UNIVERSITY
of
GLASGOW

FINAL REPORT

EMCDDA Project: Methodological Pilot Study of Local Level Prevalence Estimates

CT.96.EP.07

Centre for Drug Misuse Research:	Gordon Hay Neil McKeganey Elise Birks
EMCDDA:	Lucas Wiessing Richard Hartnoll
Other Participants:	Pierre-Yves Bello Catherine Comiskey Daniela D'Ippoliti Antònia Domingo-Salvany Sofia Freire Maria Moreira Alvar Norén Päivi Partanen Kathy Politikou Dan Seidler Filip Smit Jaap Toet Alfred Uhl Ari Virtanen

Please use the following citation:

European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).
Methodological Pilot Study of Local Level Prevalence Estimates.
Lisbon: EMCDDA, December 1997.

Contact Details

Centre for Drug Misuse Research, University of Glasgow
Glasgow, G12 8QG
United Kingdom.

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

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

Acknowledgements

This project brought together experts from various disciplines throughout the European Union. Those who attended the planning meetings of the project are listed in the introduction of this report. However, in capture-recapture studies, thanks are often due to others in the data collection and collation stages of the research and those who have given access to the various data sources. We would like to acknowledge the contributions to this project of Lucas Wiessing and Richard Hartnoll from the EMCDDA and Roland Simon from IFT in Munich. Finally, we are sure that all of those who have contributed to this research would like to thank Elise Birks of the Centre for Drug Misuse Research for her administrative support to the project.

Contents

Page No

Summary....1

1 Introduction 6

2 Capture-recapture Methods 8

3 EMCDDA Pilot Study....13

4 City Results....17

5 Discussion....24

6 Conclusions 36

Appendix 1 ....Collated Estimates

Appendix 2 ....Full Addresses of Participants

Part II City Reports

1 Introduction 1
2 Dublin 2
3 Helsinki....16
4 Rome....36
5 Rotterdam 48
6 Setúbal 69
7 Toulouse 77
8 Vienna....90

Summary

This report presents the findings from an EMCDDA funded Methodological Pilot Study of Local Level Prevalence Estimates. Following on from a joint EMCDDA / Pompidou Group meeting, comparative pilot studies to improve the prevalence estimation at local and national levels were initiated. A methodology, commonly known as capture-recapture, was considered to be of use at the local level. The aim of this Pilot Study is therefore to compare and contrast the application of the methodology in four, or more, cities.

Through the network of focal points in the 15 countries of the European Union, potential partners for this project were identified and a group of experts were invited to a meeting in Glasgow to discuss the project. Some of the participants had experience in applying the method whereas others were reacting to the need of policy makers for accurate prevalence information in their city. It soon became clear that completing a capture-recapture analysis from start to finish in the space of a year could be a problem for cities that did not have a tradition of prevalence estimation; therefore the remit of the group was broadened to include the experiences of those wishing to use this method in their own country, but who may not have been able to obtain an estimate within the stipulated timetable. This group of networks was strengthened by means of an electronic mail discussion group and a second meeting held in Utrecht later in the year.

In total, seven cities obtained a prevalence estimate during the course of this project. These are:

- Dublin, Ireland;
- Helsinki, Finland;
- Rome, Italy;
- Setúbal, Portugal;
- Rotterdam, The Netherlands;
- Toulouse, France;
- Vienna, Austria.

The participation of researchers from Athens, Malmö and Barcelona strengthened the discussions.

As one of the main objectives of the project was to achieve comparability of

the application of the method throughout Europe, several definitions were needed for the research. Although terms such as addiction or drug dependence are useful in relation to policy issues, as is injecting, it was noted that many data sources employed in a capture-recapture study give little information on the seriousness of the drug problems of the individuals. The Pilot Study therefore concentrated on the use of opiates. It was, however, noted that the case definitions used by many of the contributing sources would mean that many of the resultant estimates would be of problematic opiate use. Again for comparability, a common age range (15-54) and study period (January to December 1996) were proposed. These, and other methodological discussions, served to make the city studies more comparable; however, each individual city had particular considerations relating to the methodology.

Three cities obtained estimates of the prevalence of opiate use in 1996: Dublin, Setúbal and Rome. The estimate from Helsinki refers to 1995, the estimate from Toulouse refers to a six month period within 1995 and the estimate from Vienna used data from 1993. The prevalence of opiate use in Rotterdam was estimated using a Truncated Poisson model.

Three sources were employed in Dublin: a methadone treatment list, a database concerned with drug related hospital admissions and data from the police. In total, 6,264 individuals were identified, from these three sources, as using opiates. There was difficulty in obtaining a model which adequately fitted the overlap pattern for these three data sources. The population was therefore stratified by both age and gender; this partially solved the problem. The best estimate for the combined group suggested a hidden population of 7,335, giving a total population of 13,599, opposed to combining the stratified estimates which gave an estimate of 13,460.

The number of opiate users identified from three sources in Helsinki was very small compared to other European capitals. Estimates were, however, provided not only for opiate use, but also amphetamine use. In total, 175 individuals were identified as opiate users, from the following sources: a hospital discharge register, a criminal register, and a list of people found driving under the influence of drugs. From the analysis, there were an estimated 775 opiate users.

Two surveillance systems were employed in Rome, along with a mobile emergency unit. There were 6,896 individuals identified from these sources, from which there were an estimated 14,278 opiate users aged between 15 and 54. The nature of the data sources employed in this analysis suggested a more problematic definition of the resultant estimate.

In Rotterdam the Truncated Poisson model was used to estimate the number of opiate users. There were 2,029 individuals identified from the methadone treatment agencies and the estimates varied from 2,937 to 5,006.

Only two independent sources were available in Setúbal, the third largest city

in Portugal. One source could, however, supply data on clients which attended in the first semester, the second semester or both semesters of 1996. Thus two 'sources' were created by splitting this single source into two semesters, although they could not be assumed to be independent. The inclusion of the other data source allowed a three-sample capture-recapture analysis. In total 339 individuals were identified; 313 from the two semesters of attendances at a treatment centre and 40 from local health centres. With an overlap of 14 between these sources there was an estimated 894 opiate users.

The research in Toulouse complemented previous research undertaken in 1994. Additional sources and data which refer to 1995 were obtained, from which 799 individuals were identified. This information was used to estimate the total number of opiate users at 2,178.

The prevalence of opiate use in Vienna in 1993 has been estimated using a police source, data on drug related deaths in the following year and a combine emergency ambulance / hospital admissions data source. In total 1,028 individuals were identified, from which the number of opiate users was estimated to be 6,747.

The known and estimated total number of opiate users, the prevalence in the 15 to 54 age group, and 95% confidence intervals are presented below.

Prevalence of opiate use in seven cities.

City	Known Users	Total Users		Prevalence (%)	
		Est.	95% CI	Est.	95% CI
Dublin	6,264	13,460	10,665-14,804	2.11	1.68-2.33
Helsinki	175	775	487-1,392	0.14	0.09-0.25
Rome	6,896	14,278	12,741-16,167	0.86	0.76-0.97
Rotterdam	2,029	3,716	3,497-3,990	1.07	1.01-1.14
Setúbal	339	894	620-1,423	1.82	1.26-2.90
Toulouse	799	2,178	1,780-2,734	0.54	0.44-0.68
Vienna	1,028	6,747	4,332-11,668	0.67	0.43-1.16

We can also present the numbers of male and female opiate users estimated in each city, along with the numbers of drug opiate users stratified by age. Here, we only present the data from the cities that applied the three-sample capture-recapture methodology.

Prevalence of male and female opiate use in seven cities.

	Males		Females		
City	**Estimate**	**%**	**Estimate**	**%**	**Male : Female**
Dublin	6,831	2.2	3,179	1.0	2.15:1
Rome	12,649	1.5	2,368	0.3	5.34:1
Setúbal	696	2.9	208	0.8	3.35:1
Helsinki	536	0.2	310	0.1	1.73:1
Toulouse	1,709	0.8	466	0.2	3.67:1
Rotterdam	1,485	0.8	544	0.2	2.73:1
Vienna	5,746	1.2	554	0.1	10.3:1

Prevalence of opiate use in the young and old age groups.

	Young			Old		
	Known	**Estimate**		**Known**	**Estimate**	
City		**N**	**%**		**N**	**%**
Dublin	5,604	10,964	2.94	660	1,067	0.41
Rome	4,521	10,365	1.17	2,375	4,447	0.58
Setúbal	284	884	3.54	55	85	0.72
Helsinki	54	122	0.11	121	464	0.11
Toulouse	649	1,709	0.74	131	426	0.24
Vienna	613	3,393	0.71	94	3,354	0.73

There seems to be some difference in the prevalence of opiate use in the different cities; ranging from the low levels of use in Helsinki to the higher levels found in Dublin and Setúbal. The male to female ratio is lowest in Helsinki, and higher in Rome, Setúbal and Toulouse, whereas the highest prevalence values in the young age groups are to be found in Dublin and Setúbal; the prevalence values in the older age group are more or less equal through the cities.

We can also describe the data sources that were used in obtaining these estimates in the following table.

Comparison of the data sources employed in seven cities.

	Medical	**Social**	**Legal**	**Other**
Dublin	Methadone In-Patient		Police	
Helsinki	In-Patient		Police	Driving
Rome	In-Patient Treatment Emergency			
Rotterdam	Treatment			
Setúbal	Treatment GPs			
Toulouse	In-Patient/Treatment Low Threshold		Police/Prison	
Vienna	In-Patient/ Emergency		Police	Deaths

It is clear that most of the data sources used in the seven cities are medical based and therefore many of the estimates of opiate use will refer to problematic opiate use.

To complement the discussions contained within the city reports, a comparative discussion further explored the relevant issues of this Pilot Study. Many of the problems in applying the method could be summarised as heterogeneity; a convenient term to cover many of the issues relating to summarising drug using populations into broad categories.

This Pilot Study achieved many epidemiological and methodological advances. In relation to the methodology, it is clear that the data sources employed in a study will dictate which population will be estimated. The collation here of seven estimates using a comparable methodology is unique in the field of drug misuse prevalence estimation.

1 Introduction

While it is simple to pose the question "How many addicts are there?", the answer often takes the form of a broad generalisation or a collection of statistics followed by a series of caveats. Increasingly, the collation of data is being improved throughout Europe, as are the various methods which can be used to inform the range of policy issues relating to addiction prevalence. These methods are, however, often applied in a fragmented way at local or national levels. An extension to this process needs now to examine the comparability of these methods across national boundaries.

In June 1996, a Scientific Seminar on "Addiction Prevalence Estimation: Methods and Research Strategies" was hosted jointly by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) and the Pompidou Group of the Council of Europe. From this, a Scientific Monograph has been produced which discusses the various prevalence estimation methods and their link with policy interests (EMCDDA, 1997). From this seminar, two pilot studies were initiated by the EMCDDA to channel the existing expertise into more systematic, comparative pilot studies and secondary analyses of existing data, therefore improving prevalence estimation at local and national levels.

In the long run, it is intended to facilitate comparable prevalence estimates of addiction in all EU countries. There needs therefore to be a consensus on what methods are the most reliable, where they can be applied and which definitions are used. One method that has been increasingly used at the local level is known as capture-recapture. The scientific literature shows that this has been used in several areas of Europe, including London (Hartnoll *et al.*, 1985), Glasgow (Frischer *et al.*, 1993), urban areas of the Netherlands (Korf *et al.*, 1994), Barcelona (Domingo-Salvany *et al.*, 1995), Liverpool (Squires *et al.*, 1995), Dundee (Hay and McKeganey, 1996) and Toulouse (Bello and Chêne, 1997). A related method of prevalence estimation, known as the Truncated Poisson model (Zelterman, 1988; Chao, 1989), is also considered a worthwhile addition to the drug prevalence researcher's armamentarium.

The problems of measuring the prevalence of drug use in a population by means of a survey are well known, and these problems are particularly acute when examining more problematic forms of drug use, such as opiate addiction or injecting. A range of other methods have been developed, such as nomination techniques or applying multipliers to the number of drug related deaths. In some areas the systematic collation of data within a register or extending this process into a case finding study can give an insight in the prevalence of addiction.

The Scientific Monograph presents a concise review of the different prevalence estimation tools, including applications of the capture-recapture method. A lucid discussion of the issues surrounding drug misuse prevalence estimation in general, and capture-recapture in particular, is contained within

the monograph. However, as this Pilot Study, and the reports generated by it, focuses specifically on capture-recapture methods and the comparability of applying them in different local areas, a brief introduction to the methodology is presented in the next section.

We present in this report the results from seven cities throughout Europe, six of which used the three-sample capture-recapture methodology. A summary of the research is included in this report, as is a comparison of the findings from each city. We also include a description of the process that the research has undertaken, from a preliminary planning meeting with the EMCDDA in January to the creation of a group of experts and the production of seven reports describing the application of capture-recapture in different settings. These individual reports have been collated, with the minimum of editing, into Part II of this report.

While the city reports are of interest individually, it is the comparative element of the Pilot Study and the practical and methodological discussions that arose at the two occasions on which the experts came together as a group that form the basis of the main part of this report. Several themes arose as the main issues that need to be addressed when applying the methodology, and we attempt to address these issues individually. Many of them are related to each other, partly due to the application of a methodology which assumes that the lifestyles of drug users can be summarised by classification into different sub-groups which share common characteristics. The drug using community is as diverse as the general population and this diversity, and the problems of describing it, are often lumped together under the heading ‘heterogeneity’. We aim to describe these problems, and the steps that the contributing cities have taken to counter them, without relying excessively on the statistical terminology.

We conclude this report by examining the methodological and epidemiological advances that this Pilot Study has attained. We do not, however, see this report as an end point to the process that began at, if not before, the Joint Scientific Seminar held in 1996. A follow-up project has been initiated by the EMCDDA in which it is hoped that the expertise that has been channelled into the group responsible for this report and others in Europe can be used to promote the capture-recapture method to a larger audience, both scientific and policy orientated. Many of the studies described in this report are ongoing, and there is interest in applying capture-recapture in other areas and not just at the city level.

2 Capture-recapture Methods

There is a range of data sources which either directly or indirectly can be thought of as indicating the number of drug users in a particular locality. The indicator idea found favour in the 1980s when Hartnoll *et al.* (1985) examined how data from various sources can be pieced together, like a jigsaw, to give a fuller picture of the nature and extent of drug misuse in an area of London. There are information sources which can be referred to as indirect indicators, such as the number of people citing drug injecting as a reason for an HIV test, the number of people recorded as coming into contact treatment centres due to problems related to drug use, the number of people attending hospital for drug related problems, the number of people who come into contact with the police due to drug related crime or the number of people having a drug overdose.

On their own, each of the sources of data listed above can only give a partial insight into the prevalence of drug misuse. It cannot always be assumed that there is a direct link between an increase in the number of people recorded in a particular data source and an increase in the prevalence of drug misuse. For example, the number of people that are convicted for drug offences would depend on the operational policies of the police. However, suggestions were made as to how these types of data can be pieced together.

It was suggested that an initial method of working with such data would be to perform a multi-source enumeration. To avoid double counting, it is necessary to obtain enough information from each data source to identify each individual drug user. However, this has to be weighed against the confidentiality requirements of the agency which provides the data. Initials, sex and date of birth are often used to sift out multiple occurrences; therefore it should be noted that comparing data sources using this un-named identifier information can be imprecise, especially as the accurate collection of names and dates of birth is not the highest priority for some agencies working with drug users. However, this exercise can be useful in describing the **known** drug using population.

Information on the nature and extent of the **unknown** population of drug users is also required to give the complete picture of drug use, and in particular the size of the unknown population of drug users is needed to obtain the required prevalence estimate. Capture-recapture methods have been increasingly used to obtain estimates of the number of ‘hidden’ drug users and the prevalence of drug use in an area.

As the name alludes to, the capture-recapture methodology was originally developed by ecologists who were interested in estimating the size of animal populations. Two analogous examples are presented here, one from ecology and one of the first applications of capture-recapture applied to drug misuse.

An ecologist wants to estimate the number of fish there are in a lake, therefore

a sample of fish are caught, counted, marked in some way and then released back into the lake. At a later date the ecologist returns to catch another sample and, by checking for marks, the number of fish seen in both samples can be obtained and thus the ratio of previously caught to previously uncaught fish in the second sample can be found. As it can perhaps be assumed that the ratio of caught to uncaught in the first sample is the same as the ratio of previously caught to previously uncaught in the second sample, the total population size can be estimated by multiplying the number seen in the first sample by the inverse of the ratio. An example concerning drug misuse will perhaps illustrate the method more clearly.

In the study of Hartnoll *et al.*, data concerning opiate users who had attended a drug clinic and those that had been admitted to a hospital for infectious diseases were collected. By comparing these sources of data, they found that approximately 20%, or a fifth, of the hospital sample had also attended the drug clinic. Thus the total number of opiate users could then be estimated to be five times the number who attended the drug clinic. Therefore the size of the hidden population of drug users was estimated by employing two existing sources of data.

These two simple examples mask some of the problems of the methodology. In Hartnoll's case, if those who were attending in the clinic were more likely to have been admitted to the hospital then the resultant figure would be an underestimate. Thus if there is some kind of relationship between the data sources then the estimate will be biased. Unfortunately it is often unclear if such relationships, or interactions, are present and therefore the validity of these estimates are often questionable.

The capture-recapture methodology can compensate for some of these problems by employing three sources. The extra information present in the third sample can then be used to examine whether or not dependencies are present between data sources and, if they are, the estimate of the total population size can be adjusted accordingly. A contingency table can be constructed which describes the overlap pattern between the different sources, and these overlap data can be analysed using a statistical package. In short, the seven pieces of information in the contingency table can be used to predict the missing value, which would be an estimate of the number or drug users not present in any of the three data sources - the unknown population. This is known as log-linear analysis. Different relationships between the data sources can be described using this analysis.

The decision to include dependencies can be taken by examining how similar the observed overlap pattern is to what would be expected if such dependencies were actually present. To make this decision, a value called the deviance, or residual deviance, is examined when fitting each model. This value, known as G^2, is very similar to the values from the χ^2 distribution that are commonly used to examine contingency tables. Each model that is fitted has an associated number of degrees of freedom (df). The more interactions that are fitted in a model then the smaller the number of degrees of freedom

are. The deviance value is compared to values listed in statistical tables, taking into account the degrees of freedom; the closer the deviance is to the degrees of freedom, the better the model is. A 95% confidence interval can be attached to each estimate; this interval gives an indication of how reliable the estimate is (Cormack, 1992).

There is currently some debate as to the validity of using these methods, which were initially developed to estimate the size of animal populations, to estimate the prevalence of drug misuse. From the original application in estimating the size of animal population (for example see Seber, 1982), there are a list of assumptions inherent in the analysis and conditions that must hold before the estimates are valid. Most of the assumptions translate into our epidemiological application, for example:

- The population is closed; there is no movement into or out of the population in the period that is being studied;

- Being present in one source does not effect the probability of being in another source;

- Those that are present in more than one source - the overlap cases - are identified as such.

The first assumption translates into an assumption that the sources are samples from the same population, the third assumes that the identification of overlaps is not subject to error. This is not always the case as the data from different sources are usually subject to error, as will be the methods of matching similar records across sources.

The second assumption is often harder to describe as it relates to more than one problem. The first problem is that the study population is heterogeneous; drug users, for whatever reason, are not all equally as likely to be present in a source. The second problem is that those present in one source may be more, or less, likely, to be in another. Steps can be taken to make the first part of the assumption valid. One method would be to stratify the population, perhaps by age or sex, to give a more homogenous population on which to undertake the analysis. Where possible in this report, the analysis has also been undertaken on males and females separately and also on data split into different age categories. A fuller discussion about these caveats, and the possible violation of the assumptions used in the capture-recapture method, is presented in the discussion of this report.

Several review papers have been published over the last few years. The International Working Group for Disease Monitoring and Forecasting give an insight into the methodological issues that are raised when using capture-recapture (IWGDMF, 1995a, 1995b). They list over 50 epidemiological papers which employ the methodology; nine of which examined drug use or its consequences. The methodology is well established in disease registers such

as those concerning insulin-dependent *diabetes mellitus*. It is recognised that registers which collate data from various sources have an inherent under-ascertainment, such that there will be cases missing from the register. However in these situations the underascertainment which is to be estimated is usually a much smaller proportion of the known population than the proportion of drug misusers that are not identified from a multi-source enumeration.

Truncated Poisson Models

While the capture-recapture methodology requires three or more samples to obtain estimates of the size of a hidden population, Truncated Poisson models use data from a single source. This prevalence estimation method can perhaps be seen as an extension to the traditional capture-recapture method, and a good example of its use is the report from the city of Rotterdam, in which employs truncated Poisson estimators proposed by Zelterman (1988) and Chao (1987), along with a currently unpublished extension to the methodology.

These estimators can be applied to data generated by counts of individuals identified from within a single data source once, twice and so on. Those who are never seen fall into the zero frequency class and are missing from the observed series of frequencies. Therefore, the frequencies of the visits are incomplete and are called 'truncated below one'. Naturally, the total population size equals the number of persons ever seen plus the number of persons never seen. The estimation problem, then, becomes to estimate the number of persons never seen from the truncated series of persons ever seen by assuming that the observed series of frequencies follows a Poisson distribution which is truncated below one. This can be done by fitting a Poisson distribution to the complete series of count data, however Zelterman's and Chao's estimators both only use the information on those that are identified once or twice.

This emphasis on the lower frequencies classes makes sense. People seen rarely are likely to bear a greater resemblance to those never seen, than those seen often. In addition, the emphasis on the lower frequency classes makes the estimators robust in the presence of 'heterogeneity'; e.g. persons seen very often may form a different subgroup to persons seen rarely. The influence of the persons often seen is weighted down in both estimators and therefore heterogeneity, if present, is likely to exercise a relatively small influence. Finally, emphasis on the lower frequency classes results in another bonus as well; both estimators are known to perform rather well even when few data are available (Chao, 1989). These estimates are again subject to some of the assumptions present in the capture-recapture analysis: the closed population assumption, the homogeneous population assumption and the constant probability of being observed more than once assumption, in this case over time.

3 EMCDDA Pilot Study

In February 1997, a questionnaire was sent to key researchers identified by the national focal point of each country in the European Union to gain information from the potential partners for this project. These potential partners were assessed using several criteria, including the resources to do capture-recapture within the time period and the quality of the available data. It was also decided that the social diversity of the EU should be accounted for by working with partners from different parts of Europe. From the different member states, 21 faxes were sent out to possible partners throughout Europe. The only country from which we did not identify a possible partner was Luxembourg. Information was received back from 16 researchers from 12 countries, and from this, 13 cities were considered for involvement in the project. Some countries had many experts in capture-recapture and in other countries capture-recapture had not previously been undertaken however, to enable representations from a broad range of countries without being biased towards the north of Europe, researchers from the following cities were invited to a planning meeting in Glasgow:

- Athens, Greece;
- Dublin, Ireland;
- Helsinki, Finland;
- Malmö, Sweden;
- Rome, Italy;
- Rotterdam, The Netherlands;
- Setúbal, Portugal;
- Toulouse, France;
- Vienna, Austria.

Researchers from two of these cities (Athens and Malmö) were invited as observers as, although it would not be possible to undertake a capture recapture analysis within the time-scale of this project, the experience of these researchers undertaking similar research in their respective countries would be of benefit to the whole group. The experience of Antònia Domingo-Salvany in applying the capture-recapture methodology in Barcelona was also drawn upon.

In an attempt to facilitate comparability between the cities, a minimum set of requirements and definitions were agreed upon. For example, the ages of those in the samples employed in each city ranged from 15 to 54, enabling estimates to be stratified into 3 groups: 15-24, 25-34 and 35-54. It was also hoped that the data used in the various analyses should refer to the 12 month period from January to December 1996. This was possible in three of the seven cities. The capture-recapture methodology employed in this report assumes that the drug using population is 'closed', in other words people do not begin or cease using drugs within the study period. In the space of a year there may be movement into and out of the drug using population and steps were taken to make this assumption more realistic; for example, only including drug users who have been using drugs during the preceding months. This raised the issue of the relationship between prevalence and incidence, or indeed the effect of drug-related mortality. These issues are addressed in some of the city reports and it is indeed at the city level that the closed population assumption is best addressed, as the nature of drug use in the individual city will affect its validity. Although a direct comparison between an estimate obtained using one, three or six months of data, as opposed to one which uses data from a complete year, would be interesting, sufficient data was not available to explore the possible effects of using different time periods.

The Expert Group discussed extensively the different definitions relating to the use of drugs. These discussions were augmented by a joint meeting with the co-ordinators of the EMCDDA project which examined the prevalence of drug use at the national level. Clearly this report is not able to exhaustively discuss the semantics of drug misuse, however a discussion of the different case definitions, particularly in relation to prevalence estimation, is warranted.

The joint EMCDDA and Pompidou group Scientific Seminar took the title "Addiction Prevalence Estimation: Methods and Research Strategies", 'addiction' was, however, used only as a short-hand term to indicate that the emphasis was on methods for investigating the prevalence of heavier patterns of drug use. Indeed the title of the resultant Scientific Monograph used the different term 'problem drug use' and this was taken as a working definition in the national prevalence estimation project.

In short, there are two main factors that need to be considered when talking about definitions; the type of drug and the nature of its use. Some drugs are often viewed as non-problematic, for example cannabis, whereas there are other drugs, particularly opiates, that are usually considered more of a problem. The nature of the use of drugs also ranges from occasional use to frequent or high dose consumption, including injecting. Although the use of stimulants such as amphetamines or cocaine can be problematic, particularly in some North European countries, this study focused on opiates. In many of the cities this almost entirely referred to heroin, however methadone, prescribed or otherwise, was not excluded from this definition.

