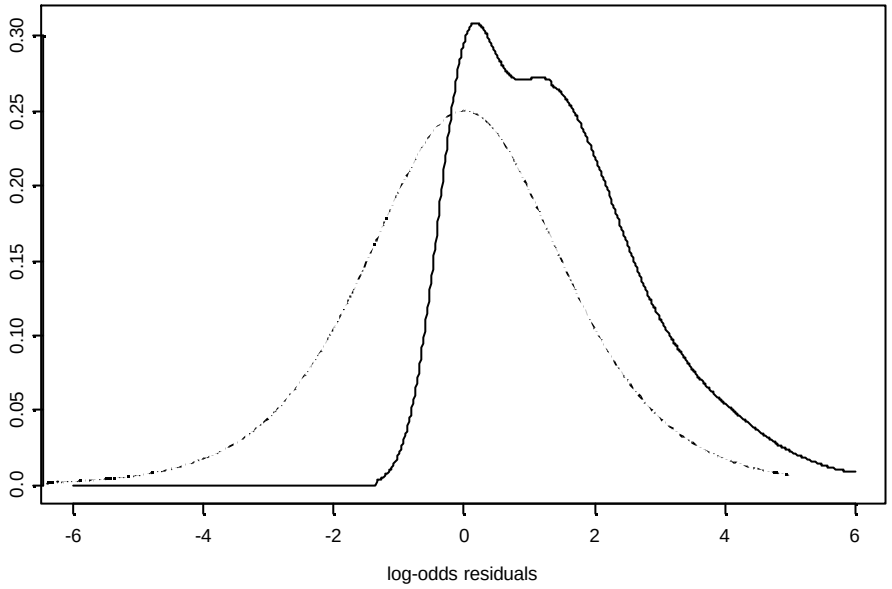
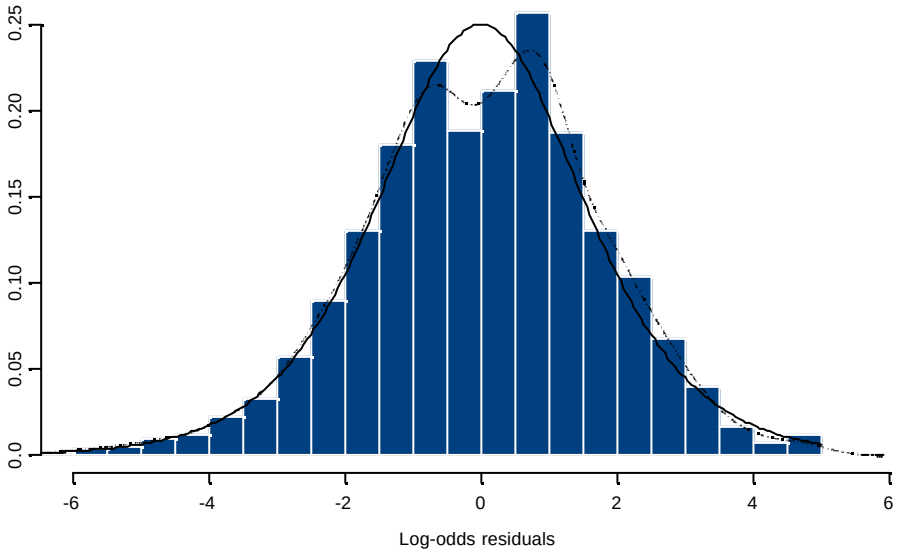


Figure 4. Residuals' empirical distribution under a misspecified model (simple Weibull instead of mixture): histogram, smoothed density (dotted line) obtained by kernel method and standard logistic distribution (continuous line).



6. Conclusions and further developments.

An extension of the previous cure rate model (Schinaia and Rossi, 2000) has been presented and the main features, which allow the application to a wide class of problems, have been outlined. The main results obtained concern the reliability of the estimate of the cure-fraction and the possibility to apply the log-odds residuals to assess the goodness of fit of the model and detect possible misspecifications.
Further work needs to be done to better investigate other important aspects both of the model and of the various procedures proposed, in particular for what concerns the general properties of the log-odds residuals. It is worthwhile noting that the mixture model can be easily extended to mirror more general situations in which different survival experiences must be taken into account either in the presence of long survivors or not. In conclusion, some final remarks can be made:

- The two class cure-rate model is general enough for a large class of different problems.
- If reliable prior information on the presence of cured individuals is available, a mixture model can be applied even with a limited follow-up. This is the case of some tumours with a favourable prognosis such as lymph node negative breast cancer.
- Conversely, whenever prior information is weak, some exploratory analysis should be performed to obtain evidence of a cure-fraction but a consistent follow-up is required (Maller and Zhou, 1994). In this case a plateau in the overall Kaplan Meier estimate is a necessary but not sufficient condition for the presence of cured individuals. Before applying a mixture model we recommend to fit different parametric models and to verify their goodness of fit. In fact long censored times can also be consistent with a non monotone risk function, i.e. a log-normal or log-logistic model. This might be the case of HIV incubation period for individuals treated with the recent triple therapy.
- It is possible to implement an efficient EM procedure, converging in few steps, to satisfactorily solve the diagnostic-prognostic problem in a cure-rate model, whenever the number of covariates is not too large, coherently including in the estimation procedure also the cure-fraction.
- The parameters are generally well estimated even with low sample size. The cure-fraction is well estimated, whatever its order of magnitude and respondents are correctly identified.
- The log-odds residuals appear to be a valid diagnostic tool to assess the goodness of fit of the specified model and detect possible misspecifications. It is worthwhile to remark that here the null hypothesis of a correctly specified model consists of different assumptions. We assume that a proportion of cured patients exists, that failure times for uncured patients follow a well specified (Weibull in the present application) distribution and that the effect of covariates on both the probability of being cured and the risk of experimenting the event is properly specified. If departures of the residuals' empirical distribution from the reference density are observed, each of these hypotheses should be reconsidered. In this respect some ad hoc procedures should be set up in order to detect departures in single specific directions.

Presently some data coming from a follow up study of patients affected by breast cancer are under study by applying the model and the estimation procedure presented in this paper, preliminary results seem encouraging. The model and the methods presented here could also be applied to the diagnostic-prognostic problem generated by the introduction of the new anti-HIV therapies for HIV infected patients which might produce an unknown fraction of long survivors and determine the impossibility to apply standard survival methods to estimate the incubation period distribution. It must be stressed that this kind of application, apart from the interest in se, would also allow again to estimate HIV incidence curves on the basis of AIDS

incidence curves, using the various Back-Calculation methods proposed and used in the past on the basis of the HIV incubation period estimated using standard survival models for homogeneous populations (Ravà et al., 1998). The estimated HIV incidence curves can then be used to estimate the prevalence of injecting drug users, by a proper calibration method, as it is shown in Ravà and Rossi (1999), providing an alternative tool to obtain this kind of estimates for such hidden population.

A different interesting application may be to data coming from follow up studies of heroin addicts in standard pharmacological (methadone, buprenorphine) treatment. In this case the aim is to early identify the patients who are not going to respond to the therapy, defined as those patients who will never become heroin-negative to the urine test. Such patients, in fact, might benefit from a different therapeutic approach, possibly by heroin prescription; it is, thus, crucial to early classify them on the basis of suitable observable covariates and then, possibly, include them in a heroin trial, where available.

References

- Brambilla C., Frontali M., Malaspina P., Rossi C., On the Estimation of the Age at Onset Distribution in Huntington's Chorea Using EM Algorithm, *Annals of Human Genetics*, 54, (1990), 225-233.
- Brambilla C., Rossi C., Schinaia G., Tree-Structured Analysis of Survival Data, *Applied Stochastic Models and Data Analysis*, 13, (1998), 333-343.
- Cantor A.B., Shuster J.J., Parametric versus Non-parametric Methods for Estimating Cure Rates Based on Censored Survival Data, *Statistics in Medicine*, 11, (1992), 931-937.
- Chen M.H and Ibrahim J.G., Maximum likelihood methods for cure rate models with missing data, *Biometrics,* 57, (2001), 43-52.
- Crowley J. and Storer B. E., Comment on paper by Aitkin M., Laird N., Francis B., *Journal of the American Statistical Association*, 78, (1983), 277-281.
- Farewell V.T., The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors, *Biometrics*, 38, (1982), 1041-1046.
- Farewell V.T., Mixture Models in Survival analysis: are they worth the risk?, *Canadian Journal of Statistics*, 14, (1982), 257-262.
- Ghitany M.E., Maller M.A., Zhou S., Exponential Mixture Models with Long-Term Survivors and Covariates, *Journal of Multivariate Analysis*, 49, (1994), 218-241.
- IbrahimJ.G., Chen M.H. and Sinha D., Bayesian semiparametric models for survival data with a cure fraction, *Biometrics*, 57, (2001), 383-388.
- Little R.J.A., Rubin D.B., *Statistical Analysis with Missing Data.* Wiley, (New York 1987).
- Maller R.A., Zhou S., Estimating the Proportion of Immunes in a Censored Sample, *Biometrika*, 79, (1992), 731-739.
- Maller R.A., Zhou S., Testing for Sufficient Follow-Up and Outliers in Survival data, *Journal of the American Statistical Association*, 89, (1994), 1499-1506.
- Maller R.A., Zhou S., Testing for the Presence for Immune or Cured Individuals in Censored Survival Data, *Biometrics*, 51, (1995), 1197-1205.
- Maller R.A., Zhou S., *Survival Analysis with Long-Term Survivors*. Wiley, (New York 1996).
- Nardi A. and Schemper M., New residuals for Cox regression and their application to outlier screening, *Biometrics*, 55, (1999), 523-529.
- Nardi A., *Discriminating between alternative parametric models for survival data via residual analysis*, (University of Vienna, Department of Medical Computer Sciences, Technical Report 2000).
- Nardi A., Schemper M., Fitting parametric models to survival data, *forthcoming*.

- Peng Y. and Dear K.B.G., A non parametric mixture model for cure rate estimation, *Biometrics*, 56, (2000), 237-243.
- Ravà L., Rossi C., Pasqualucci C. and Schinaia G., Estimating the size of the HIV/AIDS epidemic: complementary use of empirical bayesian back calculation and the mover-stayer model for gathering the largest amount of information, *SIMULATION*, 71-4, (1998), 213-227.
- Ravà L., Rossi C., Estimating the size of a hidden population involved in the HIV/AIDS epidemic: a method based on Back-Calculation and dynamical models, in *Simulation in the Medical Sciences*, Anderson & Katzper eds., 57-62, (The Society for Computer Simulation, San Diego, California, 1999).
- Rubin D.B., *Multiple imputation for non-response in surveys*, (Wiley New York 1987).
- Schinaia G. and Rossi C., EM estimation of diagnosis, *Biometrical Journal*, 42-5, (2000), 583-604.
- Tanner M.A., 1996, *Tools for Statistical Inference*, (Springer-Verlag New York 1996).

Appendix.

Proposition:
The proportion of censored times higher than a specific time τ tends to 1 for τ increasing if the population under study comprises cured individuals.

Proof:
Let us consider the independent competing processes generating failure or censoring times. Any individual of the population of interest is at risk for both processes, but only the lower time can be observed. Let us denote by C the random censoring time, by T the random event time, by $S_c(t)$ the survival function of the censoring process and by S(t) the survival function of the failure process. Let us denote by L the length of the observation period. By definition we have that $S_c(t)=0$ for $t \geq L$ and C uniformly distributed in the interval (0,L).
In order to prove the proposition we first compute:

$$P(t < \min(T,C)) = P(t < \min(T,C)/Z=0)\,p_o + P(t < \min(T,C)/Z=1)\,p_1 = S_c(t)[p_0 S_t(t) + p_1]$$

Let us now express:

$$P(C<T/t<\min(T,C)) = \frac{P[(C<T)\cap(t<\min(T,C))]}{P(t<\min(T,C))} = \frac{P(t<C<T)}{S_c(t)[p_0 S_t(t)+p_1]} = \frac{p_0\int_t^{\infty} P(t<C<t)\,f_t(t)dt + p_1 S_c(t)}{S_c(t)[p_0 S_t(t)+p_1]} = p_c(t)$$

and then, by straightforward calculations, we obtain:

$$p_c(t) = 1 - \frac{p_0 \int_t^L S_c(t)\,f_t(t)dt}{S_c(t)[p_0 S_t(t) + p_1]}$$

It follows immediately that:

$$\lim_{t<L\to\infty} p_c(t) = 1 \text{ if } p_1>0.$$

Appendix 3

Joint estimation of the latency period distribution and the onset incidence of heroin use for monitoring drug policy interventions (preprint)[4]

Carla Rossi,
Department of Mathematics,
University of Rome "Tor Vergata", Rome, Italy

Abstract

Modelling and estimating the incidence of first use (onset incidence) of drugs can be a useful tool for understanding the process of diffusion of drug use in space and time. This information is also important to evaluate current and future needs for, and effects of, services and interventions. Two indirect methods are available to estimate onset incidence and both require as a prerequisite the estimation of the latency period (LP) distribution. In some previous papers these incidence estimation methods have been presented and applied in various frameworks, separately estimating first the latency period distribution on the basis of standard methods of survival analysis by entry cohorts (i.e. by year of first treatment). This approach, though highly valuable and efficient, may cause some biases both in the estimates of LP distribution and of incidence. Such biases may be corrected on the basis of an integrated estimation approach by means of a suitable EM algorithm. This method is developed and presented below and some applications are used to show its main features.

Key words: Heroin, onset incidence, latency period, survival analysis, missing data, EM algorithm

1. Introduction.

There are similarities between the spread of drug use, in particular the use of addictive drugs such as heroin, and infectious diseases. Use of drugs is communicated, obviously not as an organic agent, but as a kind of 'innovative' social practice or custom, and not to everyone but only to those who, for whatever reason, are not immune (prone individuals). Once the basically contagious nature of drug use is accepted it becomes possible to study and to model the process of transmission (Behrens et al., 1999; Billard and Dayananda, 1993; Rossi, 2001) and to measure the extent of the phenomenon. The epidemiological concepts of incidence (the rate of new cases occurring within a certain time period) and prevalence (the number of all existing cases at a certain moment in time) are thus operationally valuable in studying illegal drug use and evaluate interventions aimed at prevention and care for drug users. In particular, epidemiological indicators may give some evidence of the effectiveness of interventions directed towards drug users (control intervention) or the at risk population (prevention intervention). Epidemiologists have suggested first use, first continuous use and first addiction as incidence indicators. Of these, the incidence of first use of drug is the best measure of the spread. Thus, incidence analysis can be used as a tool for examining the process of diffusion of drug use. In particular, incidence figures may provide an indication of

[4] The work contributes to the EU funded European Network to develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions (TSER/DGXII project ERB 4141 PL980030).

whether the number of problem drug users is growing (epidemic phase), falling or stable (endemic phase). There are two main sources of data on heroin use: the police records in which persons arrested are identified as heroin users, and the drug treatment programs. Both are samples of the total group of heroin users, but each is incomplete and biased since they focus on certain segments of a population to the exclusion of others. Therefore one must rely, rather than on the absolute incidence, on the weaker concept of relative incidence, which is the number of new cases per unit time for a specific population, as reported by some particular source. Relative incidence is commonly used in drug abuse research because it is adequate for many purposes. For example, relative incidence may show whether new drug use is increasing or decreasing about as well as absolute incidence, given a few reasonable assumptions (Hunt and Chambers, 1976). Unfortunately the population of users cannot be properly studied by standard statistical (descriptive) methods as such users, due to the actual drug laws and policies implemented in the various countries, are engaged in illegal acts, thus they resist identification and constitute a "hidden population". It is therefore necessary to use mathematical models and inferential methods to estimate the epidemiological indicators of interest. These models and methods allow, on the basis of indirect indicators, such as therapy presentations, incarcerations, non fatal overdoses, to estimate the interesting indicators, such as relative incidence of first use.

The present study, being based on treatment data, provides estimates of the relative incidence of drug users that will eventually enter treatment. This is clearly a selected population, since it excludes the individuals who will stop to use drug or die before starting any treatment, or those who will never develop problems requiring treatment. However, the estimated relative incidence, being an estimate of the number of those who will require treatment in the near future (the median lag between the start of drug use and the first treatment has been estimated in EMCDDA 2000 in 4-6 years), is appropriate for analysing "spatial" and time trends of problem drug use, for planning, monitoring and evaluating preventive interventions and for forecasting treatment needs.

Most methods to estimate relative incidence needs as a prerequisite the estimation of the time lag between first use of heroin and first treatment, the so called latency period (LP). Thus, it is necessary to set up proper estimation procedures to obtain reliable estimates for LP distribution (Rossi, 1999; EMCDDA, 1999, 2000) in order to produce reliable estimate of relative incidence.

