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

FINAL REPORT

PILOT PROJECT TO ESTIMATE TIME TRENDS AND INCIDENCE OF PROBLEM DRUG USE IN THE EUROPEAN UNION
CT.98.EP.07

University of Rome Tor Vergata
Dipartimento di Matematica:
Carla Rossi
Lucilla Ravà
Laura Re

EMCDDA:
Lucas Wiessing
Richard Hartnoll

Participants:
E. van Ameijden
M. Buster
D. De Angelis
S. Heisterkamp
M. Hickman
F. Mariani
J. Ribeiro

Please use the following citation:

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: March 1999.

Contact Details

University of Rome Tor Vergata
Dipartimento di Matematica
Via Ricerca Scientifica
00133 Rome, Italy
Tel: +39 06 72594676
Fax: +39 06 72594699

European Monitoring Centre for Drugs and Drug Addiction
Rua Cruz de Santa Apolónia 23/25
1100, Lisboa
Portugal.
Tel: +351 1 8113016
Fax: +351 1 8137943

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

Contents

1. Introduction.

1.1. Background.

There has been much work recently on developing methodologies for indirectly estimating the prevalence of problem drug use. These are important for several reasons. Firstly, because a proportion of the drug using population always remains hidden. Secondly, problem drug use is too rare and too many biases operate for general population surveys to provide reliable estimates. Thirdly, problem drug use is a chronic condition and, therefore, in the first instance estimates of prevalence are essential to enable policy makers to plan services.

However, prevalence estimates are simply a single or range of values for a certain point in time. In order to monitor and design public health prevention programmes, and consider future service planning clearer evidence on the incidence of problem drug use is required. In particular, an indication of whether the number of problem drug users are growing (epidemic), falling or stable (endemic).

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. Moreover, it is unclear in what way the available "indirect indicators" relate to the underlying occurrence of problem drug use: any change over time may be related to other factors, such as case ascertaintment, service provision, and policing policy.

Hunt and Chambers, in their early work, derived estimates of the incidence or diffusion of heroin epidemics in the U.S.. Firstly they examine the delay between "onset" of use and entry into treatment; and secondly they examine the likelihood of heroin use transferring from a person to another. These insights presage new techniques used for understanding the epidemiology of infectious disease and utilising surveillance data to estimate incidence and prevalence (e.g. back-calculation methods).

Back-calculation has been developed considerably through work on AIDS projections. Briefly, knowledge of the numbers infected with the HIV virus and the incubation period distribution are used to predict the number of AIDS cases which are expected in the future, on the other hand, if the number of AIDS cases is known and information on the incubation period distribution is available, estimates for those previously infected with the HIV virus can be obtained. Since the proportion of those infected who eventually will develop AIDS is unknown and as we are back-calculating on the basis of diagnosed AIDS cases, the method provides estimates only for those infected who will eventually develop the disease. It is this number, however, that health authorities have to finance for. It is possible to use the same approach to back-calculate from observed data on drug users presenting to treatment, or firstly recorded by some agency for any reason, (same as AIDS diagnosed cases), the incidence of starting problem drug use (same as HIV infection) at least for those who will eventually be observed.

Consider, for example, the case where information is available about users first attending treatment (observed incidence), if we know or estimate the proper "incubation period", i.e. the time lag between the first problem drug use and the first treatment demand, we can back-calculate the incidence of problem use on the basis of the known numbers of observed users. 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…; and the survival or length of drug using career). Both the incubation period and the subsequent estimates are highly sensitive to the quality and completeness of reporting data.

1.2. The objectives of the pilot project.

The pilot project focussed on two main objectives:

1) The first one was to investigate the latency period (time from first use of drug to first treatment demand) distribution, which is analogous to a disease incubation period, using data from the pilot sites and survival analysis models. The analysis should comprise the prognostic study of the latency period distribution as a function of possible covariates such as sex, age at first problem drug use, route of consumption, socio-economic and educational level etc., (Rossi, 1999).

The study of the latency period may be of interest to monitor programmes, in particular those aiming to reduce the time a drug user remains hidden to the health care services.

2) The second objective was to try to apply Back-Calculation (BC) methods for estimating incidence of problem drug use from treatment data (incidence of new cases in treatment). The investigation on the distribution of latency period is a prerequisite for this second aim.

It must be stressed that all the methods to be used either directly or indirectly rely upon surveillance data. An important subsidiary outcome of this pilot work, therefore, is the identification of key gaps and priorities for development of surveillance of problem drug use in cities and countries in Europe.

1.3. The Working Group.

The work has been carried out in three 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), and, with some limitations, the site of Casal Ventoso in Lisbon (Portugal). The limitations are due to the fact that the end point for Casal Ventoso data sets is different from the end point for the other sites. For these latter it corresponds to presentation for treatment whereas for Casal Ventoso it corresponds to the first registration in one of the agencies providing social help (not therapy) to hard core drug addicts.

The estimated distributions were then utilized to backcalculate historical trends in incidence of problem drug use from first reports of people in treatment, by using the Empirical Bayesian Back Calculation procedure described below, that was specifically set up for the present project.
The working group comprise representatives from Italy (F. Mariani, L. Ravà, L. Re, C. Rossi), The Netherlands (M. Buster, S. Heisterkamp, E. van Ameijden) and UK (D. De Angelis, M. Hickman). A special sub-project was devoted to Portugal (J. Ribeiro).

F. Mariani, L. Re, M. Buster, E. van Ameijden, and M. Hickman were involved in issues related to the analysis of the latency period, and general surveillance issues; L. Ravà and S. Heisterkamp were involved with methodological issues related to the generalization and application of the Empirical Bayesian Back-Calculation procedure, and D. De Angelis acted as methodological consultants in both parts for UK country report.

J. Ribeiro (Lisbon) provided raw data about the Casal Ventoso site to estimate the latency period distribution. All the statistical analyses to estimate the latency period were conducted in Rome by L. Re using the SPSS statistical package. Similarly, the BC estimation of incidence of problem drug use was conducted in Rome by L.Ravà, using the Empirical BC procedure, developed by S. Heisterkamp, which is described in the following.

The working group communicated via e-mail and held one plenary meeting to finalize the workplan in Rome on October 5/6 1998 (the program is attached).

A preliminary meeting was held in Rome on September 29, because UK participants could not attend the plenary meeting in October. The participants (M.Hickman and D. De Angelis from UK and C. Rossi and L. Ravà from Italy) discussed the available data and the quality and problems arising from estimating the latency period distribution for London. Some methodological issues, related to adapting the Back-calculation procedures developed for estimating the incidence of HIV infection to the present situation, were also considered.
A further small meeting was held in Amsterdam in December: L. Re and L. Ravà met the representatives of the Netherlands to discuss the preliminary results regarding both the analysis of the latency period for Amsterdam and the Back-Calculation procedure. Further small meetings in Lisbon between C. Rossi and M. Hickman and C. Rossi and J. Ribeiro and C. Rossi and F. Mariani in Pisa were held to discuss local results.

2. The study of the latency period.

2.1. Introduction.

The latency period distribution was estimated from raw treatment data.

Suitable standard statistical models were used to study such distribution. Exploratory analysis was conducted by means of the **Kaplan-Meyer** method for the global sample and for various stratifications to identify important covariates. Using the most influent covariates, the multivariate **Cox regression model** was then applied to estimate the regression parameters and evaluate the different impact (prognostic analysis) of the covariates on the latency period distribution (Collet, 1994; Marubini & Valsecchi, 1995). Finally, the best parametric models of the latency period distribution were estimated by means of P-Plot method.

2.2. The data needed for the 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

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

Data should be provided, preferably, in SPSS, or Excel format.

2.3. The exploratory data analysis.

From the exploratory analysis, the latency period appears to be remarkably similarly distributed over the different sites, with a median of between four and six years and

an average of between five and seven years (Table 2.1.). This time-lapse, however, appears to be much longer than this in young drug users and inner-city drug users.

Table 2.1. 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

The differences of the latency period, corresponding to different age classes at first use of drug, that were observed, for example, in the sub-sample related to Rome (metropolitan area), are summarised in Table 2.2..

Table 2.2. Summary statistics for Kaplan-Meyer analysis, stratified with respect to "age at first use" (Rome-metropolitan).

Age class	*Sample size*	*Mean*	*Stand.dev.*	*1° quart.*	*Median*	*3° quart.*
Less than 16	555	9.2	0.2	6	8	13
16-21	2675	7.0	0.1	3	6	10
More than 21	1426	4.7	0.1	1	3	7

Differences relating to ethnicity also were observed, whenever this covariate could be included in the analysis, in particular for the Amsterdam and London data sets (see country reports).

The multivariate prognostic analysis was conducted by means of Cox model, estimating the parameters by SPSS stepwise forward algorithmic procedure. Some results for the Lazio region are reported here only to illustrate the method. The influence of the covariates is measured by the multivariate regression coefficients and is described below.

Covariate:

***Age: reference class* "less than 20"**:

- +36% for the risk function of "20-24" with respect to **"less than 20"**;
- +74% for the risk function of "25-34" with respect to **"less than 20"**;
- +218% for the risk function of "over 34" with respect to **"less than 20"**.

This means that the entry intensity function of addicts who started their problem drug use with their age in the class "20-24" is 1.36 times the entry intensity function of addicts who started their problem drug use with their age in the class "<20". Similarly for the other age classes. In simple words, this means that:

- the expected latency period for addicts who started their drug use at an age comprised in the interval 20-24 is about 0.74 of the expected latency period of those who started in the reference age class (less than 20);

- the expected latency period for addicts who started their drug use at an age comprised in the interval 25-34 is about 0.57 of the expected latency period of those who started in the reference age class (less than 20);

- the expected latency period for addicts who started their drug use at an age over 34 is about 0.46 of the expected latency period of those who started in the reference age class (less than 20).

The same methodology was applied to the different data sets coming from the different sites. The results available for the area of Amsterdam (The Netherlands), of the Greater London (UK) and Rome, i.e. the three capital cities of the project, show an impressive agreement of the estimated means and medians and of the influence of the main covariates, specifically "age at first use".
These results give 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. Further details of the analysis appear in the country reports.
To better explain the methodology used for the exploratory data analysis and the interpretation of the results, an example is furtherly considered in Appendix 1.