The tenth revision of the International Classification of Diseases (ICD10) codes relating to the use of opiates (opiods) describe harmful use as 'A pattern

of psychoactive substance use that is causing damage to health. The damage may be physical or mental'. They also refer to dependence syndrome as 'A cluster of behavioural, cognitive and physiological phenomena that develop after repeated substance use and that typically include a strong desire to take the drug, difficulties in controlling its use, persisting in its use despite harmful consequences, a higher priority given to drug use than other activities and obligations, increased tolerance and sometimes a physical withdrawal state. The dependence syndrome may be present for a specific psychoactive substance, for a class of substances or for a wider range of pharmacologically different psychoactive substances.'

These codes are useful in a medical context, and some of the data sources used in the different cities use them as a case definition. Other data sources, particularly the police, are only concerned with people that **use** opiates. Indeed it is only an assumption that someone caught in possession of an opiate is actually an opiate user. It may therefore only be possible to attach a definition to the population that is being estimated in the light of the definitions of the contributing data sources. In a similar manner, although classifying someone as a drug injector may be less subjective, it is not possible to identify the route of administration from some data sources. We employ the term '**use** of opiates' as a broad definition.

It is worth contrasting our definition with that of the national prevalence estimation project. That project defined 'problematic drug use' to include drug injecting as long duration / regular use of opiates, cocaine or amphetamines, but they excluded ecstasy and cannabis. If we isolate 'problematic opiate use', which is common to both projects, we are left with two discrepancies between the definitions; non-problematic opiate use, and problematic use of amphetamines or cocaine.

Capture-recapture methods can be used to estimate the prevalence of the use of other drugs; a study in Liverpool examined opiate and cocaine use (Squires *et al.*, 1995) and the Helsinki project additionally examined the use of amphetamines. However, to achieve comparability throughout the different cities, we restricted the definition to opiates. The other discrepancy is more interesting.

It is possible, in some studies, to divide opiate use into problematic and non-problematic use; however, the proportion thought to be problematic users will vary throughout Europe and will depend on the data collection process. Non-problematic users may be in the minority, particularly those identified through treatment agencies. Those identified through police data will presumably be more likely to be non-problematic, although still in the minority. General population surveys will be even more likely to be including non-problematic users. Thus estimates derived from capture-recapture studies will be more aligned to problematic opiate use than those obtained from general population surveys, although the wider definitions of surveys are not always apparent due to methodological problems such as under-reporting. Care must always be taken in comparing estimates derived from different methodologies, and as

can be seen in later sections of this report, care should also be taken in comparing the results of the same method applied in different areas, due in part to the varying proportion of non-problematic users.

The experts met on two occasions; in Glasgow at the beginning of April and in Utrecht, when most of the cities had obtained estimates in October. In the intervening period meetings were held in five of the seven cities between the co-ordinating researcher and those responsible for the city report. The purpose of these meetings ranged from assisting with the collation of the data and the analyses to advising on the choice of method for obtaining confidence intervals. The network of experts was strengthened by the use of an electronic mailing list with which the discussions initiated in the Glasgow meeting were continued. These discussions often centred on practical issues such as the choice of software used to analyse the data.

4 City Results

Three cities obtained estimates of the prevalence of opiate use in 1996; Dublin, Setúbal and Rome. The report from Helsinki refers to 1995, the report from Toulouse refers to a six month period within 1995, and the report from Vienna uses data from 1993. The prevalence of opiate use in Rotterdam in 1994 was estimated using Truncated Poisson modelling techniques; ongoing research in this city is seeking to employ the three-sample capture-recapture method. In this, and the next section we briefly describe the research in the individual cities and compare the results. Part II of this report presents a more comprehensive report from each of the contributing cities.

Hook and Regal (1997) discuss the validity of methods for model selection and weighting for model uncertainty in capture-recapture estimation. In addition to describing the Akaike Information Criterion (AIC), which can be used to assess whether interactions should be included in models and therefore can be of use in choosing the 'best' model (Akaike, 1985), they discuss the Bayesian Information Criterion (BIC) which can take two different forms, the original as proposed by Schwarz (1978), the other an alteration proposed by Draper (1995). Hook and Regal denote these as SIC and DIC respectively. The formulae for these criteria are as follows:

$$\text{AIC} = G^2 - 2(\text{df}),$$
$$\text{SIC} = G^2 - (\ln N_{\text{obs}})(\text{df}),$$
$$\text{DIC} = G^2 - (\ln (N_{\text{obs}}/2\pi))(\text{df}),$$

where G^2 is the deviance and df is number of degrees of freedom associated with the model, N_{obs} is the 'known' population of opiate users and ln denotes the natural logarithm function. When using any of these criteria the model with the lowest value of the criterion would be the favoured one.

Hook and Regal go on to discuss the use of the weighted Bayesian Information Criteria, in which both the SIC and the DIC can be used to obtain a weighted average of the different estimates from the three-sample capture-recapture method; these weighted averages can also be applied to the upper and lower values of the 95% confidence intervals associated with the eight estimates. These weighted estimates can be considered as 'Bayesian' as they combine an estimate with a measure of how likely it is, to produce the weighted estimate. Hook (personal communication) recommends using all possible three-sample estimates to construct the weighted estimate, including the model which includes all two-way interactions between the three data sources and has an associated deviance of zero. There does, however, appear be to some justification for not including this 'saturated model'. Similarly, it may be questionable to include estimates in which the associated deviance suggests that the model clearly does not fit the data.

For completeness in this discussion of the various models and estimates obtained using the methodology, we can also present the three estimates (and related confidence intervals) obtained using the two-sample capture-recapture

method. In Appendix 1 we present, in tables, the 11 different estimates that have been obtained in each city, these are:

- The independence model;
- Three models with a single interaction, denoted S1xS2, S1xS3 and S2xS3;
- Three models with two interactions, denoted S1xS2+S1xS3, S1xS2+S2xS3 and S1xS3+S2xS3;
- The saturated model with the three interactions present
- Three two-sample models, denoted S1S2, S2S3 and S1S3.

We present the AIC, the DIC and the SIC for each three-sample models and also the weighted DIC and SIC estimates. We also present the associated 95% confidence intervals using a method proposed by Cormack (1992) or the asymptotic method for the two-sample method (Bishop *et al.*, 1975, p. 235). We also present, in Appendix 1, some of the various estimates and confidence intervals in graphical form, ignoring estimates which had excessively large confidence intervals. These estimates and confidence intervals can, of course, be found in the relevant tables, and more in-depth discussions about model selection can be found in the individual city reports.

Dublin

Dublin, with a population of just over one million, is the capital and largest city in the Republic of Ireland. Reports in the local media of an increase in the use of heroin sparked an interest in prevalence estimation, and a capture-recapture study was initiated to inform policy makers in the city. Working over a tight time schedule, the researcher in this project was responsible for all stages of the capture-recapture process, including collecting the data. Three sources were employed; a methadone treatment list, a database concerned with drug related hospital admissions and data from the police. In total 6,264 individuals were identified, from these sources, as using opiates.

There were 3,169 people in the methadone list, 3,787 people identified from the police and the overlap between these two sources was only 885. These sources combine to give a two-sample capture-recapture estimate of 13,560. While the latter estimate is only valid if there is no dependency between these sources, it does not seem feasible that the strength of a possible positive dependency would result in an estimated population being so close to the number suggested previously by some (6,000). The figure obtained using only the methadone and hospital sample is closer to this at 6,146, however the latter estimate may be better described as opiates users that have a ‘medically defined’ problem due to the nature of the sources.

It was difficult to adequately fit a model to the overlap pattern from these three data sources, therefore the population was stratified by both age and gender to partially solve the problem. The best estimate using the combined data suggested a hidden population of 7,335 and a total population of 13,599. This model assumed that the police source was independent of the other two sources. Examining the stratified models suggested that the interaction

between the hospital inpatient and methadone source was present in all gender / sex stratifications and that an estimate of 13,460 was preferable. Although larger than perhaps would have been expected, this estimate is comparable with other cities of that size and with the young age profile of the city. On-going research in this city is seeking to examine the reasons why there was difficulty in fitting models.

When using information criteria, the fact that none of the models, apart from the saturated model, adequately fit the data was reflected in the high values of the criteria. Thus the saturated model would be chosen with its estimate of 22,444; this is similar to both weighted Bayesian estimates. If we assume that the estimate obtained from fitting the saturated model is totally unrealistic, and therefore exclude it from the weighted Bayesian estimates, the weighted SIC and DIC estimates would be 14,125 and 14,130 respectively.

Helsinki

Alcohol has been the traditional problem substance in Finland, and in the city of Helsinki the use of amphetamines is of concern. With a population of just over half a million in Greater Helsinki aged between 15 and 54, the number of opiate users identified from three sources is very small compared to other European capitals. Estimates were, however, provided not only for opiate use, but also amphetamine use. The population was stratified by age and sex and the estimates were extrapolated to provide estimates for the whole of Finland.

In total, 175 individuals were identified as opiate users from three sources; a hospital discharge register, a criminal register and a list of people found driving under the influence of drugs. The two-sample estimates vary markedly, ranging from 244, using driving offender and drug offender data, to 1,025 when ICD9 data is used with the driving data. The latter estimate, as shown in Table 8 of Appendix 1 has a negative lower bound for the confidence interval; this problem is due to the inappropriate use of a symmetric confidence interval.

From the three-sample analysis, which suggested that there was an interaction between the driving and drug offending sources, there were an estimated 775 opiate users. This model has the lowest AIC, SIC and DIC, with the weighted SIC estimate slightly lower at 742 and the weighted DIC estimate slightly higher at 789. The confidence intervals associated with the weighted estimates are large, due to the extremely large confidence interval attached to saturated model’s estimate.

Rome

With a population of approximately 2.8 million, Rome is the largest city involved in this Pilot Study. The Italian capital city has a good reputation for epidemiological research into drug misuse and two surveillance systems were employed in the research (treatment centres and hospital discharges). A mobile emergency unit was also included, perhaps making the estimate biased towards a medical definition. There were 6,896 individuals identified from these three sources, although most came from the surveillance system.

Both of the two-sample capture-recapture analyses, which include the largest data source, produced estimates of around 13,000, however when the two, much smaller data sources are used, the resultant estimate is far smaller, less even than the number identified from the surveillance system. The three-sample capture-recapture model suggests including an interaction between these two sources. If only this interaction is included, then an estimate of 12,997 is obtained. The city researchers also included an interaction between the surveillance system and the hospital discharge data given a larger estimate of 14,278. When age / sex stratifications were explored, estimates were available for most sub-groups.

The size of the interaction between the two smaller sources is reflected in the three-sample models, with the SIC and DIC both suggesting the inclusion of only that interaction. This contrasts with the AIC which suggests the saturated model and the far larger estimate of over 24,000. Despite the size of this estimate, the weighted SIC estimate is within the bounds of the 95% confidence interval of the estimate suggested by the SIC and DIC; both weighted Bayesian estimates are close to the estimate opted for by the city researchers, which had the second lowest AIC, SIC and DIC.

Rotterdam

The Netherlands is another country with a history of epidemiological research into drug use. Rotterdam is the country's second largest city with a population of approximately 600,000. Previous attempts to estimate the number of opiate users applied the two-sample capture-recapture analysis (Korf *et al.*, 1994) and the multiplier method (Wiessing *et al.*, 1995) using data from a HIV-surveillance study and the number of registered methadone clients. The estimate from these studies were 4,000 - 4,500 for 1988 - 1990 and 3,500 - 4,000 for 1994. Because of problems with the police data from 1994, it was difficult to perform the three-sample capture-recapture (the two other data sources available for 1994 were data from methadone maintenance treat and a HIV-surveillance study). The research in this city is on-going and the inclusion of police data is being examined.

As an alternative to the three-sample capture-recapture analyses the city researchers applied seven different Truncated Poisson models to the 1994 data from the methadone treatment agencies. There were 2,029 individuals

identified from the methadone treatment agencies and the estimates varied from 2,937 to 5,006.

Setúbal

This city, the third largest in Portugal, is the smallest to be included in the Pilot Study with a population of 85,292. The city is thought to have a significant drug problem, and a study on HIV and hepatitis in the area showed that in a sample of 379 drug injectors, 85% were hepatitis C positive and 20% were HIV positive. Despite these figures, there appears to be few data sources that systematically collate data on drug misuse. Two sources were examined, one of which was a specialised drug treatment centre from which clients were recorded as attending only in the first, only in the second, or in both semesters of 1996. This single data source could therefore be split, by semester, into two 'sources', although these could not be assumed independent. The inclusion of data from health centres enabled a three-sample capture-recapture analyses. Again, the lack of a police source would make the estimate biased towards problematic opiate use. In total 339 individuals were identified; 313 from the treatment centre and 40 from the health centres. The overlap of 14 between the two samples there gives an estimated 894 opiate users. The two-sample capture-recapture estimates using the individual semesters data and the health centre data straddle the former two-sample estimate. As the preferred three-sample model included only an interaction between the two semesters of the treatment centre the three-sample capture-recapture is similar to a two-sample analysis between the combined treatment and the health centre data. Although using an information criterion in this analysis where we assume *a priori* that an interaction is present, may not be relevant, the preferred three-sample estimate does also have the lowest value of the AIC, SIC and DIC.

Toulouse

The conurbation of Toulouse, in the South West of France, has a population of 650,000. Previous research in 1994 estimated the prevalence of opiate addiction in the area. Additional sources and data which refer to 1995 have now been obtained. In total, ten data sources were available for this research. As the remits of some of these sources were comparable, these data were reduced into three sources by merging similar ones together. In total, 799 individuals were identified and this information was used to estimate the total number of opiate user at 2,178; the estimate assumed a dependency between the two 'non legal' groups of data sources. The population was stratified into two age groups and by sex and, although the stratified models did not all include the same dependencies, the results from the combined data was consistent with the results from the stratified data.

This preferred estimate happened to be the one with the lowest AIC and the lowest DIC, but the SIC suggests that the independence model may be the best, giving a lower estimate of 1,808. Both weighted Bayesian estimates lie between the estimates suggested by the criteria. The two-sample estimates again show that, when using the sources which have the strongest interaction,

the estimate is low, however neither of the other two-sample estimates differ markedly from that suggested by a three-sample analysis.

Vienna

The prevalence of opiate use in Vienna in 1993 has been estimated using four data sources; a police source, hospital admissions, an emergency ambulance data source and data on drug related deaths in the following year. The hospital and ambulance source were found to be quite similar and were therefore merged. Two-sample estimates are listed in the city report, using the hospital and ambulance sources separately and combined, and these range from 2,217 to 5,198, excluding the estimate obtained from using only the ambulance and hospital data which results in an very small estimate. In total 1,028 individuals were identified, from which the total population of opiate users was estimated to be 6,747 using a model with interaction between the ambulance / hospital source and both of the other sources.

This model was also suggested by both the AIC and the DIC, however the SIC suggested the independence model. Apart from Dublin, where the saturated model was suggested as the best by all criteria, Vienna was the only city in which two interactions were present in the best model. This could be due to the unique nature of the analysis which included drug-related deaths. As can be seen from Table 12 of Appendix 1, the estimates obtained from different models are varied, for example the estimate suggested by the SIC is less than half that suggested by the other criteria. The estimates obtained using the two-sample method are all lower than that opted for by the city researchers; this again may be due to the nature of the interactions between the three sources. Both the weighted Bayesian estimates lie between those suggested by the different criteria, however the 95% confidence intervals of these estimates appear preferable to that associated with the preferred estimate.

Summary

To summarise the various analyses undertaken in the different cities, Table 4.1 presents a comparison of the estimates, including the prevalence rate in the 15 to 54 age group, and the associated 95% confidence intervals. Although the research in Helsinki additionally estimate the prevalence of amphetamine use, we present, in the following tables, the prevalence of opiate use.

Table 4.1: **Comparison of the results from seven cities.**

City	Known Users	Total Users		Prevalence (%)	
		Est.	95% CI	Est.	95% CI
Dublin	6,264	13,460	10,665-14,804	2.11	1.68-2.33
Rome	6,896	14,278	12,741-16,167	0.86	0.76-0.97
Setúbal	339	894	620-1,423	1.82	1.26-2.90
Helsinki	175	775	487-1,392	0.14	0.09-0.25
Toulouse	799	2,178	1,780-2,734	0.54	0.44-0.68
Rotterdam	2,029	3,716	3,497-3,990	1.07	1.01-1.14
Vienna	1,028	6,747	4,332-11,668	0.67	0.43-1.16

It is clear from this table that there are large differences in the estimated prevalence values from the seven cities. This may be due to several reasons; these are discussed in the individual city reports, the summaries above, and the discussion which follows in the next section.

5 Discussion

Cross-city comparability

It was the objective of this Pilot study to compare and contrast the use of the same prevalence estimation methodology in four or more European cities. While the focus of the study has been methodological, we have presented seven estimates in Table 4.1 from studies which used the three-sample capture-recapture methodology. We can therefore begin this section by comparing these estimates.

Table 5.1 extends this prevalence information. At this stage we concentrate on the six cities which employed the three-sample capture-recapture methodology.

Table 5.1: **Comparison of the results from six cities.**

City	Total Users	Total : Known	Prevalence (%)
Dublin	13,460	2.1:1	2.11
Rome	14,278	2.1:1	0.86
Setúbal	894	2.6:1	1.82
Helsinki	775	4.4:1	0.14
Toulouse	2,178	2.7:1	0.54
Vienna	6,767	6.6:1	0.67

From this table, we can compare the total to known population ratios that were found in the cities. Comparisons between the cities can only be made with caution as the data collected from each city would come from different sources, and an exhaustive multi-source enumeration, such as that done in a case finding exercise, was not undertaken in each city. That stated, the total to known ratio seems consistent between the cities, with the exception of Vienna and Helsinki. The latter city does have the lowest prevalence value, both in size and by percentage, and as noted in the city report there traditionally is not a large opiate problem. The total to known ratio can perhaps be seen as a type of check, or validation. Using this ratio is similar to the multiplier method in which a constant multiplier is applied to treatment data in different areas; although the capture-recapture method is an improvement on the multiplier method, the fact that the ratio in Dublin and Setúbal is similar to the other cities makes these high prevalence values appear more reliable.

We can compare the cities further by examining the male to female ratio, both in the known populations, as in Table 5.2, and the estimated populations (Table 5.3).

Table 5.2: **Comparison of the known populations from six cities.**

	Males		Females		
City	**Known**	**%**	**Known**	**%**	**Male : Female**
Dublin	4,917	1.60	1,387	0.42	3.55:1
Rome	5,711	0.70	1,185	0.14	4.82:1
Setúbal	281	1.17	58	0.20	4.84:1
Helsinki	54	0.02	35	0.01	1.54:1
Toulouse	583	0.29	215	0.11	2.71:1
Vienna	819	0.17	209	0.04	3.91:1

Table 5.3: **Prevalence of male and female opiate use in six cities.**

	Males		Females		
City	**Estimate**	**%**	**Estimate**	**%**	**Male : Female**
Dublin	6,831	2.2	3,179	1.0	2.15:1
Rome	12,649	1.5	2,368	0.3	5.34:1
Setúbal	696	2.9	208	0.8	3.35:1
Helsinki	536	0.2	310	0.1	1.73:1
Toulouse	1,709	0.8	466	0.2	3.67:1
Vienna	5,746	1.2	554	0.1	10.37:1

Notes: The figures for Helsinki were not calculated by the city's researchers.
The figures for Dublin are from models that do not adequately fit the data.

It should be noted that comparing both the known populations and the estimated populations, the male to female ratio is lowest in Helsinki, and higher in Rome ans Setúbal. The much higher male to female ratio in Vienna may be due to the difficulty this city had in fitting models to sparse contingency table data.

The difference in prevalence rates between Dublin and Rome, the cities with the two highest numbers of opiate users, is not so pronounced when examining males only. Indeed, although Dublin has the largest number of known females, the size of the estimated female opiate users, coupled with the lack of fit of the model, could cast doubts on this estimate.

We can also compare 'young' and 'old' drug users in Table 5.4. In this table we use the estimates presented in the city reports where available, and combine stratified estimates when necessary.

Table 5.4: **Prevalence of opiate use in young and old age groups in six cities.**

	Young			Old		
	Known	**Estimate**		**Known**	**Estimate**	
City		**N**	**%**		**N**	**%**
Dublin	5,604	10,964	2.94	660	1,067	0.41
Rome	4,521	10,365	1.17	2,375	4,447	0.58
Setúbal	284	884	3.54	55	85	0.72
Helsinki	54	122	0.11	121	464	0.11
Toulouse	649	1,709	0.74	131	426	0.24
Vienna	613	3,393	0.71	94	3,354	0.73

Notes: In addition to the notes of Table 5.3, young refers to 15-34, apart from Helsinki which is 15-24.

Again, the demographic profile of both the known and estimated populations differ among the cities. The high percentage prevalence in the young group for Setúbal should be critically assessed, particularly as the 'young' estimate is virtually as large as the estimate obtained when modelling the combined data. Although Helsinki used a different age stratification, the constant percentage in both age groups suggests that drug use would be equally prevalence in groups stratified at 35 years old.

An interesting statistic, which has not been presented in Table 5.4, is the percentage of the total population of Dublin in the 15-24 age group that are known to be opiate users. At 1.63%, this figure is remarkably high, again suggesting that the large total prevalence value for Dublin may be due to the large number of young, male opiate users. The two preferred estimates for males aged 15 to 24 were 5,244 or 5,404 (5.4% or 5.6%) from a known population of 2,469 (2.5%).

While cross city comparisons about the known to unknown ratio or the male to female ratio can perhaps be used as a form of validation for the city researchers, it is at the city level that these ratios would need to be considered. A good example would be Helsinki, which, in contrast to the others cities, has a low, but almost constant prevalence, when stratifying the population by age or gender. This could be due to the nature and extent of opiate use in that city; for example, studies in some areas show that more males than females inject drugs, and other studies show that males are more likely to contact treatment centres. Whether or not these factors are an issue in this city is beyond the scope of this comparative report, but should perhaps be considered in capture-recapture studies.

It should also be noted that the more restrictive case definition of the Toulouse study, in which non-residents were excluded, may have, in part, resulted in a lower prevalence estimate. While the effect of only including permanent residents may be negligible in some areas of Europe, it was certainly a factor in the Rotterdam project where there were many non-residents in the police

data source; this would have led to difficulties in a three-sample capture-recapture study.

To summarise the comparison of the prevalence estimates, it is worth re-iterating that the different estimates are obtained using different data sources. The most obvious comparison, and demonstration of this point, again involves Dublin and Rome. The former uses a very large police sample (in fact the largest of the three) and, as stated in the city report, half of those were identified as opiate users because they admitted using drugs, and most were young males. The heterogeneity cause by including a source in which many, if not most, of the largest sample have a legal problem rather than a medical problem, could be the reason why there was difficulty in fitting models. This problem could be compounded with the apparently recent nature of the opiate problem in Dublin which may not have filtered through to the medical based treatment system. In contrast, the data in Rome, and indeed Setúbal, could result in the estimate being one which describes the more problematic use of opiates.

Heterogeneity

It is clear from the above section that the variation in the composition of the data sources used within a capture-recapture analysis underpins one of the main caveats related to the methodology, i.e. what is the definition of the drug using population that has been estimated. It is very simple to recognise that these variations exist, and it is convenient to label any related problems as 'heterogeneity'. From the dictionary, the word heterogeneity is derived from the adjective meaning 'composed of diverse elements.' Drug users, and their lifestyles, are diverse, and this diversity is reflected in many ways. Stratification can, in part, reduce heterogeneity by breaking the population down into more homogeneous sub groups, however the inherent diversity is often more complex.

Some of the caveats that need to be considered when applying capture-recapture are simply manifestations of heterogeneity. For example, while the debate over open / closed populations is more pertinent when examining animal populations, where samples are taken over successive time periods, when looking at human populations, this issue may simply be one of heterogeneity. There may be drug users who begin, or cease, using drugs over the study time-period and these people can be thought of as being less likely to contact services than those who had used drugs throughout the year. This may not be such a problem in areas where the average duration of an opiate user's 'career' is long in relation to the study period. Similarly, we can consider a drug using career as a lifestyle which changes over time such that older drug users have different patterns of service contact than younger drug users. Geographical movement into, and out of, the study population may, however, be more of a problem in relation to the closed population assumption; this problem was encountered in the Rotterdam study.

While the presence of heterogeneity can perhaps be assumed by the lack of fit

between the data and available models, it may be beyond the scope of this Pilot Study to completely resolve this issue. We can perhaps use simulation techniques to explore the effects that heterogeneity may have on the resultant estimate, and this is described later.

Case Definitions

Questions have arisen in this Pilot Study, as in other research projects which study the drugs phenomenon, about case definitions. The term addiction infers frequent or high dose consumption and this is often typified by drug injection. Those that are 'addicted' have a higher risk of drug related problems such as dependence on drugs or infection with HIV or hepatitis. Interventions targeted at combating dependence exist in both medical and social frameworks, whereas interventions targeted at the health consequences of drug use are usually based within medical agencies. Within the medical framework, there are different responses to drug use and the related problems; ranging from harm reduction measures such as needle exchanges for drug injectors and substitute prescribing with drugs such as methadone, through to admission to hospital. Intertwined with drug-related morbidity is the increased mortality of drug users. Some agencies may offer both a social and medical response to drug use.

Other drug users may be better classified as having only a legal problem with drug use, either due to laws restricting the possession or the supply of drugs, or the need to commit crime to finance their drug use. These drug users may be unlikely to contact medical services.