On the other hand, the analysis of LP is also important because it helps in understanding the natural history of heroin use; it can be used for planning and monitoring interventions directed at improving the attraction or appropriateness of treatment services; it is important for understanding factors influencing the progression to treatment.

Latency period can be analysed by entry cohort (ie. by year of first treatment) or by "onset" cohort (ie. by year of first use). In the second case we are estimating the latency period (LP) distribution, in the first one the backward latency period (BLP) distribution. All being equal they produce the same figures. However, if incidence changes over time, analyses by entry cohort may be biased – because it will look like LP is decreasing over time if incidence is increasing, or vice versa. In fact, the stage of the heroin use epidemic strongly influences LP. During the first years of an epidemic all observed LP would by definition be shorter, and with decreasing incidence the longer LP would be overrepresented. Unfortunately data usually available are more suitable to estimate (BLP) than to estimate LP as is shown in the following. In order to correct possible biases, due to the incidence effect, it is then necessary to properly model the mathematical relation between LP, incidence and BLP and use such model to set up an estimation procedure for both LP and incidence on the basis of BLP and treatment incidence.

This paper presents a new integrated method to estimate the latency period (LP) between first use of opiates and first treatment demand as well as the onset incidence curve, whenever LP

may be considered stationary[5]. The method allows to correct possible biases affecting previous incidence estimates (Ravà et al., 2001), which are mainly based on BLP due to incompleteness of available data.

2. The basic model.

Let us denote by t the time of entry in therapy, by τ the time of first use, thus $\nu=t-\tau$ is the LP (and BLP as well), then denote by $f_n^{(t)}(n)$ the generalized density function of LP for the cohort starting use in τ, by $g_n^{(t)}(n)$ the generalized density function of BLP for those starting therapy in t, by $I^{(t)}(t)$ the therapy incidence in t and, finally, by $I^{(t)}(t)$ the onset incidence in τ.
If $I_t(t,n)$ is the therapy incidence of those with BLP=ν, we can write:

$$g_n^{(t)}(n) = \frac{I_t(t,n)}{I^{(t)}(t)} \quad \text{and} \quad I_t(t,n) = I^{(t)}(t)\, f_n^{(t)}(n)$$

by replacing $I_t(t,n)$ we obtain:

$$g_n^{(t)}(n) = \frac{I^{(t)}(t)\, f_n^{(t)}(n)}{I^{(t)}(t)} \qquad (1)$$

It can be shown, by straightforward calculations on the basis of (1), that, if $f_n^{(t)}(n) = f_n(n)$ is stationary in τ, the mean of $g_n^{(t)}(n)$ is decreasing in t for $I^{(t)}(t)$ increasing and vice versa.
We can observe that (1) is the general equation linking therapy incidence, onset incidence, BLP density and LP density. Thus, whenever we know or can estimate three terms, we can obtain the other. Unfortunately, in general, we can only observe two terms, namely $g_n^{(t)}(n)$ and $I^{(t)}(t)$. However, an EM algorithm can be set up to iteratively estimate the two unknown terms, namely $f_n(n)$ and $I^{(t)}(t)$.

3. The EM algorithm for jointly estimating $f_n(n)$ and $I^{(t)}(t)$.

Let us suppose that LP is stationary and focus on the estimation of $f_n(n)$. Thus we can consider $I^{(t)}(t)$ as "missing" in the expression:

[5] The year of implementation of treatment services could profoundly influence LP, if the analysis would include data from the first years that treatment was available. For heroin users entering first treatment in those years, LP could never be shorter than the period since the year of their first heroin use and the year of implementation of services, resulting in an artificially longer LP (left-truncation). Changes in policy implementation and in drug laws could artificially change LP, e.g. if police activity is increased during a certain period and arrests are referred to treatment. We have no data to test this possibility, however the effect on LP is in general smoothed out when analysing a series of treatment years.

$$f_n(n) = \frac{I^{(t)}(t) g_n^{(t)}(n)}{I^{(t)}(t)}$$

obtained by inverting (1).
On the other hand, if we consider $f_n(n)$ as known, we can calculate the expected value of $I^{(t)}(t)$ by means of any estimation method, either Back-Calculation (BC) (Ravà et al., 2001) or Brookmeyer and Liao Reporting Delay Adjustment (RDA) method (Hickman, De Angelis and Seaman, 2001).
Thus, the EM algorithm proceeds as follows:

1. find an initial approximation of $f_n(n)$, namely $f_n^0(n) = g_n^{(t)}(n)$ for some fixed t;
2. calculate a first estimate for $I^{(t)}(t)$ by means of a suitable method on the basis of $f_n^0(n)$ and denote it by $I_0^{(t)}(t)$;
3. for any k>0 estimate the k-th approximation of $f_n(n)$: $f_n^k(n) = \frac{I^{(t)}(t) g_n^{(t)}(n)}{I_{k-1}^{(t)}(t)}$ (M-step);
4. calculate the expected value of $I^{(t)}(t)$ ($I_k^{(t)}(t)$) given $f_n^k(n)$ by the incidence estimation method chosen for the initial step (E-step).
5. Iterate until convergence.

The general properties of the GEM algorithms apply (Tanner, 1996). An example can be used to clarify the process.

4. Adjusting the onset incidence estimate and LP by EM algorithm.

Let us consider the data from Budapest analysed in EMCDDA (2000) divided by route of administration (injecting, non injecting). In the following Tables and Figures data and results of the estimation procedure are reported for injecting drug users (LP and onset incidence), assuming t=2000, fixed in all calculations.
Let us first consider the data available for the analysis (Table 1). It is clear that the backward estimation is easier and more informative than the forward estimation because the sample size of therapy incidence for the years 1996-2000 allow to estimate the BLP density $g_n^{(t)}(n)$ for ν=1,2,…10, whereas, if we decide to estimate the LP density $f_n(n)$, we can only use data since 1994 to 2000, thus for ν=1,2,…5. This happens in general when data from therapy services are to be used for latency analysis. Thus, we will first estimate $g_n^{(t)}(n)$ and then, by adjusting it on the basis of the EM algorithm, estimate both $f_n(n)$ and $I^{(t)}(t)$. However, due to the available data, the results only refer to the sub-population of heroin users who will eventually seek treatment before the end of the tenth year since their first use. Thus, the LP distribution being estimated is a conditional distribution: the distribution of LP conditional on it being ten years or less.
The results of the joint estimation procedure are reported below. It can be observed that the qualitative behaviour of the algorithm with respect to the incidence estimate and to the latency estimate is similar. In fact, the initial estimate of LP distribution (survival curve) is close to the final estimate and the same happens for incidence estimate. This confirms some

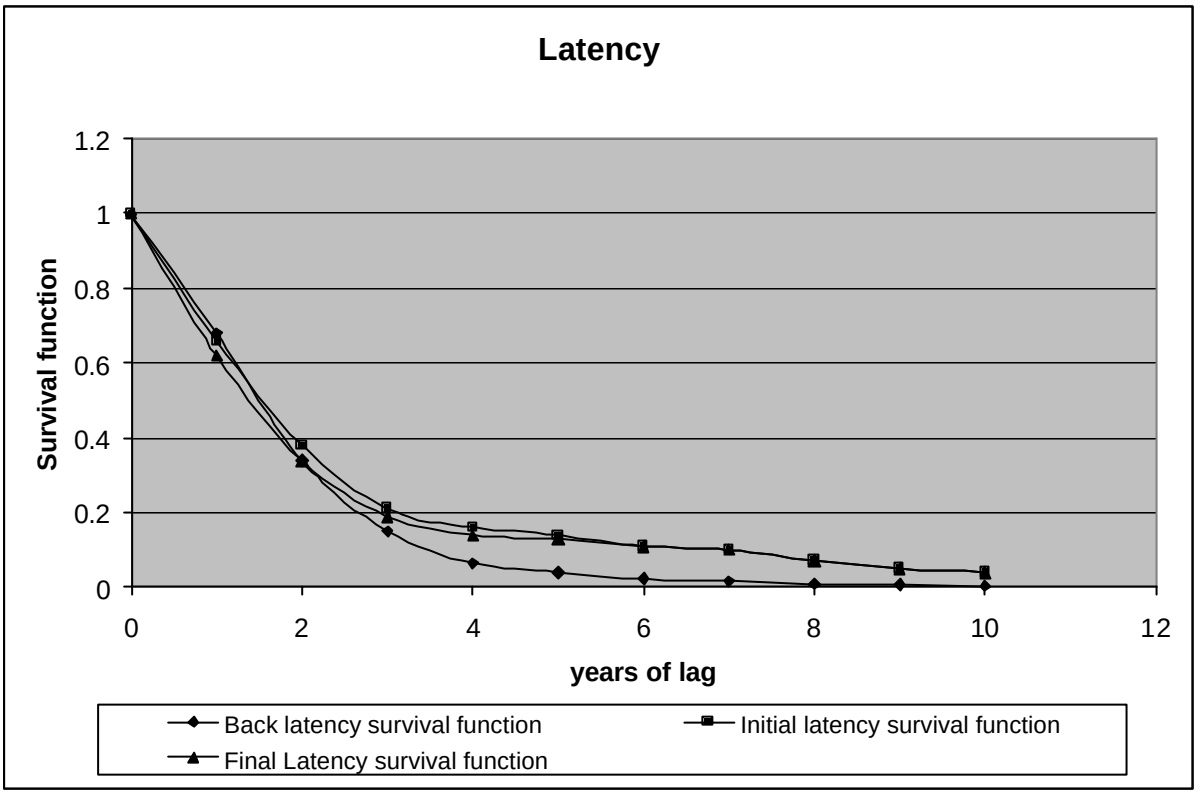
qualitative considerations reported in EMCDDA (1999) concerning robustness of the shape of incidence estimate with respect to various models adopted for LP distribution.

Table 1. Onset and therapy incidence for injectors (Budapest)		
Calendar Years	**Observed onset incidence**	**Therapy incidence**
1972	1	0
1974	1	0
1975	1	1
1980	1	1
1982	1	0
1985	0	1
1986	0	0
1987	2	0
1988	6	1
1989	5	0
1990	6	3
1991	7	2
1992	17	2
1993	34	5
1994	53	12
1995	84	16
1996	95	59
1997	68	118
1998	53	88
1999	20	77
2000	missing	68

Table 2. LP distribution estimated by EM algorithm (convergence in 3 steps).

Years (n)	**Back Latency (BLP)**	**Latency (Initial estimate)**	**Latency (Final estimate)**
0	1	1	1
1	0.68	0.66	0.62
2	0.34	0.38	0.34
3	0.15	0.21	0.19
4	0.064	0.16	0.14
5	0.04	0.14	0.13
6	0.022	0.11	0.11
7	0.015	0.1	0.1
8	0.009	0.07	0.07
9	0.007	0.05	0.05
10	0.004	0.04	0.04
11	0	0	0
mean	2.3	2.5	2.8

Figure 1. LP distribution estimated by EM algorithm (convergence in 3 steps).



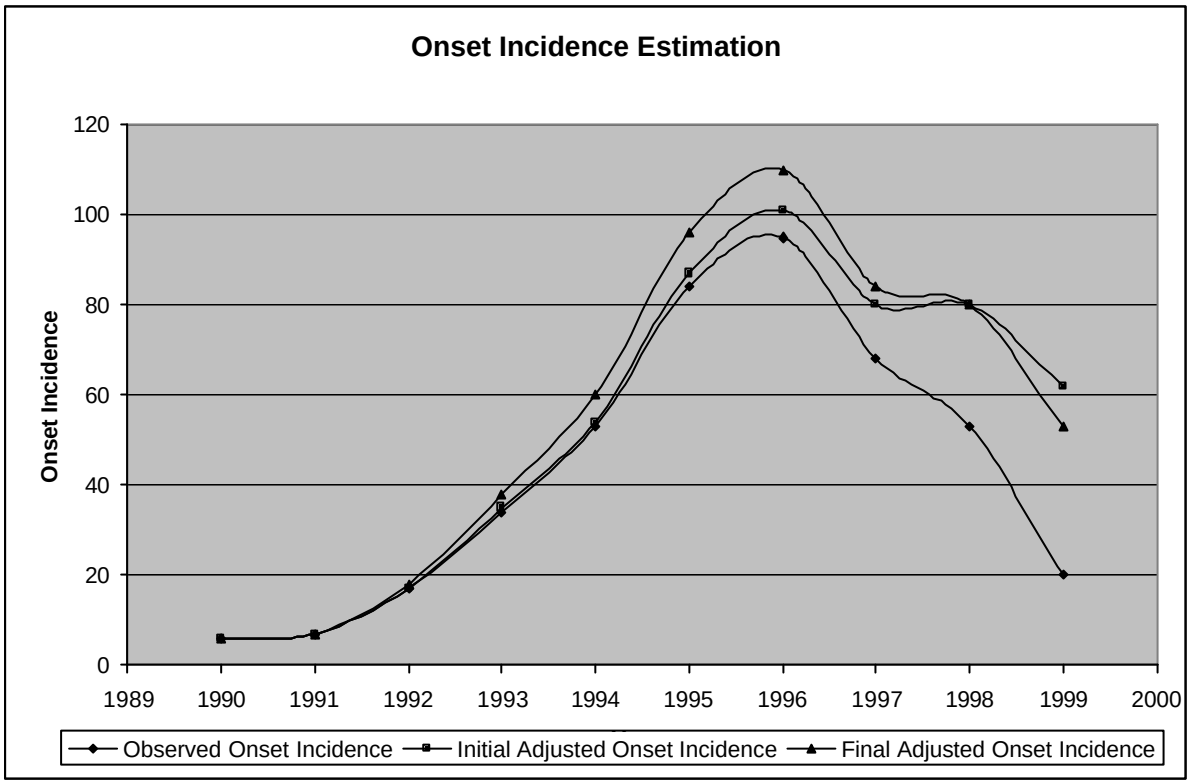
It must be observed that the final estimate of LP distribution assigns more weight to higher lags as it was expected, in fact, the onset incidence curve is mostly increasing in the 10 calendar years taken into account.

The analysis can used to better evaluate the possible effect of covariates on LP distribution. In fact, it is possible that some covariate appears more or less significant as a byproduct of different behaviours of incidence curve for different values of the covariate. In other words, the usual prognostic analysis by Cox model applied to BLP can only be accepted for LP if the incidence curves corresponding to the various levels of the covariate are proportional, which is equivalent to require stationary distributions of the covariate within each onset cohort.

Table 3. Onset incidence estimated by applying the RDA method at each E-step (convergence in 3 steps).

Calendar Year	ObservedOnset Incidence	Initial Adjusted Onset Incidence	Final Adjusted Onset Incidence
1990	6	6	6
1991	7	7	7
1992	17	17	18
1993	34	35	38
1994	53	54	60
1995	84	87	96
1996	95	101	110
1997	68	80	84
1998	53	80	80
1999	20	62	53
Total	437	529	552

Figure 2. Onset incidence estimated by applying the RDA method at each E-step (convergence in 3 steps).



In mathematical terms this corresponds to the following propositions, which can be proved by straightforward calculations.

Proposition 1:
Let us denote by $f_n(n, X)$ the LP distribution conditional to the covariate X and by $I^{(t)}(t, X)$ the onset incidence function conditional to the covariate X. Then, the X-effect on LP is equivalent to the X-effect on BLP if and only if $f_n(n, X)$ can be expressed as a proportional hazard model and $I^{(t)}(t, X) = k(x)I^{(t)}(t)$.

Proposition 2:
$I^{(t)}(t, X) = k(x)I^{(t)}(t)$ if and only if the distribution of X within each onset cohort starting drug use in τ does not depend on τ.

5. Some further example of analysis.

Let us consider some of the analyses reported in EMCDDA (1999) regarding the estimation of LP in Amsterdam. Analysing BLP by Cox model, there seem to be differences in latency period between drug users originating from different countries. These differences mostly reflect differences of the onset of the heroin epidemic among different subgroups. In this case the heroin epidemic among those originating from Surinam would be the oldest, followed by the epidemic among Dutch. The epidemic among the Moroccans would be the youngest.
To better highlight this fact, it is valuable to study this issue in a more prospective way. We can return in time to the year of onset of drug use and look forward until treatment demand, as the available data allows this analysis. Figure 3 shows the LP survival estimates by country of

birth based on Kaplan Maier method. The figure does not show any difference between the different groups of countries. So, the differences between the different countries of birth observed as result of the Cox-proportional hazard model for BLP are likely to reflect the period of the heroin epidemic for different subgroups.