2.4. The parametric estimation of the baseline distribution of the latency period.

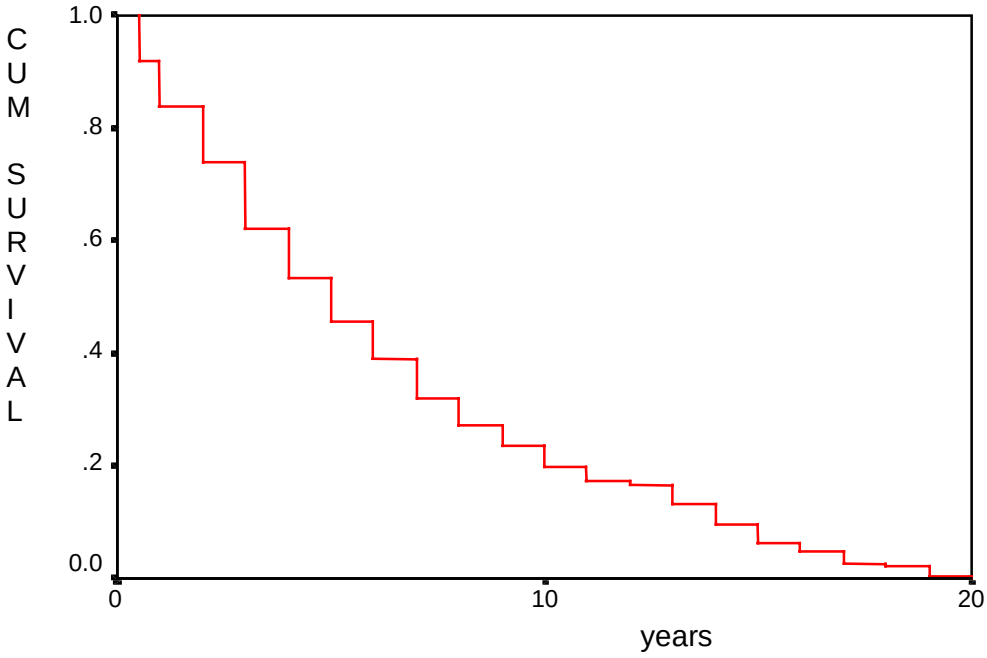
In order to use the distribution of the latency period in the Back-Calculation procedure, described below, it is useful to obtain a suitable parametric form, such as a gamma or a Weibull distribution, for the baseline latency density, that is to say for the distribution of the latency period for the global sample without covariates. The covariates can then be included using proper (standard) models

(Collet, 1994; Marubini & Valsecchi, 1995), as possible further developments.

In the following, an example of the methodology used to estimate a suitable parametric distribution of the latency period is reported. The estimation was performed by the SPSS P-Plot standard procedure: the estimation of the parameters of the chosen density (gamma, Weibull…) is conducted using linear regression for the expected versus the empirical quantiles. In simple words, the expected quantiles are expressed as functions of the unknown parameters of the density to be estimated, such parameters are then estimated by optimization of the fit with respect to the empirical quantiles computed from the sample. If the fit is good, then the plot of the expected versus the empirical quantiles shows data close to the diagonal of the square (0,1)x(0,1) on a Cartesian plane. For better understanding the procedure, the results obtained for the Amsterdam data are reported in the following.

First, the Kaplan-Meyer curve for the global sample is reported in Figure 2.1., with the summary statistics, then the P-Plot obtained for the study of the goodness of fit related to various parametric models are reported in Figures 2.2.-2.4.

Figure 2.1. Survival function of the variable "latency period" (Amsterdam), estimated by the Kaplan-Meyer method.



Summary statistics.

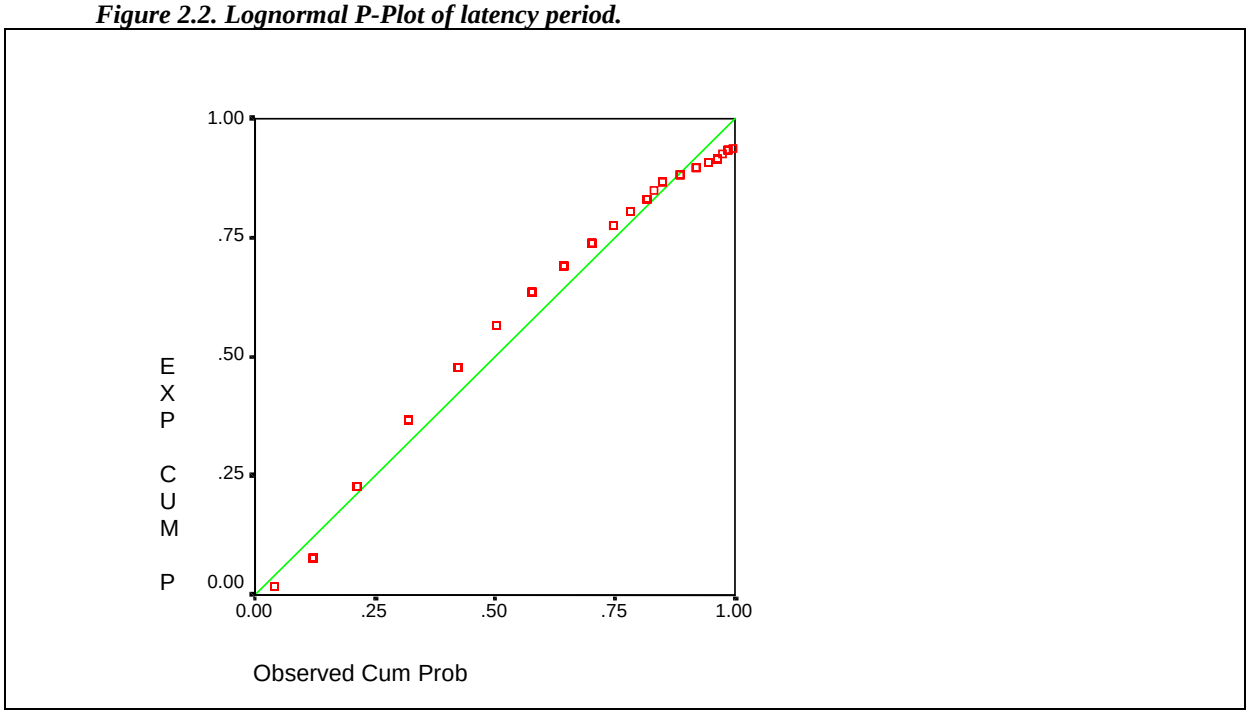
	Survival Time	Standard Error	95% Confidence Interval	
Mean:	6.3	.4	(5.6,	7.0)
Median:	5.0	.5	(4.1,	5.9)

	Percentiles		
	25.00	50.00	75.00
Value	2.00	5.00	9.00
Standard Error	.29	.48	.76

P-Plot (lognormal)

Expected Lognormal quantiles calculated using Blom's estimation formula.
Lognormal distribution parameters estimated: scale=4.22 shape=1.006.

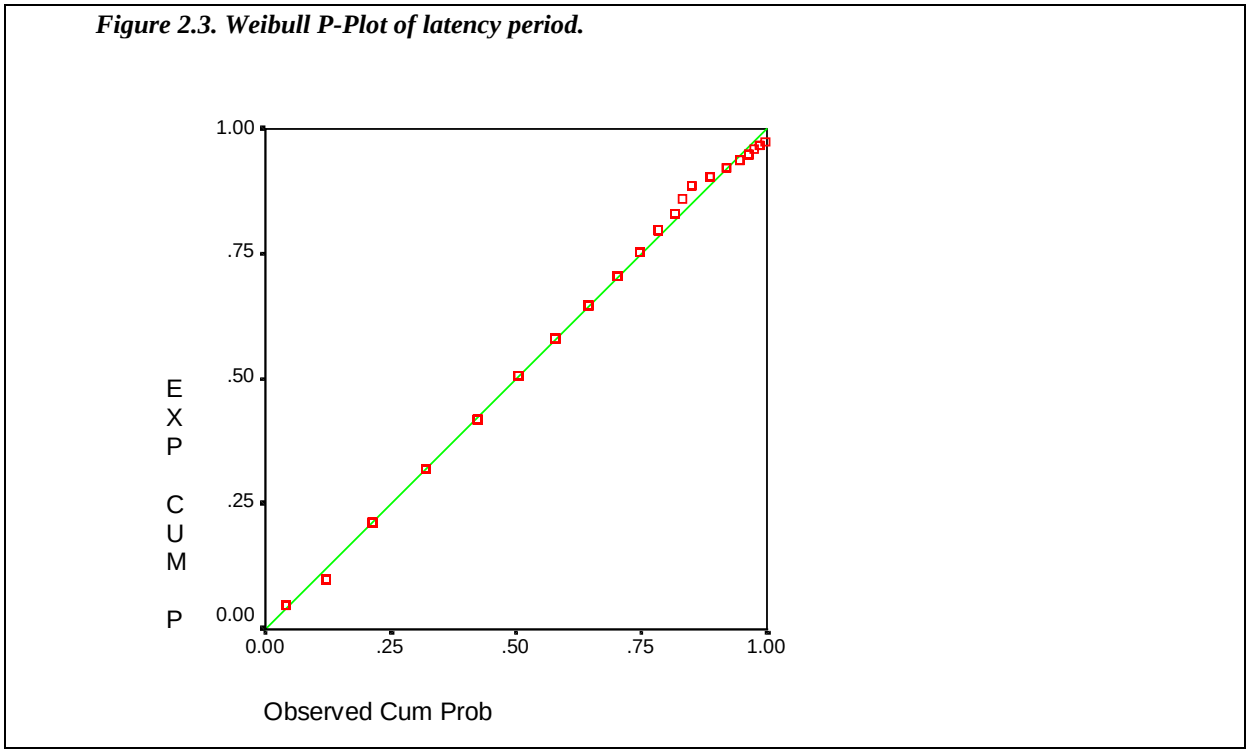
Figure 2.2. Lognormal P-Plot of latency period.



P-Plot (Weibull)

Expected Weibull quantiles calculated using Blom's estimation formula.
Weibull distribution parameters estimated: scale=7.45 shape=1.16.

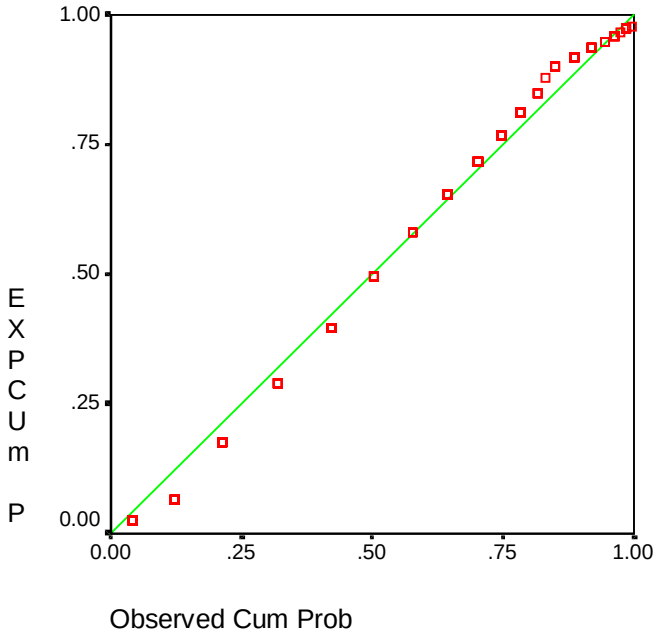
Figure 2.3. Weibull P-Plot of latency period.