Drug use is often a transient activity, both in the frequency that people use drugs and the type and route of administration of the drugs used. The border between problematic use and non-problematic use can therefore be vague. An individual's drug use may not initially cause a problem, save for the legality of the possession of drugs. This may evolve from a legal into medical problem. Other drug users may never contact services, however, those that use certain drugs may be more likely to seem as problem drug users in the medical sense of the term, in particular opiate users.

This Pilot Study employed the broad definition 'opiate use', principally as it is difficult to ascertain how problematic this use may be for individuals within samples. In some cities non-problematic use of opiates, as identified from data sources, would be rare; in others, larger sections of the known opiate using population are not yet thought of as problematic. For comparability with the EMCDDA funded project which examines drug use at the national level, a case definition such as problematic opiate use would have been preferable, although it is unclear as to how many of the cities included in this report would meet this definition. Table 5.5 presents a summary of the sources used in this study into broad categories, and the percentage identified from each.

Table 5.5: **Comparison of the data sources employed in seven**

cities.

	Medical	Social	Legal	Other
Dublin	Methadone (51%) In-Patient (9%)		Police (60%)	
Helsinki	In-Patient (37%)		Police (60%)	Driving (13%)
Rome	In-Patient (9%) Treatment (93%) Emergency (5%)			
Rotterdam	Treatment (100%)			
Setúbal	Treatment (92%) GPs (12%)			
Toulouse	In-Patient/Treatment (63%) Low Threshold (33%)		Police/Prison (21%)	
Vienna	In-Patient/ Emergency (24%)		Police (70%)	Deaths (14%)

It is clear from Table 5.5 that most of the data sources used in the seven cities are medical based. Indeed in Rome, Rotterdam and Setúbal the 'known' opiate users can all be thought of as problematic and therefore the estimate obtained would refer to the number of problem opiate users within each city; in Dublin and Helsinki is harder to attach a definition to the estimate. The problems in fitting models to the Dublin data might suggest that even a hybrid definition of 'medical-legal' problematic drug users may not be valid. In Vienna however, a more relaxed definition of opiate use may be preferred as many of those identified come from a legal source. The approach taken in Toulouse, where initially ten sources were reduced to three, resulted in models fitting the data well. This perhaps made the problem of assigning a definition to the estimate easier to solve as some of the data from that city came from specialised centres which were mainly working on a psychotherapeutic basis, and therefore may be more easily classified as intervening with those who have a 'social' problem. The estimates may therefore not be restricted to the medical or legal definitions of other contributing sources.

It appears, therefore, that the definition of the population that has been estimated depends on the case definition of the samples that are used. In addition, the relative numbers in each source and the goodness of fit of the models which employ these sources may be important. In the case of cities in which data sources have similar definitions then it appears, as was the case in Setúbal models fit the data better.

Model Selection

The selection of the models used to describe the overlap pattern and to produce an estimate of the size of the hidden population is, on one level, intrinsically linked with the preceding discussion about case definitions and the data sources that are employed. On another level the concept of parsimony is also important such that simple models without complex interactions may be preferable. All of the possible models that can fitted to the data and the results form Appendix 1.

It was not possible, in some of the cities, to fit certain models to the stratified data due to zero values in some cells. While none of the participating cities had structural zeros, the size of samples used, especially when stratified, did sometimes result in small values in some cells. Related to the problem of structural zeros, where a mechanism exists which ensures that no overlap can occur between certain sources is the problem of artificial overlaps where inclusion in one source will automatically mean inclusion in another. It is interesting to note that in the two cities with smaller samples; Helsinki and Setúbal, it was easier to fit models, and in some cases the independence model was satisfactory, even though it may have been sensible to include an interaction between sources.

An additional benefit of the three-sample capture-recapture method, as opposed to the two-samples, is that a specific interaction can be included in the model, even when a criterion such as the AIC does not make the suggestion. In many of the cities, the interactions that were included were consistent with what professionals in the field would have expected, i.e. ‘medical’ sources being dependent on each other but independent of a ‘legal’ source. Without returning to the discussion on the case definitions used in the different lists, this is sensible when you consider the differences between those who have ‘medical’ problems due to drugs and those who only have legal problems. A good example of including interactions which were not always identified statistically was Setúbal. In this small city there was difficulty in collecting data and this resulted in small samples, two of which were different semesters of the same treatment centre. Within the unstratified analysis, including a positive interaction between these two semesters greatly improved the fit of the model. This, of course, was to be expected, but when examining the overlap patterns after stratifying the population, the independence model was usually sufficient. As the interaction between the two semesters was to be expected, it was included to obtain stratified estimates.

The debate between including interactions because they seem sensible, and being led by statistics such as the change in deviance or the AIC, is made more interesting because log-linear modelling, which capture-recapture modelling essentially consist of, is the preferred method of examining categorical data, and thus discovering relationships between sources. It is also possible, as was the case in Setúbal, that the sample sizes, in this case the data from the sources may not be large enough to show significant interactions, even when they are

present. If these interactions are thought to be present between data sources then they should perhaps be included.

Confidence Intervals

The estimates obtained from the application of a method such as capture-recapture need to be assessed in conjunction with the statistical uncertainty that is inherent in any estimate. Those that successfully undertake capture-recapture studies are in some way fortunate in that the estimates produced can usually be compared with ball-park estimates derived by those working in the field. The number of 'current' drug users within a city will be variable and, as trends in drug use can fluctuate, the best any prevalence study can hope for when using retrospective data is a historical estimate of drug use; this usually is still of relevance to policy makers at the present time. Additionally, any statistical estimate needs to be interpreted with caution for the reasons detailed below.

The estimate produced from a capture-recapture project is, as shown in the city reports, the end point of a process in which error can be introduced at several stages. Mistakes can be made during the collation of an individual data source; identifiers such as the date of birth can be recorded erroneously either by mistake or because falsified information has been recorded, particularly in a low-threshold agency. The matching process is not always guaranteed to correctly identify all matches, or eliminate any possible false matches. It is only once a contingency table has been produced that the statistical modelling can commence, and even then, the relevance of the confidence interval depends on how accurately the model portrays the relationships between the data sources.

Various methods for producing a confidence interval can be used, however each city used a method favoured by Cormack (1992) which recognises that the estimate for the hidden population is derived from an asymmetric distribution. Thus the possible problem of producing confidence intervals, in which the lower bound of the total drug using population is less than that which has been identified from the multi-source enumeration, does not arise. Some of the upper limits of the confidence intervals reported in the in analyses from the different cities are not feasible, and these usually occur when the model has included more than one interaction. That is not to say that a preferred estimate should be chosen because of its small confidence interval.

An interesting addition to the discussion about confidence intervals is some of the analyses performed in Toulouse. Here the BMDP statistical package was used which offered the option of adding 0.5 to each cell of the contingency tables because the observed values were judged, by the package, to be small. While this had the effect of slightly changing the estimate, it also decreased the size of the confidence interval. As the statistical package uses this correction to improve the power of statistical tests, questions should perhaps be raised about its use to produce a point estimate.

The three-sample capture-recapture analysis can be carried out on a range of statistical software, often with the aid of pre-written macros or subroutines. While the point estimates derived are usually exactly similar, the methods for obtaining confidence intervals sometimes differ therefore there needs to be a consensus about which packages are of most use, and perhaps an inventory of available macros should be developed; this is being pursued in a follow-up to this project.

Validation of Estimates

Capture-recapture is only one method of estimating drug use prevalence at the local level and the results should be considered in the context of other estimates, including surveys. Methods which use only one source of data such as mortality multipliers are attractive, particularly when only a few data sources are available. Within this Pilot study, the Truncated Poisson method has been shown to be worthwhile, and where possible the minor alterations to the data collection process should be tried in order to replicate this method in other localities.

By splitting a single sample into two, as the research in Setúbal did by comparing consecutive semesters within a treatment agency, three samples can be employed. Where this has been done, and the obvious interaction between the two semesters has been included, the analyses do, in fact, reduce to a two-sample capture-recapture. However, the benefit of using both semesters is that the presence of a dependency between the treatment source and the other can be seeked. As was the case in Setúbal, where this dependency was not shown to always be statistically present, a valid estimate could be produced; care had to be taken, however, in the interpretation of the result.

Steps can be taken to ascertain how realistic the obtained estimates are. An initial check would be to examine the ratio of known to unknown opiate users as identified in the analysis and to compare these values stratified by age and gender with the demographic profile of the city. This comparison could take the form of deciding whether the results are sensible, particularly from the perspective of those working with drugs users in the city. Again, this needs to be done in the context of the case definitions of the data sources that have been employed.

Simulations

Many of the different scenarios that occur within a capture-recapture analysis can be explored further using simulation methods; for example, dependency between sources and the inherent heterogeneity of the population. In order to do this, a hypothetical population can be created and this population can be split into different cohorts. Members of each cohort can then be assigned a probability of being identified from each source, and this probability can be adjusted to include dependencies between the sources.

In conjunction with the research in Vienna, that is presented within the city reports, different scenarios were simulated. This was done by creating three samples; S_A, S_B and S_C. Each of the samples could be split up into five cohorts C_1 through C_5, each having a certain probability p of being in each of the three samples.

There are two kinds of dependency that can be introduced. One dependency results if the 'catchability' for different cohorts is different in each sample; the other kind happens if the probability of capture in a certain sample is dependent on having been captured in another. The first kind of dependency can be introduced by choosing the probabilities for the different cohorts to reflect the different catchability pattern. The second kind of dependency is introduced by multiplying the probability p, by a factor f, in the samples S_B and S_C in the cases where the person has been captured in sample S_A already.

The size of the cohorts are known, as is the total simulated population, and these simulated samples can be used within an analysis and the resultant estimate can be then compared with the known total population size. The issue of model selection is still pertinent in considering simulated population, and indeed it can be more acute as the artificial heterogeneity or dependencies often result in a difficulty in fitting non-saturated models to the data. Often the saturated model is the only model that fits simulated data. From the different scenarios that had been modelled, it could be shown that heterogeneity constituted a problem but dependency between the samples is not a major problem; indeed, it may help to reduce the bias caused by heterogeneity.

While simulation methods may not completely answer some of the more pertinent questions relating to the application of capture-recapture to drug misuse prevalence estimation, it should be seen as a useful tool in examining different scenarios.

Practical Implementation of the Method

There are various practical issues that are raised when undertaking a study which uses the capture-recapture methodology and it is prudent to discuss these within this report.

A study which aims to estimate the prevalence of drug misuse, whether it is a general population survey to examine the use of cannabis or a capture-recapture study to explore opiate use, needs to be undertaken within the framework of the policy issues relevant to the city. This was evident in Dublin where the funding of the project was provided to answer a specific question, and in Rome where the prevalence study complemented the existing research into HIV and drug-relate mortality. Other, more local, issues may be pertinent to the decision to estimate prevalence, not just at the city level but at the level at which funding of drug related services is decided.

Clearly the success of obtaining an estimate of the size of a drug using

population relies on the data that is available within a city. It may not always be possible to obtain data from three or more distinct sources and a method of prevalence estimation, when this is the case, is discussed elsewhere in this report. Where data from a sufficient number of sources are available, there are certain considerations that need to be explored before inclusion within a capture-recapture study.

The first consideration is another issue which is often dismissed as heterogeneity; the geographical coverage and distribution of the individual samples across the city that is being studied. This issue is perhaps more pertinent when capture-recapture is applied outwith cities and access to drug services may be constricted, however, in larger cities where agencies only take clients form particular areas, the coverage of a samples needs to be assessed. Capture-recapture was not feasible in Athens mainly because it was not possible to obtain data from two of the drug treatment centres in the city; one of which is the largest in the Athens area. Although the remaining six treatment centres took clients from all areas of the city, there could have been difficulties in using this data if it excluded parts of the city. If that was the case, there would be difficulties in combining such data area with other city-wide samples to provide a meaningful capture-recapture estimate. This problem would not just be restricted to geographical coverage; in Setúbal, data on drug users being conscripted into the army may have been available but they would not have included females. Another obvious example would be a data source which only included drug users within a particular age range. While this is not immediately apparent from the data sources included in these studies, questions should arise when the age distribution is markedly different between different samples.

The confidential nature of the data employed in this method of prevalence estimation is countered by the need for accurate identification of the overlap between data sources. This issue was highlighted within some of the different projects, for example in Helsinki the identification of overlaps was done in strict accordance with the instructions from the Office of the Data Protection Ombudsman, whereas the difficulty in obtaining data in Athens was due to the confidentiality of the data. Many analyses rely on un-named identifiers such as initials and dates of birth. It may be rare to have people with similar identifiers in a small city such as Setúbal, or a city with a small number of opiate users such as Helsinki, however the potential for false matches between sources needs to be borne in mind in larger cities.

In some instances, complete data is not available from all sources. This was the case in Vienna, however the city researchers examined different matching procedures and evaluated how many matches may be false. They noted that 17 out of the 727 relevant cases within one of their samples could be thought of as erroneous if they only used date of birth and sex; this was 2.3% of the cases. Considering that there were 69 matches between this sample and the others, they expected something between one and three false matches. There did not appear to be an exact way to estimate the influence of insufficient identifiers based on the distribution of the existing identifiers so they could not

do anything to correct for this source of bias. Since they possibly overestimated the number of matches because insufficient identifier information, the estimation of the hidden population could be treated as a slight underestimation. This problem will not be confined to the Vienna project.

The more methodological considerations of undertaking a drug use prevalence estimate at the local level have been discussed above, however two closing comments are worth considering. Once estimates have been produced, they need to be placed back into the framework of the policy issues, critically assessing what population has an estimate been attached to, and how this may inform policy, or other research. Coupled with this is a recommendation that the results of a study should be disseminated back to the agencies from which the data was obtained. Not only does this help to promote the uses of the method, it also makes additional data collection easier.

Finally in this discussion it is worth considering the reasons why capture-recapture may not be the most suitable method of estimating the prevalence of drug misuse. Participants from two cities, Athens and Malmö, were involved in this research project even though a capture-recapture research project was not feasible. In the case of Athens, difficulty was found in obtaining access to personal data from the police and some of the city's treatment agencies. Setúbal also had problems in collecting data, but this problem was more readily overcome due to the size of the city. It cannot be assumed that adequate data, even where it is available, can be included in a capture-recapture model. Similarly, in larger cities, data may be available but held in several different sources, however the benefits of using surveillance systems which collate information from throughout a city was demonstrated in Rome; the largest city to take part in this study.

6 Conclusions

Clearly the issues concerning the capture-recapture methodology and drug misuse prevalence in general, could be discussed further. Indeed, the next stage of research initiated by the EMCDDA is to disseminate methodological guidelines on the applicability of capture-recapture, and to produce a comprehensive scientific review of capture-recapture studies of problem drug use in Europe with a critical assessment of methods, sources and definitions. Discussions pertinent to the individual cities are contained within the city reports and there are other excellent reviews of the capture-recapture methodology (IWGDMF, 1995a, 1995b), and drug misuse prevalence estimation such as the Scientific Monograph described in the introduction. The discussion generated from the city reports needs now to be put into context with the rationale for this Pilot Study.

Estimates of the prevalence of opiate use have been obtained in seven cities of the European Union. Subtle differences exist in the application of the capture-recapture method in these cities, ranging from definitional differences between data sources to applying the related method known as Truncated Poisson modelling. From one viewpoint, these differences make the comparative elements of the Pilot Study more difficult, but it is these variations on the common theme which enable a more comprehensive assessment on how well the methods work. Clearly, the differences in applying the method in differing settings reflects differences in the nature and extent of drug use in each city and the policy related priorities these create. For some areas, the prevalence of drug injecting may be the more relevant issue and as demonstrated by the Helsinki report, the use of drugs other than opiates may be of more immediate interest.

Even if a ‘gold standard’ version of the methodology existed, its applicability would vary throughout Europe. Pertinent issues to the research in one city may be irrelevant to those planning a project in another; for example the problems of matching confidential records is eliminated where unique identifying numbers are used by several sources within a city. The uses of the derived estimate will also vary; in at least three cities included in this Pilot Study the estimates complement existing prevalence information and gives an idea how, perhaps, the prevalence of drug use has changed. In other cities, information on the level of drug use is scarce therefore possibly imprecise estimates can be useful. The multi-source enumeration forms an integral part of the capture-recapture process and gives an absolute minimum level of drug use; in Dublin, merging two large data sources together and removing the overlap demonstrated that existing ball-park estimate were comparatively quite low.

Although it is desirable to obtain the related confidence interval of an estimate, obtaining very large upper bounds may cast doubts on the estimate. But this uncertainty belies the use of the lower bounds of the interval. Even though the three-sample capture-recapture may not offer a reliable estimate, it can make it easier to interpret the estimates when using two samples. As stated

in the discussion, the reasons why the capture-recapture has failed in some areas will be of interest too.

The application of the methodology needs not to be restricted just to the level of a city. In many cities, it is possible to isolate data from specific areas in which a capture-recapture analysis can inform policy makers at the local level. Capture-recapture can also be extended to larger areas, either by extrapolation, as was the case in the Helsinki report, or by collecting data at a national level.

The extension of the method to other areas would, in part, bring the expertise present in this research project closer to that of the EMCDDA national prevalence estimation project. The prevalence information obtained when using capture-recapture will also inform other research projects, for example those that examine the health consequences of drug use, such as the nature and extent of HIV or hepatitis infection, but also the EMCDDA funded project which examines the potential uses of dynamic models. In the EMCDDA funded national prevalence project, estimates obtained by capture-recapture were employed in other methods, for example using them as anchor points in a multi-indicator regression model.

Methodological advances have been made within this course of this project, more so in the cities where the application of the method was not straightforward. The research in Dublin gave rise to many methodological questions, some of which may still not be completely answered. Indeed it was the nature of the sources used in this city that informed the interesting discussion about the effects that different definitions have on the ability to fit models where heterogeneity exists, and what this means for the definition of the estimate.

The epidemiological advances of this Pilot Study have also been clearly demonstrated within the city reports. Prevalence estimates have been obtained in seven cities and, although the researchers in each city collated the data and analysed them individually, the discussions at both meetings and, in particular the input from people who had experience in applying the methodology, have been invaluable. These ranged from suggesting different ways of interpreting the results, placing the obtained values in context, and promoting the use of certain computer packages. The use of a common set of definitions and how they relate to the estimates within the seven cities has been an important epidemiological advance. While comparative tables, such as those presented in the discussion, are not entirely free from footnotes giving reasons why comparisons should be done with caution, this report presents, for the first time, prevalence estimates not only attained over the same time period and within the same age range (something which is still rare even for surveys), but with the same methodology.

As stated previously, this report is not the end point of the methodological discussions which surround the use of capture-recapture methods in the field of drug use epidemiology. Many of the city reports will be extended into a format for publication in peer reviewed academic journals. It is hoped that

such submissions should be viewed, not only in respect of the estimates that they include, but for the methodological advances attained, thus giving the methodology the credence it deserves in both the fields of epidemiology and addiction. It should also be a priority for the epidemiological advances of the comparative elements of the pilot study to be discussed within the wider scientific community.

As stated in Section 4, where the results from the various cities were briefly discussed, many of the research projects are ongoing. While we have been fortunate in producing estimates using the capture-recapture methodology in sevem cities, these estimates should perhaps be re-evaluated in the light of the methodological discussions contained herewith. Additionally, the process of applying capture-recapture does not always end once an estimate has been produced as the assumptions used within the analysis can be examined and the data employed in the study can be adjusted to, for example, reduce heterogeneity by excluding people that may not be comparable to the definitions of the other sources. These adjustments could take the form of isolating data from a smaller time period, or focusing on a certain age range and can, if sufficient supplementary data has been collected, be undertaken retrospectively.

To conclude, the methodological and epidemiological advances of this Pilot Study are not limited to this report. The experience of being a member of the expert group will be apparent as capture-recapture will be applied again by the researchers in the different countries, perhaps extending its use into other cities. The experience of this group will also be invaluable to the next EMCDDA project which seeks to promote the uses of the methodology to both the wider scientific and policy oriented audience.

References

Akaike, H. (1985) Prediction and entropy. In *A Celebration of Statistics* (eds. A.C. Atkinson and S.E. Fienberg) pp. 1-24. New York: Springer.

Bello, P-Y, Chêne, G, GREATT (1997) Tentative d'estimation de la taille de la population toxicomane et de la prévalence de la toxicomanie à Toulouse, en 1994, par modélisation log-linéaire. Psychotropes. *Revue Internationale des Toxicomanies*, **3**, 97-110.

Bishop, YMM, Fienberg, SE and Holland, PW. (1975) *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge MA.

Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics*, **45**, 427-438.

Cormack, R. (1992) Interval estimation for mark-recapture studies of closed population. *Biometrics*, **48**, 567-576.

Domingo-Salvany, A, Hartnoll, RL, Maguire, A, Suelves, JM and Antó, JM. (1995). Use of capture-recapture to estimate the prevalence of opiate addiction in Barcelona, Spain, 1989. *American Journal of Epidemiology*, **141**, 567-74.

Draper, D. (1995) Assessment and propogation of model uncertainty. *Journal of the Royal Statistical Society (B*), **57**;78-9.

EMCDDA (1997) *Estimating the prevalence of problem drug use in Europe*. EMCDDA Scientific Monograph Series No 1, Lisbon.

Frischer M, Leyland A, Cormack R, Goldberg DJ, Bloor M, Green ST *et al*. (1993) Estimating the population prevalence of injection drug use and infection with Human Immunodeficiency Virus among injection drug users in Glasgow, Scotland. *American Journal of Epidemiology* **138**,170-181.

Hartnoll R, Daviaud E, Lewis R, Mitcheson M. (1985) *Drug problems: assessing local needs. A practical manual for assessing the nature and extent of problematic drug use in a community*. Drug Indicators Project, London

Hay, G and McKeganey, N. (1996) Estimating the prevalence of drug misuse in Dundee, Scotland: an application of capture-recapture methods. *Journal of Epidemiology and Community Health*, **504**, 469-472.

Hook, EB and Regal, RR. (1995) Capture-recapture methods in epidemiology: Methods and Limitations. *Epidemiologic Reviews*, **17**, 243-264.

Hook, EB and Regal, RR. (1997) Validity of methods for model selection, weighting for model uncertainty and small sample adjustment in capture-recapture estimation. *American Journal of Epidemiology*, **145**, 1138-1144.

International Working Group for Disease Monitoring and Forecasting (1995a) Capture-recapture and multiple-record systems estimation 1: History and theoretical development. *American Journal of Epidemiology*, **142**,1047-1058.

International Working Group for Disease Monitoring and Forecasting (1995b) Capture-recapture and multiple-record systems estimation 1: Applications in human diseases. *American Journal of Epidemiology*, **142**, 1059-1068.

Korf, D, Reijneveld, S and Toet, J. (1994) Estimating the number of heroin users, a review of methods and empirical findings from The Netherlands *International Journal Of The Addictions* **29**, 1393-1417

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-4.

Seber, GAF. (1982) *The estimation of animal abundance and related parameters.* 2nd ed. Charles Griffin & Co., London

Squires NF, Beeching NJ, Schlecht BJM, Ruben SM. (1995) An estimate of the prevalence of drug misuse in Liverpool and a spatial-analysis of known addiction. *Journal of Public Health Medicine*, **17**,103-109.

Wiessing, LG, Toet, J, Houweling, H *et al.* (1995) *Prevalentie en risicofactoren van HIV-infectie onder druggebruikers in Rotterdam.* RIVM, Bilthoven / GGD Rotterdam.

Zelterman, D. (1988). Robust estimation in truncated discrete distributions with applications to capture-recapture experiments. *Journal of Statistical Planning and Inference,* **18**, 225-237.

Appendix 1

Collated Estimates, Confidence Intervals and Information Criteria

In this appendix, we present analyses of the data from the six cities which employed the three-sample capture-recapture methodology. We intially present, in Tables 1 to 6, the data on opiate users and the overlap between the three data sources in a common format for each city, and then, in Tables 7 to 12, the results from the range of different models that can be fitted to the data. Although we briefly describe the data sources within this appendix, the reader is directed to the individual city reports for more comprehensive descriptions. In Tables 7 to 12, we highlight the lowest values of the AIC, DIC and SIC in bold; the estimates suggested by the use of these criteria were not, however, always the ones preferred by the city researchers; these are indicated with an arrow. The reader is also directed to Section 5 to the discussion of these results.

Table 1: **Dublin**

		S1			
		Present		Absent	
		S2			
		Present	Absent	Present	Absent
S3	Present	121	764	71	2831
	Absent	160	2124	193	*

Source 1: Methadone List
Source 2: Hospital Inpatient List
Source 3: Police Arrests

Table 2: **Helsinki**

		S1			
		Present		Absent	
		S2			
		Present	Absent	Present	Absent
S3	Present	1	1	9	71
	Absent	8	15	70	*

Source 1: Those suspected of driving under the influence of drugs.
Source 2: Drug related offenders.
Source 3: Hospital Patient Discharge Register (ICD-9).