Figure 3a (LP survival Amsterdam)

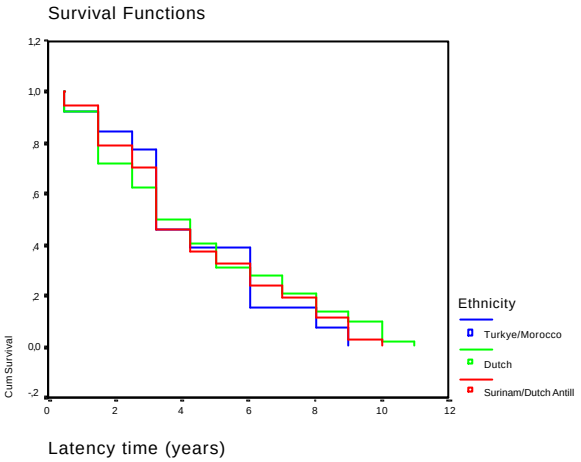
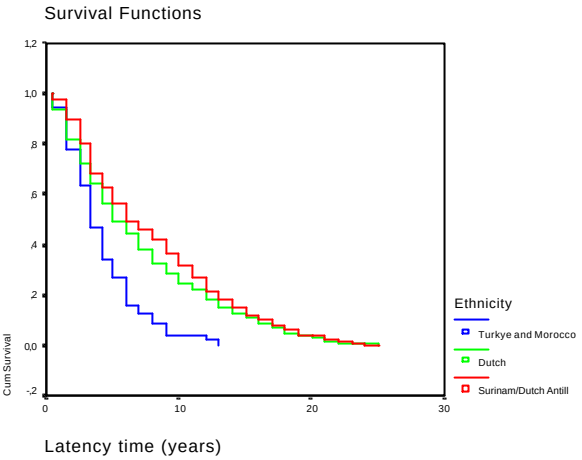


Figure 3b (BLP survival Amsterdam)



The EM algorithm described above would allow to immediately find the same results even using more incomplete data (shorter time series), such as those for Budapest. We can then complete the analysis for Budapest, taking into account non injectors, as the route is the only significant covariate for BLP. The adjusted estimates will show whether the route remains significant for LP as well.

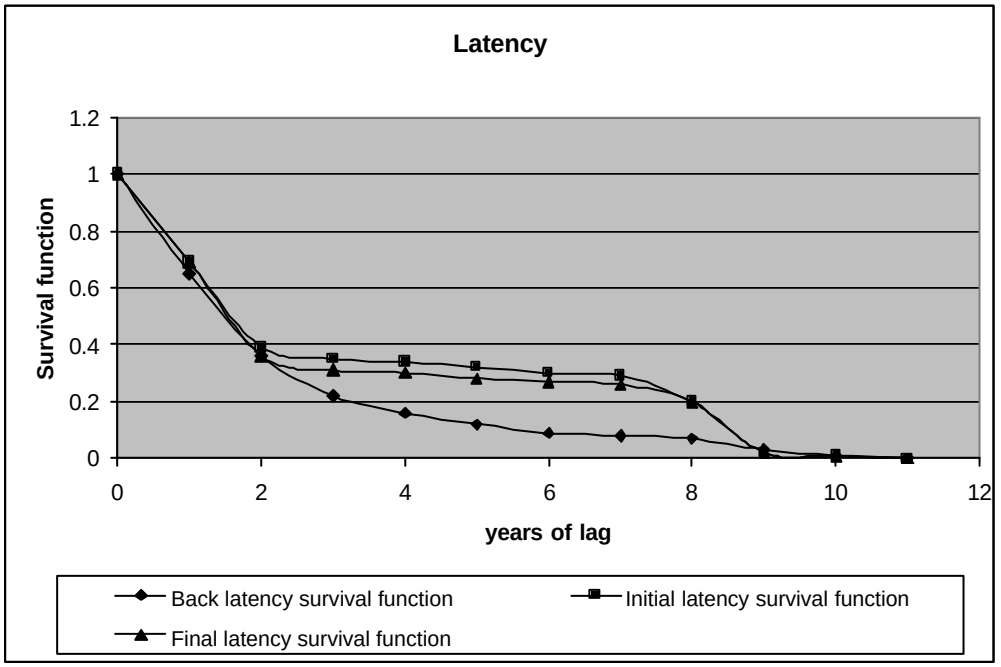
Table 4. Onset and therapy incidence for non injectors (Budapest)		
Calendar Years	**Observed onset incidence**	**Therapy incidence**
1972	0	0
1974	0	0
1975	0	0
1980	2	0
1982	1	0
1985	1	0
1986	1	0
1987	1	0
1988	2	1
1989	0	0
1990	7	0
1991	2	0
1992	2	2
1993	9	4
1994	14	15
1995	12	8
1996	26	17
1997	18	26
1998	5	24
1999	3	15
2000	missing	8

In Table 4 observed data are reported and in the following Figures the estimation results are presented. As can be observed in Figures 4 and 5, the algorithm produces significant modifications with respect to the initial estimates based on BLP both for LP distribution and for onset incidence. The expected value of the lag between first use and first treatment, for example, grows from 2.76 to 3.53 years and the estimated total number of non injectors, starting during the ten years period taken into account, from 129 to 141. Thus, LP is significantly different for the two routes of administrations, even more than BLP.

However, it must be observed that in all the situations analysed, the median of LP is much closer to the median of BLP than the mean of LP to the mean of BLP, as it is expected due to the robustness property of the median. This allows to use the median of BLP as a suitable indicator for comparisons.

Several analyses on BLP, aimed at measuring the influence of the covariate "age at first use", are reported in EMCDDA (1999 and 2000). The results show that young heroin users appear to progress much slower to treatment than older ones. Age at first use is significant showing an adjusted odd ratio 1.03 for each year of age. It can be easily verified that the distribution of this covariate in each onset cohort is similarly distributed in the various years and in the various sites. Thus, the covariate maintains its significance for LP too. This suggests to carefully interpret the results because, in general, the covariates that appear to influence LP may give important information for planning / targeting interventions. It is known that drug users in treatment are relatively protected from infection or overdose death by having less instances of risk. If young drug users take longer to get into treatment this implies that they are longer at risk for infection or overdose, and this would be similar for the other subgroups of users identified by the analysis of covariates.

Figure 4. LP distribution estimated by EM algorithm (convergence in 3 steps) for non injectors.



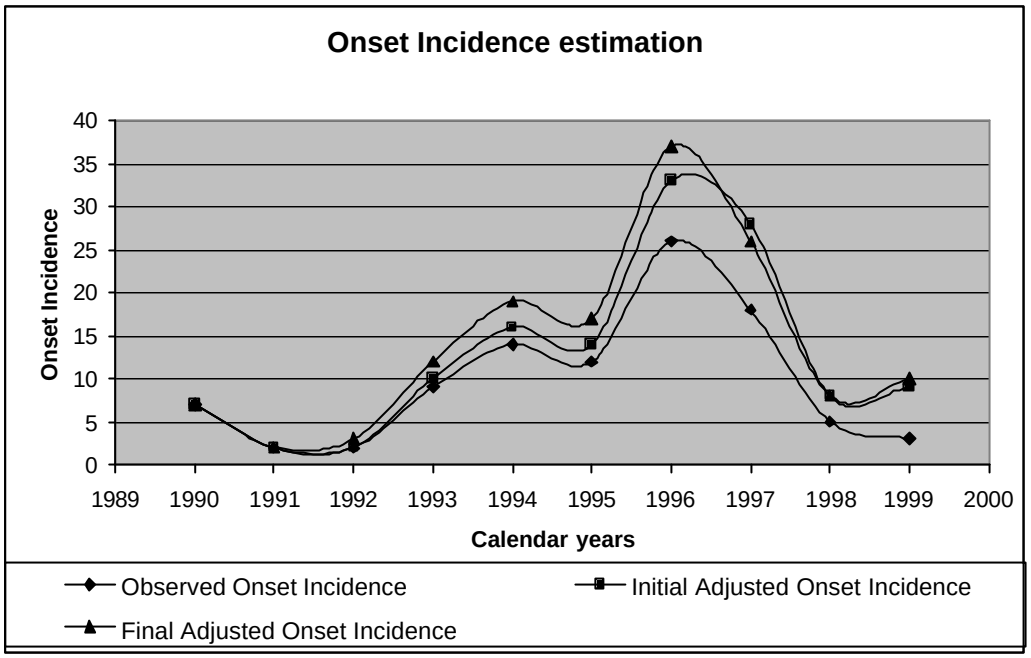


Figure 5 . Onset incidence estimated by applying the RDA method at each E-step (convergence in 3 steps).

6. Final remarks.

In this paper a joint estimation method has been proposed in order to estimate: the number of heroin users that will present for treatment and will be reported within a specified time period and the LP distribution, possibly depending on observable covariates. The results of these analyses offer a description of sub-groups of heroin users that may need to be targeted through special programmes, and provide a baseline to monitor the effect of drug policy.

The analysis of LP has interesting public health implications for drug policy, given that one of its aims is to encourage heroin users to seek treatment. Careful consideration must be given to the assumptions on which the method relies (stationarity of LP). This may not be the case, for example, if treatment capacity increased and managed to increase the proportion of heroin users seeking treatment. However, the similarities between sites and over time (EMCDDA, 1999 and 2000) suggest that LP is not strongly related to external factors, such as treatment availability, but that it may be part of the natural history of problematic heroin use. This result corroborate the hypothesis of stationarity of LP, which is the necessary prerequisite for the EM algorithm set up in the present paper.

As a final remark, it must be underlined that the EM algorithm presented above is quite easy to implement and allows to produce useful estimates for both LP and onset incidence even with short time series of therapy data. Such estimates can offer a valuable support to policy makers.

References

1. Behrens D.A., Caulkins J.P., Tragler G., Haunschmied J.L., Feichtinger G., A dynamic model of drug initiation: implications for treatment and drug control, *Mathematical Biosciences*, 159, 1-20, 1999.

2. Billard L., Dayananda P.W.A., Drug addiction-pushers generated from addicts, Biometrical Journal, 35, 227-244, 1993.
3. Brookmeyer R and Liao J. The analysis of delays in disease reporting: methods and results for the Acquired Immunodeficiency Syndrome. American Journal of Epidemiology 1990; 132: 355--365.
4. Brookmeyer R and Damiano A. Statistical methods for short-term projections of AIDS incidence. Statistics in Medicine 1989; 8: 23--34.
5. EMCDDA, *Pilot project to estimate time trends and incidence of problem drug use in the European Union*, Final report, Lisbon, 1999.
6. EMCDDA, *Study on incidence of problem drug use and latency time to treatment in the European Union*, Lisbon, 2000.
7. Hickman M., Seaman S., De Angelis D., Estimating the relative incidence of heroin use: application of a method for adjusting observed reports of first visits to specialized drug treatment agencies, *American Journal of Epidemiology*, 2001 Apr 1; 153(7), 632-641.
8. Hunt L.G., Chambers C.D., *The Heroin Epidemics*, SPECTRUM PUBLICATIONS INC, NY, 1976.
9. Ravà L., Calvani M.G., Heisterkamp S., Wiessing L., Rossi C., Incidence indicators for policy making: models, estimation and implications, *UN Bulletin on Narcotics*, 2001, in press.
10. Rossi C., Monitoring drug control strategies: hidden phenomena, observable events, observable times, *International Journal of Drug Policy*, 10-1, 131-144, 1999
11. Rossi C., A Mover-Stayer type model epidemics of problematic drug use, *UN Bulletin on Narcotics*, 2001, in press.
12. Tanner M.A., *Tools for Statistical Inference*, Springer-Verlag New York, 1996.

Appendix 4

A Mover-Stayer type model for problem drug use epidemic

Carla Rossi,
Department of Mathematics,
University of Rome "Tor Vergata", Rome, Italy

Summary:
A modified version of a model for the HIV/AIDS epidemic, proposed at the beginning of the 90s and recently generalized, is presented to mirror the problem drug use epidemic. The model can be used both to estimate interesting epidemic macro-parameters and to make scenario analyses. The model is a Mover-Stayer type model (Morgan et al., 1983; Rossi, 1991; Albert, 1999) and allows for heterogeneous risk behaviour among the susceptibles. Such kind of models considers the susceptible population as subdivided into two groups: the group of stayers, that is the group of individuals who are considered not at risk of „infection", and the group of movers, who are at risk. Due to the interactions between infectious individuals (pushers) and the susceptibles, some of these latter may pass to the drug user compartments and begin a „drug user career". The present model comprises two different stages of hidden drug use. The first (light use) stage, which can be more strictly defined, is the initial (or non problematic) stage of drug use, then light drug users can either stop using drug or pass to hard drug use (or death). Other interesting stages taken into account comprise: therapy stage, recidivist use (either light or hard) stage and temporary non-use stage (these latter are the visible stages). The simpler version of the model is studied using a Markov hybrid approximation and some "what if" scenario analyses are obtained by simulation. A more complex and realistic model is outlined as possible further development.

1.Introduction: what mathematical models tell us that we would not know otherwise.

The drugs problem and its consequences for society represent a complicated research field. Policy makers and researchers are seeking answers to a number of questions concerning drug use, its consequences and related costs. They focus on questions such as: 'how much drug use?', 'what are the consequences of this use?', which policies are effective?', what are the costs of the policies and of the consequences?'. The consequences comprise, among many others, the adverse effect of infectious diseases and the costs imposed to society by the drug-related criminality. Hence it is important to understand and "measure" drug use and how it responds to drug control interventions (Behrens et al., 1999). The present paper contributes to this objective by introducing a simple compartmental epidemic model of drug use, which incorporates various effects on initiation into new use or relapse into recidivist use of drugs.
Compartmental models represent a powerful mathematical tool well established in modelling the spread of "diseases" in a population. They provide a framework in which numbers of people in different compartments (each one homogeneous with respect to some specified characteristic) and the relationships between such compartments, modelling the dynamics of the population, can be described in mathematical terms.
Dynamical compartmental modelling of epidemic processes, either producing operational or transmission models, occurs through the usual representation of the dynamics involved by a system of stochastic or deterministic differential (or difference) equations; this is the case for both the operational and the transmission models. The main difference between the two kind of models is in the fact that transmission models take into account the dynamic processes at

micro level, modelling the interactions between individuals belongings to the different sub-groups involved in the epidemic, whereas operational models work on macro-variables or indicators suitable to be used to estimate the size of the phenomena or monitoring the impact of various interventions, modelled by suitable scenario-parameters. Many models of the two types have been developed to study the HIV/AIDS epidemic and are suitable to be used, with some modifications, to model the problem drug use epidemic as well.

Some examples of transmission models are in Dietz, 1988; Hadeler, 1989; Kretzschmar and Dietz, 1997 for what concerns the HIV/AIDS epidemic and in Billard and Dayananda, 1993 and in Beherens et al., 1999, for problem drug use epidemic. The complexity of these models is due to the intention of the researcher to introduce in the modelling process a detailed formalization of the interactions existing, or supposed to exist, between a large number of subjects involved in the epidemic process. This is the case of the analysis of HIV transmission across risk groups, such as drug users and heterosexuals, when some specific hypothesis on the contact pattern is formulated (Hadeler, 1989; Jacquez et al., 1988) or when the time at which the contact occurs is supposed to somehow affect the probability of transmission of the infection (Kretzschmar and Dietz, 1997;). As already mentioned, such models can be very helpful when specific contact or transmission patterns need to be thoroughly analyzed, but may be extremely cumbersome when the whole picture of the epidemic is under study. In fact, these detailed model structures contain all sorts of parameters regulating every single interaction in the simulation process and very seldom enough data is available for robust estimation of such a large number of parameters involved.

A more efficient way of obtaining a simulated epidemic is by using a “simple” operational model. As opposed to transmission models, simple models do not attempt to include all the possible group or individual interactions into the modelling structure, but summarize the dynamics of the epidemic by some non-linear interaction term and sum up all the infected individuals in chains of compartments. Most of the parameters controlling the dynamics in such systems are derived from epidemiological studies, external to the model, and their values simply come from specialized studies or from the literature; only a limited number of “internal” parameters is usually left to be estimated by fitting the existing data or to be used for scenario analyses. Typically, the set of internal parameters includes some form of control of the transmission and of the size of the core group; other internal parameters may have different origins and interpretations, depending on the design of each single model. The ability of correctly describing the epidemic of the simple models is theoretically much more limited when compared to the complex transmission models; however, in general, they turn out to be much more tractable, both because of the limited number of parameters required for their functioning and because of the quality of their output. In fact, while the large number of parameters in the transmission models often results into their unidentifiability, parameter estimates in simple models are usually robust enough for the model to be successfully used in complex scenario analyses and under other strained conditions.