P-Plot (Gamma)

Expected Gamma quantiles calculated using Blom's estimation formula.
Gamma distribution parameters estimated: shape=1.47 scale=0.21.

Figure 2.4. Gamma P-Plot of latency period.



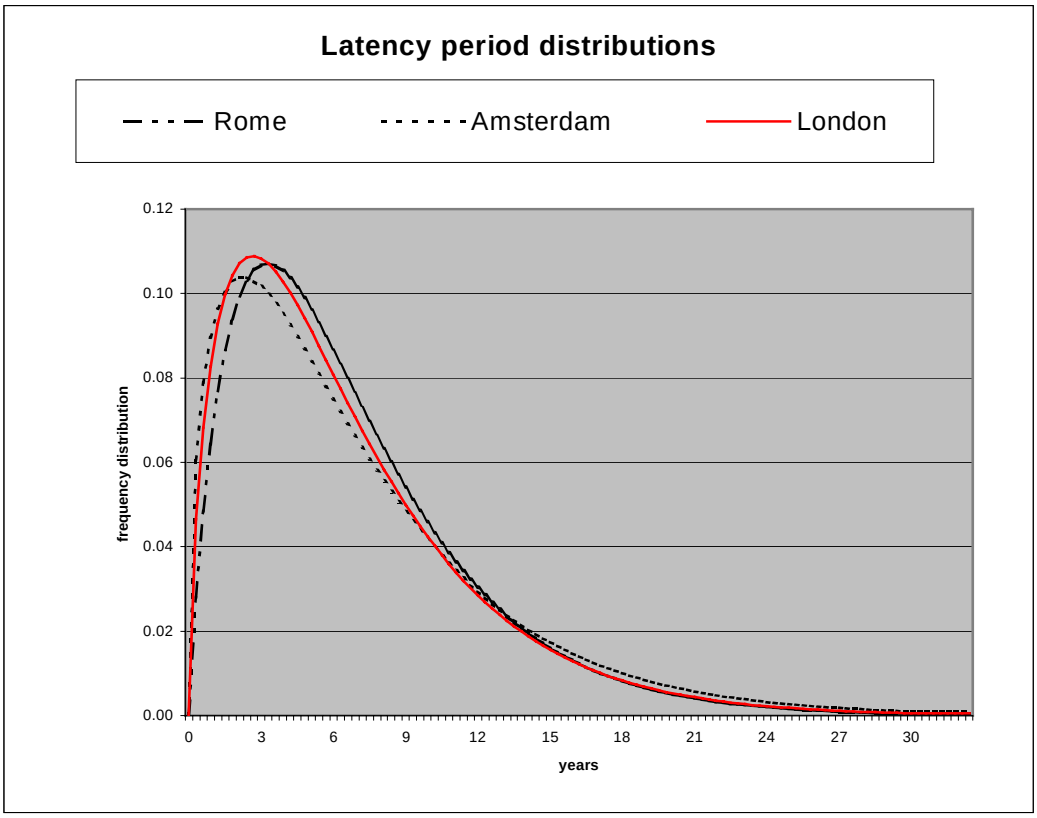
Comparing the graphical results, reported in Figures 2.2.-2.4., we obtain that:

- the lognormal density is not suitable to represent, in parametric form, the latency period distribution, as the behaviour of the data does not fit the linear trend of the diagonal;

- both the Weibull and the Gamma can be used to represent, in parametric form, the latency period distribution, but the fit of the Weibull is better than that of the Gamma.

Both the Weibull and Gamma densities were used in the BC estimation procedure for the Amsterdam sample.
For all the other sites the Weibull and Gamma densities appear to be the best parametric models of the latency period distribution. Figure 2.5. shows the three Gamma functions related to the latency period distributions of the capital cities.

Figure 2.5. Gamma functions related to the latency period distributions of the capital cities.



The parameters of the distributions shown in the above graph 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.

In Table 2.3. the parameters obtained by the P-Plot for the Gamma and Weibull models to be used in the BC procedure are shown. It must be observed that for Italy the latency period models corresponding to Latina were chosen to represent the national distribution as they appear to be the most suitable to adapt the empirical therapy incidence curve to the projected one obtained by BC, as explained below (see also country report).

Table 2.3. Latency period distributions used in the BC estimation (parametric models).

ITALY (Latina)		**NETHERLANDS (Amsterdam)**		**UNITED KINGDOM (London)**	
Gamma	***Weibull***	***Gamma***	***Weibull***	***Gamma***	***Weibull***
Shape: 1.51 Scale: 0.30	Shape: 1.28 Scale: 5.33	Shape: 1.47 Scale: 0.21	Shape: 1.16 Scale: 7.45	Shape: 1.66 Scale: 0.25	Shape: 1.29 Scale: 7.16

3. Estimating the incidence of problem drug use through the Back-Calculation.

3.1. Introduction.

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 in section 2.

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's $A(t)$ 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, $s \le t$ 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 $h(s)$, $s \le t$.

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.

3.2. 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) 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 one time interval. The AIDS incidence in each

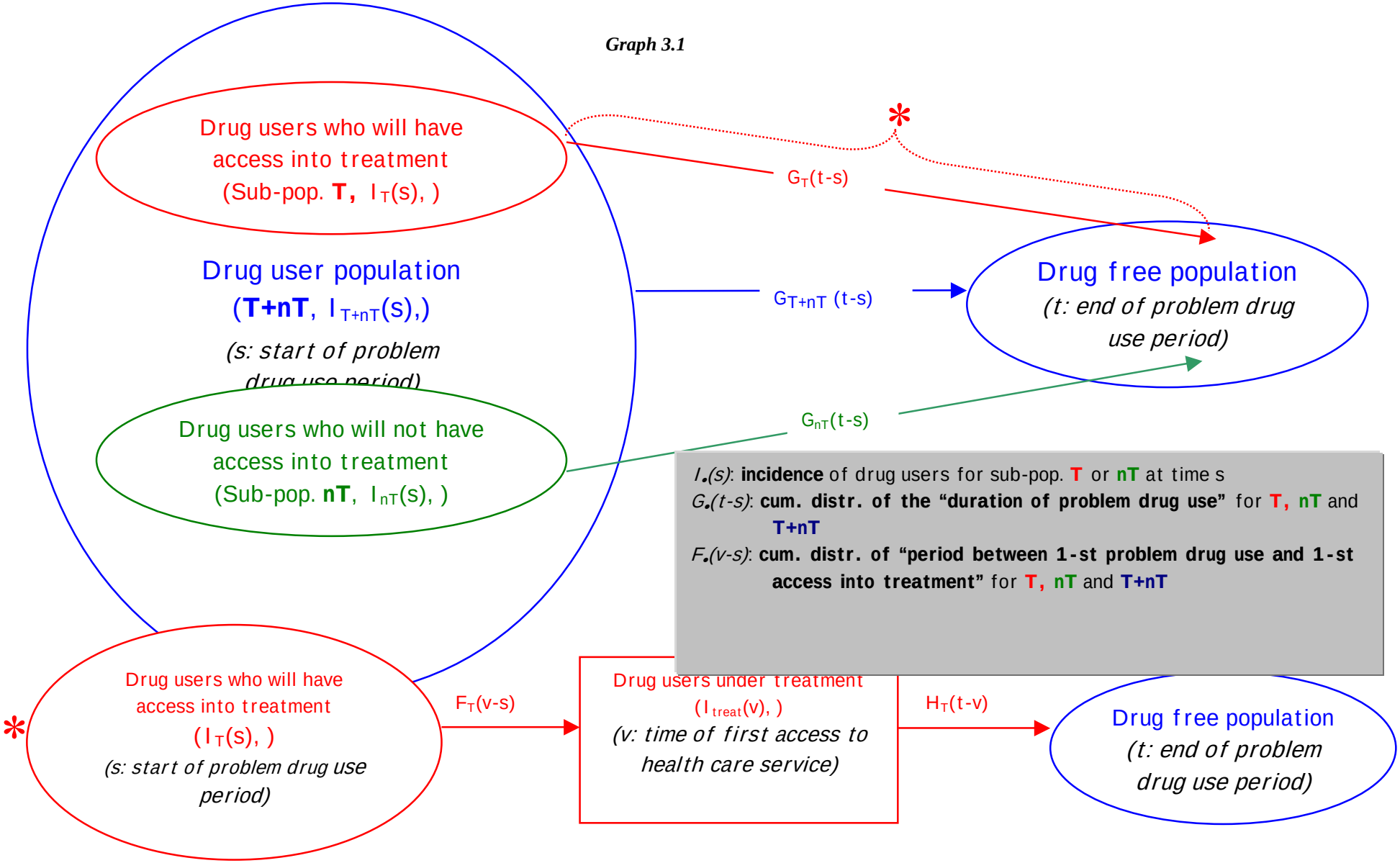
interval of diagnosis is assumed to be independently Poisson distributed. Without any constraint the estimated HIV incidences might be highly variable, therefore a smoothness restriction is adopted by placing a prior distribution for the infection parameters to be jointly estimated. The advantage of 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.

The implementation of the present method allows the inclusion of covariates and also age-specific incubation period distributions.