Table 3: **Rome**

		S1			
		Present		Absent	
		S2			
		Present	Absent	Present	Absent
S3	Present	27	134	11	166
	Absent	302	5959	297	*

Source 1: Surveillance System on Drug Addiction
Source 2: Hospital Discharges (ICD-9)
Source 3: Mobile Emergency Unit

Table 4: **Setúbal**

		S1			
		Present		Absent	
		S2			
		Present	Absent	Present	Absent
S3	Present	6	2	6	26
	Absent	91	92	116	*

Source 1: Treatment Centre (Semester 1)
Source 2: Treatment Centre (Semester 2)
Source 3: Health Centres

Table 5: **Toulouse**

		S1			
		Present		Absent	
		S2			
		Present	Absent	Present	Absent
S3	Present	6	79	34	389
	Absent	13	162	116	*

Source 1: Low Threshold
Source 2: Repressive
Source 3: Sanitary

Table 6: **Vienna**

		S1			
		Present		Absent	
		S2			
		Present	Absent	Present	Absent
S3	Present	5	13	11	113
	Absent	49	658	179	*

Source 1: Police
Source 2: Hospital Admission and Emergeny Ambulance Data
Source 3: Drug-related Deaths

Table 7: **Dublin**

Model	df	G^2	AIC	SIC	DIC	Estimate	95% CI
Independent	3	225.5	219.5	199.3	204.8	12,338	11,863 - 12,845
S1xS2	2	32.6	28.6	15.2	18.8	13,599	12,990 - 14,257
S1xS3	2	155.4	151.4	137.9	141.6	9,399	8,867 - 10,005
S2xS3	2	219.5	215.5	202.0	205.7	12,555	12,035 - 13,113
S1xS2+S1xS3	1	32.5	30.5	23.8	25.6	13,960	12,108 - 16,414
S1xS2+S2xS3	1	15.7	137.7	7.0	8.8	14,134	13,422 - 14,900
S1xS3+S2xS3	1	145.8	143.8	137.1	138.9	8,826	8,316 - 9,435
Saturated	0	0.0	**0.0**	**0.0**	**0.0**	22,444	7,381 - 29,687
2 sample S1, S2						6,146	5,669 - 6,624
2 sample S2, S3						10,750	9,557 - 11,942
2 sample S1, S3						13,560	12,896 - 14,224

Weighted SIC = 22,195 (7,562 - 29,244)
Weighted DIC = 22,344 (7,454 - 29,509)

In Dublin, the preferred estimate (13,460) was obtained by combining stratified estimates.

Table 8: **Helsinki**

Model	df	G^2	AIC	SIC	DIC	Estimate	95% CI
Independent	3	8.7	2.7	-6.8	-1.3	557	398 -833
S1xS2	2	0.4	**-3.6**	**-9.9**	**-6.2**	775	487 - 1,392 ←
S1xS3	2	7.4	3.4	-2.9	0.8	513	365 - 777
S2xS3	2	6.2	2.2	-4.1	-0.5	400	268 - 685
S1xS2+S1xS3	1	0.2	-1.8	-5.0	-3.2	727	445 - 1382
S1xS2+S2xS3	1	0.0	-2.0	-5.2	-3.3	1,240	368 - 18,125
S1xS3+S2xS3	1	1.8	-0.2	-3.4	-1.6	306	224 - 494
Saturated	0	0.0	0.0	0.0	0.0	1,210	181 - 28,285
2 sample S1, S2						244	123 - 365
2 sample S2, S3						722	327 - 1116
2 sample S1, S3						1,025	-321 - 2371

Weighted SIC = 742 (446 - 2405)
Weighted DIC = 789 (431 - 4186)

Table 9: **Rome**

Model	df	G^2	AIC	SIC	DIC	Estimate	95% CI	
Independent	3	32.2	26.2	5.7	11.2	12,712	11,990 - 13,512	
S1xS2	2	31.3	27.3	13.6	17.3	13,225	11,975 - 14,726	
S1xS3	2	30.5	26.5	12.8	16.5	12,363	11,543 - 13,297	
S2xS3	2	10.3	6.3	**-7.4**	**-3.7**	12,997	12,229 - 13,854	
S1xS2+S1xS3	1	30.1	28.1	21.3	23.1	11,378	9,350 - 15,448	
S1xS2+S2xS3	1	6.3	4.3	-2.5	-0.7	14,278	12,741 - 16,167	←
S1xS3+S2xS3	1	9.6	7.6	0.8	2.6	12,756	11,857 - 13,787	
Saturated	0	0.0	**0.0**	0.0	0.0	24,716	15,263 - 45,521	
2 sample S1, S2						12,434	11,524 - 13,344	
2 sample S2, S3						5,666	4,020 - 7,312	
2 sample S1, S3						13,482	11,994 - 14,970	

Weighted SIC = 13,350 (12,330 - 14,727)
Weighted DIC = 14,475 (12,631 - 17,668)

Table 10: **Setúbal**

Model	df	G^2	AIC	SIC	DIC	Estimate	95% CI	
Independent	3	21.8	15.8	4.3	9.8	494	466 - 553	
S1xS2	2	2.1	**-1.9**	**-9.5**	**-5.8**	894	620 - 1,423	←
S1xS3	2	14.3	10.3	2.6	6.3	474	430 - 529	
S2xS3	2	17.7	13.6	6.0	9.7	477	431 - 533	
S1xS2+S1xS3	1	2.1	0.1	-3.8	-1.9	842	543 - 1,654	
S1xS2+S2xS3	1	0.2	-1.8	-5.7	-3.8	1,535	667 - 7,378	
S1xS3+S2xS3	1	9.0	7.0	3.2	5.0	456	416 - 507	
Saturated	0	0.0	0.0	0.0	0.0	1,864	569 - 1,539	
2 sample S1, S2						431	386 - 476	
2 sample S2, S3						730	394 - 1,066	
2 sample S1, S3						995	376 - 1,534	

Weighted SIC = 973 (621 - 2,144)
Weighted DIC = 1,070 (621 - 2,831)

Table 11: **Toulouse**

Model	df	G^2	AIC	SIC	DIC	Estimate	95% CI	
Independent	3	6.6	0.6	**-13.5**	-8.0	1,808	1,599 - 2,065	
S1xS2	2	4.8	0.8	-8.6	-4.9	1,750	1,542 - 2,009	
S1xS3	2	0.1	**-3.9**	-13.2	**-9.6**	2,178	1,780 - 2,734	←
S2xS3	2	3.9	-0.1	-9.5	-5.8	1,690	1,474 - 1,965	
S1xS2+S1xS3	1	0.0	-2.0	-6.7	-4.8	2,126	1,691 - 2,786	
S1xS2+S2xS3	1	0.7	-1.3	-6.0	-4.2	1,597	1,389 - 1,866	
S1xS3+S2xS3	1	0.1	-1.9	-6.6	-4.8	2,245	1,609 - 3,496	
Saturated	0	0.0	0.0	0.0	0.0	2,055	1,159 - 4,248	
2 sample S1, S2						2,313	1,369 - 3,256	
2 sample S2, S3						2,146	1,589 - 2,704	
2 sample S1, S3						1,554	1,307 - 1,801	

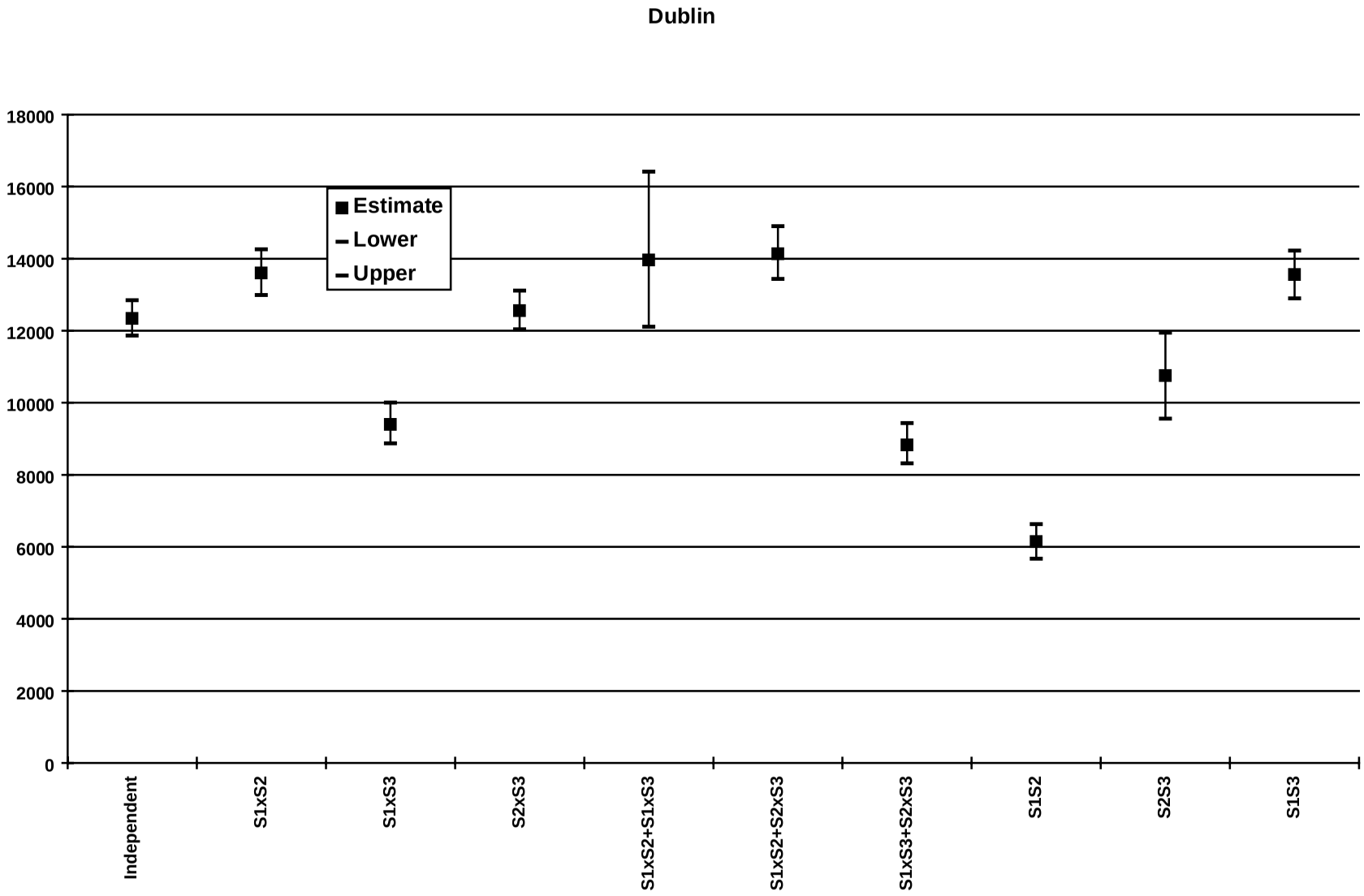
Weighted SIC = 1,956 (1,661 - 2,356)
Weighted DIC = 2,014 (1,675 - 2,499)

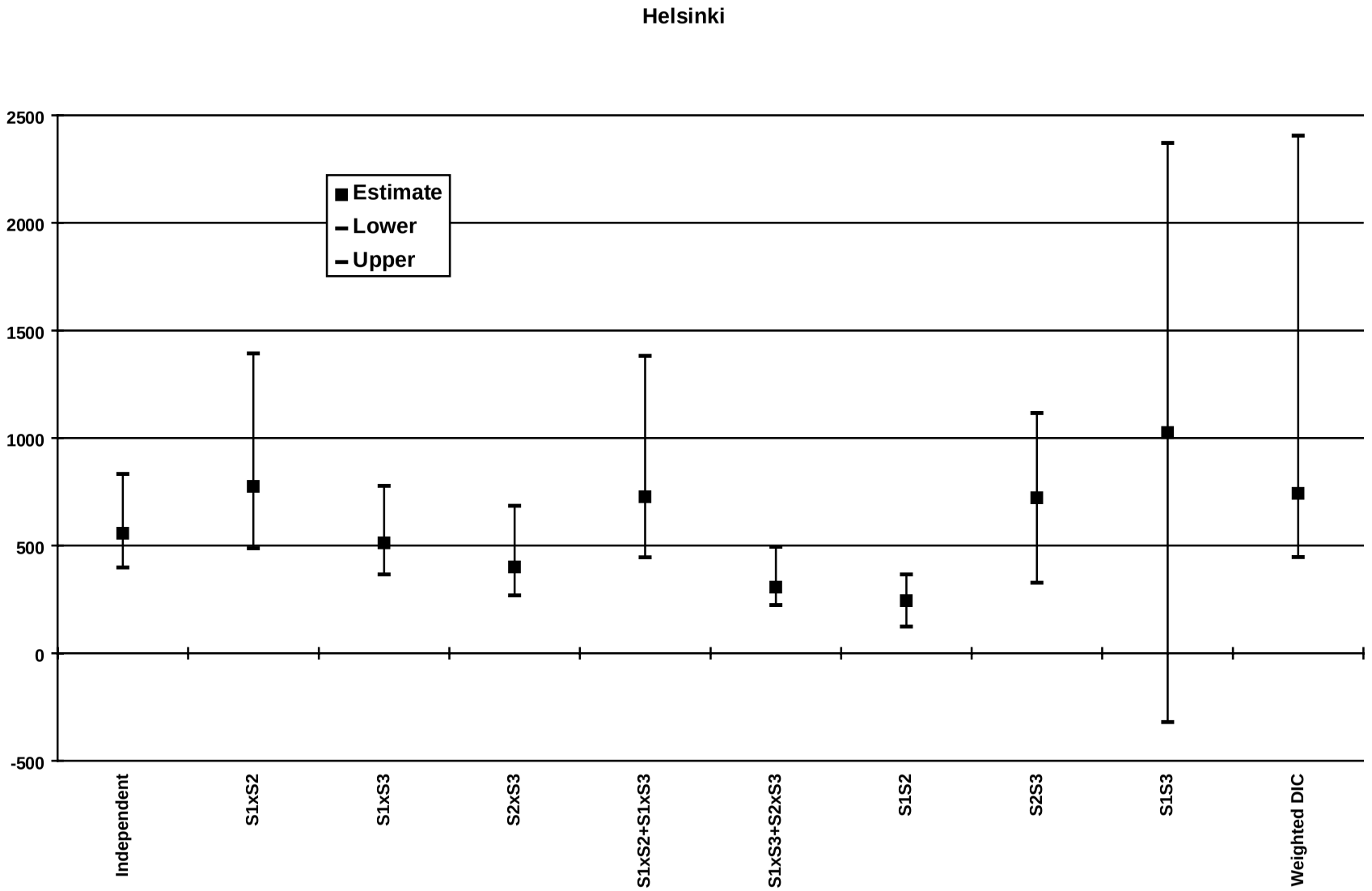
Table 12: **Vienna**

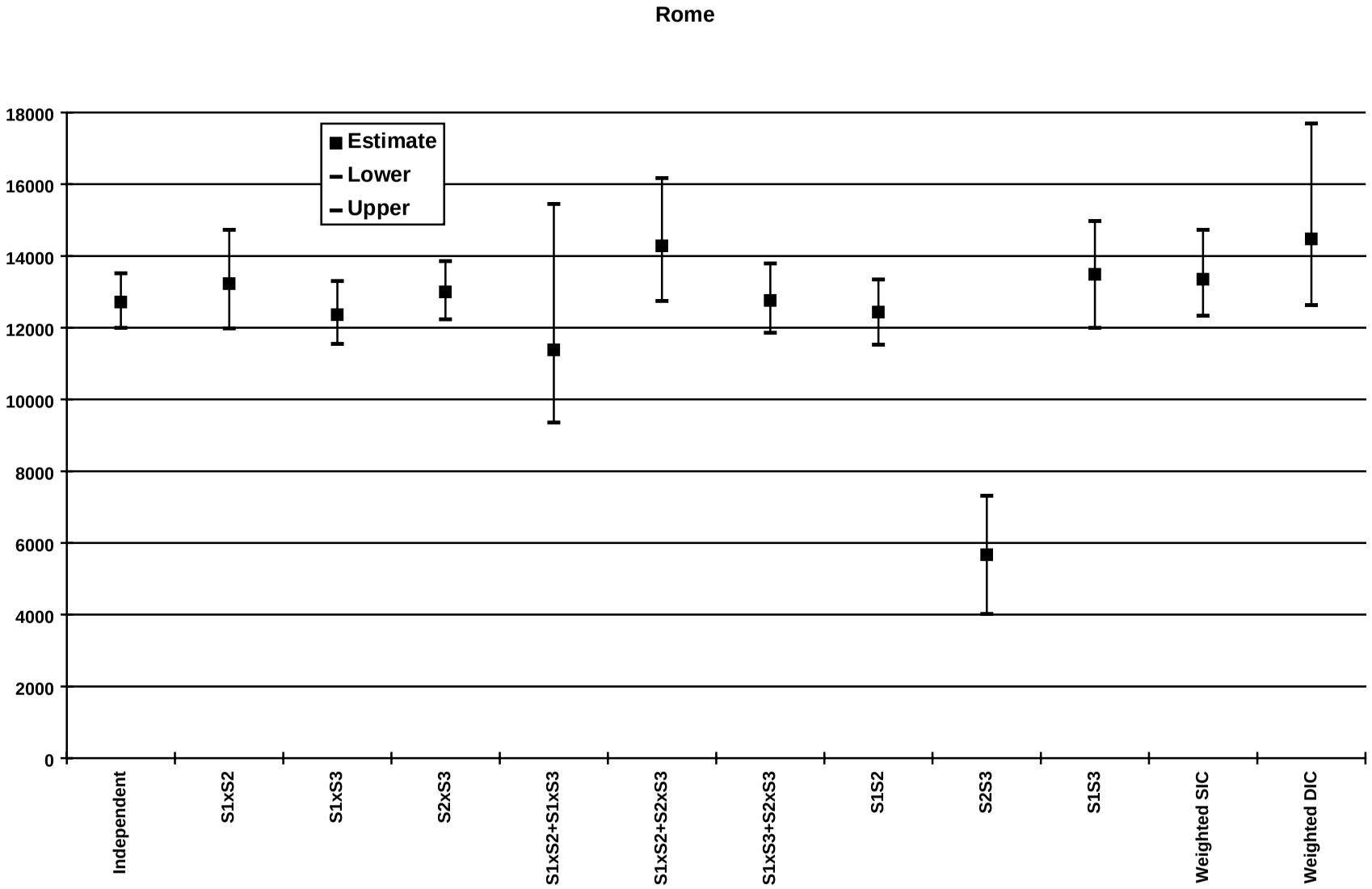
Model	df	G^2	AIC	SIC	DIC	Estimate	95% CI	
Independent	3	14.2	8.2	**-6.6**	-1.1	3,708	3,105 - 4,497	
S1xS2	2	11.6	7.6	-2.2	1.4	4,480	3,328 - 6,318	
S1xS3	2	7.3	3.3	-6.6	-2.9	3,187	2,636 - 3,935	
S2xS3	2	9.2	5.2	-4.7	-1.0	4,004	3,289 - 4,972	
S1xS2+S1xS3	1	7.0	5.0	0.1	1.9	2,867	2,023 - 4,635	
S1xS2+S2xS3	1	0.8	**-1.2**	-6.2	**-4.3**	6,748	4,332 - 11,668	←
S1xS3+S2xS3	1	4.4	2.4	-2.5	-0.7	3,432	2,765 - 4,379	
Saturated	0	0.0	0.0	0.0	0.0	10,525	3,479 - 32,933	
2 sample S1, S2						3,285	2,541 - 4,049	
2 sample S2, S3						2,196	3,177 - 1,215	
2 sample S1, S3						5,234	3,135 - 7,334	

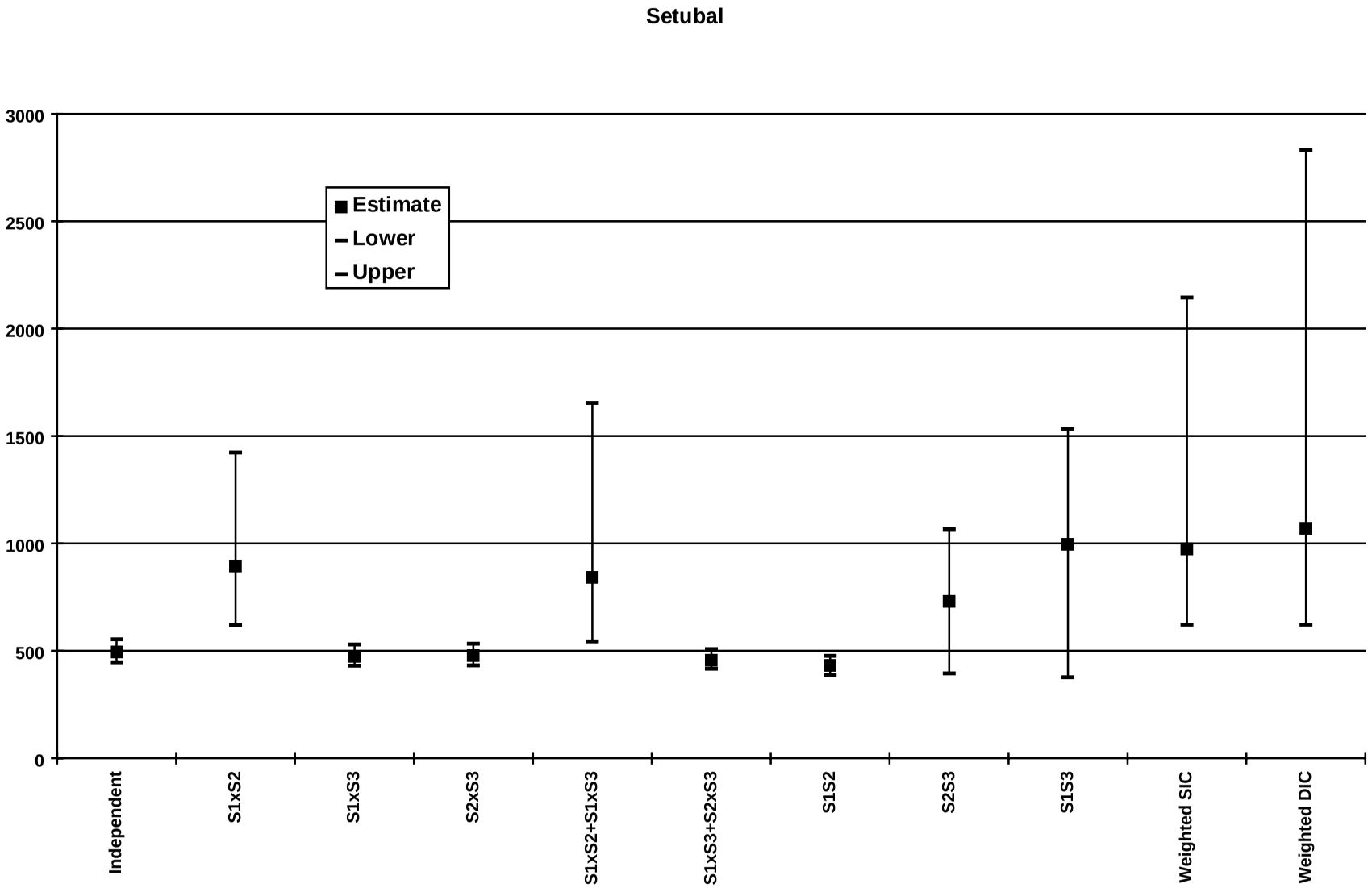
Weighted SIC = 4,372 (3,263 - 6,400)
Weighted DIC = 5,300 (3,542 - 9,087)

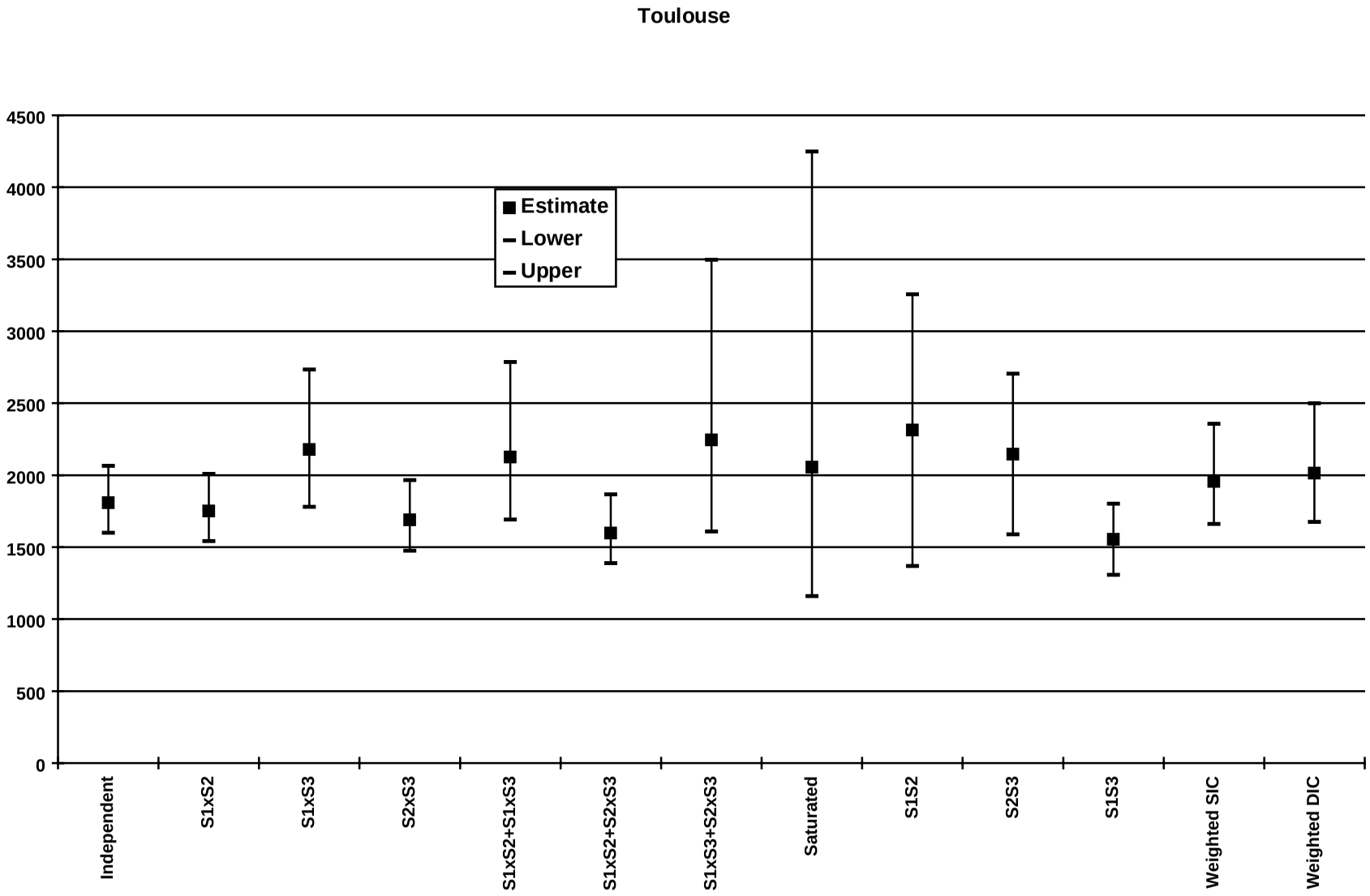
We also present the selection of the various estimates and confidence intervals in Figures 1 to 6. We omit those estimates that have large corresponding confidence intervals for ease of presentation. These can be found in the above tables.