Most of the simple models have a stochastic basis, although their output results into some deterministic forecast curves. Such models, defined as hybrid models in Bailey (1991, 1993, 1994), are the result of the deterministic approximation of stochastic models, using mean times of stay in compartments with conditional Poisson arrivals and Maximum Likelihood (ML) or numerical estimates of internal parameters. Their use, however theoretically disputable, is more and more frequent among applied researchers because of their technical simplicity and the clarity and immediate possibility of use and interpretation of their results: their output consists of deterministic curves of epidemic indicators (incidence, prevalence, dynamics of the core group, etc.), while the parameter ML estimation procedure is directly derived from the distributional hypotheses of the original stochastic model. One of these models for the HIV/AIDS epidemic has been proposed at the beginning of the ‘90s and has been recently generalized. It allows to easily obtain scenario analyses (Rossi, 1991; Rossi and

Schinaia, 1998). Such model, when used jointly with suitable Back-Calculation methods, allows to reduce uncertainties in the estimates of incidence curves (Ravà et al, 1998) and it has also been used for indirectly estimating the prevalence of injecting drug use in Italy (Rossi, 1999a). A modified version of this model is presented and developed in the following to mirror the problem drug use epidemic. In Section 2, the model is presented. Section 3 is devoted to discuss the hypotheses about the parameters of the model. Section 4 deals with the simulation program and the scenario analyses. Section 5 addresses issues related to discussion and further developments. The model will be presented from a mathematical point of view elsewhere (Rossi, in preparation).

2. The operational model for problem drug use epidemic.

The graph reported in Figure 1 describes the main features of the proposed model. The model is a Mover-Stayer type model and allows for heterogeneous risk behaviour among the susceptibles. Such kind of models considers the susceptible population as subdivided into two groups: the group of stayers that is the group of individuals who are considered not at risk of „infection“ (these models are suitable to make scenario analyses in order to assess the impact of various proportions of vaccinated persons on the probability of extinction of a given epidemic) and the group of movers who are at risk. Due to the interactions between infectious individuals (for our problem we can imagine these are the problem drug users who are also pushers[6]) and the susceptibles, or due to the pressure of the black market on the susceptibles, some of these may pass to the drug user compartments and begin a „drug user career“. Similarly to the model proposed by Beherens et al. (1999), the present model comprises two different stages of hidden drug use. The first (light use) stage, which can be more strictly defined, is the initial (or non problematic) stage of drug use, then light drug users can either stop using drug or pass to hard drug use (or death). The other arrows in the graph completely describe the other possible transitions in a drug user career. The curves connecting the drug use (infectives) compartments and the susceptible (or temporary no-use) compartments indicate the possible interactions which may produce transitions from susceptibles (or temporary no-use) to infectives (non-linear terms in the equations), the other possible transitions from susceptibles (or temporary no-use) to infectives are induced by the pressure of the black market and are supposed to be represented by a linear terms in the equations.
In order to write the corresponding equations (either deterministic or stochastic) some further hypotheses have to be explicitated and the known and unknown parameters have to be described. A first approximation may be through a Markov model, possibly a marked Markov process. In such a case the length of stay in each compartment is assumed to be exponentially distributed and the results are useful to get a first qualitative insight in the epidemic process. A more realistic approximation is by using semi-markov processes. In such a case the length of stay in each compartment may be assumed to be distributed differently with respect to an exponential variable, possibly depending also on covariates describing the individual experience of drug use. The study is much more complex but suitable mathematical and simulation techniques can be used. In the present paper only the Markov model will be considered and used to make some scenario analyses.

[6] From the surveys conducted among military conscripts, reported in the Annual Report on the state of the drugs problem in Italy for the year 1999, published by the National Focal Point, it results that, in terms of the reason for drug use, the two most mentioned factors were curiosity (more than 40%) and peer group pressure (more than 30%).

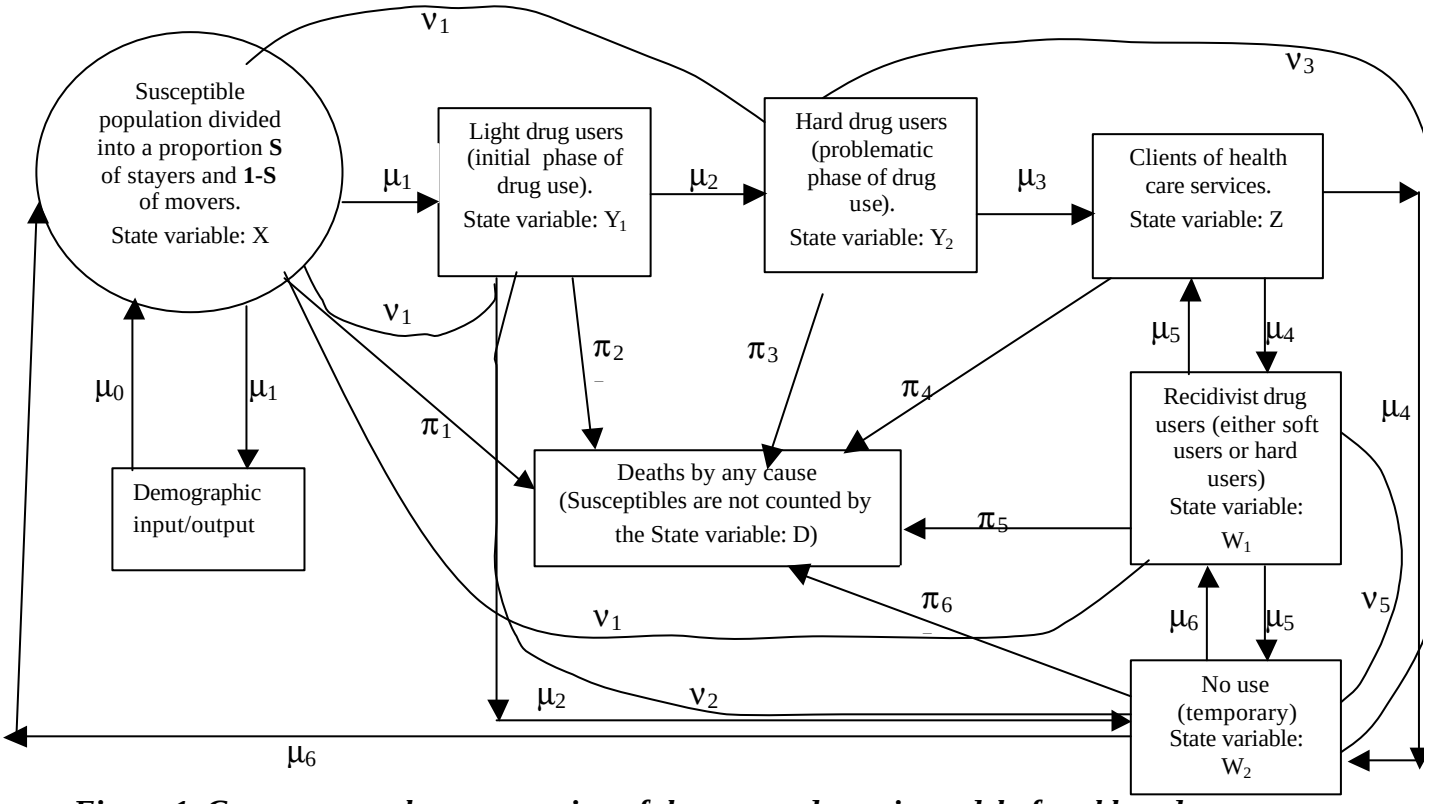


Figure 1. Compartmental representation of the system dynamic model of problem drug use epidemic.

From the graph reported in Figure 1, it is immediate to write the equations of the model either in the form of deterministic (continuous or discrete) equations or in the form of stochastic (continuous or discrete) equations. The discrete stochastic equations will be reported elsewhere (Rossi, in preparation); only the simulation program and some scenarios aimed at evaluating the impact of different parameters on the epidemic are considered in the present paper. The state variables used in the model (with the exception of S(t), which is the proportion of stayers at time t) represent the incidence per million inhabitants, the unit time for the simulation runs is taken equal to one week.

The equation for the proportion of stayers *S* is derived under the hypothesis (Rossi,1991) that the new entrances in the susceptible compartment are divided into Stayers and Movers according to constant proportions S_0 and $1-S_0$ (stationarity), with $0< S_0<1$, even if other hypotheses can be trivially included in the model as well as possible transitions from the Movers to the Stayers due to prevention campaigns or to the effect of law enforcement activities. This further development is briefly outlined in Section 5.

The qualitative analysis (transient and asymptotic behaviours) of the epidemic, in particular for what concerns the incidence of drug use, that is the transitions from susceptibles to drug users, can be conducted analysing jointly the equation for *X* and the equation for *S*. The method is analogous to that used in Rossi (1991) and will be considered elsewhere (Rossi, in preparation).

3. The "crucial" parameters to be externally estimated and the scenario parameters.

For what concern the parameters and the distributions of the lengths of stay, some are already available from the study of the latency period (EMCDDA, 1999a). Some can be derived using therapy data already available in some sites. The demographic parameters regulating the dynamics of the susceptible population, namely μ_{01}, μ_{10} and π_{17}, are supposed to be known and are country-specific. The other parameters π can be externally estimated using the information from mortality studies among drug users which are available for most countries in the EU (EMCDDA, 1999b). The parameters μ_{23} and μ_{34} (natural history parameters) can be estimated on the basis of some data available for the study of the latency period (Rossi, 1999b, Ravà et al., in preparation). The parameters μ_{45}, μ_{46}, μ_{54}, μ_{56}, μ_{65} and μ_{61} (therapy parameters) can be obtained, at least for what concerns their order of magnitude, from therapy data available in most countries, the values of all these parameters for Italy (order of magnitude) are reported in Table 1. All the other parameters, namely μ_{12}, μ_{26}, ν_{12}, ν_{13}, ν_{15}, ν_{26}, ν_{36} and ν_{56} can be used as scenario parameters, as well as the parameter "initial proportion of Stayers", S_0.

In the present paper it is assumed, (following Billard and Dayananda, 1993) that most pushers are hard drug users (basic ν-scenarios). Elsewhere, it was suggested that most pushers are soft drug users (Hunt and Chambers, 1976), thus, some simulation runs will be devoted to obtain scenarios under this latter hypothesis (alternative ν-scenarios) by modifying the order of magnitude of the parameters ν reported in Table 1 (10^{-5} instead of 10^{-6} and viceversa).

Table 1. Transmission parameters "estimated" for Italy. Scenario parameters are in bold character (order of magnitude).

Between Compartments	m	p	n (order of magnitude)
0-1	0.00025		
1-0	0.00002		
1-2	**10^{-5}/10^{-6}**		**10^{-6}**
1-3			**10^{-5}**
1-5			**10^{-6}**
6-1	0.0096		
1-7		0.00023	
2-3	0.009		
2-6	**0.004 / 0.0004**		**10^{-6}**
2-7		[0.0002-0.0008]$^{+}$	
3-4	0.004		
3-6			**10^{-5}**
3-7		[0.0002-0.0008]$^{+}$	
4-5	[0.014-0.018]*		
4-6	[0.007-0.009]*		
4-7		[0.0002-0.0008]$^{+}$	
5-4	0.001		
5-6	[0.05-0.1]$^{\#}$		**10^{-6}**
6-5	0.001		
5-7		[0.0002-0.0008]$^{+}$	
6-7		[0.0002-0.0008]$^{+}$	

**Average length of treatment (hypothesis): 27-36 weeks.*

$^{+}$ Estimates from mortality studies used within the Consensus Conference on AIDS (Italy, 1998), reported in Ravà et al. (1998) and EMCDDA (1999b).

$^{\#}$ Average length of stay in compartment 5 (hypothesis): 5-10 weeks.

All the parameters μ and π represent transition rates per person of the origin compartment per week, the ν parameters are rates per week per pair. Using the simulation procedure described below, some impact analysis can be conducted to evaluate the influence of the scenario parameters on the course of the epidemic. It is also possible to make some further scenario analysis to evaluate the impact of an increased efficacy of the therapy services, by taking the scenario parameters fixed and using the therapy parameters as variable. Similarly, the natural history parameters can be used to make further "what if" scenario analyses (the results of such analyses will be presented elsewhere).

4. Some scenario analyses.

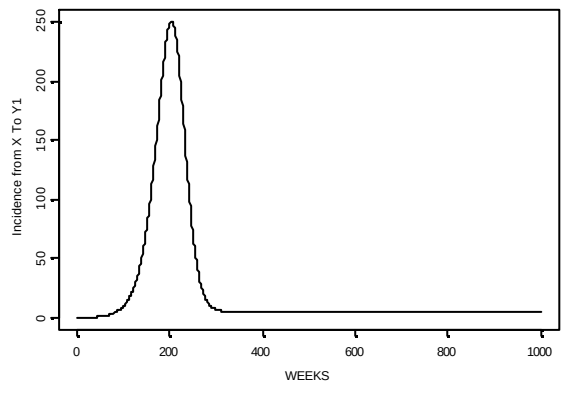
The simulation procedure, used to obtain scenario analyses, is written in S-plus for PC. All the parameters can be modified at the beginning of each run. The standard output comprises graphs of the prevalence curves in each compartment and of the incidence curves of major interest. It is possible to choose the total simulation time, which is measured in weeks.
Some scenario analyses have been conducted to study the impact of S_0 and some other parameters on the qualitative behaviour of the epidemic. The results are reported in graphic form in the following. As expected, the effect of S_0 is higher with respect to the other parameters (Rossi and Schinaia, 1998) considered for the simulation runs reported in this paper.

Basic n-scenarios

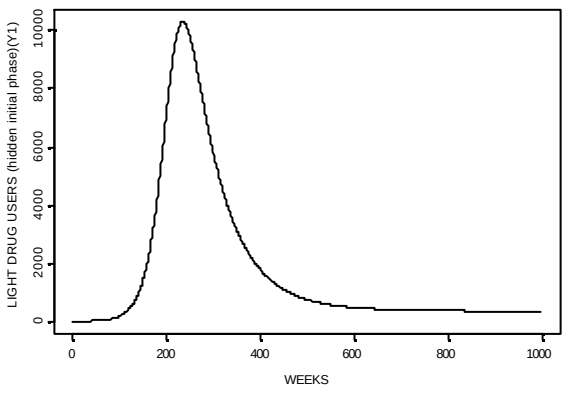
Some scenarios have been obtained using various hypotheses on the parameters. The results are reported in graphic form and a summary table (Table 2) describes the macro-characteristics of the different epidemics using the locations and the sizes of the peaks for the incidence and prevalence curves, the value of the cumulative curve of deaths at the end of the period taken into account, and the location of the maximum of the proportion of stayers, which represents the time of the beginning of the endemic phase (saturation effect).

Scenario 1.

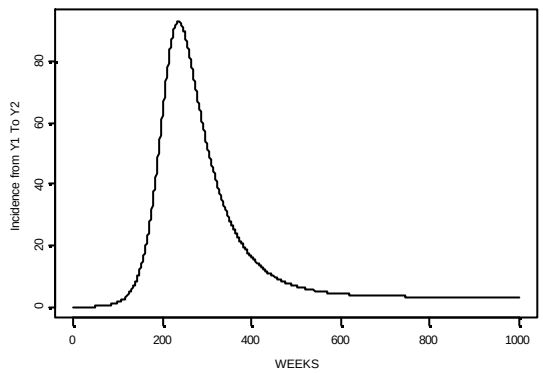
The present scenario is obtained using the parameters reported in Table 1 with S_0=0.98, $\mu_{12}=10^{-5}$, μ_{26}=0.004. Various graphs representing the curves of interest are reported below.