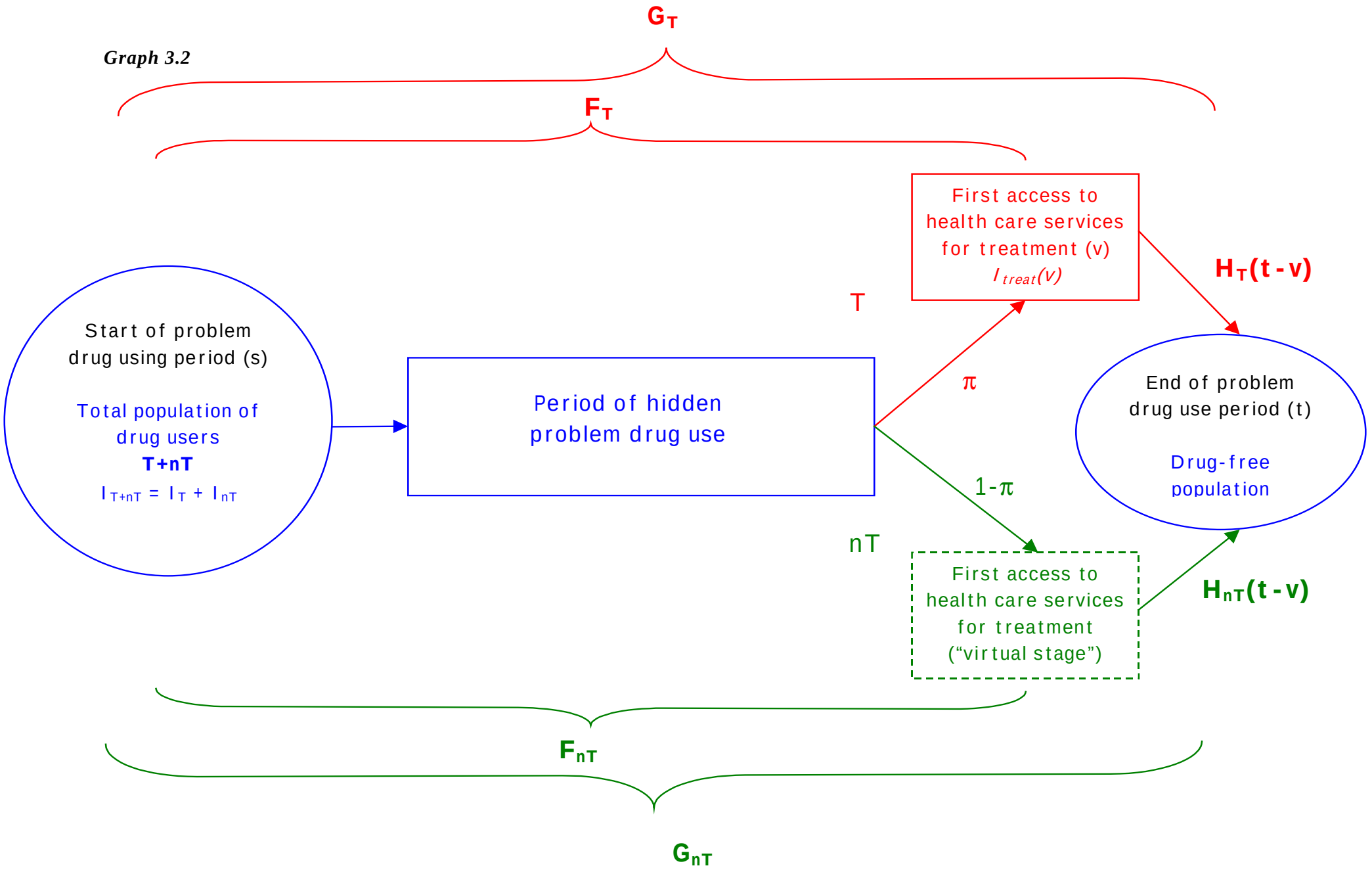
The EB-BC was applied in this context by defining the figures and the functions involved in the analysis of problem drug use as described in the following of this section and reported in graphs 3.1 and 3.2.

The total population of problem drug users (DUs) can be split into the two different subpopulations: **T** individuals who, after a period of hidden problem drug use, will eventually present for treatment, and **nT** individuals who will never present for treatment, according to the proportions π and $1 - \pi$, in other words, π is the probability of a DU ever presenting for treatment.

Graph 3.1


Drug users who will have access into treatment (Sub-pop. T, $I_T(s)$,)
Drug user population (T+nT, $I_{T+nT}(s)$,)
(s: start of problem drug use period)
Drug users who will not have access into treatment (Sub-pop. nT, $I_{nT}(s)$,)
$G_T(t-s)$
$G_{T+nT}(t-s)$
$G_{nT}(t-s)$
Drug free population (t: end of problem drug use period)
I.(s): incidence of drug users for sub-pop. T or nT at time s
G.(t-s): cum. distr. of the "duration of problem drug use" for T, nT and T+nT
F.(v-s): cum. distr. of "period between 1-st problem drug use and 1-st access into treatment" for T, nT and T+nT
Drug users who will have access into treatment ($I_T(s)$,)
(s: start of problem drug use period)
$F_T(v-s)$
Drug users under treatment ($I_{treat}(v)$,)
(v: time of first access to health care service)
$H_T(t-v)$
Drug free population (t: end of problem drug use period)

Graph 3.2


G_T
F_T
First access to health care services for treatment (v)
$I_{treat}(v)$
$H_T(t-v)$
T
Start of problem drug using period (s)
Total population of drug users
T+nT
$I_{T+nT} = I_T + I_{nT}$
Period of hidden problem drug use
π
End of problem drug use period (t)
Drug-free population
1-π
nT
First access to health care services for treatment ("virtual stage")
$H_{nT}(t-v)$
F_{nT}
G_{nT}

The following quantities can be defined:

$I_{T+nT}(s)$: the total incidence of problem drug use at time s, $s = 1,\ldots,S$;

$I_T(s)$: the incidence of drug users who present to treatment (at least once) at time s, $s = 1,\ldots,S$. These individuals pass through a period of hidden drug use before they become visible by having their first contact with some health care service;

$I_{nT}(s)$: the incidence of drug users who will never present for treatment at time s, $s = 1,\ldots,S$. For the present analysis, this proportion of drug users population will remain always hidden;

$G_{T+nT}(t\text{-}s)$: the cumulative distribution of the "duration of problem drug use", that is the period between the time s of the first problem use of drug, and the time t of the exit from the total drug user population **T** + **nT** (end of problem drug use period);

$G_T(t\text{-}s)$ and $G_{nT}(t\text{-}s)$: are the cumulative distributions defined as above, but for the sub-populations **T** and **nT**;

$F_T(v\text{-}s)$: 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 "Latency period distribution" defined and estimated in section 2.

$H_T(t\text{-}v)$: the cumulative distribution of the period between the time v of the first presentation for treatment to some health care service and the end of problem drug use period, t.

Virtually, analogous cumulative distributions, $F_{nT}(v\text{-}s)$ and $H_{nT}(t\text{-}v)$, might be defined for the proportion of DU population who will always remain hidden and will exit from the problem drug use for other causes rather than treatment, as well.

Clearly such distributions are not observable, but it can be assumed, as a first approximation in order to analyse some basic scenario, that:

$$F_T(v\text{-}s) = F_{nT}(v\text{-}s) = F_{T+nT}(v\text{-}s)$$

$$H_T(t\text{-}v) \neq H_{nT}(t\text{-}v).$$

The goal of the present pilot project is to estimate the incidence of DUs eventually seeking for treatment $I_T(s)$ using data on the incidence of DUs in treatment, $I_{treat}(v)$ and estimate the latency period distribution $F_T(v\text{-}s)$ using data collected by the Focal Point and data provided by the health care services offering treatment (of any kind) to the drug users. The possibility of using the incidence of this sub-population to estimate the incidence for the total population also will be considered from a methodological point of view as a further development.

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).

The EB-BC was applied, in order to study the extent of problem drug use in Italy, The Netherland, and United Kingdom, through a software programme written in S-plus language (see appendix 2)

3.3. Application to the problem drug use.

3.3.1. 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, which will be presented in detail in the following of this section.
The data-file needed for the EB-BC could be derived or by 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 under 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 3.1., for a detailed description of records.
Note that for this Pilot project 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 next 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 3.1. 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	

3.3.2. The models allowed 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 (as originally in the ECMAS study (see Appendix 2), 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 λ_i of which the first k-1 have an equal rate λ, 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.

3.3.3. 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. Such results are summarised in 6 graphs and 2 ASCII tables (see appendix 2)
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.

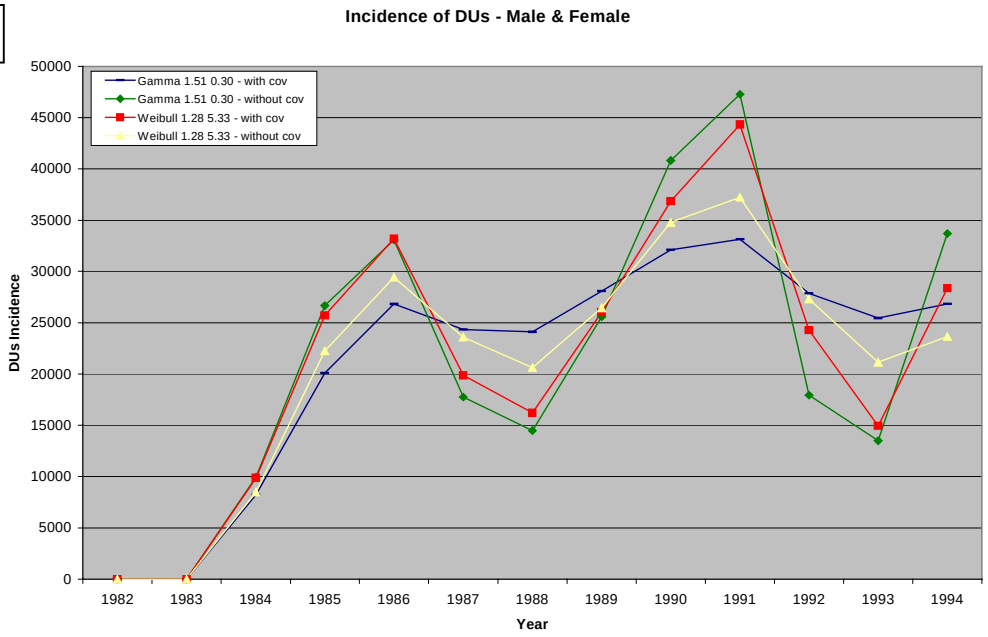
3.4. Results

The EB-BC was applied separately for Italy, The Netherland and United Kingdom, by using the treatment data, which were provided by the health care services of each Country as mentioned in section 3.3.2, 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 and with a Weibull densities, whose parameter are those estimated in section 2 (Table 2.3.). Clearly the results obtained are local, and therefore they are shown in detail in each country report. Nevertheless in this section an overview of such results is given through the Figures 3.1 and 3.2. 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 estimates used.

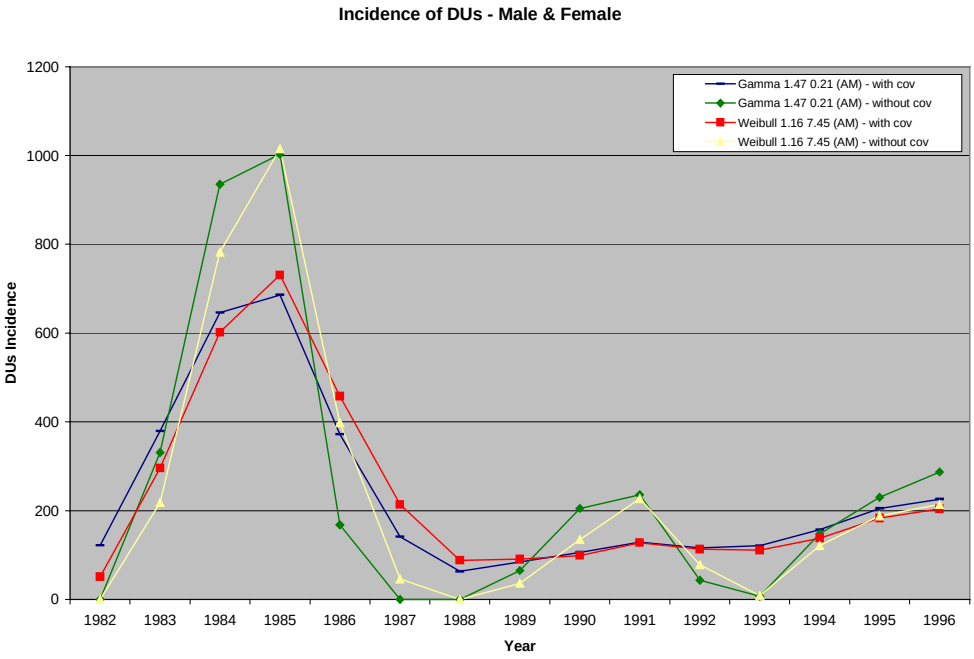
It is important to stress that such results must be considered just as preliminar. Nevertheless, they can provide some qualitative comparisons about the extent and the dynamic of problem drug use in the three countries. Clearly, the incidence curves represented in the figures do not allow us to compare 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 scope of the problem drug use in the three countries.