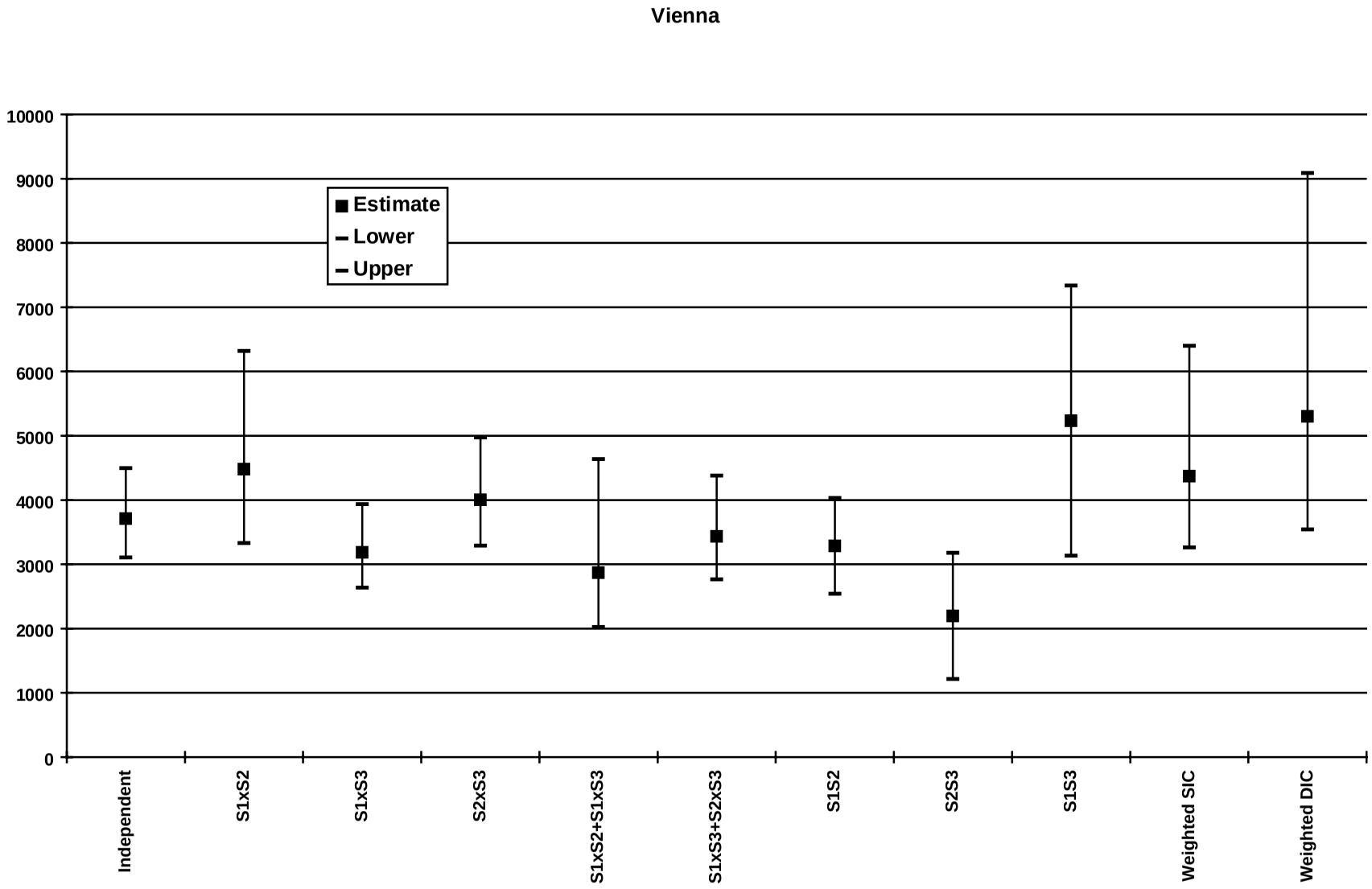

Dublin
Estimate
Lower
Upper
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Independent
S1xS2
S1xS3
S2xS3
S1xS2+S1xS3
S1xS2+S2xS3
S1xS3+S2xS3
S1S2
S2S3
S1S3


Helsinki
Estimate
Lower
Upper
2500
2000
1500
1000
500
0
-500
Independent
S1xS2
S1xS3
S2xS3
S1xS2+S1xS3
S1xS3+S2xS3
S1S2
S2S3
S1S3
Weighted DIC


Rome
Estimate
Lower
Upper
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Independent
S1xS2
S1xS3
S2xS3
S1xS2+S1xS3
S1xS2+S2xS3
S1xS3+S2xS3
S1S2
S2S3
S1S3
Weighted SIC
Weighted DIC


Setubal
3000
2500
2000
1500
1000
500
0
Estimate
Lower
Upper
Independent
S1xS2
S1xS3
S2xS3
S1xS2+S1xS3
S1xS3+S2xS3
S1S2
S2S3
S1S3
Weighted SIC
Weighted DIC


Toulouse
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Estimate
Lower
Upper
Independent
S1xS2
S1xS3
S2xS3
S1xS2+S1xS3
S1xS2+S2xS3
S1xS3+S2xS3
Saturated
S1S2
S2S3
S1S3
Weighted SIC
Weighted DIC


Vienna
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Estimate
Lower
Upper
Independent
S1xS2
S1xS3
S2xS3
S1xS2+S1xS3
S1xS3+S2xS3
S1S2
S2S3
S1S3
Weighted SIC
Weighted DIC

Appendix 2

Full details and addresses of the Expert Group which participated in the project.

Pierre Yves Bello
Observatoire Régional de la Santé
Midi-Pyrenées
37 allés Jules Guesde 31073
Toulouse Cedex

Tel: +33 561 53 11 46
Fax: +33 562 26 42 40
Email: bello@cict.fr

Catherine Comiskey
Regional Technical College
Tallaght
Dublin 24

Tel: ++ 353 1 4042875
Fax: ++ 353 1 4042700
Email: ccom@staffmail.rtc-tallaght.ie

Daniela D'Ippoliti
Osservatorio Epidemiologico
Via di S. Costanza, 53
00198 Roma

Tel: ++39 6 5168 6469
Fax: ++39 6 5168 6463

Antònia Domingo-Salvany
IMIM
Dr. Aiguader, 80
E-08003
Barcelona

Tel: ++ 34 3 221 1009
Fax: ++ 34 3 221 3237
Email: adomingo@imim.es

Sofia Freire
Observatório VIDA
Av. Columbano Bordalo Pinheiro
87-2
1050 Lisboa

Tel: ++351 1 721 0270
Fax: ++351 1 727 3803
Email: sofia@obvida.pt

Maria Moreira
Observatório VIDA
Av. Columbano Bordalo Pinheiro
87-2
1050 Lisboa

Tel: ++351 1 721 0270
Fax: ++351 1 727 3803
Email: maria@obvida.pt

Alvar Norén
Valfards-och Folkhalsoenheten
Room 7008 Stadshuset
205 08 Malmö

Tel: ++46 40 34 5972
Fax: ++46 40 34 3390
Email: alvar.noren@mailbox.malmo.se

Päivi Partanen
University of Jyväskylä
Dept of Statistics
Box 35
40351 Jyväskylä

Tel: ++358 14 602 980
Fax: ++358 14 602 981
Email: ppartane@cc.jyu.fi

Kathy Politikou
University Mental Health
Research Group, Eginition
Hospital

72-74 Vas. Sophias Ave.
115 28 Athens

Tel: ++30 1 722 5109
Fax: ++30 1 723 3690

Dan Seidler
Emergency Dept. General Hospital
AKH Waehringer Quertel 18-20
A-1090 Wien

Tel: ++43 1 404 001964
Fax: ++43 1 404 001965
Email: dan.seidler@akh-wien.ac.at

Filip Smit
Methods Section
Trimbos Institute
PO Box 725
3500 AS Utrecht

Tel: ++31 30 297 1100
Fax: ++31 30 297 1111
Email: fsmit@trimbos.nl

Jaap Toet
GGD Rotterdam
Dept of Health Promotion
Postbus 70032
3000 LP Rotterdam

Tel: ++31 10 433 9215
Fax: ++31 10 433 9493
Email: toetj@ggd.rotterdam.nl

Alfred Uhl
LBIS
Mackgasse 7-11
A-1237
Wien

Tel: ++43 1 888 2533158
Fax: ++43 1 888 2533138
Email: lbi@api.or.at

Ari Virtanen
STAKES
PO Box 220
SF-00531 Helsinki

Tel: ++358 9 3967 2378
Fax: ++358 9 3967 2324
Email: ari.virtanen@reitox.net

Lucas Wiessing
EMCDDA
Rua Cruz da Santa Apolonia 23-25
1100 Lisboa

Tel: ++ 351 1 811 3016
Fax: ++ 351 1 813 7943
Email: Lucas.Wiessing@emcdda.org

FINAL REPORT

EMCDDA Project: Methodological Pilot Study of Local Level Prevalence Estimates

CT.96.EP.07

Part II City Reports

Contents

		Page No
1	Introduction	1
2	Dublin	2
3	Helsinki	16
4	Rome	36
5	Rotterdam	48
6	Setúbal	69
7	Toulouse	77
8	Vienna	90

Introduction

The collection of reports contained in this document form the basis of the report of the EMCDDA Methodological Pilot Study of Local Level Prevalence Estimates. In Part I the results from the different cities have been drawn together and the discussion of the issues that arise from these city reports leads on to the conclusions of the study.

A range of expertise has been employed in the production of these reports, not just from the individuals listed as contributors to the Pilot Study but from others in the respective countries. Some of the report authors are experienced in applying the capture-recapture method and scientifically describing the results. For others the method was new, although it is clear that all contributors have benefited from the expert group meetings and the discussions that arose from them.

The experience of those authors who had experience of the capture-method was complemented by that of Antònia Domingo-Salvany from Barcelona, and the input from the EMCDDA by Lucas Wiessing was invaluable in creating a framework in which these reports could be compared. Finally the contributions from the cities of Malmö and Athens, where it was not feasible to apply the method this year, are acknowledged.

The city reports have been adapted to comply with a common style as the contributions were received in various formats. Any errors in a city reports may therefore be the responsibility of the co-ordinating research team.

2 Dublin

ESTIMATING THE PREVALENCE OF OPIATE DRUG USE IN DUBLIN, IRELAND DURING 1996.

Catherine Comiskey, Regional Technical College, Tallaght, Dublin

...,the evidence of increasing numbers in treatment leads to the question as to whether the increases are artefacts of better reporting and a greater provision of services or is the number of drug misusers in the community actually increasing? Without some estimation of overall prevalence, the answer to that question must remain in the realms of speculation."

O'Higgins (1995).

2.1 Introduction

In this study we aim to address this issue raised in the concluding paragraph of the Health Research Board Report, Treated Drug Misuse in the Greater Dublin Area: A Review of five years 1990-1994. In order to do this we examine the nature and history of drug misuse in Dublin.

2.2 Description of the city.

Dublin is the capital city of The Republic of Ireland. It is situated at the mouth of the river Liffey on the east coast. The 1996 census of the population identified 3,626,087 inhabitants. This represented a 2.8 per cent increase from the previous census in 1991. The population of Dublin is 1,058,714. This represents over 29 per cent of the total population. In addition, of those living in Dublin 430,385 (or 41 per cent) are under the age of 25 years. A detailed age profile of the Dublin population is provided in table 2.1 below.

Table 2.1: **Age profile of the Dublin population**.

Age (years)	**0 - 14**	**15 - 24**	**25 - 34**	**35 - 44**	**45 - 54**	**> 54**
Males	120,444	97,222	84,350	69,540	56,682	81,178
Females	112,700	100,009	91,891	75,188	60,447	109,063
Total	233,144	197,231	176,241	144,728	117,129	190,241

2.3 Description of drug misuse in the city.

Johnson et al (1994) in a study of the risk behaviour in attendees at a Dublin needle exchange program speaks of the high level of unsafe injecting and sexual activity. The authors point out the need for more effective health promotion among drug users in Dublin. Comiskey (1991) and Comiskey et al (1992) in a 2 year survey of drug users estimated that a total of 375 people enter the drug using population each year with 198 of these being in the Dublin region. O Higgins (1996) in a five-year review found that the numbers seeking treatment for the first time had almost doubled from 624 in 1990 to 1150 in 1994. In addition the mean age over the five year period was seen to decrease from 25.2 years in 1990 to 23.8 years in 1994. The review also finds that the most commonly used primary drug is heroin with over 82% of those attending treatment citing it as their primary drug of misuse. This is also reflected in the police statistics. Keogh (1996) finds that opiates represent 93% of those arrests where drugs were noted. While these studies provide a significant and valuable contribution to our understanding of the drug misuse profile in Dublin, they are primarily indicators of the number of drug

misusers in the population and there is to date no comprehensive study on the true prevalence of drug misuse in Dublin.

2.4 Description of the sources used in the analysis.

We applied the capture-recapture methodology to estimate the prevalence of opiate drug users in Dublin in 1996. We used three sources of information on drug users living in Dublin. Firstly, the Central Patient Methadone Treatment List, secondly, the Hospital Inpatient Enquiry database (known as H.I.P.E.) and thirdly the police (known in Ireland as The Garda Siochana) database of arrests in 1996.

The Central Patient Methadone Treatment List

This list is maintained by The Department of Health. It records those who have ever received methadone from either a Department of Health clinic or a general practitioner. The variables recorded in the database are surname, Christian name, date of birth, age, methadone card number, date of issue of the card, date of expiry of the card, current status of the card (expired, void or current) and the clinic or doctor that the client attends. From the list it is possible to determine how many times, when and for how long the client was on methadone. This list was received in paper format and was not received on computer disk. From the list those on methadone in 1996 were isolated and entered onto the computer. We found that 3170 individuals were in receipt of methadone in Dublin in 1996. No sex is recorded on the list by the Department of Health so this had to be entered manually following a visual analysis of the clients names. The mean age of those receiving methadone in 1996 was found to be 27.24 years, with a standard deviation of 6.87 years and a range of 15 to 60 years. There were 2225 (70.2%) males, 920 (29.0%) females and for 25 (0.8%) clients the sex was unknown.

The Hospital Inpatient Enquiry Database (H.I.P.E.)

The Economic and Social Research Institute, Dublin, maintains the central HIPE database. This database records all discharges from Irish hospitals and the primary and secondary diagnoses of those patients discharged. Those patients who used opiates were identified from the ICD IX classification code 304.0, opioid type dependence, code 304.7, combinations of opioid type drug with any other and code 305.5, opioid abuse. As of March 1997 the central HIPE database identified 603 patients in the Dublin area with these codes as a primary or secondary diagnosis. These 603 patients were identified in 12 different Dublin hospitals. However, 92% of these cases were in

4 of the Dublin hospitals. As the central HIPE database does not have access to patients initials it was decided to contact these 4 hospitals individually and seek their permission for access to these patients initials. Permission was granted and information on 545 opiate drug users was obtained. The variables recorded were surname initial, sex, date of birth and Dublin postal code. Of the 545 patients identified, 353 (64.77%) were male and 192 (35.23%) were female.

Police Arrests

In 1996 the Assistant Police Commissioner, Mr. T. King commissioned a study on illicit drug use and related criminal activity in the Dublin Metropolitan Area (DMA). After an extensive search of all Garda records held at police station level, a database consisting of 4105 individuals identified with drug use was constructed. The majority of these were male, unemployed and living at home. Males accounted for 3467 (84.46%) of cases and females accounted for 638 (15.54%) of cases. 80% were in the 15 to 30 year age group, with the youngest user being 12 years and the eldest being 60 years. The principal drug used was opiates (heroin and methadone) with 3817 (93%) users identified. Not all of those arrested were arrested under The Misuse of Drugs Act. Most were arrested for other crimes, for example shop lifting and while in custody they asked to see a medical doctor so that methadone may be prescribed. In some cases those arrested were found with needle marks on their body. A list of seven reasons why drug users were known to the police is provided in table 2.2 below.

Table 2.2: **Reasons why drug users were known to the police.**

Drug User Identified By:	**Number of Users**	**%**
Possession	407	10
Admits	2098	51
Treatment	151	4
Paraphernalia	501	12
Physical Signs	285	7
Custody Methadone	473	12
Other	190	5
Total	4105	100

2.5 Other information about the analysis.

To remove duplicates within the three data sources, three different

procedures were used. For the Central Patient Methadone Treatment List individuals appearing on the list more than once were identified from their Christian name, surname and date of birth. Data supplied from the HIPE database was provided with duplicates removed. Data from the police arrests also had the duplicates removed but included some drug users who were living outside of the Dublin area. These were easily removed after a visual inspection of the data.

The variables that were common to all three of the data sources were surname initial, date of birth and sex. In order to identify overlaps between sources these variables were used. For example if C 16/11/1962 F was observed in say the methadone treatment list and the HIPE database then it was assumed that this was the same person and an overlap was noted.

Finally of the 4105 cases noted in the police data, 3817 (93%) were opiate drug users and 288 (7%) were non opiates. Due to the nature of the data supplied we were unable to remove these non opiate users from the data set.

2.6 Results

A summary of the number of contacts within each age group and within each data source is provided in table 2.3 below.

Table 2.3: **1996 Data Sources By Sex and Age.**

	Males	Males	Males	Females	Females	Females
Age in years	15-24	25-34	35-54	15-24	25-34	35-54
S1: Methadone List	925	986	338	450	375	95
S2: HIPE Database	133	165	55	99	73	20
S3: Police Arrests	1820	1164	218	371	179	35
Total Contacts	2878	2315	611	920	627	150
Individuals	2469	1874	534	755	506	126

A table detailing the overlaps between each source for all age groups is provided below.

Table 2.4: **Data from the 3 samples illustrating the overlaps between data sources. (This includes those outside the 15-54 year age range or where age was unknown).**

		Source 1

		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	121	764	71	3015
	Absent	160	2125	193	*

From table 2.4 above and table 2.5 below we can see that a total of 6449 individual opiate users were identified from the three data sources and that 6264 of these were within the 15-54 year age range. It is also interesting to note that of the 185 outside the age range 184 were identified within data source 3, the police arrests. In many of these cases the age of the individual was unknown. The figures in tables 4 and 5 do however provide us with a minimum estimate of the prevalence of opiate drug use in Dublin in 1996.

Table 2.5: **Data from the 3 samples illustrating the overlaps between data sources (Note this includes only those users within the 15-54 year age range).**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	121	764	71	2831
	Absent	160	2124	193	*

Using the methods of Bishop, Fienberg and Holland (1995) and the statistical modelling package GLIM together with macros devised by Cormack and Comiskey and Hay, loglinear models for capture recapture studies were fitted The following results were obtained,

Table 2.5a: **Prevalence estimates for the hidden population of opiate drug users in Dublin in 1996 within the 15-54 year age range.**

Model	G^2	d.f.	p value	N	95% CI
Independence	225.54	3	0.00	6074	5599-6581
+S1xS2	32.64	2	0.00	7335	6726-7993
+S1xS3	155.37	2	0.00	3135	2603-3741
+S2xS3	219.52	2	0.00	6291	5771-6849
+S1xS2+S1xS3	32.52	1	0.00	7696	5844-10150
+S1xS2+S2xS3	15.73	1	0.00	7870	7168-8636
+S1xS3+S2xS3	145.81	1	0.00	2562	2052-3171
Saturated	0.00	0	1.00	16180	11117-23423

We can see from table 2.5a. above that the data does not fit the models well and hence does not provide us with an estimate of prevalence. To overcome this problem we stratify the data by gender and by age within gender. Tables 2.6, 2.6a, 2.7 and 2.7a below provide an analysis of the data by gender.

Table 2.6: **Data from the 3 samples illustrating the overlaps between data sources. Males only (Note this includes only those users within the 15-54 year age group).**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	79	621	54	2448
	Absent	94	1455	126	*

Table 2.6a: **Prevalence estimates for the hidden population of male opiate drug users in Dublin in 1996 within the 15-54 year age range.**

Model	G^2	d.f.	p value	N	95% CI
Independence	139.34	3	0.00	4680	4259-5135
+S1xS2	17.52	2	0.00	5438	4926-5996
+S1xS3	100.17	2	0.00	2511	1993-3118
+S2xS3	136.21	2	0.00	4816	4360-5311
+S1xS2+S1xS3	17.42	1	0.00	5712	4135-7892
+S1xS2+S2xS3	9.26	1	0.00	5736	5160-6369
+S1xS3+S2xS3	92.07	1	0.00	1950	1470-2555
Saturated	0.00	0	1.00	11248	7127-17560

Table 2.7: **Data from the 3 samples illustrating the overlaps between data sources for females only (Note this includes only those users within the 15-54 year age range).**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	42	143	17	383
	Absent	66	669	67	*

Table 2.7a: **Prevalence estimates for the hidden population of female opiate drug users in Dublin in 1996 within the 15-54 year age range.**

Model	**G^2**	**d.f.**	**p value**	**N**	**95% CI**
Independence	65.82	3	0.00	1150	967-1359
+S1xS2	23.29	2	0.00	1521	1253-1836
+S1xS3	44.16	2	0.00	640	461-863
+S2xS3	58.61	2	0.00	1245	1036-1486
+S1xS2+S1xS3	23.29	1	0.00	1509	879-2630
+S1xS2+S2xS3	7.94	1	0.00	1792	1447-2212
+S1xS3+S2xS3	43.70	1	0.00	679	468-958
Saturated	0.00	0	1.00	4494	2240-9010

Again it is evident that the models do not fit the data in spite of gender stratification.

Tables 2.8 to 2.13 provide a breakdown of the data by age within gender together with the results from the models fitter.

Table 2.8: **Data from the 3 samples illustrating the overlaps between data sources. Males aged 15-24 Years.**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	29	288	28	1475
	Absent	35	573	41	*

Table 2.8a: **Prevalence estimates for the hidden population of male opiate drug users in Dublin in 1996 within the 15-24 year age range.**

Model	**G^2**	**d.f.**	**p value**	**N**	**95% CI**
Independence	68.06	3	0.00	2428	2109 - 2786
+S1xS2	4.69	2	0.10	2775	2397 - 3204
+S1xS3	42.43	2	0.00	1041	696-1485
+S2xS3	65.98	2	0.00	2513	2166-2907
+S1xS2+S1xS3	3.58	1	0.06	2160	1311-3485
+S1xS2+S2xS3	0.30	1	0.58	2935	2511-3422
+S1xS3+S2xS3	34.54	1	0.00	671	411-1048
Saturated	0.00	0	1.00	3561	1701-7072

Table 2.9: **Data from the 3 samples illustrating the overlaps between data sources. Males aged 25-34 Years.**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	43	287	22	812
	Absent	46	610	54	*

Table 2.9a: **Prevalence estimates for the hidden population of male opiate drug users in Dublin in 1996 within the 25- 34 year age range.**

Model	**G^2**	**d.f.**	**p value**	**N**	**95% CI**
Independence	56.64	3	0.00	1402	1208-1618
+S1xS2	9.89	2	0.01	1638	1402-1904
+S1xS3	44.09	2	0.00	831	582-1144
+S2xS3	55.41	2	0.00	1441	1232-1675
+S1xS2+S1xS3	9.22	1	0.00	1993	1201-3305
+S1xS2+S2xS3	6.52	1	0.01	1726	1463-2028
+S1xS3+S2xS3	42.62	1	0.00	716	467-1062
Saturated	0.00	0	1.00	3960	2002-7670

Table 2.10: **Data from the 3 samples illustrating the overlaps between data sources. Males aged 35-54 Years.**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	7	46	4	161
	Absent	13	272	31	*

Table 2.10a: **Prevalence estimates for the hidden population of male opiate drug users in Dublin in 1996 within the 35- 54 year age range.**

Model	**G^2**	**d.f.**	**p value**	**N**	**95% CI**
Independence	9.13	3	0.03	780	579-1045
+S1xS2	5.31	2	0.07	893	641-1239
+S1xS3	8.17	2	0.02	619	350-1058
+S2xS3	8.63	2	0.01	808	589-1101
+S1xS2+S1xS3	4.84	1	0.03	1248	464-4050
+S1xS2+S2xS3	4.27	1	0.04	952	667-1359
+S1xS3+S2xS3	8.11	1	0.00	649	330-1267
Saturated	0.00	0	1.00	3973	923-17755

Table 2.11: **Data from the 3 samples illustrating the overlaps between data sources. Females aged 15-24 Years**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	24	80	12	255
	Absent	25	321	38	*

Table 2.11a: **Prevalence estimates for the hidden population of female opiate drug users in Dublin in 1996 within the 15-24 year age range.**

Model	**G^2**	**d.f.**	**p value**	**N**	**95% CI**
Independence	35.92	3	0.00	680	538-849
+S1xS2	17.90	2	0.00	844	654-1081
+S1xS3	27.49	2	0.00	409	260-611
+S2xS3	29.42	2	0.00	759	592-965
+S1xS2+S1xS3	17.88	1	0.00	808	414-1589
+S1xS2+S2xS3	6.77	1	0.01	1023	770-1353
+S1xS3+S2xS3	26.07	1	0.00	488	282-814
Saturated	0.00	0	1.00	3110	1245-7710

Table 2.12: **Data from the 3 samples illustrating the overlaps between data sources. Females aged 25-34 Years.**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	15	54	5	105
	Absent	32	274	21	*

Table 2.12a: **Prevalence estimates for the hidden population of female opiate drug users in Dublin in 1996 within the 25-34 year age range**.