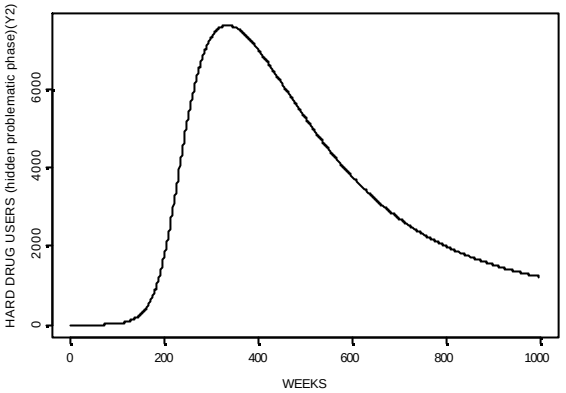
a) Incidence curve from susceptibles to light users



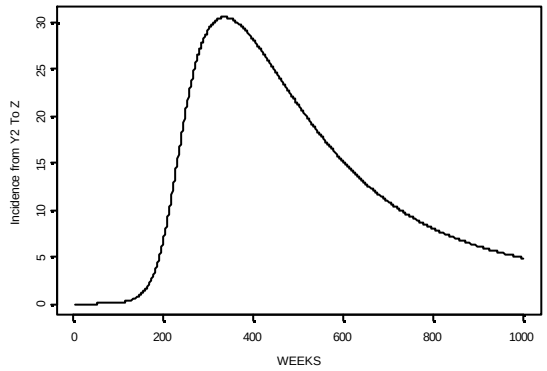
b) Prevalence curve of light users



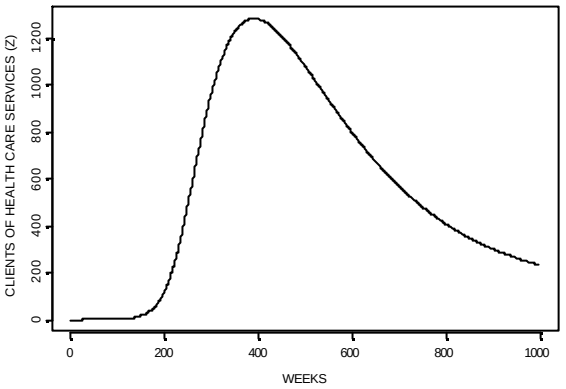
c) Incidence curve from light users to hard users



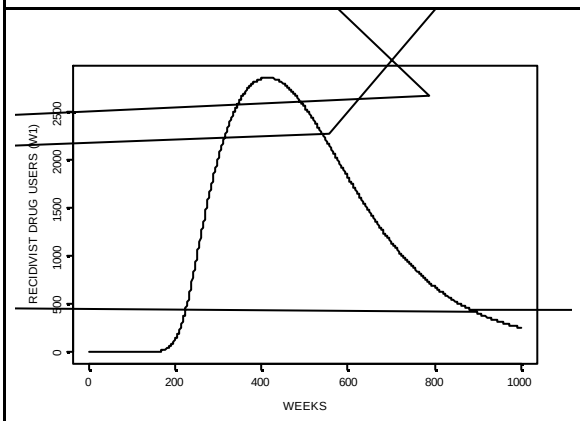
d) Prevalence curve of hard users



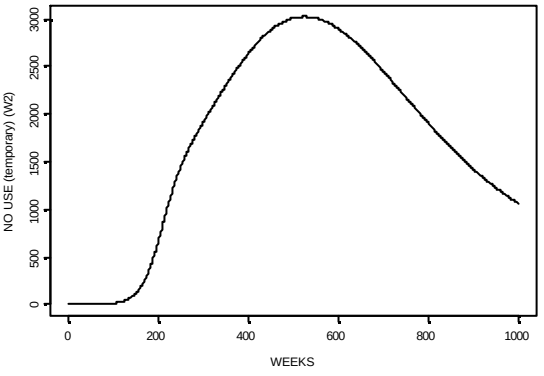
e) Incidence curve from hard users to therapy



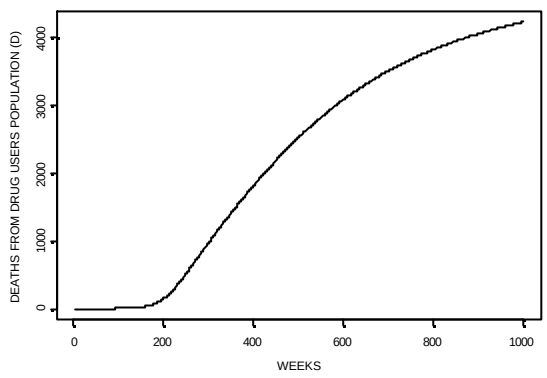
f) Prevalence curve of clients of therapy services



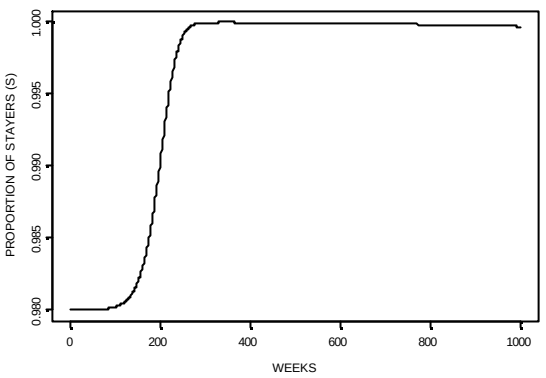
g) Prevalence curve of recidivist use



h) Prevalence curve of no use (temporary)



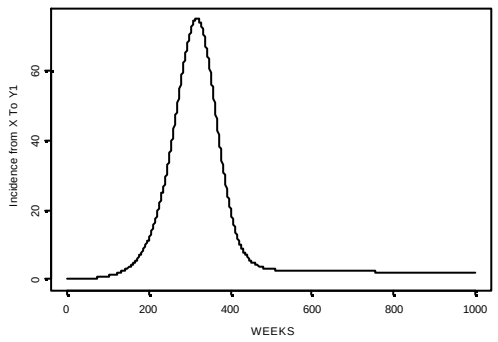
i) Cumulative curve of deaths



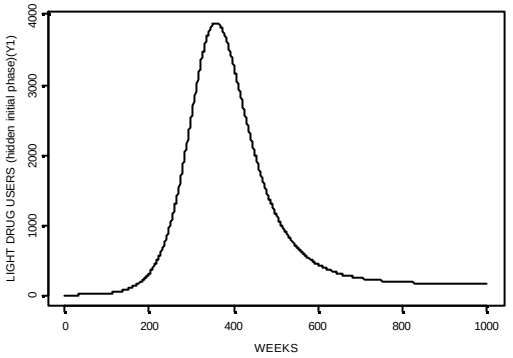
j) Proportion of stayers

Scenario 2.

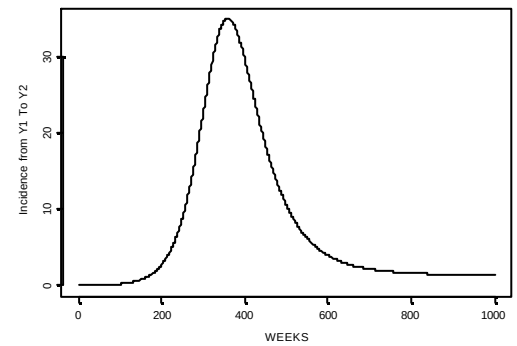
The present scenario is obtained using the parameters reported in Table 1 with S_0=0.99, $\mu_{12}=10^{-5}$, μ_{26}=0.004. The graphs representing the curves of interest are reported below.



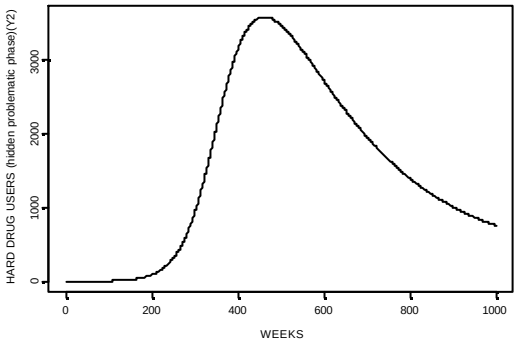
a) Incidence curve from susceptibles to light users



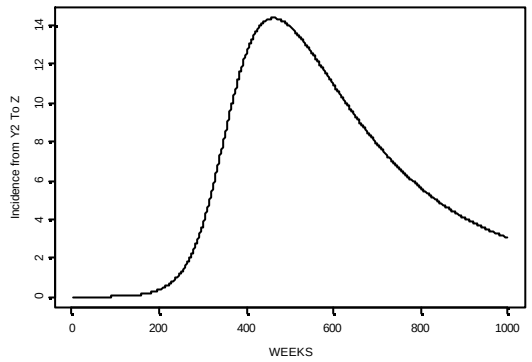
b) Prevalence curve of light users



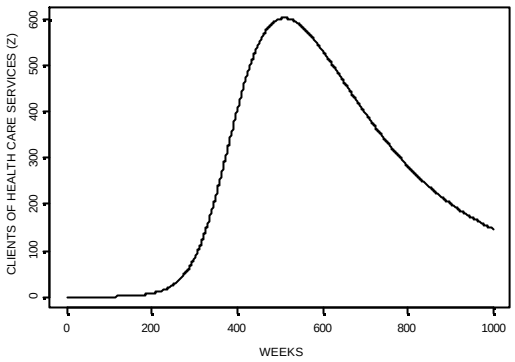
c) Incidence curve from light users to hard users



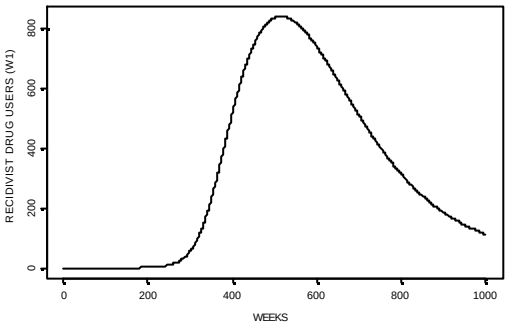
d) Prevalence curve of hard users



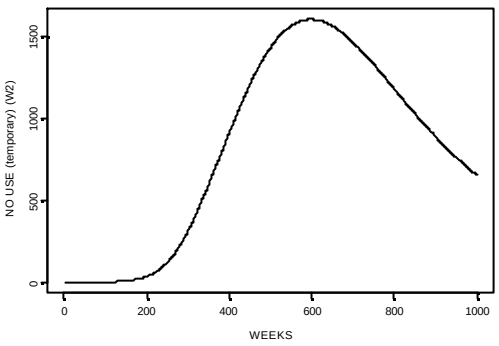
e) Incidence curve from hard users to therapy



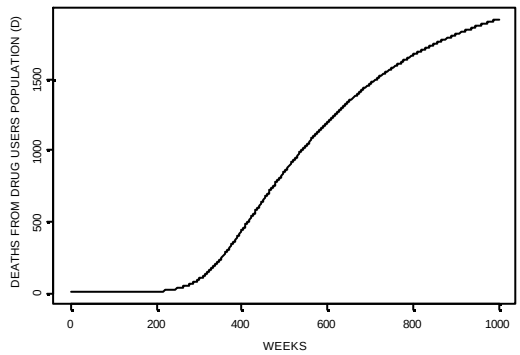
f) Prevalence curve of clients of therapy services



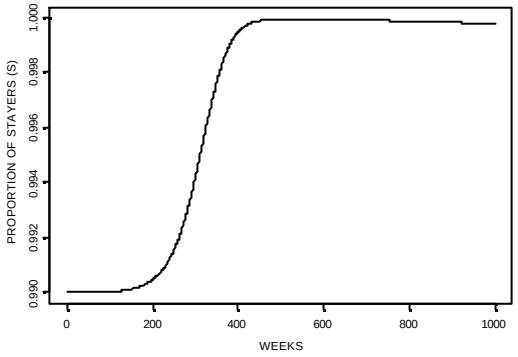
g) Prevalence curve of recidivist use



h) Prevalence curve of no use (temporary)



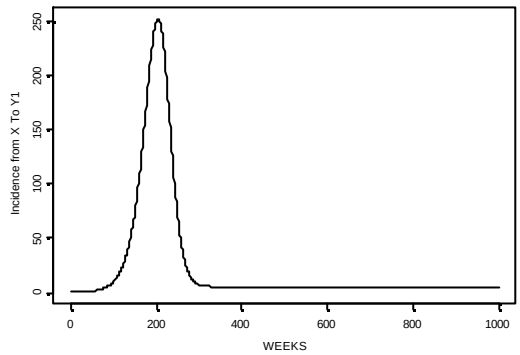
i) Cumulative curve of deaths



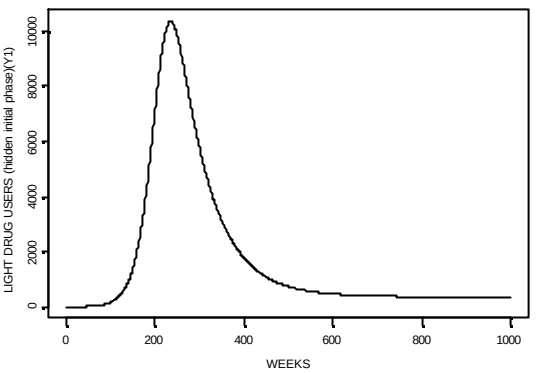
j) Proportion of stayers

Scenario 3.

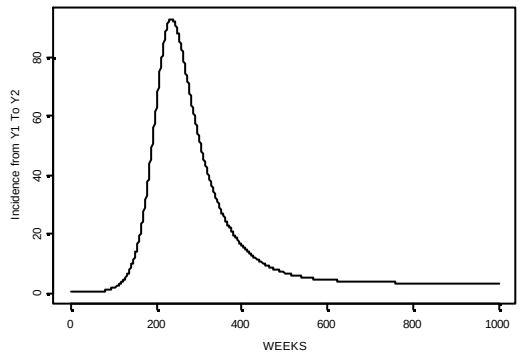
The present scenario is obtained using the parameters reported in Table 1 with S_0=0.98, $\mu_{12}=10^{-6}$, μ_{26}=0.004. The graphs representing the curves of interest are reported below.



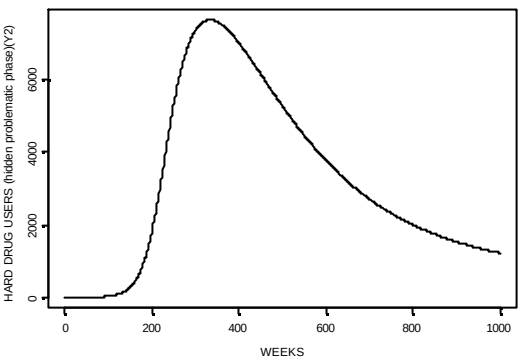
a) Incidence curve from susceptibles to light users



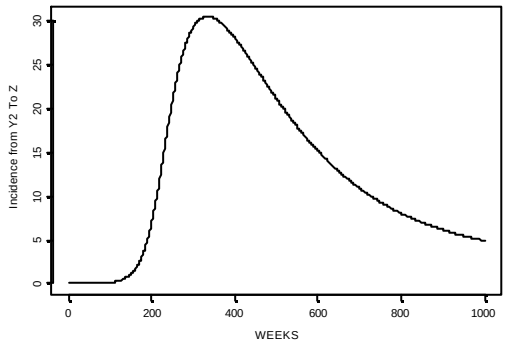
b) Prevalence curve of light users



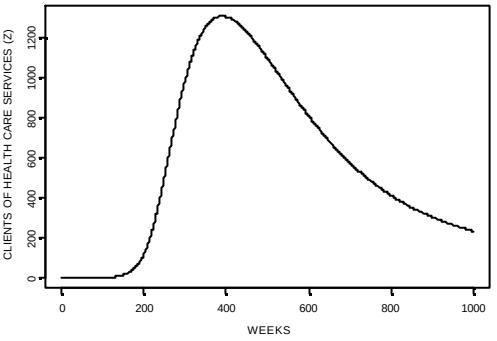
c) Incidence curve from light users to hard users



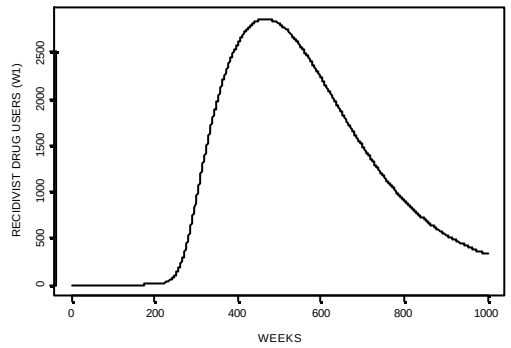
d) Prevalence curve of hard users



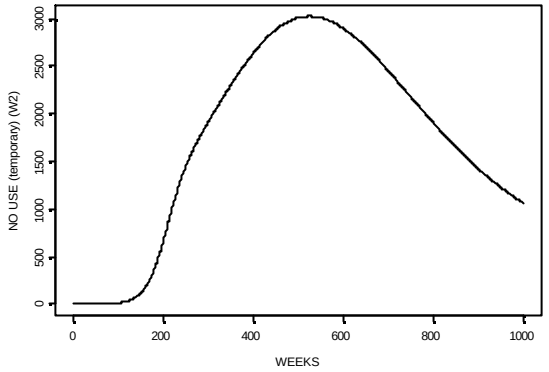
e) Incidence curve from hard users to therapy



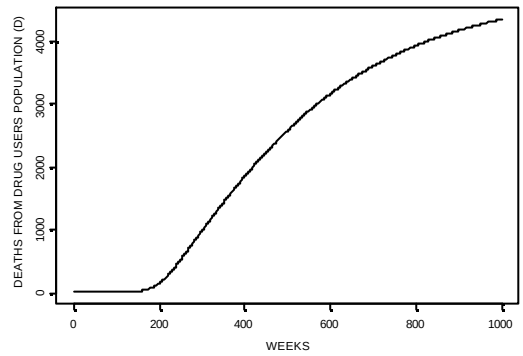
f) Prevalence curve of clients of therapy services



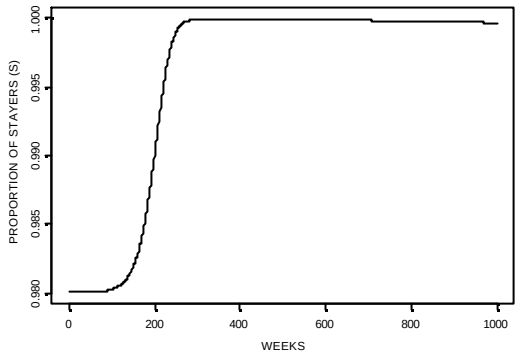
g) Prevalence curve of recidivist use



h) Prevalence curve of no use (temporary)



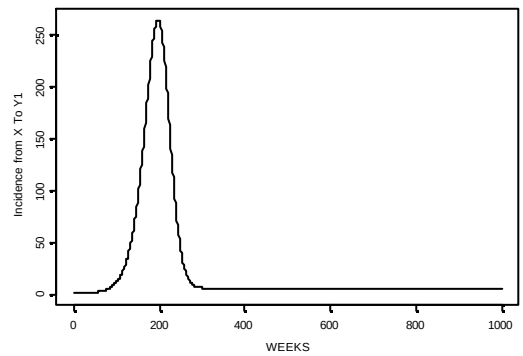
i) Cumulative curve of deaths



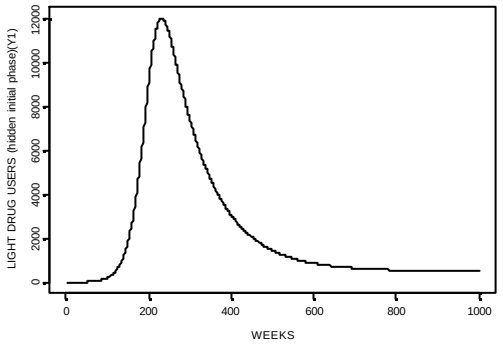
j) Proportion of stayers

Scenario 4.