Figure 3.1

ITA



NET



UK

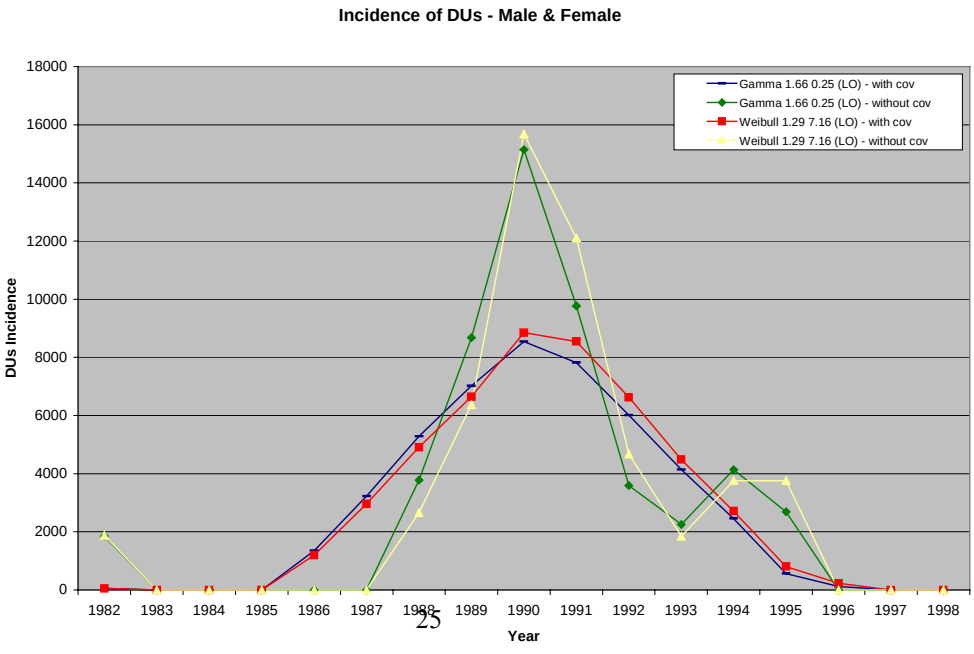
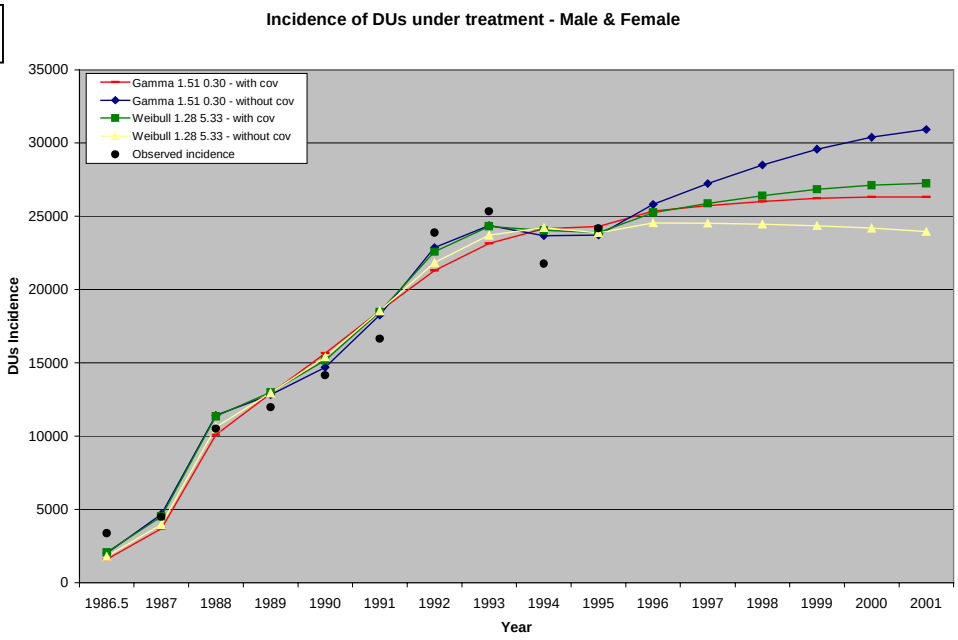
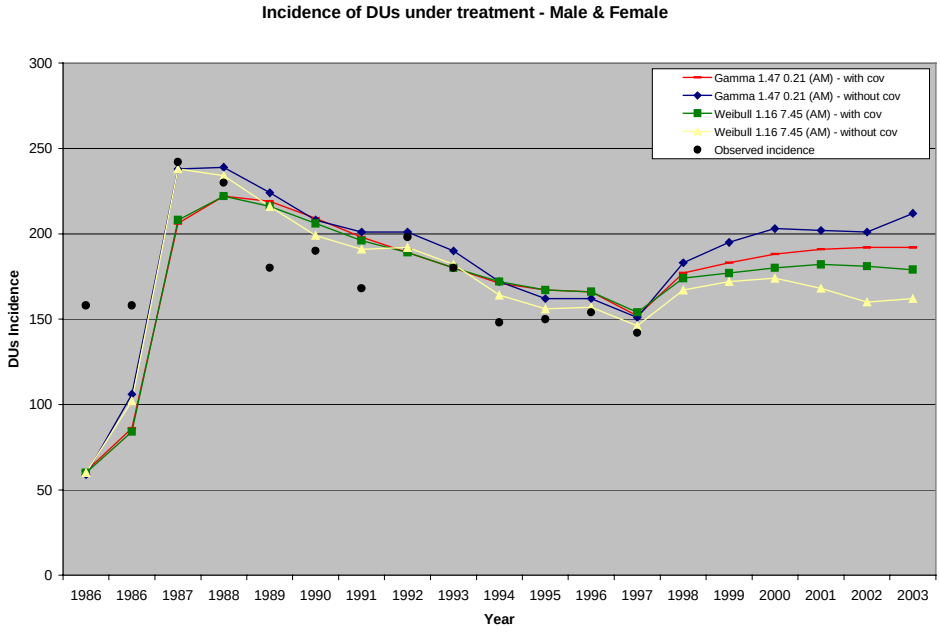


Figure 3.2

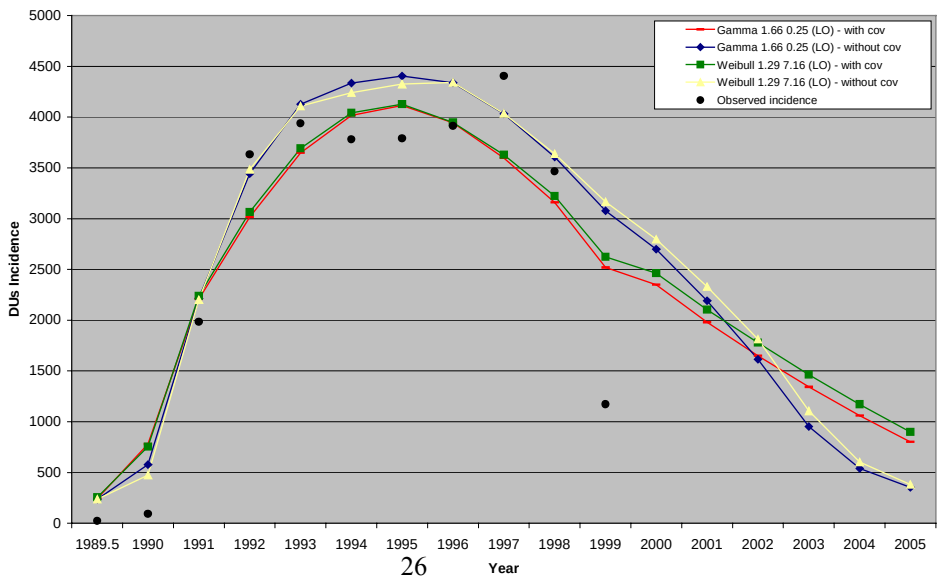
ITA



NET



UK



3.5. 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.

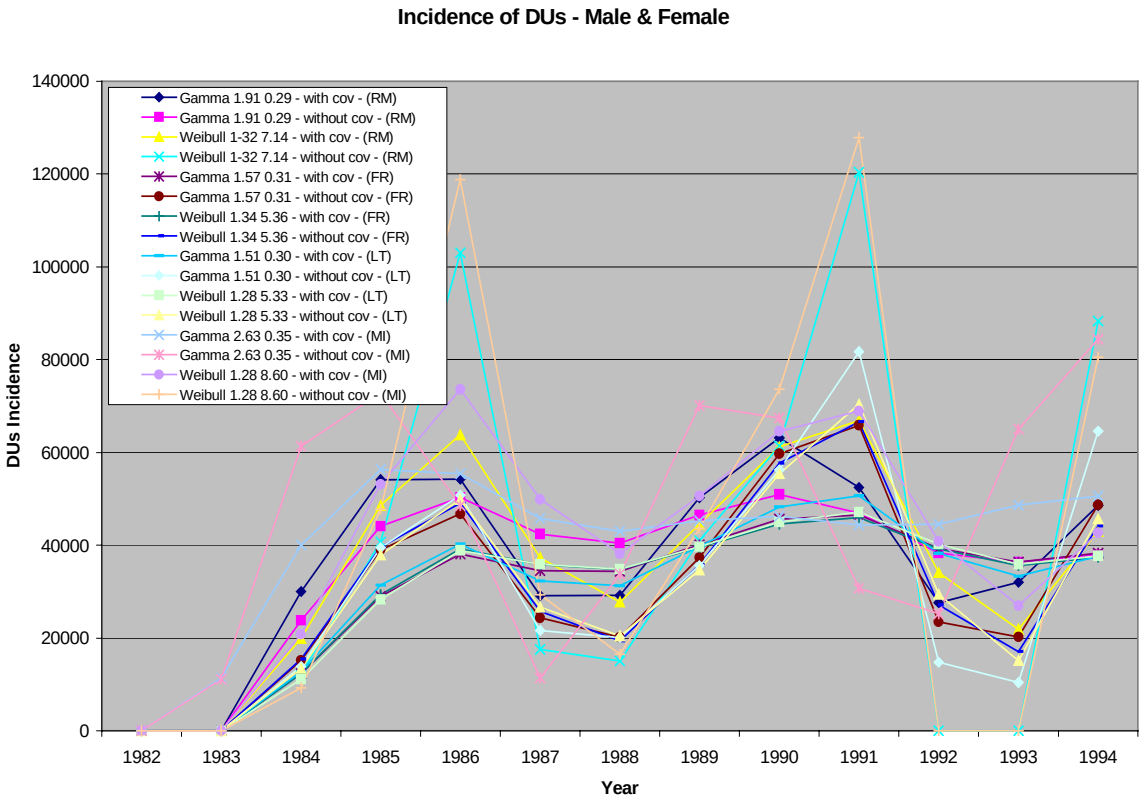
The Italian data about DUs in treatment in the public institutions, without any inflation for the proportion of double counting and for the proportion of drug users in treatment who are not heroin users (see country report), were used; 8 estimates of latency period distribution (4 Gamma and 4 Weibull 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.