Model	**G^2**	**d.f.**	**p value**	**N**	**95% CI**
Independence	22.12	3	0.00	335	247-443
+S1xS2	5.79	2	0.06	464	330-640
+S1xS3	12.47	2	0.00	175	95-286
+S2xS3	20.29	2	0.00	355	259-477
+S1xS2+S1xS3	5.78	1	0.02	441	162-1250
+S1xS2+S2xS3	1.40	1	0.24	533	369-759
+S1xS3+S2xS3	12.41	1	0.00	180	94-308
Saturated	0.00	0	1.00	1049	300-3605

Table 2.13: **Data from the 3 samples illustrating the overlaps between data sources. Females aged 35-54 Years.**

		Source 1			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 3	Present	3	9	0	23
	Absent	9	74	8	*

Table 2.13a: **Prevalence estimates for the hidden population of female opiate drug users in Dublin in 1996 within the 35-54 year age range.**

Model	**G^2**	**d.f.**	**p value**	**N**	**95% CI**
Independence	8.62	3	0.03	109	57-194
+S1xS2	3.67	2	0.12	174	80-365
+S1xS3	7.12	2	0.03	71	22-167
+S2xS3	8.62	2	0.01	109	55-199
+S1xS2+S1xS3	****	1	****	****	****
+S1xS2+S2xS3	3.41	1	0.06	189	81-432
+S1xS3+S2xS3	6.98	1	0.01	66	19-167
Saturated	****	0	****	****	****

It is evident from tables 2.8 to 2.13a above that the models in 2 of the 6 stratifications did not fit and estimates can only be provided for 4 of the groupings. In table 2.14 below we provide a summary of the estimates obtained. We include the known numbers of cases, the estimated hidden population, the overall estimates and 95% confidence intervals for these estimates.

Table 2.14: **Estimated prevalence of opiate drug use in specific population in Dublin in 1996.**

	Known Number	**Estimated Hidden Number**	**Ratio of Known to Hidden**	**Estimated Total**	**95% CI for the total estimate**	**Prevalence per 1000 of population**
Males 15-24 yrs	2469	2935	1:1.19	5404	4980-5891	56/1000
Males 35-54 yrs	534	893	1:1.67	1427	1175-1773	11/1000
Females 25-34 yrs	506	533	1:1.05	1039	875-1265	11/1000
Females 35-54 yrs	126	174	1:1.38	300	206-491	2/1000

2.7 Discussion

In spite of the poor performance of the models fitted to the total population in table 5a and to the gender specific populations in tables 6a and 7a we can derive some useful estimates from the summary of results in table 14 above. If we average over the population prevalence estimates we arrive at an estimate of 21/1000 or 13,460 opiate drug users in Dublin between the ages of 15 and 54 inclusive. This estimate agrees with the crude estimates derived using 2 sample capture recapture methods (5,795 estimated from the two medical data sources, methadone and HIPE, 13,560 estimated from police and methadone data sources and 10,750 estimated from police and HIPE data sources). On a European scale this estimate is in accordance with estimates published by other European cities. In Barcelona, Domingo-Salvany et al (1995) estimate an opiate drug user population prevalence of 9.2/1000, Frischer (1992) estimates the prevalence of injecting drug use in Glasgow to be 13.5/1000 with this rising to 43.45/1000 in males aged between 20 to 24 years. High prevalence amongst young males is noted in our study also, as can be seen from table 14 where a prevalence rate of 56/1000 is estimated for males aged 15-24 years. In a similar study in Setúbal, Portugal prevalence of opiate use among males aged 15-24 years was estimated to be 53/1000 (private communication). Finally in Dundee, Scotland, Hay and McKeganey (1996) estimated the prevalence of opiate use to be 30/1000.

References

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1995). Discrete Multivariate Analysis, Theory and Practice. MIT Press, Cambridge, Massachusetts. 12th. Edition.

Comiskey, C.M., Ruskin, H.J. and Wood, A.D. (1992). Mathematical models for the transmission dynamics of HIV in Ireland. Report for The AIDS Fund. Dublin City University, Ireland.

Comiskey, C.M. (1991). Mathematical models for the transmission dynamics of the Human Immunodeficiency Virus (HIV) in Ireland. Proceedings of the sixth European conference on mathematics in industry. 125-128. BG Teubner, Stuttgart, Germany.

Johnson, Z., O Connor, M., Pomeroy, L., Johnson, H., Barry, J., Scully, M., Fitzpatrick, E. (1994).Prevalence of HIV and associated risk behaviour in attendees at a Dublin needle exchange. Addiction, May; 89(5): 603-7.

Keogh, E. (1997). Illicit Drug Use and Related Criminal Activity In The Dublin Metropolitan Area. Research Report No. 10/97, Garda Research Unit.

O Higgins, K. (1995). Treated drug misuse in the greater Dublin area. A review of five years 1990-1994. The Health Research Board, Baggot St. Dublin, Ireland.

3 Helsinki

ESTIMATING THE NUMBER OF AMPHETAMINE AND OPIATE USERS IN THE GREATER HELSINKI AREA

Olavi Kaukonen; STAKES

Aarne Kinnunen; National Council for Crime Prevention

Päivi Partanen; University of Jyväskylä, Department of Statistics

Timo Seppälä; National Public Health Institute

Jussi Simpura; STAKES

Ari Virtanen; STAKES

Acknowledgements

The authors are grateful to the persons who have aided this study in its different stages:

Data Protection Ombudsman

Jorma Kuopus, Risto Järveläinen; The Province of South Finland Police Department

Jaakko Salonen; TT-Valtionpalvelut Oy

Jari Ollila; Tietogroup

Mikael Albrecht; QA-Lan Vision,

Juha Alho; University of Joensuu, Department of Statistics

Antti Penttinen; University of Jyväskylä, Department of Statistics.

3.1 Introduction

Alcohol has been the traditional problem substance in Finland. Drug use in Finland has been stable and of much lower prevalence as compared to in other European countries in the 1980s. For this reason only a few separate drug researches have been carried out. From the beginning of the 1990s the situation has become more serious, and the epidemiological research has focused its interest on studying the prevalence of (experimental) drug use, while drug addiction has still not been studied to a larger extent. It is not an easy task to get a reliable estimate of the prevalence of drug use, and there is a great deal of difference between the separate study results. According to two national surveys (Partanen, 1994; Kontula, 1995), in the beginning of the 1990s about 0.1-0.5 %, or 4,000 - 20,000, of persons in the adult population were more than casual users (source: Poikolainen, 1997). These estimates concern cannabis products, amphetamine, cocaine and opiates. However, a traditional sampling survey suffers from several sources of error, like underreporting and false answers, so that the results include a considerable uncertainty, especially concerning the use of heavy drugs.

The prevalence of drug use can also be studied indirectly by utilising different sources of information such as records from hospitals, police, courts of justice or cause of death registers. On the basis of cause of death register information there are an estimated 7,000-14,000 drug users in Finland (Vuori et al., 1997), and based on the number of hepatitis-C infections there are approximately 2,000 - 10,000 persons who have injected drugs at some time (Poikolainen, 1997).

It is not, however, possible to make generalisations referring to the entire drug field because of the biased information given by a single register and because of the limited use of heavy drugs (amphetamine derivatives, opiates). The problem caused by the biased information derived from one data source can be avoided by applying the capture-recapture method to provide an estimated population of addicts. In this approach the lists of drug users collected from different registers are thought of as samples drawn only from the population of interest. The method has recently been applied in several European cities to estimate the prevalence of drug addiction (mainly the use of heroin) and the samples or combined registers vary according to local circumstances (e.g. Bloor et al., 1991; Brecht and Wickens, 1993; Domingo-Salvany et al., 1995; Frischer et al., 1991; Hay and McKeganey, 1996). The development of the statistical theory and applications of the capture-recapture method has a long tradition on the one hand in the fields of biology and ecology, and on the other hand in demography and epidemiology.

In biology the method is used to estimate the size and dynamics of animal populations (e.g. Seber, 1992). With human populations this technique is applied in situations where traditional probability sampling tends to miss the target population, such as, for example, when the interesting cases are extremely rare or when people in the target population tend to not have a permanent residence (e.g. Sekar and Deming, 1949; Hook and Regal, 1992; McCarty et al., 1993).

The aim of this study was to carry out for the first time in Finland an estimation of the prevalence of drug addiction by applying the capture-recapture method and data collected from three official registers.

The use of heavy drugs as a measure of the extent of addiction was defined in this study as the use of amphetamine (and its derivatives) and opiates (morphine, heroin etc.). In many cases this means also intravenous drug use. Considerations here were limited to include only Helsinki and the surrounding environs. A rough estimate is given also for the whole country based on the information in one national register.

With this research Finland is participating in a larger ongoing project initiated by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), an EU agency based in Lisbon. The intention of this project is to improve the methodology in a practical sense (to improve data collection and validation using other methods) and to generate more comparable local estimates. Six countries from those member states that belong to the EMCDDA were considered for involvement in the project, the aim of which was to undertake drug research this year in some of the cities in the participating countries. These countries were invited in April to a planning meeting in Glasgow, as the Centre for Drug Misuse Research at the University of Glasgow was chosen to co-ordinate this research. In the meeting the common statistical methods and the target population of the study were agreed upon. The project is primarily concerned with opiate users in the age range from 15 to 55 years and the data was to preferably refer to year 1996. Preliminary results will be discussed in the next meeting in autumn, and the final reports are to be presented in December 1997.

3.2. Materials and methods

3.2.1. Data

Information from three different registers were applied in this study concerning cases of amphetamine or opiate (or their derivatives) use in 1995 in Helsinki and in the surrounding areas of Espoo, Vantaa and Kauniainen. The data included cases of Finnish citizens who had been given medical care in hospitals, who were suspected of drug-related offences or who had been sent for tests because of suspicion of driving under the influence of drugs (DUID). As far as possible, the cases were classified according to the drug used, but if the person's name (social security number) appeared several times in the register with varying principal drugs, or if it was otherwise impossible to distinguish amphetamine from opiates as a principal drug, the person was regarded as an opiate user. In addition, the information on sex and age needed as covariates for the statistical analysis were recovered for every case. The data covered persons aged 15 to 55 years divided into two classes: 15-25 years old (or born in 1970-1980) and 26-55 years old (or born in 1940-1969). Table 3.1 summarises the cases identified from the registers sorted according to persons sex, age class and principal drug.

The study applied the following official data sources:

The Hospital Patient Discharge Register (HILMO) maintained by STAKES

Hospital Discharge Register data concern patients who have got treatment for drug-related diseases (outpatient clients are not included in the register). Sampling was based on the highest order amphetamine or opiates diagnosis (encompassing opiate diagnoses 304.0, 305.5, 965.0, 965.8 and amphetamine diagnoses 304.4, 305.7, 969.7 excluding caffeine). Cases were divided into 3 categories of drug use: amphetamine, opiate and "both" if the person has had two separate treatment periods where the highest order diagnosis related to both amphetamine and opiate. There were 216 individuals who were included in the study.

Criminal Report Register (RIKI) maintained by the police

As for the police records the sampling concerned persons registered for opiate- or amphetamine-related offences (use or possession) in 1995. If the person had (separate) reports of an offence for both drugs he/she was recorded in the class of "both". The data from the police register was identified according to the date of the report - regardless of the true date of committing the offence. If a drug code was not directly connected to the case, it was not identified from any other background report material, the case was automatically excluded. The same concerned such cases where the suspected person had no

identity number (like foreigners). 353 cases were obtained from this source.

Laboratory of Substance Abuse Database of persons suspected of driving under the influence of drugs (DUID) maintained by the of the National Public Health Institute

In total there were 78 cases of people in the Laboratory of Substance Abuse Database (DUID) who were arrested for driving under the influence of drugs and whose urine was tested and found to contain opiates. The principle for classification of cases was the most important finding (derivatives of amphetamine or opiates), or if it was impossible to put the findings in order of priority or if the findings were of different drugs on separate occasions, the cases were classified as "both".

Table 3.1: **Drug users identified from the registers of STAKES (HILMO), the police (RIKI) and the National Public Health Institute (DUID) according to their sex, age class and principal drug.**

	HILMO (216)		RIKI (353)		DUID (78)	
	Number	%	Number	%	Number	%
Men	156	72.2	289	81.9	76	97.4
Women	60	27.8	64	18.1	2	2.6
Age 15-25 Years	80	37.0	118	33.4	27	34.6
Age 26-55 Years	136	63.0	235	66.6	51	65.4
Amphetamines	138	63.9	273	77.3	55	70.5
Opiates	72	33.3	48	13.6	9	11.5
Both	6	2.8	32	9.1	14	17.9

Because of the delicacy and confidentiality of the register data used in this research all the personal identification codes were destroyed according to the instructions from the Office of Data Protection Ombudsman. First the sample cases were identified by the personal identification code separately in every register. Secondly the institutions responsible for the registers agreed on the common cipher system (DES-3 routine) to encode the identification codes. The key for encrypting the identification codes was generated in the same process as the codes were encrypted. Encrypting of the data was done separately for every register so that no identification data was transferred physically from the computer of the register holder. After the encrypting process the encryptkey was destroyed. The integration of the data was then done by an institute that was independent of the register holders. The integration of the register data was done on the basis of information that consisted of encrypted identification codes,

codes for the name of the register (HILMO/RIKI/DUID), for the drug (amphetamine/opiate/both), for sex and for the age group (15-25 years old/26-55 years old). The researcher received the data combined from three register in a format where the once encrypted identification codes were re-encrypted a second time with different key - and as in the first case the second key was also destroyed. After this process neither the researcher, the information producers, the institute responsible for the first encrypting nor the institution responsible for the integration of the data were in the position to retrieve the genuine identification codes.

The final data consisted of the specifications for every case and the information on sex, age class, drug class and register(s). This enabled sorting of the data as the capture-recapture method requires, i.e. identification of cases belonging to different registers and the register overlap cases, which can be stratified according to the drug type, age class and sex.

3.2.2. Statistical method

The capture-recapture method requires at least two separate samples from the population under study. In the simplest type of this kind of experiment concerning wild animals, a sample is taken from the population and the captured animals are marked and released. The estimation of population size is based on the ratio of recaptured marked individuals in the second sample. The study dealt with a human population (namely amphetamine and/or opiate users and addicts) and the sample was taken from the register data concerning different harmful effects of drug use for individuals. In our case, the "mark" was a personal identifier, the identity number common in the registers under comparison, and the statistical inference was made on the basis of overlaps between samples (i.e. registers).

The major disadvantages of the traditional capture-recapture analysis (e.g. Seber, 1992) are the assumptions of independent samples and equal probabilities (homogeneity) of all individuals in the population of being captured in a sample. In field experiments these assumptions are seldom valid, leading to a biased estimate of population size. One of the recent developments has proved that this kind of study design can be interpreted as a special case of the general loglinear framework (Fienberg, 1972, Bishop et al., 1980, Cormack, 1981, 1989, 1993). This approach is based on modelling a contingency table formed from the data, and it allows for the handling of the dependencies between samples and heterogeneous individual sampling probabilities in the population.

With a set of k samples the data can be expressed in the form of a k-dimensional contingency table which compacts information on individuals present or absent in different combinations of samples. However this table is incomplete since one entry, the number absent from all samples, is unknown. Table 3.2 represents the data of this study with cases cross-tabulated according to the source of information. The last cell represents the unobserved number of drug users not recorded in any of the three registers. The aim of loglinear analysis is to find a model that fits the observed data well, being as simple as possible and reasonably interpreted. This model is then used to predict the missing count.

Table 3.2: **Cross-tabulation of data denoting presence and absence of 591 individuals (647 reported cases) in each of the three registers.**

	DUID				
	Present		Absent		
	RIKI		RIKI		
HILMO	Present	Absent	Present	Absent	Total
Present	4	1	23	188	216
Absent	24	49	302	-	375
Total	28	50	325	188	591

HILMO - Hospital Patient Discharge Register
RIKI - Criminal Report Register
DUID - Database of persons suspected of driving under the influence of drugs

The statistical package GLIM (Francis et al., 1994) was used to perform loglinear analysis in this study. Once an adequate model has been selected, the expected value of the missing cell is readily extracted from GLIM, thus yielding an estimate of the population size when summed with the number of observed individuals. The 95% confidence intervals of the unobserved cell and furthermore of the population size can be obtained by using the likelihood interval approach (Cormack, 1992). This procedure assumes that the model which describes the observed data also describes the count of unobserved individuals.

Loglinear modelling allows the assessment and analysis of possible dependencies between samples. The analysis begins by fitting the simplest loglinear model assuming independent samples, and adding interaction terms until an acceptable model is found. From the statistical point of view, model selection is guided by examination of residual deviance and deviance differences between models compared with the χ^2-distribution to assess the significance. In addition, the modelling can include subjectivity since the model is supplemented by interaction terms for dependencies of those samples

which can be presumed to have some connection on the grounds of the knowledge of the present application. In the case of drug addicts, dependencies between samples arise, for example, when persons entering one register (like hospital patients) are sent for a procedure that acts as the basis of another register (like HIV-testing), or when being in one register (like prison) reduces the possibility of entering other registers.

Heterogeneity in the population due to varying probabilities among individuals of appearing in samples was controlled for by using post-stratification (Doscher and Woodward, 1983). This involves breaking down the data into subgroups on the basis of relevant covariates, each having more homogeneous sampling probabilities within. In this study the probabilities of drug addicts entering different registers are assumed to depend on a person's age and sex, and possibly on the drug of choice. It can be hypothesised, for example, that older persons having used drugs for a long time enter the registers in a different manner than younger users. The effect of heterogeneity on estimates is assessed by examining if the model for the unstratified data can be simplified by removing some terms. Stratified population estimates are compared with the estimate derived from the unstratified data. If the results are approximately of the same magnitude the results are not biased by heterogeneity.

3.3. Results

Estimates for greater Helsinki

The results of loglinear analysis of the three samples are set out in Table 3.3. First trying a model with the assumption of independence of the samples yielded a poor fit as shown by the χ^2 value. Adding to the model the interaction term for the RIKI and DUID samples gave the biggest decrease in the χ^2 value (χ^2=4.42 with df=2). This model was selected as an adequate description of the data. Assuming that the model describes the observed counts of the cross-tabulated data and also the number of addicts not recorded in any of the three sources, the estimate of the number of unknown users was 2518 (95% confidence interval (CI) 1687-3858). As the samples included 591 known drug users the total population of addicts in the area of greater Helsinki in 1995 was estimated to be 3109 (95% CI 2278-4449).

Table 3.3: **Results of loglinear models used to estimate the number of amphetamine and opiate users in greater Helsinki in 1995.**

Loglinear model	χ^2	df	p value	$\hat{u}$	$\hat{N}$
Independence	28.19	3	0.000	1513 (1108,2076)	2104 (1699,2667)
S1xS2	19.70	2	0.000	867 (535,1387)	1458 (1126,1978)
S1xS3	26.53	2	0.000	1409 (1017,1962)	2000 (1608,2553)
S2xS3	4.42	2	0.110	2518 (1687,3858)	3109 (2278,4449)

Sample 1: HILMO - Hospital Patient Discharge Register
Sample 2: RIKI - Criminal Report Register
Sample 3: DUID - Database of persons suspected of driving under the influence of drugs
$\hat{u}$ - fitted value for the missing cell with 95% confidence interval.
$\hat{N}$ - total estimate of users with 95% confidence interval.

The estimate of the interaction term for the RIKI and DUID samples is positive, indicating that the overlap for these two samples is proportionally greater than for other combinations of samples. One reason for this is probably that a drug finding at National Laboratory of Substance Abuse can as well result in a report of an offence related to driving under the influence of drugs as in a drug use offence report. Another possibility is that some persons in the police register suspected of drug-related offences were caught driving drunk wherefore they were systematically recorded in the DUID database. Thus the same occurrence may cause a record in both registers. On the basis of information from police, reports in the RIKI register concerning suspicions of drug-related offences included 8 such cases where there existed at the same time a suspicion of driving under the influence of drugs. There is, however, no information available on how many of these 8 cases were in the HILMO register maintained by STAKES. Thus of the 24 cases common to the RIKI and the DUID registers (see Table 3.2) 8 can be assumed to form a systematic (artificial) overlapping, or at most 4 of them can be included in all three registers. Although the different variations of the original contingency table were re-analysed the elimination of this known overlap did not result in changes in the loglinear model nor in the final estimate.

In an additional analysis the model for the unstratified data was fitted to stratified samples based on sex, age class and drug. As Table 4 shows, the same model described as well most of the partial data sets with the possible exception of the group of amphetamine users. Attempts to simplify the model for the unstratified data by removing the interaction term proved a good fit only with regard to the data for female users (χ^2=2.91 with df=3).

Table 3.4: **Estimated number of amphetamine and opiate users in greater Helsinki in 1995, based on the loglinear model including a dependence**

between the RIKI and the DUID registers.

	χ^2 (2 df)	No. of known cases	$\hat{u}$	$\hat{N}$
All cases	4.42	591	2518 (1687,3858)	3109 (2278,4449)
Males	4.50	470	1727 (1111,2756)	2197 (1581,3226)
Females	0.26	121	854 (328,2834)	975 (449,2955)
Total			2581 (1439,5590)	3172 (2030,6181)
Age 15-25 years	3.53	206	790 (416,1595)	996 (622,1801)
Age 26-55 years	2.51	385	1743 (1050,3027)	2128 (1435,3412)
Total			2533 (1466,4622)	3124 (2057,5213)
Amphetamine	6.48	416	1941 (1171,3367)	2357 (1587,3783)
Opiates	0.42	175	600 (312,1218)	775 (487,1393)
Total			2541 (1483,4585)	3132 (2074,5176)

$\hat{u}$ - estimate of the number of unknown users with 95% confidence interval

$\hat{N}$ - total estimate of users with 95% confidence interval.

There in no doubt is heterogeneity within the population of drug addicts, which in regard to sex and age does not seem to affect the model selected, and the estimates from the stratified samples do not differ in magnitude from the total result. In the case of amphetamine use the statistical goodness-of-fit would require adding an interaction term also for the HILMO and RIKI samples. However this makes the estimation impossible as one of the observed cells has a zero entry. If the zero were replaced, for example, with unity, the estimated number of unknown amphetamine users would be 3978. As a result of the poor goodness-of-fit the estimate for this subgroup is more speculative than those for others. This uncertainty reflects also on the reliability of the total estimate.

Further stratification by sex and age within the data sets of amphetamine and opiate users makes the cell sizes smaller and the estimation more unstable. However, a quick overview of these strata shows little deviation from the previous.

Table 4.5: The estimated prevalence of amphetamine and opiate use in greater Helsinki in 1995.

	Known cases		**Total Estimate**	**Population**	**Prevalence**
	Number	**%**	**of users**		**(%)**
All cases	591	19.0	3109	550000	0.57
Males	470	21.4	2197	267000	0.82
Females	121	12.4	975	283000	0.34
Aged 15-25 years	206	20.7	996	109000	0.91
Aged 26-55 years	385	18.1	2128	441000	0.48

Amphetamines	416	17.7	2356	550000	0.43
Opiates	175	22.6	775	550000	0.14

The estimate of 3109 users of heavy drugs from the unstratified data represents approximately 0.6% (95% CI 0.4-0.8%) of the greater Helsinki population aged 15-55 years during 1995 (Table 4.5).

The ratio of known to unknown cases was about 1:4. Looking at the subgroups showed that female users were relatively more hidden than males. Comparing the results with the information in Table 1 showed that the females seem to enter the RIKI and the DUID registers less frequently than men. No equally clear difference appeared between the two age classes or between amphetamine and opiate users. Among both amphetamine users and opiate users females represented roughly 1/3 of the group. As for the age distribution in these groups, about 1/3 of amphetamine users and 1/4 of opiate users were 25 years of age or younger.

Estimates for the whole country

Application of the same estimation procedure as above to the national registers is not straightforward, and such data sources were not available for this study. For a rough national estimation, information on drug users registered in hospitals outside greater Helsinki was obtained in the form shown in Table 4.6. These numbers may have included some persons in common with the data concerning greater Helsinki. The first figures (and per cent intervals) indicate which ratio the known cases registered at the hospitals in greater Helsinki represent of the total estimate in greater Helsinki. These estimated ratios were based on the 95% confidence intervals in Table 4.4. Assuming that the same ratios would be valid also for the cases recorded at hospitals in other parts of Finland, rough intervals describing the amphetamine and opiate use in the whole country in 1995 were obtained. On the basis of these calculations the number of hard drug users in Finland would be in round figures 5,000-10,500 (about 0.2-0.4% of the population), of which over 40% would be concentrated in the greater Helsinki area. Let us emphasise that these figures must be considered with extreme reserve. They were based on information from hospital registers only and on an assumption that drug users elsewhere in the country enter hospitals at the same rate as in greater Helsinki. Applying the same multipliers led to the estimated upper limit of 10500, which is obviously too high.