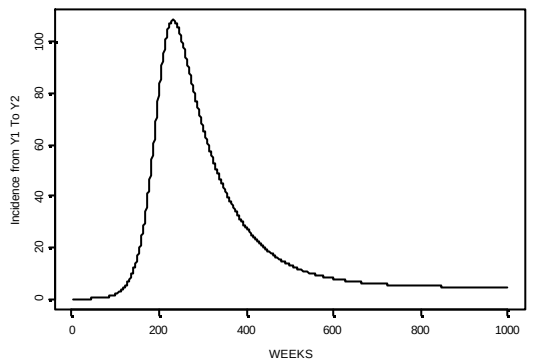
The present scenario is obtained using the parameters reported in Table 1 with S_0=0.98, $\mu_{12}=10^{-5}$, μ_{26}=0.0004. The graphs representing the curves of interest are reported below.



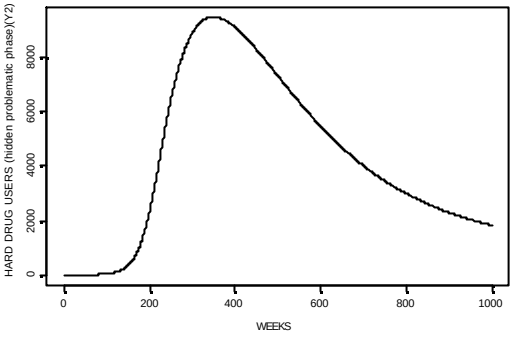
a) Incidence curve from susceptibles to light users



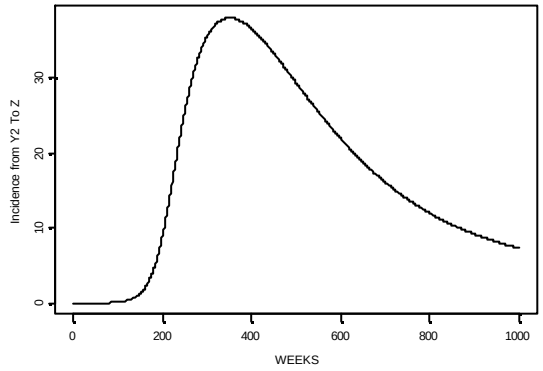
b) Prevalence curve of light users



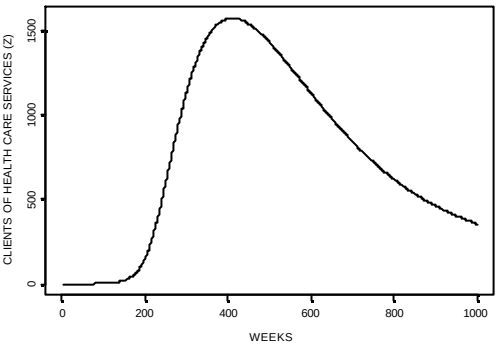
c) Incidence curve from light users to hard users



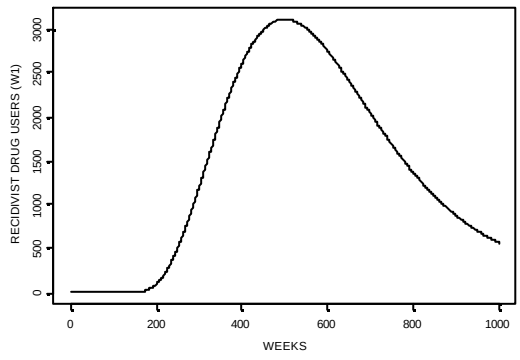
d) Prevalence curve of hard users



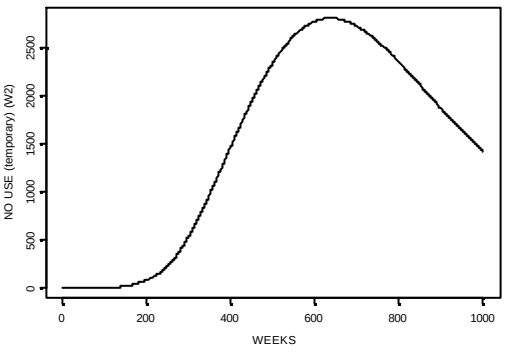
e) Incidence curve from hard users to therapy



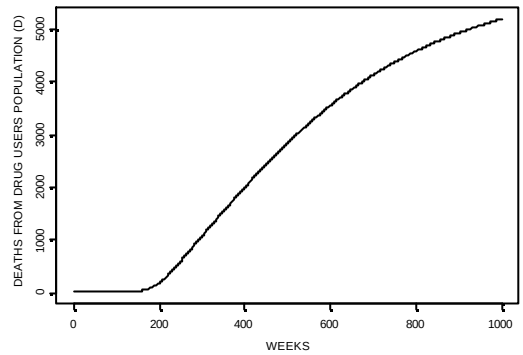
f) Prevalence curve of clients of therapy services



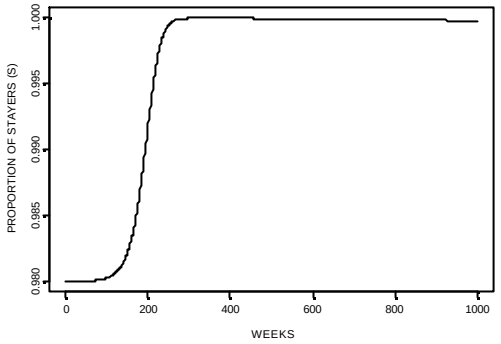
g) Prevalence curve of recidivist use



h) Prevalence curve of no use (temporary)



i) Cumulative curve of deaths

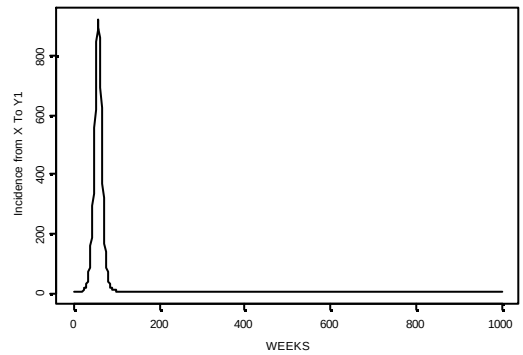


j) Proportion of stayers

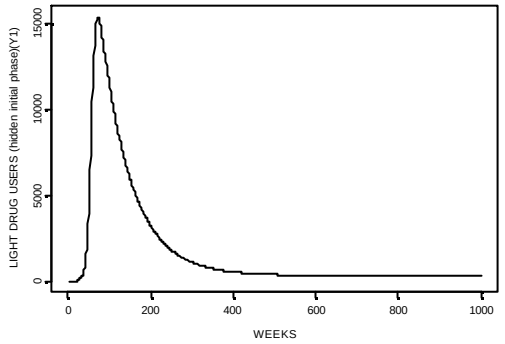
Alternative n-scenarios

Scenario 1.

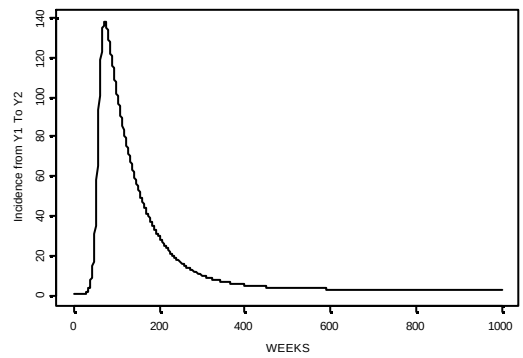
The present scenario is obtained using the parameters reported in Table 1, but modifying the order of magnitude of the parameters ν (10^{-5} instead of 10^{-6} and viceversa) and with $S_0=0.98$, $\mu_{12}=10^{-5}$, $\mu_{26}=0.004$. The graphs representing the curves of interest are reported below.



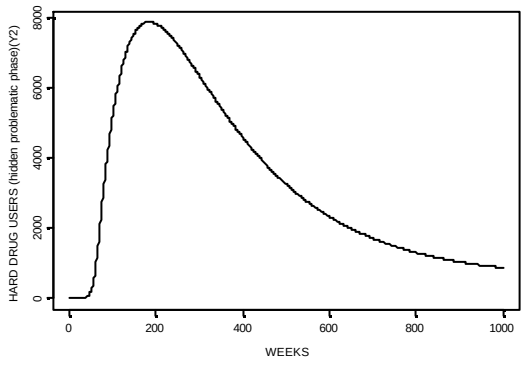
a) Incidence curve from susceptibles to light users



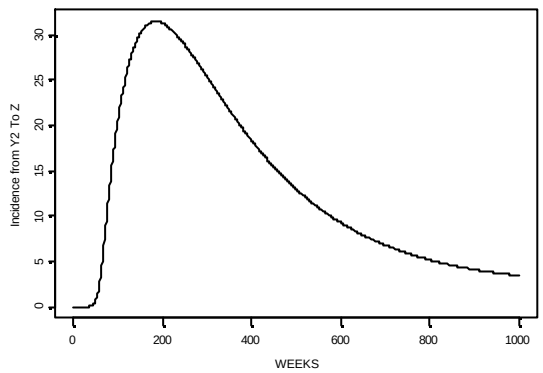
b) Prevalence curve of light users



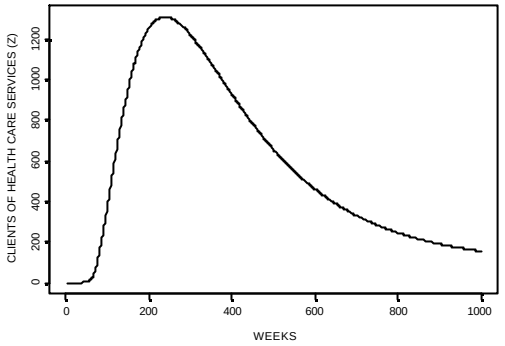
c) Incidence curve from light users to hard users



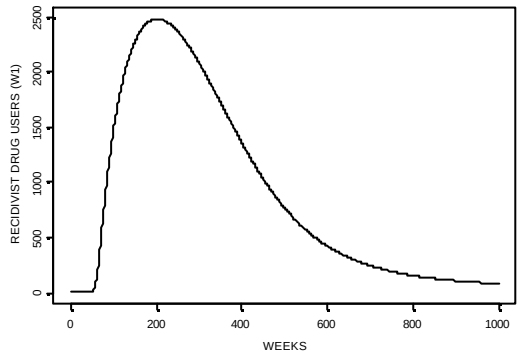
d) Prevalence curve of hard users



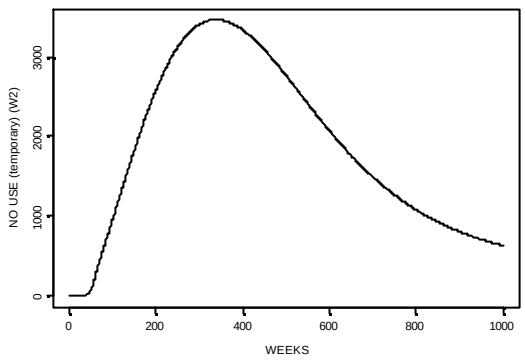
e) Incidence curve from hard users to therapy



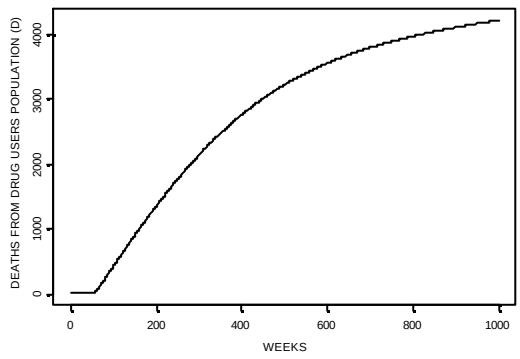
f) Prevalence curve of clients of therapy services



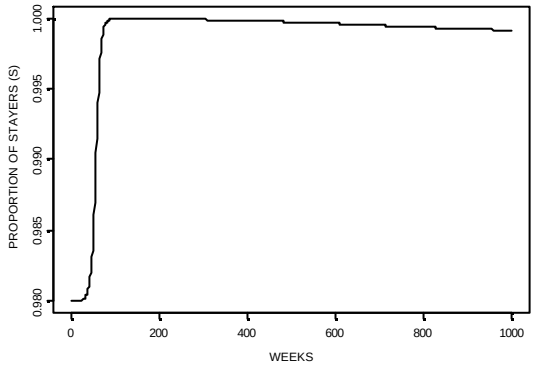
g) Prevalence curve of recidivist use



h) Prevalence curve of no use (temporary)



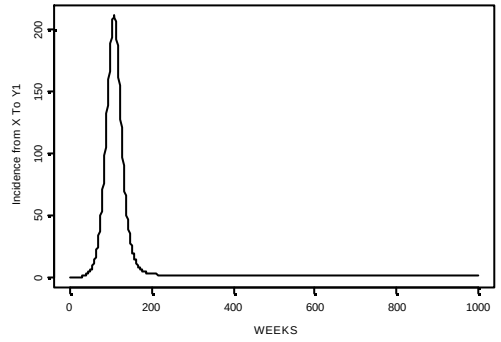
i) Cumulative curve of deaths



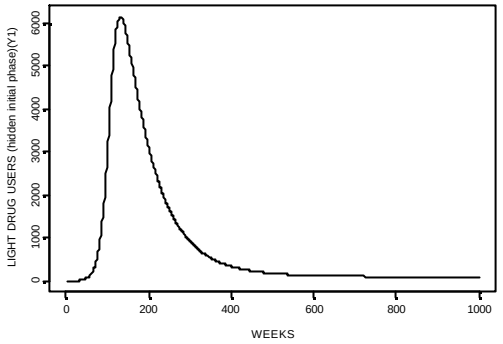
j) Proportion of stayers

Scenario 2.

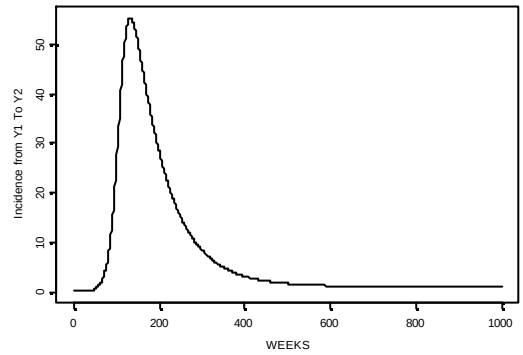
The present scenario is obtained using the parameters reported in Table 1, but modifying the order of magnitude of the parameters ν (10^{-5} instead of 10^{-6} and viceversa) and with $S_0=0.99$, $\mu_{12}=10^{-5}$, $\mu_{26}=0.004$. The graphs representing the curves of interest are reported below.