It is important to note that the fitting of the estimated latency period distributions while was generally good for Rome, Frosinone and Latina, was quite poor for Milan (see country report).

Figures 3.3 and 3.4 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 3.3 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 3.3 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).

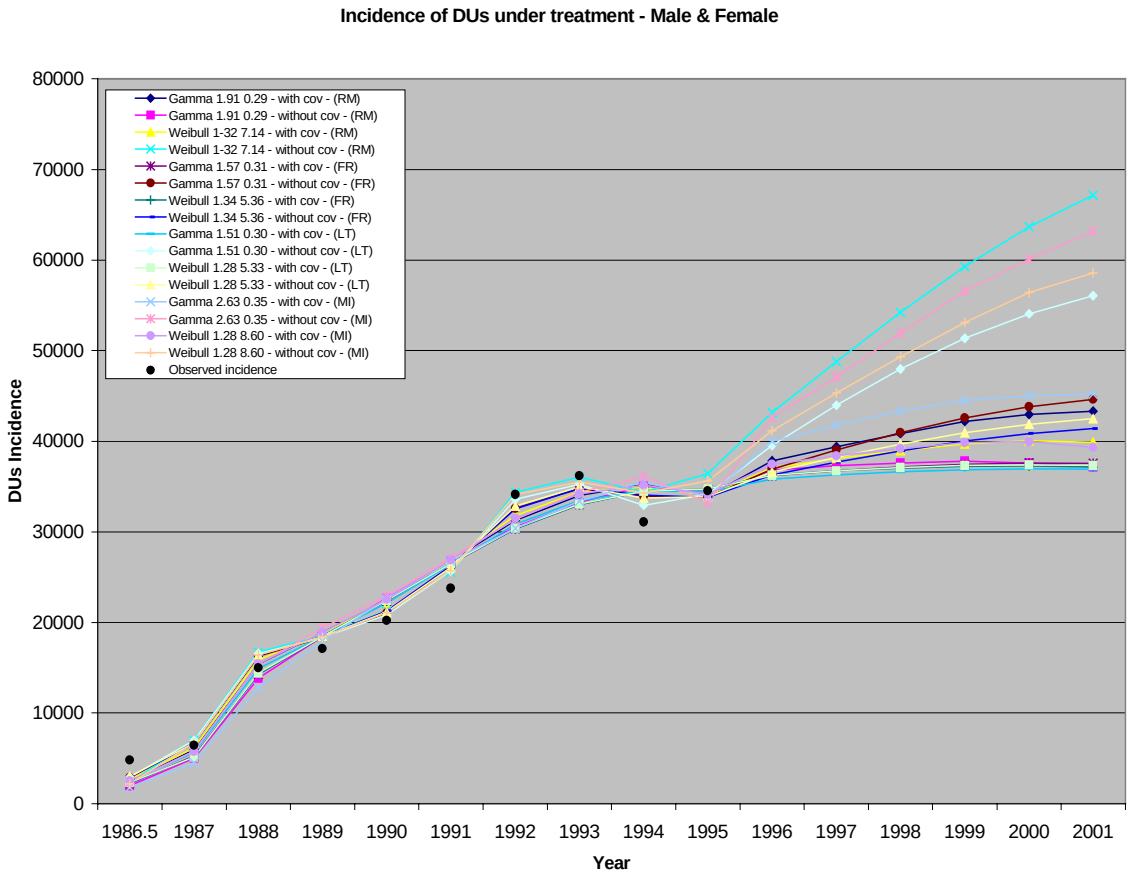
From Figure 3.4. it can be observed that, the fitting of the expected to the observed treatment data is quite satisfactory, and similar to each other, for any one of the latency models. The same figure provides evidence that, for each latency period distribution, the right tails of the expected incidence of DUs (the predicted incidence) in treatment are heavily determined by the corresponding estimated curves of incidence of problem drug use reported in Figure 3.3.

Figure 3.3. Incidence curves for the total population estimated by the Empirical Bayesian Back-Calculation procedure with different models of the latency period distribution.



The Figure 3.3 shows null incidence values for years 1982 and 1983: this is very likely an effect of some constrains used in the EB-BC. Such constrains assumes in fact a null incidence of problem DUs before 1982, but such date could be easily changed according to some plausible hypothesis

Figure 3.4. 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.



4. Conclusive remarks and further developments.

From the results of the analysis of the different data sets provided by the partners, both for the estimation of the latency period distribution and of the incidence of problem drug use, reported above and, more extensively, in the country and local reports, 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 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 only 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.
- 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. Further details on this subject are reported in the country and local reports.
- There seem to be differences in latency period between drug users originating from different countries (see Amsterdam, Portugal and UK reports). 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" (see country reports).

2) *Back-Calculation:*

- the main problem related to the data used to apply the BC procedure is represented by the double counting (see Italy country report) 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 a 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 representing the latency period. Figures 4.1 and 4.2 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 changes depending 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 4.3 and 4.4, respectively for the EB-BC performed with and without the age-covariate. From Figure 4.3, i.e. EB-BC with age covariate, 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 4.4), 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 4.5 and 4.6 report respectively the incidence curves of problem drug use and of DUs presenting to treatment and the Chi-square values.

Figure 4.1. Sensitivity analysis for double counting: Incidence of DUs estimated through the EB-BC with age covariate.

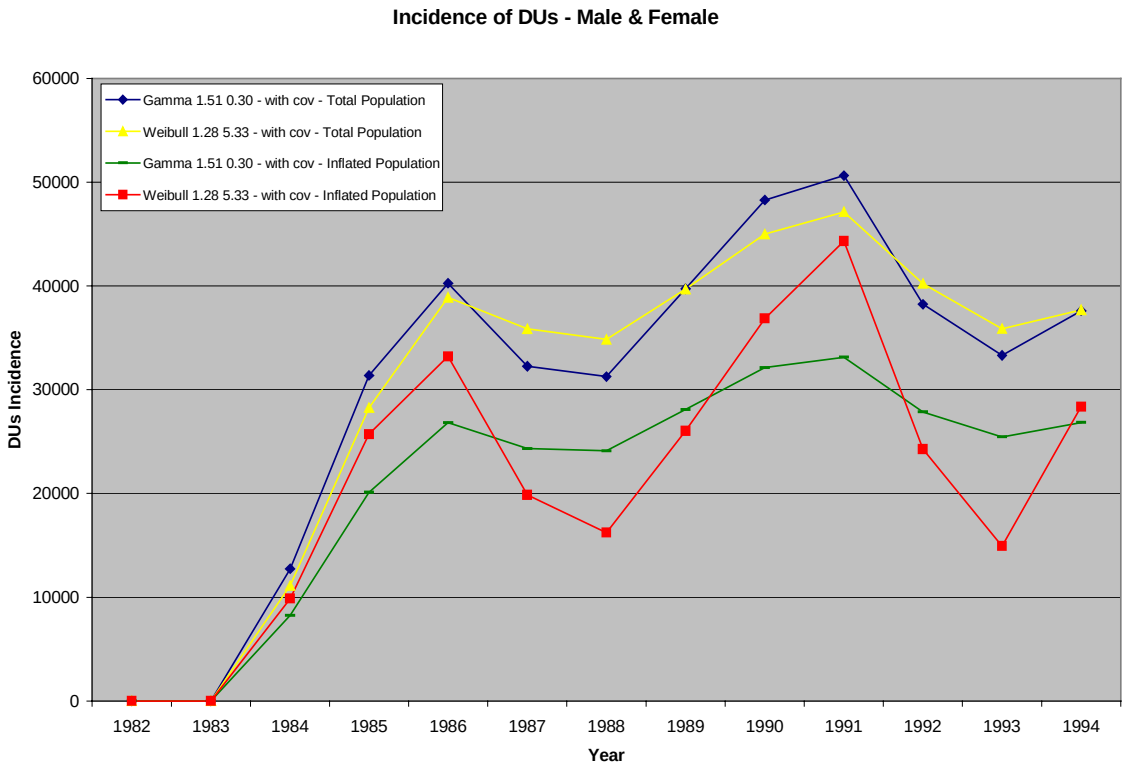


Figure 4.2. Sensitivity analysis for double counting: Incidence of DUs estimated through the EB-BC without age covariate.