Table 4.6: **Registered cases in the Hospital Patient Discharge Register of STAKES in greater Helsinki and in other parts of Finland and the**

rough estimates of the number of hard drug users outside greater Helsinki and in the whole country.

	Greater Helsinki			Other parts of Finland		Whole Finland
	Known cases		Total estimate	Known	Total estimate	Total estimate
	Number	%	of users	cases	of users	of users
All cases	216	6.95	2278-4449	294	3100-6050	5380-10500
Male	156	7.10	1581-3226	207	2100-4280	3680-7510
Females	60	6.15	449-2955	87	650-4290	1110-7250
Total			2030-6181		2750-8570	4780-14760
Age 15-25 years	80	8.03	622-1801	112	870-2520	1490-4320
Age 26-55 years	135	6.39	1435-3412	182	1920-4560	3360-7970
Total			2057-5213		2790-7080	4850-12290
Amphetamines	138	5.86	1587-3783	193	2220-5290	3810-9070
Opiates	78	10.06	487-1393	101	630-1800	1120-3190
Total			2074-5176		2850-7090	4930-12260

3.4. Discussion

Registers

In order to ensure the coverage and representativeness of the data it would be desirable, that the notifications on drug users be obtained from as diverse registers as possible, because among other things users are probably recorded in different registers in different periods of their life.

Selecting the sources of information is not, however, an easy task as different combinations of data sets may result in inconsistent estimates. One reason for this is, for example, that some registers can be geographically heterogeneous, so that they do not necessarily reflect the true extent of drug use in the study area. In addition, it should be assessed whether there is artificial overlapping between some sources as is the case between police register (RIKI) and the Laboratory of Substance Abuse register (DUID) in this study. The mechanism with which a drug finding in the Laboratory of Substance Abuse leads to a report of an offence related to drug use should be assessed more carefully in future studies.

If mortality data is being used, the time period to which it refers should be considered in relation to other registers. The official data on mortality (such as ICD-9 or ICD-10 classification) do not accurately reflect the true mortality of drug users within a city. This topic is being studied in another EMCDDA project. An interesting

question is also how applying multipliers to mortality data could be a simple validation of capture-recapture results. Usually these statistics contain too few drug users to be incorporated as a sample in a capture-recapture model.

Besides the data quality also the number of registers used in the analysis affects the reliability of results. In order to distinguish independence between samples three registers is an absolute minimum, but more reliable estimates are obtained from at least four data sets. In the future the combination of three registers in this study can be supplemented with information concerning hepatitis-C infections.

Hay (1997), for example, has studied the differences that occur when employing different combinations of data sets in estimating the population size. He has found that the estimates obtained when using three registers are often markedly different from the estimates which use four sources.

In this examination the three-source method often resulted in substantial underestimates, and several confidence intervals did not include the value which the best fitting four-source model suggested to be the population size. This again reminds one of the need to consider a 95% confidence intervals with care. The same concerns especially the case when dealing with several registers and interactions between them. These confidence intervals are always connected with the particular model, and model selection again often includes subjectivity.

Loglinear modelling

As mentioned in Section 3.2., the loglinear method leads to a model that fits the data well and furthermore leads to an estimate of the population size. It should be remembered, however, that a statistical model is always a simplification of a complex phenomenon. A merely adequate fit of the model to the data does not guarantee the reliability of the estimate, as one of the essential assumptions underlining the analysis is that no hidden subgroups are allowed in the population. In other words, individuals not observed in any of the registers should have the same probability of appearing in them as the observed cases. There is no way of checking this assumption.

Closed population

So far, in all capture-recapture studies that survey the prevalence of

drug use the population has been thought of as closed, in that the population remains unchanged and its size constant during the study. This, however, is not a realistic assumption in practice. Cases had usually been collected from registers covering such a long time period (from 6 months to one year) that some of the persons may have stopped using drugs, or may have moved out of the area of interest, or died, whilst new users join the population. One solution is to collect the cases over a shorter period, but this may give rise to the problem of sparse data, leading to unstable estimation because the contingency table might have many small entries, possibly several zeros. Yet, special methods for addressing the problems of sparse data have been suggested (e.g. Chao, 1989).

Another possibility is to apply the methods for open populations developed to control for the changes in the population. These have not yet been proved to work in practice, but they are being studied. The problem with loglinear models describing an open population is the increasing complexity of the fitted models and the difficulty in interpreting them. On the other hand it has been supposed (e.g. Larson et al., 1994), that even if the assumption of a closed population is not quite valid, it would not seriously effect the estimate.

The definition of the population is closely connected to the question of how its members - the drug users - are defined. As noted in the EMCDDA first project meeting, the analysis can be restricted to persons who have used a certain drug during the study interval (e.g. in 1995). Alternatively the estimate can be thought of as the number of potential clients of a city's treatment agencies. Thus although a person may not have been identified as using a particular drug that year, it serves to consider them still as users because of the complex nature of drug use.

Heterogeneity of the population

It is clear that heterogeneity appears in the population of drug users (addicts) so that overlapping cases of certain subgroups in different samples is markedly high and some groups can be missing.

It is necessary for the accuracy of the estimation procedure that the data can be stratified (post hoc) for the analysis at least on the basis of principal factors causing heterogeneity. For example, age correlates with one such factor, namely the length of drug use. Stratifying the population is useful also because the estimated sizes of different subgroups give useful information for making decisions related to the drug use situation.

According to the most pessimistic opinions the capture-recapture methodology is not applicable to this kind of situation where being entered in the registers depends on the lifestyle of individuals and on their ability to control their life. Among humans - and especially among heavy drug users - there are different kinds people in character and moreover people vary in their willingness to seek help for their problems or to avoid consequences that might be caused by entering into a certain type of register. Thus the population of drug users manifests heterogeneity by subgroups and also on an individual level. People can, for example, stop using drugs temporarily or leave the study area, whereupon they do not belong to the target population and their probability of entering the applied data sources is zero. When the people begin to use drugs again or return to the study area they join the population and their probability of appearing in the registers grows higher. In the worst case these phenomenon can invalidate the entire capture-recapture analysis and it would be better not to give any estimate.

All the registers applied in this study can be thought to express problems in the life of a drug user, so that the estimate based on these data sources concerns mostly the group of problem users.

It can be supposed that there are also quite a lot of people in greater Helsinki who use these drugs in a more controlled manner and do not so easily enter the official registers. This assumption is also supported by the estimated ratio of 1:4 of known to unknown users, which seems too small: in other words, the total estimate does not cover the whole group of users not observed in the registers. Especially the estimated prevalence of amphetamine use is speculative. The use of amphetamine should clearly be more common than opiate use: For example, on the basis of quantities of drugs sent to the Laboratory of Substance Abuse, the number of amphetamine users should be about sevenfold that of the number of opiate users. According to Table 5 the difference is only threefold. It may be that there is a hidden group among amphetamine users which stays outside all registers.

Study area

As mentioned in Section 3.3, the capture-recapture method is for the present a questionable technique with which to consider the national level drug use situation. So far the method has been applied only in urban centres to estimate drug use prevalence, and there is no evidence of how it serves in estimates for sparsely populated areas or for the whole country.

National prevalence estimates are under development in another EMCDDA project. Their preliminary suggestions involve utilisation of demographic multipliers, consideration of the ratio of known to unknown users or application of multivariate regression analysis in the different parts of the country where reliable local estimates are available at least from three districts.

A national survey gives fairly reliable estimates on the use of cannabis and on the mixed use of sedatives and tranquillisers with alcohol, but it easily fails to reach the users of heavy drugs or to get truthful information from them. The capture-recapture method and a survey can be regarded as mutually complementary approaches in estimating the prevalence of drug use. If, for example, it is possible to carry out a survey reliably, it can offer information among other things about a relevant stratification which then can be utilised in modelling capture-recapture data.

Other statistical methods

In estimating the number of drug users it is preferable to apply some other complementary approach along with the capture-recapture method for validating the results. This could be a survey as mentioned above or an analysis of multipliers with the Cause of Death Register. Some studies in the EMCDDA-project utilise truncated Poisson models (Zelterman, 1988; Chao, 1989) which use data from only one source. During the observation interval some individuals are recorded there only once, others twice and so on. This phenomenon is interpreted as a Poisson process and the aim is to estimate the number of those users who have not once entered the register on the basis of the number of observed users. Truncated Poisson models are not very particular concerning how the data have been collected and the assumptions behind these models are fairly reasonable with regard to this kind of application.

To conclude, the potential of capture-recapture methods is limited in estimating the prevalence of drug use, and more accurate examination of its applications together with development of other approaches still a statistical challenge in this field. It is possible to develop the registration systems and practices in such a way so as to make this kind of study easier.

References

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1980). Discrete multivariate analysis: Theory and Practice. MIT Press. Cambridge, Mass.

Bloor, M., Leyland, A., Barnard, M. and McKeganey, N. (1991). Estimating hidden populations: A new method of calculating the prevalence of drug-injecting female street prostitution. Br. J. Addict. 86, 1477-83.

Brecht, M.L. and Wickens, T.D. (1993). Applications of multiple-capture methods for estimating drug use prevalence. J. Drug Issues 23, 229-50.

Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. Biometrics 45, 427-438.

Cormack, R.M. (1981). Loglinear models for capture-recapture experiments on open populations. In R.W. Hiorns and D. Cooke (Eds.). the Mathematical Theory of the Dynamics of Biological Populations II. Academic Press, London.

Cormack, R.M. (1989). Log-linear models for capture-recapture. Biometrics 45, 395-413.

Cormack, R.M. (1992). Interval estimation for mark-recapture studies for closed populations. Biometrics 48, 567-76.

Cormack, R.M. (1993). The flexibility of GLIM analyses of multiple recapture or resighting data. In J.D. Lebreton and P.M. North (Eds.). Marked Individuals in the Study of Bird Populations. Birkhäuser-Verlag, Basel.

Domingo-Salvany, A., Hartnoll, R.L., Maguire, A., Suelves, J.M. and Antó, J.M. (1995). Use of capture-recapture to estimate the prevalence of opiate addiction in Barcelona, Spain, 1989. Am. J. Epidemiol. 141, 567-74.

Doscher, M.and Woodward, J.A. (1983). Estimating the size of subpopulations of heroin users: Applications of log-linear models to capture-recapture sampling. Int. J. Addict. 18, 167-82.

Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2k contingency tables. Biometrika 59, 591-603.

Francis, B., Green, M., Payne, C. (Eds.) (1994). The GLIM System, release 4. Clarendon, Oxford.

Frisher, N., Bloor, M. and Finlay, A. et al. (1991). A new method of estimating prevalence of injecting drug use in an urban population: Results from a Scottish City. Int. J. Epidemiol. 20, 997-1000.

Hay, G. and McKeganey, N. (1996). Estimating the prevalence of drug misuse in Dundee, Scotland: An application of capture-recapture methods. Journal of Epidemiology and Community Health 50 (4), 469-472.

Hay, G. (1997). The selection from multiple data sources in epidemiological capture-recapture studies. The Statistician 46 (to be published).

Hook, E.B. and Regal, R.R. (1992). The value of capture-recapture methods even for apparent exhaustive surveys. Am. J. Epidemiol. 135, 1060-67.

Kontula, O. and Koskela, K. (1992). Huumeiden käyttö ja mielipiteet huumeista. Suomi ja Eurooppa vertailussa. (Drug use and opinions about drugs: Comparing Finland and Europe) Sosiaali- ja terveysministeriön jukaisuja 8.

Kontula, O. (1995). The prevalence of drug use with reference to problem use in Finland. Int. J. Addict. 30, 1053-66.

Larson, A., Stevens, A. and Wardlaw, G. (1994). Indirect estimates of hidden populations - capture-recapture methods to estimate the number of heroin users in the Australian Capital Territory. Soc. Sci. Med. 39, 823-31.

McCarty, D.J., Tull, E.S., Moy, C.S., Kwoh, C.K. and LaPorte, R.E. (1993). Ascertainment corrected rates: Applications of capture-recapture methods. Int. J. Epidemiol. 22, 559-65.

Partanen, J. (1994). Märkä pilvi. (Wet Cloud) Alkoholipolitiikka 59, 397-411.

Poikolainen, K. (1997). Huumeet Suomessa: Esiintyvyys jakehityssuunta. (Drugs in Finland: the prevalence and the trends) Tiimi 2, 4-11.

Seber, G.A.F. (1992). A review of estimating animal abundance II. Int. Statistical Review 60, 129-66.

Sekar, C.C. and Deming, W.E. (1949). On a method of estimating birth and death rates and the extent of registration. J. Am. Stat. Assoc. 44, 101-15.

Vuori, E., Ojanperä, I. and Rasanen, I. (1997). Oikeuskemiallisesti todetut myrkkykuolemat vuosina 1993-1995. (Deaths by poisoning according to the forensic medical analysis) Suom. Lääkäril. 52, 421-5.

Zelterman, D. (1988). Robust estimation in truncated discrete distributions with applications to capture-recapture experiments. Journal of Statistical Planning and Inference 18, 225-237.

4 Rome

PREVALENCE ESTIMATION OF DRUG ADDICTION IN ROME

Daniela D'Ippoliti, Osservatorio Epidemiologico, Rome

4.1 Introduction

Rome is the capital of the Lazio Region (population: 4,154,554) with a population of 2,775,250 residents as of census 1991. The population of Rome is as follows:

Table 4.1: **Population of Rome.**

	0-14	**15-24**	**25-34**	**35-44**	**45-54**	**>54**
Males	191611	217929	231847	185725	181220	317268
Females	181821	209027	226785	199983	202822	429222
Total	373432	426956	458632	385698	384042	746490

Current information on drug addiction in Rome is available from a Surveillance System implemented in our Region by our Department. This system collects individual data on drug users attending public treatment centres and non-governmental organisations. Comparisons of the data from this system, from 1992 through 1994, show that in our city drug users attending treatment centres are predominantly males (more than 80%), 29 years old on average, more than 90% are heroin dependent and the injecting use is the primary route of administration for more than 75% of clients.

4.2 Data Sources

We applied the capture-recapture methodology to estimate the prevalence of drug users in the city of Rome. We used three sources of information on drug users living in Rome: the Surveillance System on Drug Addiction, the Hospital Discharge System, and the Mobile Emergency Unit. All data collected refer to subjects who approached these services in 1996.

Surveillance System on Drug Addiction (SSDA)

The SSDA gathers data on drug users attending public treatment centres. The system collects information about socio-demographic and toxicological characteristics of patients and type of treatment offered.

In this system, the “case definition” refers to persons who have used drugs in the last 30 days and have a contact with treatment centres for their state of drug dependence. From this database 6422 people were included in the study. More than 80% were males, 95% were heroin users and, among these, 82% were injectors.

Hospital Discharge System (HDS)

The Hospital Discharge System is not a specific system collecting data on drug addiction. Drug users have been identified as those having been discharged from hospitals with primary or secondary diagnosis of drug dependence.

We considered subjects with diagnosis of drug dependence codified according to the ICD-IX classification. There were 637 persons who satisfied the criteria for inclusion in the study, as shown in the following scheme:

Table 4.2: **Codes related to drug dependence (ICD-IX)**

ICD-IX	definition	n
304.0	drug dependence for morphine type drugs	396
304.1-304.8	drug dependence for other drugs	96
304.9	unspecified drug dependence	68
965.0	poisoning by opiates and related narcotics	77

About 74% of them were males.

Mobile Emergency Unit (MEU)

The Mobile Emergency Unit is a component of a Harm Reduction Programme for drug users started in Rome in 1994. The Unit deals with emergencies related to drug dependence (withdrawal, overdose...). From this source 338 people were included, 81% of whom males.

4.3 Matching Process

To remove duplicates within sources and to identify overlaps among the three sources, we used a linkage procedure based on four types of identification variables available for each source: the first three letters of name and surname, sex and date of birth.

4.4 Case Definition

The “case definition” used in each of our sources is different among them. Cases reported by the SSDA are known to be predominantly heroin users, commonly injecting, while notifications from the HDS and the MEU certainly identify individuals with problems related to drug dependence but no detailed information is available about their habit of using drugs. Despite this limitation, data from our sources

allow to estimate the prevalence of "problematic drug abuse" in Rome, where opiate use is likely to represent the great majority.

4.5 Capture-recapture analysis

For our analysis, we considered people 15-54 aged, altogether 7397 cases were collected from the three sources. After the linkage procedure, we identified 6896 individuals who had at least one report in one of the three sources during 1996.

Log-linear models were used according to the likelihood approach, and the estimation procedure was applied separately to males and females and to two age groups (15-34, 25-34 and 35-54 years).

Data used from each source and the preliminary estimates of the total population of drug users, obtained from different models, are shown in the following tables.

Table 4.3: **Sources of data, Rome 1996.**

	Males			**Females**		
Sources	15-24	25-34	35-54	15-24	25-34	35-54
Source 1: SSDA	543	2918	1882	130	583	366
Source 2: HD	56	253	161	24	100	43
Source 3: MEU	57	163	53	18	38	9
Total of contacts	656	3334	2096	172	721	418
Individuals	601	3128	1982	150	642	393

Table 4.4: **Overlaps between each source for all age groups.**

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	27	134	302	5959
	Absent	11	166	297	*

Table 4.5: **Overlaps between each source for males, aged 15-54.**

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	14	105	234	4990
	Absent	8	146	214	*

Table 4.5a: **Overlaps between each source for males aged 15-24.**

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	5	21	22	495
	Absent	2	29	27	*

Table 4.5b: **Overlaps between each source for males, aged 25-34.**

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	7	57	131	2723
	Absent	4	95	111	*

Table 4.5c: **Overlaps between each source for males, aged 34 - 54.**

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	2	27	81	1772
	Absent	2	22	76	*

Table 4.6: Overlaps between each source for females, aged 15-54.

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	13	29	68	969
	Absent	3	20	83	*

Table 4.6a: Overlaps between each source for females, aged 15-24.

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	4	6	8	112
	Absent	0	8	12	*

Table 4.6b: Overlaps between each source for females, aged 25-34.

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	9	16	42	516
	Absent	3	10	46	*

Table 4.6c Overlaps between each source for females, aged 34-54.

		Source 3			
		Present		Absent	
		Source 2			
		Present	Absent	Present	Absent
Source 1	Present	0	7	18	341
	Absent	0	2	25	*

Table 4.7: **Males aged 15-54 years.**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	16.54	3	0.001	10759	10051-11559	
+S1xS2	1.60	2	0.003	11963	10607-13641	
+S1xS3	10.15	2	0.006	10092	9334-10971	
+S2xS3	8.96	2	0.011	10913	10175-11715	
+S1xS2+S1xS3	10.05	1	0.002	9617	7656-14081	
+S1xS2+S2xS3	**1.11**	**1**	**0.292**	**12649**	**11080-14634**	←
+S1xS3+S2xS3	3.88	1	0.049	10275	9473-11211	
Saturated	0.00	0	1.000	16816	9990-34006	

Table 4.7a: **Males aged 15-24 years.**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	7.55	3	0.056	1145	965-1396	
+S1xS2	7.52	2	0.023	1164	922-1546	
+S1xS3	7.41	2	0.025	1108	887-1454	
+S2xS3	**2.18**	**2**	**0.336**	**1199**	**996-1486**	←
+S1xS2+S1xS3	7.29	1	0.007	993	694-2627	
+S1xS2+S2xS3	1.77	1	0.184	1285	977-1804	
+S1xS3+S2xS3	2.17	1	0.140	1209	933-1668	
Saturated	0.00	0	1.000	2968	911-16216	

Table 4.7b: **Males aged 25-34 years.**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	13.84	3	0.003	5998	5469-6617	
+S1xS2	4.90	2	0.086	7270	6138-8790	
+S1xS3	4.29	2	0.117	5375	4857-6008	
+S2xS3	11.53	2	0.003	6060	5516-6706	
+S1xS2+S1xS3	4.18	1	0.041	5764	4140-11139	
+S1xS2+S2xS3	**0.39**	**1**	**0.530**	**7666**	**6376-9447**	←
+S1xS3+S2xS3	2.86	1	0.091	5435	4896-6099	
Saturated	0.00	0	1.000	9858	4833-28890	

Table 4.7c: **Males aged 35-54 years**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	**1.17**	**3**	**0.760**	**3568**	**3175-4060**	←
+S1xS2	0.80	2	0.670	3351	2739-4295	
+S1xS3	1.06	2	0.588	3610	3156-4199	
+S2xS3	0.19	2	0.908	3593	3191-4098	
+S1xS2+S1xS3	0.37	1	0.542	2818	2188-6062	
+S1xS2+S2xS3	0.00	1	0.959	3426	2769-4468	
+S1xS3+S2xS3	0.04	1	0.844	3645	3177-4254	
Saturated	0.00	0	1.000	3337	2052-11634	

Table 4.8: **Females aged 15-54 years.**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	26.87	3	0.000	2029	1819-2284	
+S1xS2	21.31	2	0.000	1683	1459-2004	
+S1xS3	22.10	2	0.000	2191	1911-2554	
+S2xS3	9.82	2	0.007	2119	1883-2414	
+S1xS2+S1xS3	21.28	1	0.000	1738	1349-3213	
+S1xS2+S2xS3	8.08	1	0.004	1853	1541-2339	
+S1xS3+S2xS3	**2.76**	**1**	**0.096**	**2368**	**2026-2824**	←
Saturated	0.00	0	1.000	4720	2043-17665	

Table 4.8a: **Females aged 15-24 years.**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	9.63	3	0.022	257	201-350	
+S1xS2	9.63	2	0.008	255	184-409	
+S1xS3	9.30	2	0.010	276	199-430	
+S2xS3	**6.53**	**2**	**0.038**	**275**	**208-390**	←
+S1xS2+S1xC3	7.80	1	0.005	*	*	
+S1xS2+S2xC3	6.35	1	0.012	299	194-583	
+S1xS3+S2xC3	5.61	1	0.018	318	211-569	
Saturated	0.00	0	1	*	*	

*Estimates can not be computed because of zero values of the expected counts.

°All models don't fit data well

Table 4.8b: **Females aged 25-34 years**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	19.58	3	0.000	1038	913-1204	
+S1xS2	15.53	2	0.000	858	731-1060	
+S1xS3	17.53	2	0.010	1104	940-1329	
+S2xS3	4.21	2	0.122	1096	949-1292	
+S1xS2+S1xC3	15.20	1	0.000	795	678-1208	
+S1xS2+S2xC3	3.28	1	0.070	965	770-1318	
+S1xS3+S2xC3	**0.68**	**1**	**0.409**	**1207**	**1001-1505**	←
Saturated	0.00	0	1.000	1702	832-6049	

Table 4.8c: **Females aged 35-54 years**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	5.06	3	0.168	769	600-1043	
+S1xS2	1.65	2	0.439	503	400-785	
+S1xS3	**0.91**	**2**	**0.633**	**879**	**647-1290**	←
+S2xS3	4.00	2	0.045	761	596-1029	
+S1xS2+S1xS3	0.71	1	0.398	*	*	
+S1xS2+S2xS3	0.00	1	1	490	399-739	
+S1xS3+S2xS3	0.00	1	1	867	640-1240	
Saturated	0.00	0	1	*	*	

*Estimates can not be computed because of zero values of the expected counts.

Table 4.9: **Males aged 15-34 years**.

Model	G^2	d.f.	p-value	N	95% CI	
Independence	18.91	3	0.000	7149	6586-7801	
+S1xS2	11.50	2	0.003	8262	7175-9651	
+S1xS3	10.38	2	0.006	6489	5909-7185	
+S2xS3	12.16	2	0.002	7278	6686-7969	
+S1xS2+S1xC3	10.37	1	0.081	6581	5012-10684	
+S1xS2+S2xC3	**1.38**	**1**	**0.239**	**8845**	**7559-10535**	←
+S1xS3+S2xC3	5.29	1	0.021	6632	6009-7385	
Saturated	0.00	0	1.000	12957	6890-30879	

Table 4.10: **Females aged 15-34 years.**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	25.50	3	0.000	1302	1158-1483	
+S1xS2	22.87	2	0.000	1141	975-1386	
+S1xS3	23.51	2	0.000	1379	1191-1631	
+S2xS3	7.51	2	0.023	1376	1207-1593	
+S1xS2+S1xS3	22.87	1	0.000	1140	891-2075	
+S1xS2+S2xS3	7.30	1	0.007	1360	1050-1731	
+S1xS3+S2xS3	**3.60**	**1**	**0.058**	**1520**	**1277-1862**	←
Saturated	0.00	0	1.000	3375	1392-13224	

Table 4.11: **All cases°.**

Model	G^2	d.f.	p-value	N	95% CI	
Independence	32.24	3	0.000	12712	11990-13512	
+S1xS2	31.31	2	0.000	13225	11975-14726	
+S1xS3	30.51	2	0.000	12363	11543-13297	
+S2xS3	10.28	2	0.006	12997	12229-13854	
+S1xS2+S1xS3	30.11	1	0.000	11378	9350-15448	
+S1xS2+S2xS3	**6.30**	**1**	**0.012**	**14278**	**12742-16167**	←
+S1xS3+S2xS3	9.61	1	0.002	12756	11857-13787	
Saturated	0.00	0	1	24716	15263-45521	

°All models don't fit data well

4.6 Results

We chose as the best models those consistent with our a priori knowledge about sources and according to statistical criteria based on the goodness-of-fit ($G^2 < 3.84$, p-value >0.05) and with the fewest possible parameters (see ←).