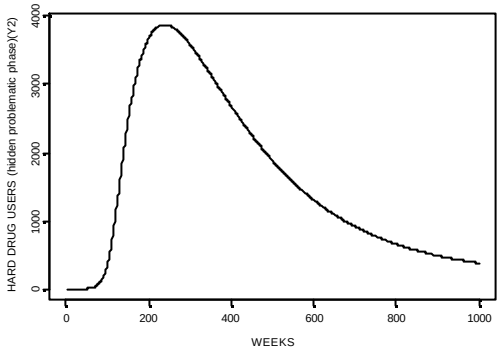
a) Incidence curve from susceptibles to light users



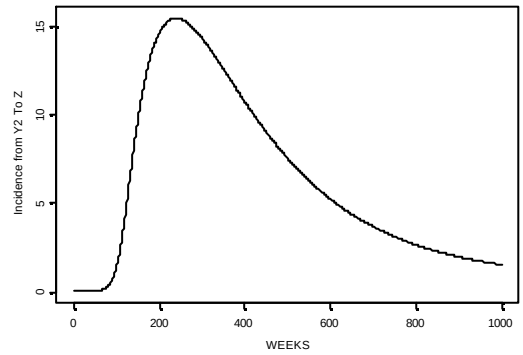
b) Prevalence curve of light users



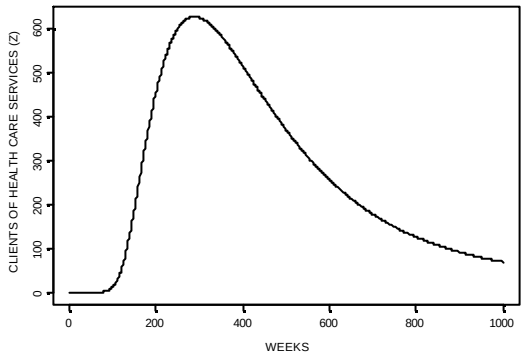
c) Incidence curve from light users to hard users



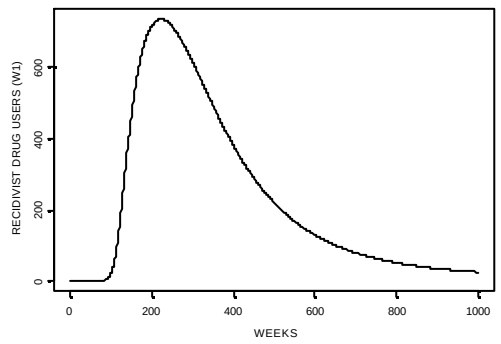
d) Prevalence curve of hard users



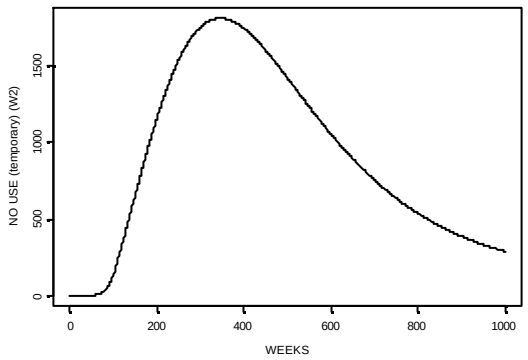
e) Incidence curve from hard users to therapy



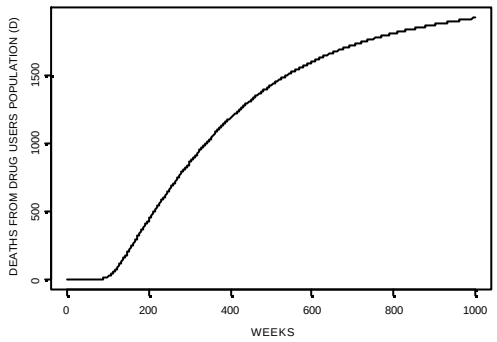
f) Prevalence curve of clients of therapy services



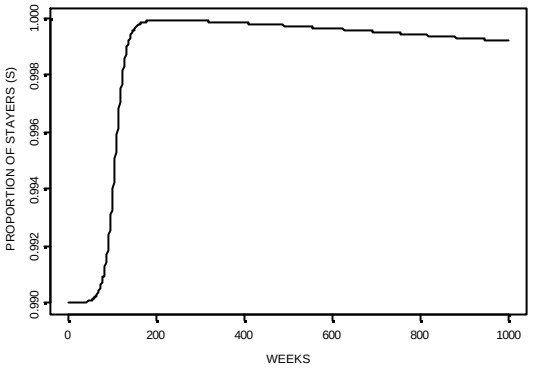
g) Prevalence curve of recidivist use



h) Prevalence curve of no use (temporary)



i) Cumulative curve of deaths



j) Proportion of stayers

Table 2. Macro-indicators describing the results of the various scenario analyses.

	Basic scenarios								Alternative scenarios			
	1		2		3		4		1		2	
Prevalence and incidence Curves♣												
	L•	S♦	L	S	L	S	L	S	L	S	L	S
a	**50**	**250**	**75**	**80**	**50**	**250**	**50**	**250**	**15**	**900**	**30**	**210**
b	**62**	**10000**	**90**	**3800**	**62**	**10000**	**62**	**12000**	**20**	**1500**	**40**	**6200**
c	**62**	**90**	**90**	**35**	**62**	**90**	**62**	**120**	**20**	**140**	**40**	**60**
d	**80**	**7500**	**125**	**3500**	**80**	**7500**	**80**	**9000**	**50**	**8000**	**60**	**3800**
e	**90**	**30**	**110**	**15**	**90**	**30**	**90**	**38**	**50**	**30**	**60**	**16**
f	**100**	**1200**	**125**	**600**	**100**	**1200**	**100**	**1600**	**62**	**1300**	**64**	**620**
g	**100**	**2700**	**125**	**800**	**100**	**2700**	**100**	**3000**	**62**	**2500**	**64**	**700**
h	**130**	**3000**	**150**	**1500**	**130**	**3000**	**150**	**2600**	**80**	**3200**	**80**	**1700**
Cumulative death curve: i												
	size		size		size		size		size		size	
	4100		**1800**		**4100**		**5000**		**4100**		**1800**	
Proportion of stayers: j♠												
	Location of the maximum		Location of the maximum		Location of the maximum		Location of the maximum		Location of the maximum		Location of the maximum	
	70		**100**		**70**		**62**		**18**		**40**	

Analyzing the graphs it is immediately possible to realize that the parameter with the highest impact on the course of the epidemic is S_0, which is a measure of the size of the group of susceptibles who are at risk of infection (core group). The bigger the core group the faster the

♣ The unit time for the incidence curves is one week, the prevalence curves represent the sizes concerning cross-sectional observations.

• The location, wherever reported, is measured in months since the beginning of the epidemic.

♦ The size, wherever reported, is measured as a rate per million inabithants.

♠ The proportion of stayers is an increasing function until the epidemic reaches the endemic phase, a decreasing function afterwards (Rossi, 1991; Rossi and Schinaia, 1998).

evolution of the epidemic and the higher the prevalence and incidence curves. The influence of the parameter which measures the pressure of the black market, namely m_{12}, appears to be less important. For what concerns the comparison between the basic hypothesis on the n parameters, that is that pushers are mainly hard users, and the alternative hypothesis, that is that pushers are mainly light users, on the basis of the results of the simulations, it appears that the alternative scenarios show faster evolutions of the epidemics. Table 2 summarizes the results in terms of macro-indicators (locations and sizes of the peaks for the incidence and prevalence curves, value of the cumulative curve of deaths at the end of the period taken into account, location of the maximum of the proportion of stayers) as described above. All the values related to sizes represent rates per million inhabitants.
It can be observed that the initial proportion of stayers S_0 is the parameter with the highest impact: there are higher differences in the macro parameters between scenario 1 and scenario 2 than between scenario 1 and scenario 3 or between scenario 1 and scenario 4.
Higher values of S_0 correspond to bigger and faster epidemics, whereas the parameter related to the pressure of the market is not so important, at least in the range of values presently investigated. A minor impact is also related to the parameter m_{26} which models transitions from ligth use to non use (spontaneous cessation). As already mentioned above, alternative scenarios correspond essentially to faster epidemics, ceteris paribus.

5. Further developments.

The use of suitable markers (marked processes) might allows to incorporate further descriptions of each individual involved in the epidemics, for instance marks may take into account the numbers of incarcerations, or the numbers of non-fatal overdoses, or the numbers of failed therapy interventions and so on. As a matter of fact, possible repeated therapy interventions should be incorporated in the model in order to get a more realistic picture of the problem drug user career; there is a general agreement that the time spent in the therapy compartment for the first therapy episode is differently distributed with respect to those related to the following episodes. These latter, for sake of simplicity, can be supposed identically distributed. In Figure 2, such kind of generalized model is presented. Writing the equations is straightforward on the basis of the graphical representation and, thus, the qualitative study can be easily obtained. In order to make scenario analyses further information is required on the “in” and “out” processes concerning the therapy compartments. Unfortunately, such data is not presently available, thus the only possibility is to use such transition parameters describing recidivist use as scenario parameters. The complexity of such analysis is evident from the graphical representation of the model and the interpretation of the results might be quite uncertain and unreliable. Thus, it is much better to wait until the therapy data-sets comprise reliable and complete information to allow the external estimation of the parameters of interest.

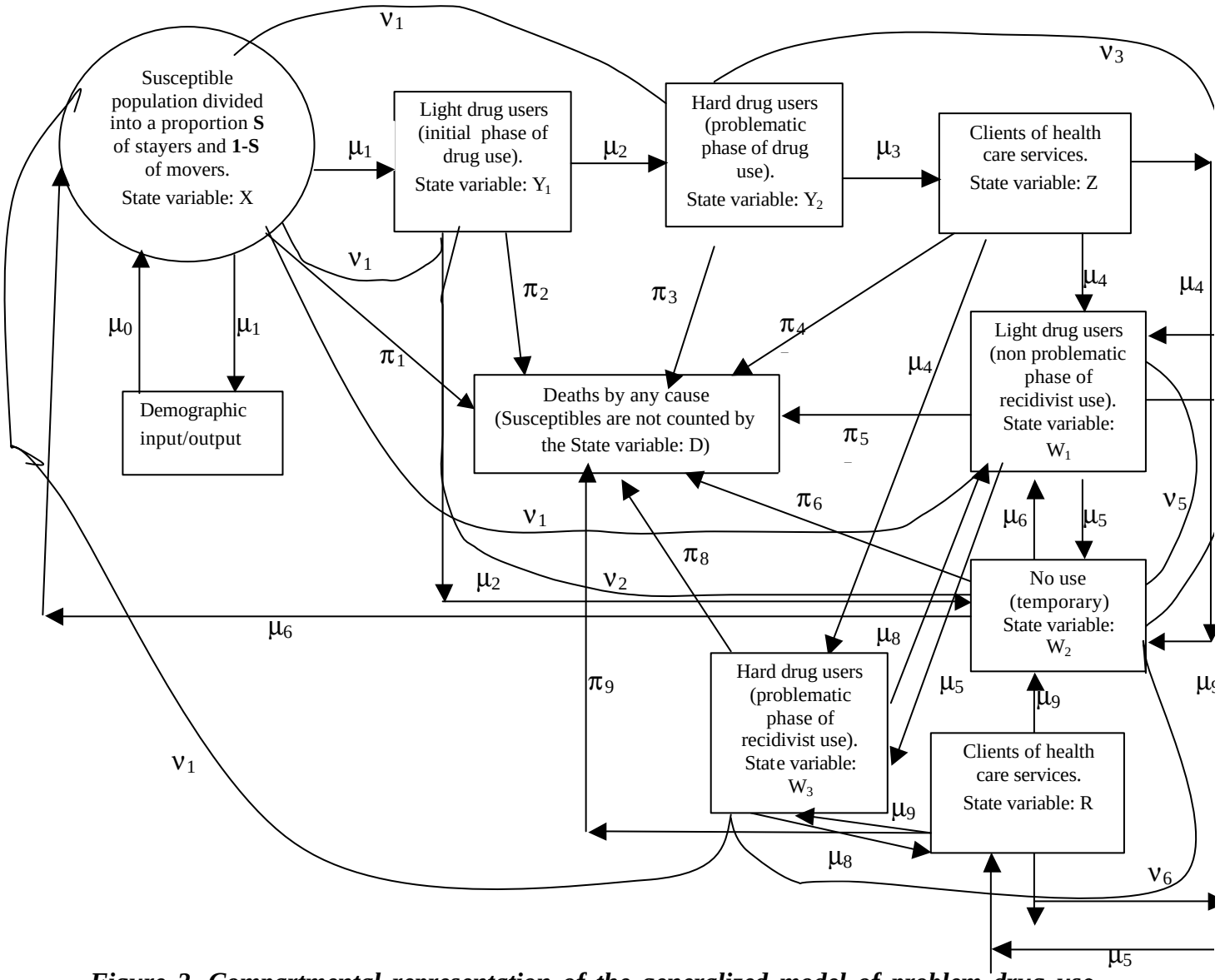


Figure 2. Compartmental representation of the generalized model of problem drug use epidemic.

In order to obtain more realistic models, the transition parameters should not be taken as constant but should be represented as functions taking into account the history of drug use for any individual, represented by statistical variables supposed known (covariates), and the history of policy interventions (availability of services, law enforcement activities …), represented by other covariates (time dependent) and, possibly, by latent variables. This would results in a very realistic but, at the same time, unreliable and untractable transmission model of no use. Some years ago, Edward Kaplan published a paper with the title "Can bad models suggest good policy?" where it was clearly explained how simple models, taking into account only the main peculiarities of the phenomenon of interest and neglecting minor effects, which may just mask the relevant behaviours, are much more useful in order to assess the main consequences of policy interventions than complex models. I have always thought that this title is the best description of the usefulness of simple mathematical models for addressing complex policy issues. Just to make a final example, consider to be interested in evaluating the impact of some kind of prevention campaign or intervention on problem drug use addressing susceptibles.

Using the Mover-Stayer model proposed above (simpler version), this issue can be easily dealt with by introducing the possibility that a mover become a stayer during the period of the campaign (or soon after). This can be simply modelled as follows. Let T_1 be the prevention campaign starting time, τ the duration time of the campaign (in weeks) and $0<K<1$ the impact parameter (overall proportion of susceptibles passing from the Mover population to the Stayer population).
Let us consider $T_2 = T_1 + \tau$ and T^* in the time interval (T_1 , T_2), then the instruction of the simulation procedure that calculates the proportion of stayers is completed, at each simulation step for T*, as follows[7]:

$$S(T^*) = S(T^*) + K\frac{\left[\,1 - S(T^*)\,\right]}{t}.$$

If we want to include in the model the effect of law enforcement, which may influence the movers pushing a proportion to become stayers, due to the information about the adverse consequences of using illegal drugs (Beherens et al., 1999), K may be considered dependent on the numbers of addicts assisted by the health care services or the numbers of incarcerated addicts or both. Further generalizations might concern a more realistic approach to modelling the length of stay in the various compartments taking into account the heterogeneity of individual behaviours. Most of these issues will be addressed in future contributions.

References

1.Albert P.S., A Mover-Stayer Model for longitudinal marker data. *Biometrics*, 55-4, 1252-1257, 1999.
2.Bailey N.T.J., The use of operational modelling of HIV/AIDS in a system approach to public health decision making, *Mathematical Biosciences*, 107, 413-430, 1991.
3.Bailey N.T.J., An improved hybrid HIV/AIDS model geared to specific public health data and decision making, *Mathematical Biosciences*, 117, 221-237, 1993.
4.Bailey N.T.J., HIV/AIDS Core-group dynamic and public health action, in *Modeling the AIDS epidemic*, Kaplan E.H. & Brandeau M.L. eds., Raven Press, New York, 29-52, 1994.
5.Beherens D.A., Caulkins J.P., Tragler G., Haunschmied J.L., Feichtinger G., A dynamic model of drug initiation: implications for treatment and drug control, *Mathematical Biosciences*, 159, 1-20, 1999.
6.Billard L., Dayananda P.W.A., Drug addiction-pushers generated from addicts, Biometrical Journal, 35, 227-244, 1993.
7.Dietz K., 1988, On the Transmission Dynamics of HIV *Mathematical Biosciences*, 90, 397-414.
8.EMCDDA, 1999a, "Pilot project to estimate time trends and incidence of problem drug use in the European Union", Final report, Lisbon.
9.EMCDDA, 1999b, "Co-ordination of implementation, follow-up and analysis of cohort studies on mortality among drug users in European Union Member States", Final report, Lisbon.
10.Hadeler K.P., Modeling AIDS in Structured Populations in *Proceedings of I.S.I. 47th Session*, Paris, 83-99, 1989.