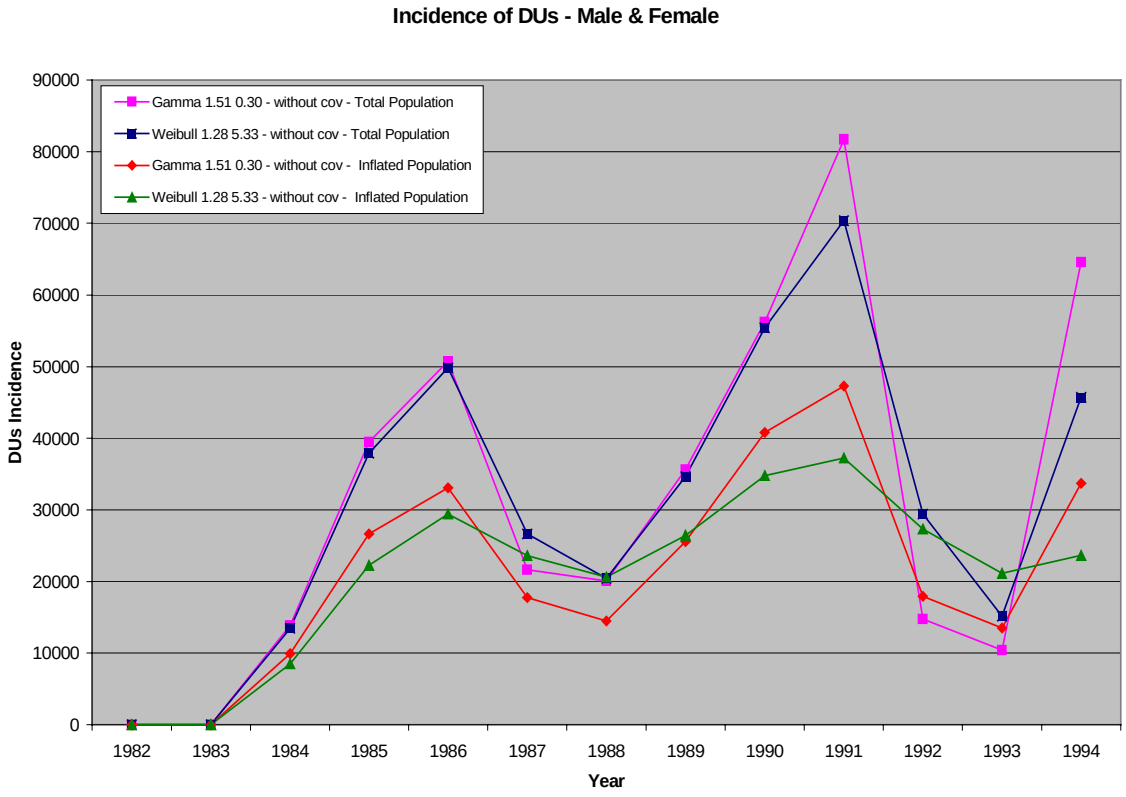


Figure 4.3. Sensitivity analysis for double counting: Incidence of DUs presenting to treatment estimated through the EB-BC with age covariate.

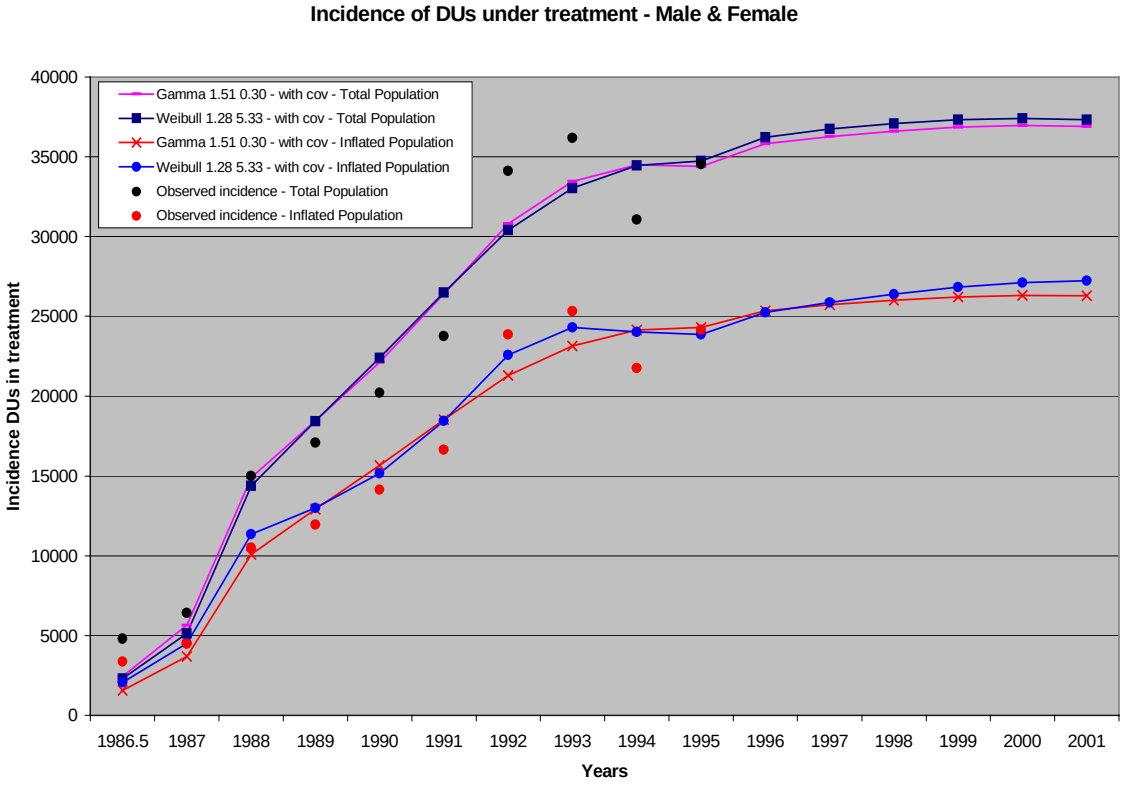


Figure 4.4. Sensitivity analysis for double counting: Incidence of DUs presenting to treatment estimated through the EB-BC without age covariate.

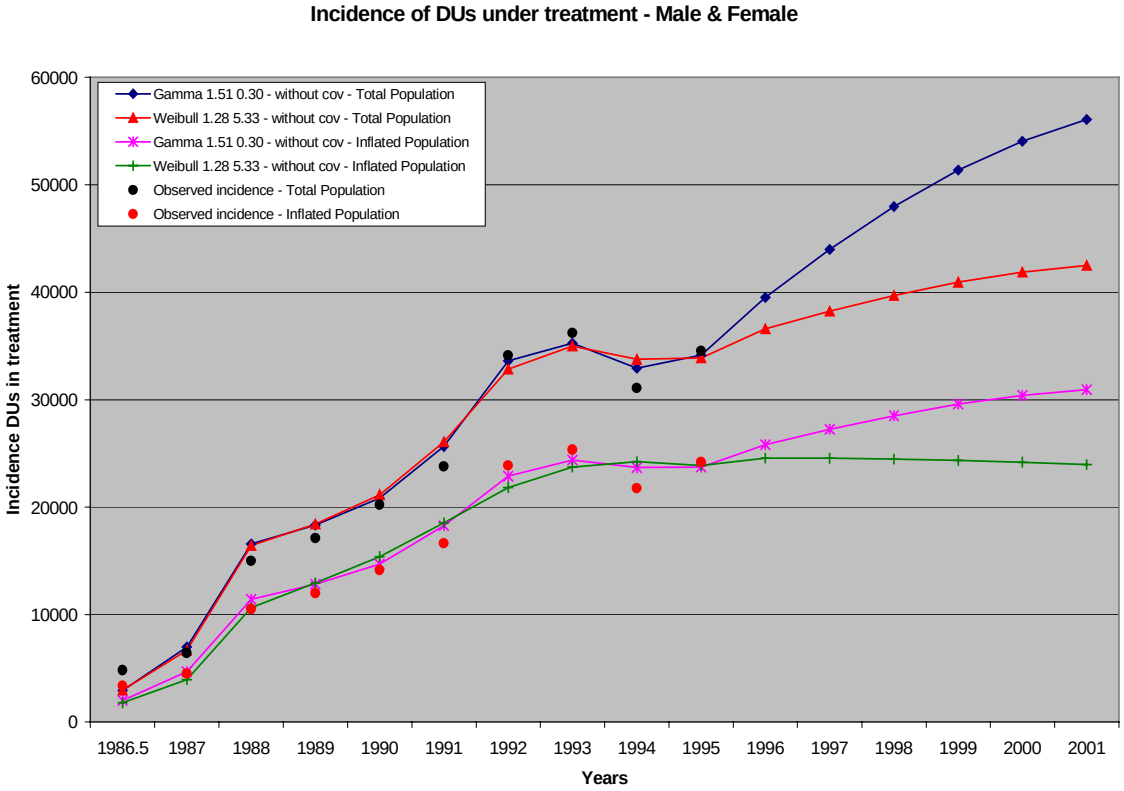


Figure 4.5. Sensitivity analysis for epidemic starting year: Incidence of DUs estimated through the EB-BC with age covariate.

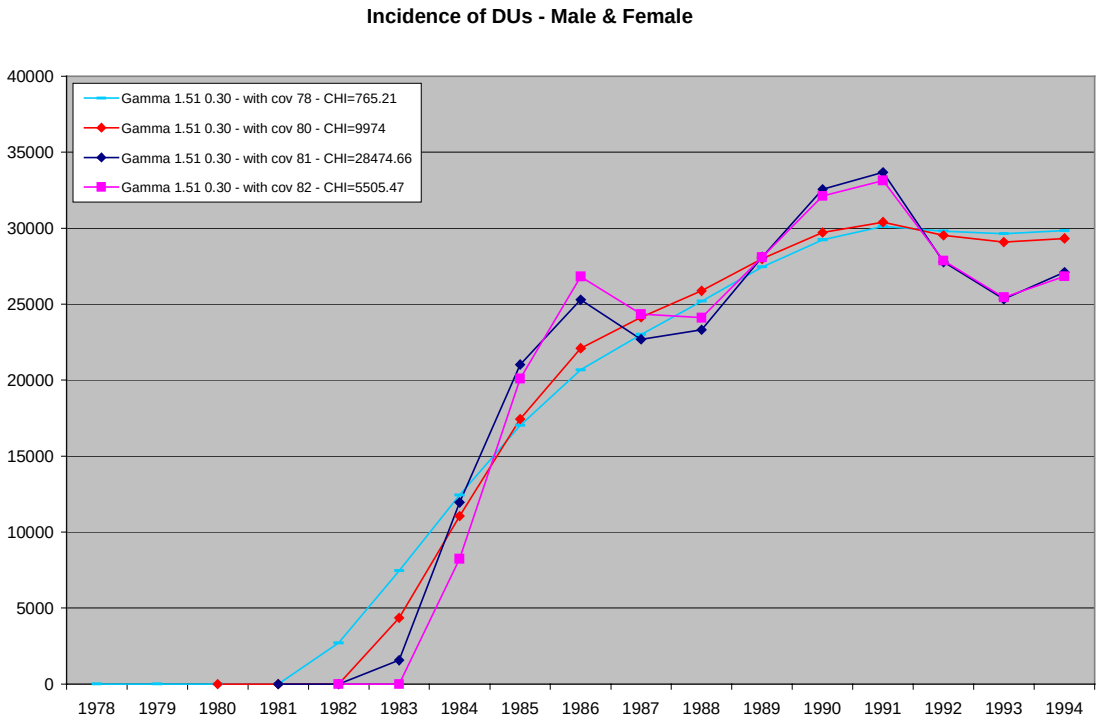
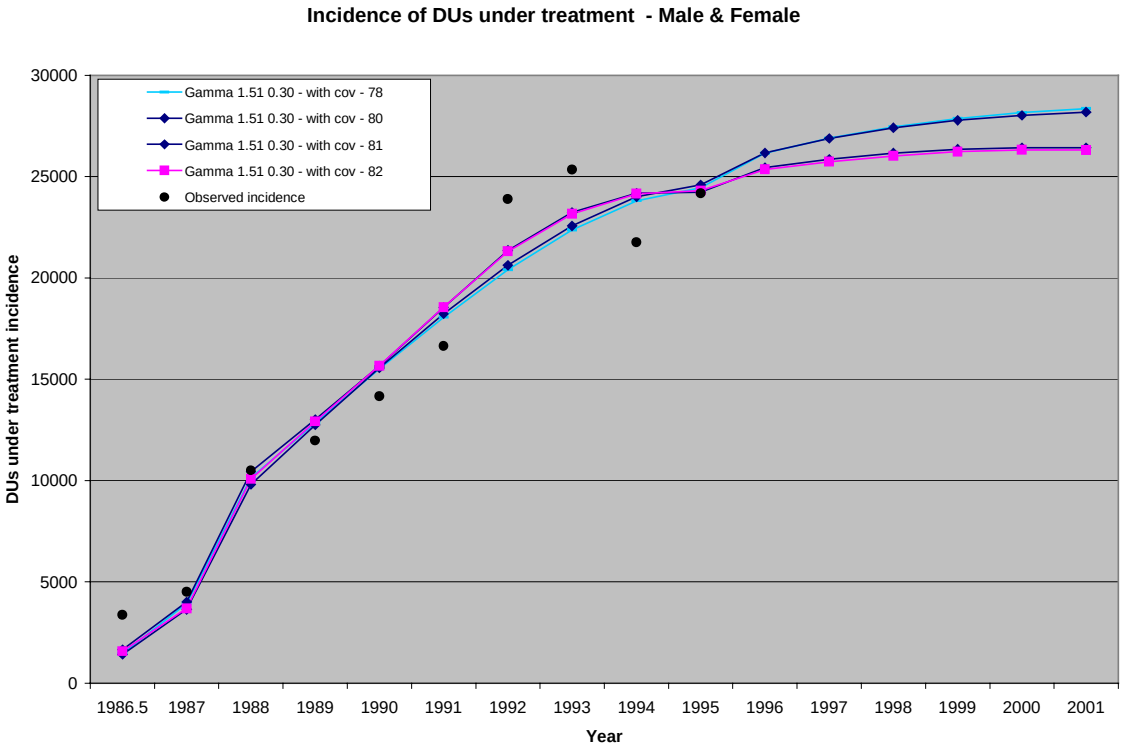