[7] The simulation procedure for the AIDS epidemic can be dowloded from: www.mat.uniroma2.it/~rava/tovaids_home.html

11. Hunt L.G., Chambers C.D., *The Heroin Epidemics*, SPECTRUM PUBLICATIONS INC, NY, 1976.
12. Jacquez J.A., Simon C.P., Koopman J., Sattenspiel L., Perry T., Modeling and Analyzing HIV Transmission: The Effect of Contact Patterns, *Mathematical Biosciences*, 92, 119-199,1988.
13. Kretzschmar M., Dietz K., The Effect of Pair Formation and Variable Infectivity on the Spread of an Infection without Recovery, *Mathematical Biosciences*, 1997.
14. Morgan T.M., Aneshensel C.S., Clark V.A., Parameter estimation for Mover-Stayer models: analysing depression over time. *Sociological Methods and Research*, 11, 345-366, 1983.
15. Ravà L., Rossi C., Pasqualucci C. and Schinaia G., Estimating the size of the HIV/AIDS epidemic: complementary use of empirical bayesian back calculation and the mover-stayer model for gathering the largest amount of information, *SIMULATION*, 71-4, 213-227, 1998.
16. Ravà L. et al., The heroin epidemic in Italy: space-time modelling, data analysis and forecasts, *in preparation*.
17. Rossi C., A stochastic mover-stayer model for HIV epidemic, *Mathematical Biosciences*, 107, 521-545, 1991.
18. Rossi C., Schinaia G., The Mover-Stayer Model for the HIV/AIDS Epidemic in Action, *Interfaces*, vol. 28, no. 3, 1998, 127-143, 1998.
19. Rossi C., "Estimating the prevalence of injecting drug users on the basis of Markov models of the HIV/AIDS epidemic: applications to Italian data", *Health Care Mangement Science*, 2, 173-179, 1999a.
20. Rossi C. "Monitoring drug control strategies: hidden phenomena, observable events, observable times", *International Journal of Drug Policy*, 10-1, 131-144, 1999b.
21. Rossi C., Operational models for problem drug use epidemic: the Mover-Stayer approach to heterogeneity, *submitted*.

Acknowledgements

This work was partially funded by the European Monitoring Centre on Drugs and Drug Addiction in the framework of the European Network to develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions (TSER/DGXII project ERB 4141 PL980030). The present paper was presented at the Workshop on "Dynamic drug policy: understanding and controlling drug epidemics", Vienna International Centre (UNDCP), 22nd-24th May, 2000; the present version is improved, with respect to the previous one, due to the discussions during the meeting. The simulation procedure used to obtain scenario analyses has been developed by Dr. Alessandro Ghizzoni.

Appendix 5

The SEM model for Italy

INTRODUCTION

Drug addiction is only one aspect of the "drug system", a complex universe not completely explored because of the impossibility of direct surveys, non-homogeneous terminology, data collection and method of analysis that do not allow significant comparisons of available information.
This work does not give an answer, e.g. to 'why does an individual take drugs', but represents a rigorous multivariate analysis of indicators that have an influence on the complex drug universe.
For a preliminary application, the statistical units are the twenty Italian regions and the collected data, for each variable, are referred to the period 1991-1996.
We separately assessed six structural equation models referred to each year of the period 1991-1996 by a statistical software package, AMOS (equivalent to LISREL) and we obtained six "good" final models, in the sense that the probability level, p, is higher than 0.05.

STRUCTURAL EQUATION MODEL (SEM)

The structural equation models (SEMs) are multivariate regression models.
SEM can be graphically represented by means of path diagrams.
Observed variables are assumed to be imperfect measures of latent variables, which are theoretic constructs not directly observed but which have implications for the relationships among variables (observed). In a path diagram, the observed variables are represented by a square or rectangle and the latent variables by a circle or ellipse.
All variables, both latent and observed, can be of two different types: exogenous and endogenous. The first ones are variables that are exclusively influenced by factors lying outside the model while variables that are influenced from inside the model are called endogenous.
In a SEM, regression coefficients are called structural coefficients and they are represented by unidirectional straight lines. Covariances between variables are represented by bi-directional curved lines.

DATA AND MODELS FOR ITALY

The structural equation model considered here is a "macro" model. The statistical units are aggregates of territory-homogeneous people. The selected territorial disaggregation is regional (20 Italian regions) and the period refers to six years (from 1991 to 1996). Data are normalized by resident population. We examined only OPIATE USE which represents the most problematic illegal substance of abuse in Italy.
The statistical variables used are:
Div Sep = incidence of divorces and separations;
Area = Residence area. Variable with three values, 1=South, 2=Centre, 3=North;
Unemployment = incidence of unemployed as compared with the work force;
Funds = regional funds for recovery activities;
Risk Pop = incidence of people between fifteen and twenty-four years old;

Income = monthly mean pro-capite income;
School dropout = incidence of students who leave high school.
We consider the hypothesis that these first seven observed variables directly influence an endogenous latent variable OPIATE USE, measured by only one observed variable: Reported = incidence of young people who use drugs and then are reported to prefectures. Then, we introduced two structural relations between OPIATE USE and HEALTH CONSEQUENCES, that is health condition of drug users, and between OPIATE USE and CRIMINALITY, that is delinquency linked to drug use.
The latent variable HEALTH CONSEQUENCES is measured by the following variables:
Deaths = incidence of deaths due to drug abuse;
Aids = incidence of Aids cases among drug users;
Treated = incidence of new treatments.
The second latent variable, CRIMINALITY, is measured by the following indicators:
Dealer inmates = incidence of incarcerations for dealing drugs;
Addict inmates = incidence of incarcerations of addicts.
Moreover, we suppose that Area directly influences CRIMINALITY.
Therefore, the initial model is the following (Figure 1):

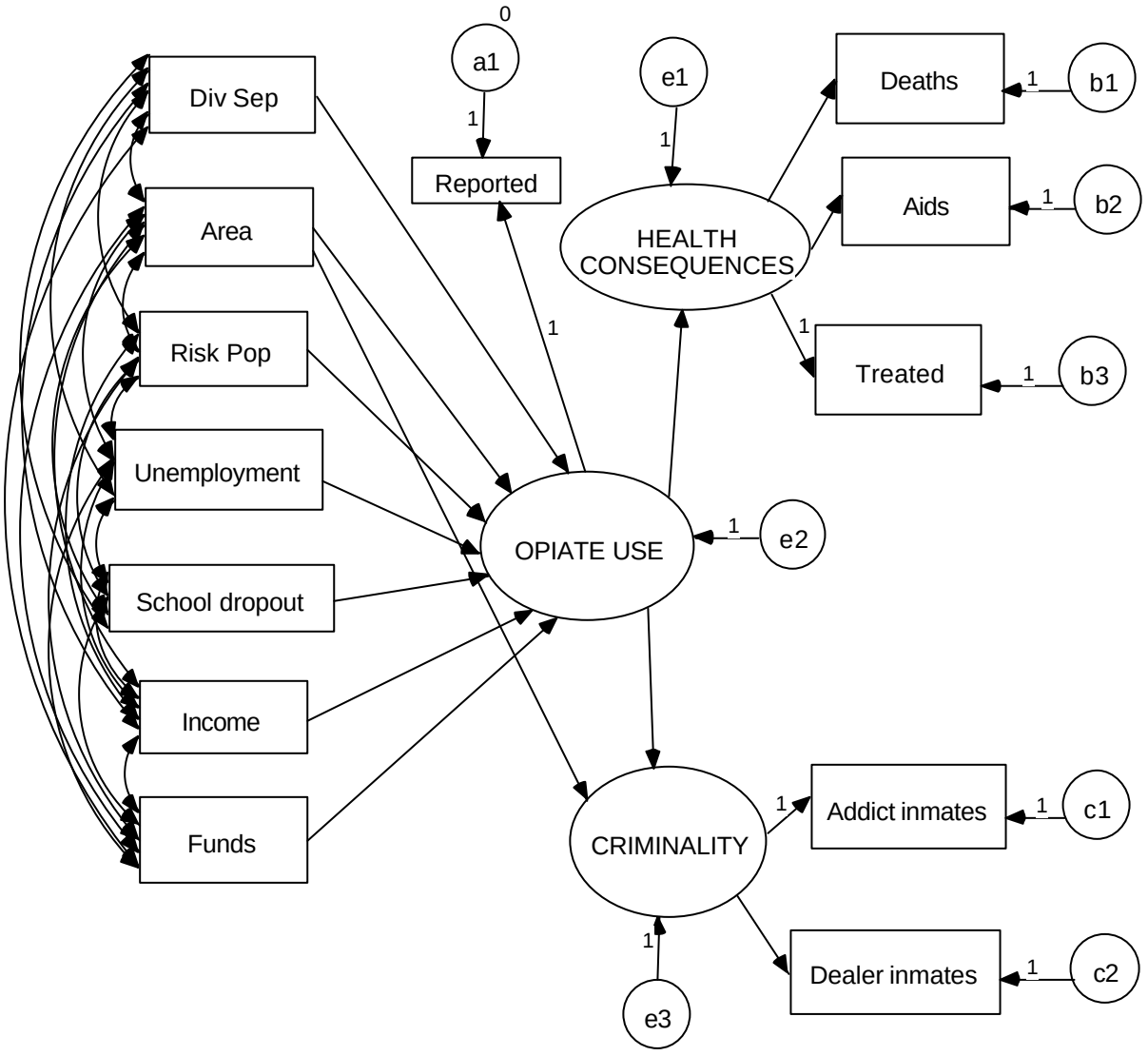


Figure 1 Initial model on drug use

Each latent variable in a model must be given a unit of measurement. This can be accomplished in this way: for a given latent variable, specify its scale to be equal the scale of one of its indicator variable, that is, fix one of the structural coefficients, leading from the latent variable to an associated indicator variable, to 1.0. The unit of measurement of OPIATE USE has been set equal to unit of measurement of Reported, the unit of measurement of HEALTH CONSEQUENCES has been set equal to that one of Treated, the unit of measurement of CRIMINALITY has been set equal to that one of Addict inmates. Moreover, if there is only one observed indicator variable for a given latent variable, we assume the accuracy of the indicator to be perfect, that is, we assume that the observed variable perfectly measures the latent construct and, thus, we set its measurement error variance to 0. Hence, since Reported is the only indicator of OPIATE USE, its measurement error variance is equal to 0.

Initial model has been modified according to AMOS outputs until data fits the a priori hypothesised model.

For each modification, AMOS provides tables reporting the value of chi-square, degree of freedom, probability level and goodness of fit indices to be used to choose the best modification step.

The final model obtained is the following (Figure 2):

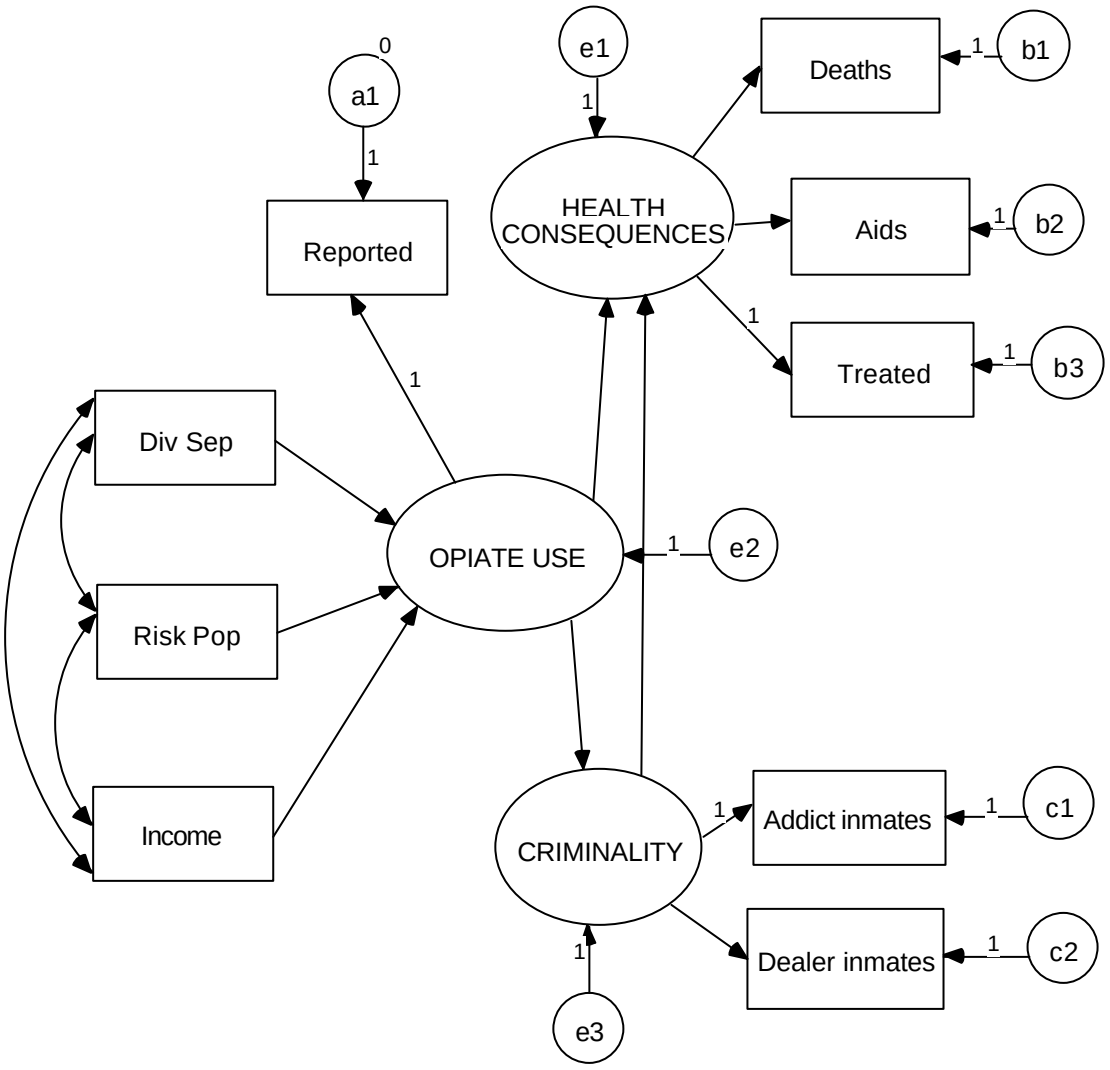


Figure 2 Final model on drug use

CONCLUSIONS

Let us justify the steps done.
We excluded the variable School dropout, because, maybe, the incidence of dropouts does not supply a correct measurement of what we wanted to introduce in the model, that is, the educational level in the different Italian regions. In fact, a more appropriate variable could be the mean rate of people that continue studying and then enter at the university.
Also variable Funds is taken away from model. Even if, when government grants funds to certain associations, their bureaucratic process is followed, the results that they reach are not controlled and reported, so they are not decisive for evaluating drug addiction phenomenon and drug system.
As the coefficient between OPIATE USE and Risk Pop is negative, a greater incidence of 15-24 years old is a protective factor with respect to the use of illegal substances. Perhaps, this also means that it is more timely to consider people of 15-39 years old as population at risk.
Another important structural coefficient, which appears in all of the six models, is that between HEALTH CONSEQUENCES and CRIMINALITY. To explain this coefficient, we have to observe that our study concerns the "drug system" of which drug addiction is only one aspect. An additional aspect is drug market of which drug user delinquency, that is, CRIMINALITY, is an indicator. After these introductory statements, it is easy to realise how drug market can have an influence on health, that is, on the variable HEALTH CONSEQUENCES. If great quantities of "harmful" drug are distributed, incidence of deaths from overdose would increase. Thus, to better understand this coefficient, one could develop a further model in which there are other market indicators to know what variable mostly influences health.
The structural coefficient between Income and OPIATE USE is negative, so a greater income leads to a lesser opiate use, and vice versa.
In the final model, we did not consider the variable Area. If we forced the entry of this variable, we obtained a negative coefficient between Area and OPIATE USE and this means that opiate use is more widely spread in the North than in the South of Italy.
Finally, we do not really conclude that data fits the model because all the analyses are based on a small sample size (twenty statistical units). The ratio between the number of parameters and the sample size should be at least 1:5, preferably 1.10 or 1:50. Therefore, a further analysis based on provinces or towns should be performed.