Figure 4.6. Sensitivity analysis for epidemic starting year: Incidence of DUs presenting to treatment estimated through the EB-BC with age covariate.



. It is important to stress that the choice of the epidemic starting year should not be determined just on the basis of the Chi-square value, since external epidemiological information on the DUs epidemic must be taken into account as well. In particular it would be important to consider any available information on the extent and trend of the DUs incidence curve, possibly provided by observational study.

- Several developments can be proposed to improve the results obtained within the present pilot project, 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.
- Drug surveillance systems tend to collect many data items. The reliability of these would be increased if a core data set was identified. For epidemiological purposes age of first use is an essential data item, and needs to be collected better.
- 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.

Other possible methodological developments should be aimed at reducing the uncertainties of the estimate of the drug use incidence in recent years, which constitute the main difficulties in applying BC estimation procedures. Two approaches can be proposed for this important task:

- The development of a dynamic (compartmental) model of the drug user career to estimate the size of the drug epidemic by a different approach and gather further information on the behaviour of the incidence curve in recent years. This approach, based on the joint use of the BC method and the dynamic model, was recently applied in the AIDS epidemic framework with very promising results (Ravà et al., 1998).
- The second possibility is to develop a method based on the snapshot estimation procedure proposed by Kaplan and Brookmeyer (1999), specifically to estimate "recent HIV incidence rates", which was successfully applied to estimate the HIV incidence in Israel (Kaplan, Kedem and Pollack, 1998). In simple words, the snapshot estimation method to estimate recent HIV infections uses a single cross-sectional sample of HIV infected individuals and a model of a marker related to the progression of the disease. Let us consider, for example, the CD4 cell counts; if a region R of the possible values of the marker is fixed, say, for example, R={values > 900}, then it is possible to calculate the expected time spent by the marker in the fixed region for an HIV infected individual. If, out of a sample of n individuals, a proportion π shows the marker in the fixed region, and if a proper model of the marker dynamics is available, then the proportion of infections occurred in the recent past can be estimated. The implementation of the method for the actual problem drug use analysis can be conducted by choosing different

markers related to the "typical" drug user career. Let us consider, for example, a sample of drug users in treatment, ask how many people are there for the first time (or for the second). This, combined with a model of people moving in and out of drug treatment over time, allows the use of the snapshot method: the probability that the marker is in the fixed region is replaced by the probability that a person is in drug treatment for the first time. Suppose that a fraction *F* of all drug users never enter drug treatment. For those who do, suppose that D is the average duration of time spent in treatment in the first episode, then the expected time τ spent by the marker in the region is given by $\tau=(1-F)D$. If p is the estimated prevalence (proportion) of problem drug use in the population and π is the proportion of drug users in treatment for the first time, then the recent incidence rate of problem drug use can be estimated by:

$$incidence = \frac{p\pi}{(1-p)\tau}$$

which is the general formula of the snapshot estimator reported in Kaplan and Brookmeyer (1999).

A possible alternative marker may be related to the first phase of (non-problematic) drug use which was studied for some italian sites (see Italy country report). In this situation the sample of drug users should be taken from the total population of users and not only from the clients of therapy services, the probability that the marker is in the fixed region is replaced by the probability that a person is in the first phase of drug use, the rest is quite similar to the example outlined above.

Further methodological developments can be proposed:

- The first one, explained in the following, is aimed at estimating the total incidence of problem drug use $I_{T+nT}(s)$ by using proper models and assumptions. As it has been shown, the incidence $I_T(s)$ can be estimated through the Empirical Bayesian Back-Calculation, by using the following equation:

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

Let's define:

$P_T(v)$: the prevalence of DUs belonging to the sub-population **T,** at time *s*, *s=1,…, S*

$$P_T(v) = \int_0^v I_T(s)(1-F_T(v-s))ds \quad v = 1,\ldots, V$$

and the analogous, yet "virtual", figures:

$P_{nT}(v)$: the prevalence of DUs belonging to the sub-population **nT** at time *s*, *s=1,…, S*

$$P_{nT}(v) = \int_0^v I_{nT}(s)(1 - F_{nT}(v-s))ds \quad v = 1,\ldots, V$$

$P_{T+nT}(v)$: the prevalence of DUs belonging to the total population **T + nT** at time *s*, *s = 1,…, S*

$$P_{T+nT}(v) = \int_0^v I_{T+nT}(s)(1 - F_{T+nT}(v-s))ds \quad v = 1,\ldots, V$$

Then, we can set up a basic scenario by assuming that the ratio of the incidences of DUs belonging to each of the two sub-populations **T** and **nT**, is constant over time and $F_{nT}(s)=F_T(s)$ for each s, thus the total incidence of drug users can be easily estimated on the basis of prevalence estimates and of $I_T(s)$.

$$I_{T+nT}(s) = I_T(s) + I_{nT}(s) = I_T(s) + kI_T(s)$$

can be estimated easily through the following calculations:

$$P_{T+nT}(v) = \int_0^v I_{T+nT}(s)(1 - F_{T+nT}(v-s))ds = \int_0^v (I_T(s) + kI_T(s))(1 - F_{T+nT}(v-s))ds =$$

$$= (1+k)\int_0^t I_T(s)(1 - F_T(v-s))ds = (1+k)P_T(v)$$

$$\Rightarrow k = \frac{P_{T+nT}(v)}{P_T(v)} - 1 \Rightarrow k = \frac{P_{T+nT}(v) - P_T(v)}{P_T(v)} \Rightarrow k = \frac{P_{nT}(v)}{P_T(v)} \Rightarrow k = \frac{1-\pi}{\pi}$$

$$I_{T+nT}(s) = \left(1 + \frac{1-\pi}{\pi}\right) I_T(s) \Rightarrow I_{T+nT}(s) = \frac{I_T(s)}{\pi}$$

$$I_{T+nT}(v) = (1+k)I_T(v) = \left(1 + \frac{P_{T+nT}(v) - P_T(v)}{P_T(v)}\right) I_T(s) = \frac{P_{T+nT}(v)}{P_T(v)} I_T(s)$$

The idea is that, if the basic scenario states that the dynamic of the drug user population is the same for those who will eventually present for treatment and for the others who will remain hidden, then the ratio between the observable and hidden incidences is the same as between observable and hidden prevalences. This last proportion can be estimated on the basis of prevalence studies, thus the overall incidence of problem drug use, corresponding to the basic scenario, can be estimated. Alternative scenarios can be set up by considering different hypotheses: for example, we could consider proportional hazard models to express $F_{nT}(s)$ as a function of $F_T(s)$. The formal developments are straightforward but heavy and are left to a future work.

- Another interesting analysis relates to further sensitivity studies on the results of the Back-calculation procedure, possibly using boostrap methods (Efron and Tibshirani, 1986).

- A simulation study could be useful to assess the performance of the EB-BC in evaluating the "best" starting point of the epidemic.
- Finally, 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.

References

Andersen P.K., Borgan O., Gill R.D., Keiding N., "Statistical models based on counting processes", Springer Verlag, Berlin-Heidelberg, 1993.

Brookmeyer R., Gail H.G., Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States, *Lancet, 2*, 1320-1322, 1986.

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.

Brookmeyer R., Reconstruction and Future Trends of the AIDS Epidemic in the United States. *Science Articles*, 37, 37-253, 1991.

Collet D. "Modelling survival data in medical research", Chapman and Hall, London, 1994.

Efron B. and Tibshirani R., Bootstrap methods for standard errors, confidence intervals, and other methods of statistical accuracy, *Statistical Science*, 1, 54-77, 1986.

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.

Kaplan E.H., Kedem E. and Pollack S., HIV incidence in Ethiopian immigrants to Israel, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 17, 465-469, 1998.

Kaplan E.H., Brookmeyer R., Snapshot estimators of recent HIV incidence rates, *Operations Research*, no. 41, 1, in press, 1999.

Marubini E., Valsecchi M.G., "Analysing survival data from clinical trials and observational studies", Wiley, NY, 1995.

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;

Rosemberg P., Gail M.H., Backcalculation of Flexible Linear Models of the Human Immunodeficiency Virus Infection Curve. *Applied Statistics* 40, 269-282, 1991.

Rossi C., Monitoring drug control strategies: hidden phenomena, observable events, observable times, *International Journal of Drug Policy*, 10-1, in press, 1999.

SPSS manual, 1997.

Tanner A.M.,"Tools for statistical inference", Springer Verlag, Heidelberg, 1996.

UNDCP, World Drug Report, Oxford University Press, 1997.