Because of sparse data in the females group, it is better to consider estimates obtained from the two age groups: 15-54 and 35-54 years. However, also the estimates from the three age groups (15-24, 25-34, 35-54) combine to a similar figures to the estimates from the two age groups, as shown in the following table:

Table 4.12: **Estimates of the number of drug users in Rome, 1996.**

	G^2	No of known DUs	Estimate No of DUs	CI 95%
All cases	6.30			
		6896	14278	12742-16167
Males	1.11	5711	12649	11080-14634
Females	2.76	1185	2368	2026-2824
Total		6896	15017	
Males, age group				
15-24	2.18	601	1199	996-1486
25-34	0.39	3128	7666	6376-9447
35-54	1.17	1982	3568	3175-4060
total (1)		5711	12433	
Females, age group				
15-24	6.53	150	275	208-390
25-34	0.68	642	1207	1001-1505
35-54	0.91	393	879	647-1290
total (2)		1185	2361	
Males and Females (1+2)			14794	
Males, age group				
15-34	1.38	3729	8845	7559-10535
35-54	1.17	1982	3568	3175-4060
total (3)		5711	12413	
Females, age group				
15-34	3.60	792	1520	1277-1862
35-54	0.91	393	879	647-1290
total (4)		1185	2399	
Males and Females (3+4)		6896	14812	

4.7 Discussion

To estimate the total population of drug users, we selected models as the simplest and with a good fit between observed and expected counts, allowing for various dependencies among sources. The accepted models include the interactions SSDA*HDS, SSDA*MEU, HDS*MEU and these dependencies are consistent with the nature of

our sources.

Both the total estimates less then 15000 DUs around and the stratified estimates seem sensible, since they are about twice the number of known DUs. More over, as shown in table 4.12, the stratified estimates combine to a similar figure to the total estimates; being (turning out) sex and age heterogeneity variables.

4.8 Conclusions

We are confident of our estimates since we already estimated prevalence of DUs in 1992, obtaining similar results. In 1992, we produced estimates using two different methods: capture-recapture and multiplier formula.

Estimates from capture-recapture combine to a total of 24513 DUs residents in the Lazio Region. Data from the SSDA show that DUs residents in Rome, and in treatment, in 1992 were about 60.5% of the total. Then we could infer from the Lazio estimate that DUs residents in Rome were 14800 around.

Estimates from multiplier formula, applied to the overdose mortality rate observed in a cohort of DUs, produced a total of 13721 DUs residents in Rome in 1992.

These results seem to suggest that the number of drug users in our city remain stable from 1992 through 1996. Also data from SSDA confirm that number of DUs in treatment at public treatment centres in constant from 1992 through 1996 in our city.

Our estimates allow to refer to the prevalence of "**problematic drug abuse**" in Rome, because of heterogeneity of case-definition used in the different sources, even if the opiate use is likely to represent the great majority.

5 Rotterdam

ESTIMATING THE NUMBER OF OPIATE USERS IN ROTTERDAM USING STATISTICAL MODELS FOR INCOMPLETE COUNT DATA

Filip Smit, Methods Section, Trimbos Institute, Netherlands Institute of Mental Health and Addiction

Jaap Toet, Dept. of Health Promotion, Municipal Health Service, Rotterdam

Peter van der Heijden, Dept. of Methodology & Statistics, Faculty of Social Sciences, University of Utrecht

Abstract

Often, the number of opiate users in a city is not known. This lack of knowledge hampers adequate policy making. However, the unknown size of the opiate using population can be estimated with help of statistics. We review seven statistical models and apply them to a real data-set. The estimation problem can be described as follows. All models are based on a single data-set. This is remarkable since most population size estimators require multiple data-sources. This single data-set is incomplete in the sense that only opiate users who enter a low threshold methadone maintenance programme are registered while all others are not. The only information which is available in this data-set is the number of treatment episodes of uniquely identifiable persons who are on methadone maintenance, plus a set of covariates like AGE, SEX and so on. Still, from this single and incomplete data-set inferences about the total population size can be made. The estimates are compared and the pro's and con's of the estimators are evaluated. In principle, these estimators can be used for unmet needs assessment and health service demand estimation, resource allocation and health service performance evaluation.

Key words Population size estimators, prevalence, opiate use, capture-recapture, truncated Poisson, truncated Negative Binomial, regression.

5.1 Introduction

Background

The total number of persons in a city who are dependent on opiates is usually not known. Some opiate users might be known through health services, police contacts and the like, but others are never seen. Such a population is said to be 'partially observed' just like the proverbial ice berg. Not knowing the total size of a population hampers developing, implementing and evaluating health policies. Therefore, it is important to accurately estimate the unknown size of a population. Statistics can play a role here.

The utility of population size estimation has long been recognised. As a consequence there are many population size estimators; see Seber (1982, 1986, 1992), Pollock (1991) and Khorrazaty *et al*. (1977) for reviews. These estimators have by and large been developed in biometrics (fisheries and wild life studies) and their validity, when applied to human populations, is not well established. This is one reason to exercise some care and good judgement when selecting an estimator for studying human populations, specifically a population of opiate users (Simeone, Nottingham and Holland, 1993; EMCDDA, 1997). Another reason presents itself when new estimators are developed and have not yet been tried very often. This is the case with estimators based on a single data-set (Chao, 1988, 1989; Zelterman, 1988; Wilsson and Collings, 1992; Van der Heijden *et al*, in progress) when these estimators are applied on a population of drug users. Either way, one likes to know how competing estimators perform and what they 'cost' in terms of data collection and computational effort. Predecessors in this form of methodology comparison are, for example, Simeone, Nottingham and Holland (1993), Brecht and Wickens (1993), Wickens (1993) and Korf, Reijneveld and Toet (1994). This is not to say that all methodological issues have been resolved (EMCDDA, 1997).

Purpose

In this study we want to compare different population size estimators when applied to a partially observed population of opiate users. To this end, we use a real life data-set which contains information on opiate users in Rotterdam, The Netherlands in 1994. We will compare these estimators in terms of the 'quality' of their outcomes (congruence, criterion validity), assumptions (realism, validity) and 'costs' in terms of data requirements and computational effort. The rationale for this comparison is that each of these estimators is based

on different assumptions and violation of these assumptions may render the estimates invalid. Also, some estimators require very little information whereas other estimators can only be calculated when more information is available. Possibly, there is a trade off between validity and the effort to obtain the required data. The purpose of this study, therefore, is to evaluate the balance between the 'quality' of the results and the 'costs' of the required data for these estimators.

Outline

In the methods paragraph we will describe the population, the data source, the estimators and their underlying assumptions. The estimators are

1. two truncated Poisson estimators independently developed by Daniel Zelterman (1988) and Anne Chao (1987, 1989), and their stratified counterparts.

2. three estimators based on the truncated Negative Binomial regression and special cases thereof, which have recently been developed by Peter Van der Heijden *et al* (in progress).

We will not deal exhaustively with the related estimation procedures and refer the interested reader to the relevant literature. In the results paragraph we will present the outcomes of each of the estimators when applied to the data-set. Finally, in the discussion paragraph we will evaluate the pro's and con's of each of the estimators.

5.2 Methods

Data

Data on opiate users who apply for treatment are routinely collected and entered in the Rotterdam Drug Information System (RODIS). These data are collected through three health service centres where opiate users have entered a low threshold methadone maintenance programme. The close co-operation between the health centres effectively makes it one single institute which happens to have offices at three different locations. We therefore treat the RODIS-data as coming from a single 'data collecting agent'. The visitors of the centres can be classified as 'problematic opiate users'. In 1994, the case register contained information on 2029 persons. Per person a number of variables is known, like: sex, age, date of first contact, number of visits, marital status, educational level, source of income. A unique number is assigned to each person. This number serves as

an identifier of that person. This identification helps to tally the number of visits of that particular person. In the following analyses we use the frequencies of these visits in the year 1994 as a key variable. We postpone a more detailed discussion of this key variable. We will now discuss the estimators and their underlying assumptions in more detail.

Zelterman's and Chao's truncated Poisson estimators

Zelterman's (1988) and Chao's (1987) estimators, but see also Chao (1989) and Wilson and Collins (1992), can be applied on data generated by counts of individuals who have been seen once, twice and so on. In our study, the health centre's administrator tallies the number of treatment episodes brought by persons who are on methadone maintenance. Persons who are never seen fall into the zero frequency class and are missing from the observed series of frequencies. Therefore, the frequencies of the visits are incomplete and are called 'truncated below one'. Naturally, the total population size equals the number of persons ever seen plus the number of persons never seen. The estimation problem, then, becomes to estimate the number of persons never seen from the truncated series of persons ever seen. Both Zelterman's and Chao's estimators are based on this idea and both assume that the observed series of frequencies follows a Poisson distribution which is truncated below one. Since the calculations are so easy, we give the equations. Zelterman's (1988) estimator of the unknown population size, est(n), is given by

$$\text{est}(n) = S / [1-\exp(-2f_2 / f_1)]$$

and Chao's (1989) estimator is given by

$$\text{est}(n) = S + f_1^2 / (2f_2)$$

where,

f_1	=	the number of persons falling in the first frequency class
f_2	=	the number of persons falling in the second frequency class
S	=	the sum of all frequencies

We refer the interested reader to the cited literature for the calculations of the 95% confidence intervals.

Note, both estimators are primarily based on the lower frequencies (f_1

and f_2). This emphasis on the lower frequencies classes makes sense. People seen rarely (only once or twice) are likely to bare a greater resemblance with persons never seen, then what would have been the case with persons seen very often. In addition, the emphasis on the lower frequency classes makes the estimators robust in the presence of 'heterogeneity', e.g. persons seen very often may form a different subgroup as compared to persons seen rarely. The influence of the persons often seen is weighted down in both estimators and therefore heterogeneity, if present, is likely to exercise a relatively small influence. Finally, emphasis on the lower frequency classes results in another bonus as well: both estimators are known to perform rather well even when we have few data (Chao, 1989).

Assumptions

To be valid, both estimators assume that

1. the population is 'closed'

2. the population of interest is homogenous (no heterogeneity across individuals)

3. the individual probabilities to be observed and re-observed are constant over time

We will now discuss each of these assumptions in turn.

Closure assumption

The first assumption, known as the 'closure assumption', asserts that the true population size, N, is unaffected by migration, birth and death during the period of interest. In this particular study, we have chosen a period of one year because we want to estimate the one year prevalence of opiate use in Rotterdam. We, therefore, must hope that the true population size is not too much affected by in-migration and out-migration. Keeping the study period short (say, one month) is one way of meeting the closure assumption. Evidently, it is hard to see how the population size of opiate users can change dramatically in a single month. Note, that a shorter period will result in fewer observations, but then again, both estimators are known to perform well when the data are sparse.

Homogeneity assumption

The second assumption -the homogeneity assumption- dictates that the probabilities of being observed and re-observed should not differ too much across groups of individuals. In theory this assumption should not cause too much worries. Both estimators are known to be fairly robust in the sense that both will *under*estimate the true population size in the presence of heterogeneity (Chao, 1989; Wilson and Collins, 1992). So, if heterogeneity is suspected, then one may reason that the estimates are lower bounds of the true population size. Alternatively, one may prefer to stratify the data-set and then carry out subgroup analysis on groups that are more homogeneous and finally pool these estimates into a single estimate of N. In the Results paragraph we will present both the stratified and the estimates for comparison.

Constant (re)capture probability

The third assumption -about constant capture probability- effectively denies the possibility that individuals show a behavioural response to the treatment they receive. Whatever their experience with the methadone treatment, their probability to become a second time or a third time visitor is assumed to depend on a constant individual probability of being observed one time, two times, three times and so on. Clearly, with respect to the data generating process this assumption is a worrying one. At any rate, we do not find it particularly realistic. Only a cynic would say that methadone maintenance is so ineffective that this assumption is not at risk of being violated anyway. Again, one way of dealing with this assumption is to keep the time period of interest short, and bring it down from one year to, say, a single month. The influence of a single month of methadone maintenance is likely to be small and this may help to decrease the behavioural response problem.

Interpretation of f_0

We need to elaborate a little more on the constant (re)capture assumption. The assumption of a constant (re)capture probability is fundamental to the idea that one can safely extrapolate from the observed series of f_1, f_2, ..., f_K to the group of persons falling into the zero frequency class, f_0. One way of understanding this assumption is to think in terms of persons having a latent probability -a propensity- of being seen as a visitor at a health service centre. Persons who have a latent probability of generating a contact with the health services are either actually seen at such a centre during a specific period, or not. So, one may picture the whole population of opiate users as consisting of three different segments, or, groups:

1. a group which has actually been seen (their latent propensity to generate a contact with an health service has become manifest in an observable event)

2. a group which has not *yet* been seen, but which has a non-zero latent probability of generating a contact with an health centre (in due time they may be observed as visitors of the health centre)

3. and, finally, a group which has a latent probability equal to zero (and this group of users will *never* be seen at the health centre).

We will give examples of each of these groups. Clearly, the first group consist of visitors. These visitors have been observed at least once and so they are easy to recognize. The second group could be perceived as belonging to the target group of the methadone programme, but for some or another reason they have not entered the methadone programme yet. Therefore, we could call them 'potential clients' of the methadone programme. They are, for example, opiate users who may feel they need help, but have reservations about the kind of treatment which is on offer. The third group, are opiate users but, as said, have a latent probability equal to zero of becoming visitors of the methadone programme. We could think of them as persons who are, for example, sufficiently rich, or successfully engaged in crime, and therefore can sustain their habit without the need to go to a methadone programme.

Conclusions

From the previous section two important conclusions can be drawn: truncated Poisson estimators are only able to estimate the size of the group of persons who have a latent non-zero probability of being 'captured'. So we are, in principle, only able to estimate the size of the first and the second group combined, to the exclusion of the third group. Therefore, it is not safe to generalise research findings based on these estimators to the whole opiate using population. The other conclusion is that the size and composition of the estimable f_0 group is likely to be interesting from the perspective of those who manage the methadone programme: we are talking about their target group, their potential clients. This is interesting in terms of health service performance evaluation, unmet needs assessment and service demand estimation.

In sum, both estimators of Zelterman and Chao appear to be fairly realistic with respect to the underlying assumptions, but we are not

sure about the constant recapture assumption. The logistics of the data collection are easy to manage as only counts of visits are required. This is also an advantage with regard to privacy regulations. Finally, the estimators are computationally easy and these computations will therefore not result in appreciable costs.

Van der Heijden's et al (in progress) truncated regression models

Peter van der Heijden *et al* (in progress) developed three related models. We begin our description of these models with a brief outline, then continue our description with a discussion of each of these models and their assumptions. At the end of this section we present some conclusions.

Outline

Like Zelterman (1988) and Chao (1987, 1989) before him, Van der Heijden *et al* reasoned that a truncated Poisson (or related) distribution might be useful for estimating the unknown population size, f_0, from an incomplete series of observations f_1, f_2, ..., f_K. In addition Van der Heijden *et al* realised that a truncated Poisson regression analysis might be one step towards estimating f_0. By using a regression model as the core of his population size estimator, it can easily be extended as to include auxiliary information on the group of observed opiate users. This auxiliary information can be included in the regression model as covariates and this would help to explicitly model observed between-subject heterogeneity. As with all regression-type models, the need to enter covariates in the regression equation can be formally tested. Without covariates the truncated Poisson regression model reduces to an 'intercept only' model and this gives rise to a model which is called the 'homogeneous truncated Poisson model'. So, the simplest estimator is based on a truncated Poisson regression model without covariates. When covariates are entered into the truncated Poisson regression equation we get a more complex model. These covariates help to account for observed heterogeneity. In the presence of observed heterogeneity, the truncated Poisson model with covariates is a better choice because it will result in less underestimation of the true population size. There is one problem left. Some heterogeneity might still be unobserved, e.g. the covariates may not be able to capture all observed between-subject heterogeneity and some amount of unobserved heterogeneity may still be present and bias results. However, if an additional 'dispersion parameter' is entered into the regression equation, then we get the most complex model, known as the 'truncated Negative Binomial regression model'. This model can deal with observed heterogeneity (through the covariates) and, in addition to that, with

unobserved heterogeneity (through the dispersion parameter). Whether or not it is necessary to model dispersion can be formally tested. So, a researcher shouldn't be in doubt about what model should be selected.

Homogenous truncated Poisson model

The first model makes no use of auxiliary information on the respondents and assumes that all respondents have the same probability of being seen never, once, twice, et cetera. For this reason it is appropriate to call this model 'homogeneous truncated Poisson model'. Knowing that the population of opiate users is heterogeneous (cf. EMCDDA, 1997), we find the homogeneous truncated Poisson model not particularly relevant for our purposes and will use it only as an illustrative bench mark.

Truncated Poisson model with covariates

As said, a truncated Poisson regression analysis forms the core of Van der Heijden's *et al* (in progress) population size estimation model and covariates can be included in the model as a matter of course. In this way observed between-subject heterogeneity, if any, can be taken into account. The model allows formal testing whether or not sources of heterogeneity -and interactions between these sources- should be included in the model. Once an appropriate model has been fitted to the data, one can compute the probability that a person with a set of covariates has never been observed. Finally, summing the numbers of persons that have been seen plus the estimated numbers of persons never seen produces an estimate of N. The corresponding estimation of the Poisson coefficients and computations for est(N) are somewhat involved and can not be detailed here. The interested reader is referred to Van der Heijden *et al* (in progress).

The model allows that for each group of individuals who share the same characteristics -in terms of the covariates- the unobserved number of persons can be computed. This is in a sense equivalent to the subgroup analysis on a stratified data-set that can be carried out using Zelterman's and Chao's estimators, but it is more efficient since we do not have to assume that all possible interaction terms are present.

Clearly, if no covariates are used, then the truncated Poisson regression model reduces to the homogeneous truncated Poisson model. It is advised to formally test the appropriateness of this simplification. If this is indeed the case, than, naturally, a simple model is preferred over an unnecessary complex one.

Truncated Negative Binomial regression model

The truncated Negative Binomial estimator with covariates is based on a similar strategy, only this time a truncated Negative Binomial regression model is used. The latter is a generalisation of Poisson models. Occasionally, there is an advantage in this approach. Under the Poisson model it is assumed that the variance equals the Poisson mean and this assumption may be violated. Under the Negative Binomial model a separate parameter, δ, is estimated which captures variance in excess of what is expected under the Poisson model, i.e. over-dispersion. So, the Negative Binomial regression model also accounts for that part of the heterogeneity that could not be modelled explicitly using the covariates. The presence of dispersion can be formally tested, and if present, can be accounted for in the model.

Assumptions

The three estimators assume that

1. the population is 'closed', and

2. the individual probabilities to be observed and re-observed are constant over time.

Absence of the homogeneity assumption

Note, the homogeneity assumption has been relaxed now. The presence or absence of observed and unobserved heterogeneity can be formally tested and if present can be accounted for by choosing the appropriate model. Failing to account for heterogeneity will result in underestimation of the true population size.

Closure assumption

The closure assumption remains in full force since in- and out-migration of the population can not be accounted for by the model. The effect of migration on the true population size is smaller, of course, during a brief period of time and so it might be advisable to keep the time period under consideration short with respect to the population dynamics. Deciding what 'short' means, falls outside the reign of statistics and remains a matter of intelligent judgement.

Constant recapture probability

The 'constant capture probability over time' assumption is still the most worrying one when studying opiate users who are observed in the context of a methadone maintenance programme. The very treatment they receive may exercise an influence on the probability of being seen again. In addition, the f_0 group is only a segment of the unobserved population -recall, the segment which has a non-zero latent probability of entering the methadone programme.

Conclusions

In the presence of observed and unobserved between-subject heterogeneity, regression-type estimators are possibly an improvement over Chao's and Zelterman's estimators. The choice between the models can be based on formal tests of significance. Also, the regression-type estimators produce some insight in the factors which are associated with the frequency of contacts, and this may be interesting in its own right. Further, the presence or absence of interactions between the covariates can be formally tested, which is an improvement over subgroup analyses. However, we must pay a price for these benefits. More data, in the form of covariates on all observed individuals, are required. Finally, the estimation procedures are 'expensive' in the sense that they are complicated and non-standard statistical software must be used.

Hypotheses

We have no way of knowing the true number of opiate users in Rotterdam. There is, however, another estimate for the number of opiate users in Rotterdam in 1994 (Wiessing *et al.*,1995). This estimate is based on the multiplier method and indicates a population size in the range of 2400 - 3500 persons (three-month period prevalence of 2400, extrapolated to a one-year prevalence of at least 3500). We must bear in mind that this is an estimate and that the true number is, in fact, unknown. In the absence of a gold standard it is impossible to say how accurate a particular estimator is. All we can do is compare one estimator with an other and use these comparisons as the base for some inferences about the 'quality' of the estimators. To that end we will formulate several hypotheses. The hypotheses, or rather expectations, are:

1. Both Zelterman's and Chao's estimators will produce about the same estimates of N. This is a well known result and it is only likely that it will be reproduced here.

2. Both Zelterman's and Chao's estimators will produce higher estimates than the estimate of N based on the homogeneous

Poisson estimator. This hypothesis is based on the assumption that heterogeneity in the population will not severely affect Zelterman's and Chao's robust estimators, but will result in an underestimation of the true N by the homogenous Poisson estimator because this estimator can not handle heterogeneity.

3. The pooled estimators of Zelterman and Chao will be about the same and will be higher than the respective unpooled estimates. This expectation is motivated as follows: the pooled estimators will capture heterogeneity in the population better than the unpooled ones, and this, in turn, will result in less underestimation.

4. The truncated Poisson regression estimator will be higher than the pooled estimators of Chao and Zelterman respectively, because it is better in accounting for heterogeneity than the pooled estimators of Chao and Zelterman.

5. Finally, the truncated Negative Binomial regression estimator is likely to come up with the highest estimate since it is not only able to deal with observed heterogeneity but also with unobserved heterogeneity.

We will now present the results. With the results in hand we will return to these hypotheses in the Discussion paragraph.

5.3 Results

Results from Zelterman's and Chao's estimators

For the 1994 RODIS-data we obtain an estimate of 3727 opiate users in Rotterdam when using Zelterman's (1988) Truncated Poisson estimator. The 95% confidence interval ranges from 3497 to 3990. Table 5.1 gives estimates for the population when it is stratified by sex and age groups (10 year bands).

Table 5.1: **Observed numbers, estimated numbers within 95% CIs using Zelterman's (1988) Truncated Poisson estimator of N by sex and age groups.**

	Males				**Females**			
	obs(n)	**low**	**est(n)**	**high**	**obs(n)**	**low**	**est(n)**	**high**
15-24 yrs	94	134	177	252	65	77	101	148
25-34 yrs	708	1138	1258	1408	306	448	518	614
35-44 yrs	585	961	1080	1233	153	279	368	540
45-54 yrs	90	142	195	311	20	29	61	∞

55-64 yrs	8	6	11	35	0	0	0	0
all ages	1485	2524	2716	2940	544	894	1011	1162

With regard to Table 5.1 we like to make the following remarks. When all est(N) are summed over all strata, then the sum total is 3769, which is only marginally higher than the direct estimate of 3727 and well within the confidence interval of 3497 - 3990. So the pooled estimate does not differ significantly from the direct estimate. This, in part, reflects the robustness of the estimator in the presence of heterogeneity. It may also reflect that SEX and AGE are variables which are not strongly associated with capture-recapture frequencies of opiate use. Note that in one instance (the males in the 55-64 age group) the lower bound of the 95% CI is lower than the observed number. Conceptually, this does not make sense, but is the result of an obs(N) which is too small for the asymptotic estimation of the 95% CIs. Note also that in an other occasion the upper bound explodes into infinity (the females in the 45-54 year band). These freak results are known to happen when using Zelterman's CIs.

Characteristically, Chao's (1987) estimator compares well with those of Zelterman, but her CIs behave better. Using her estimate we find 3565 opiate users in Rotterdam in the year 1994 as a best guess. This is well within the 95% CI of Zelterman's (1988) estimator. The 95% CI of Chao's estimator is 3348 - 3818, which has, of course, a substantial overlap with Zelterman's 95% CI. In table 5.2 we present the observed and estimated numbers under Chao's model.

Table 5.2: **Observed numbers, estimated numbers within 95% CIs using Chao's (1989) Truncated Poisson estimator of N by sex and age groups.**

	Males				**Females**			
	obs(n)	**low**	**est(n)**	**high**	**obs(n)**	**low**	**est(n)**	**high**
15-24 yrs	94	130	167	241	65	83	104	151
25-34 yrs	708	1080	1192	1338	306	441	506	601
35-44 yrs	585	929	1041	1189	153	274	359	504
45-54 yrs	90	138	187	285	20	33	73	226
55-64 yrs	8	8	11	33	0	0	0	0
all ages	1485	2402	2583	2799	544	875	985	1131

The pooled estimate under Chao's model is 3640 persons which is slightly higher than her direct estimate of 3565, but stays well with in the latter's CI-bounds. This, again underscores that both Zelterman's and Chao's estimators are fairly robust in the presence of heterogeneity. Table 5.2 shows that Chao's CIs do not produce lower bounds lower than the observed number, while her upper limits do not explode into infinity as Zelterman's CIs sometimes do. Finally,

note that Chao's 95% CIs are symmetric around est(N) while Zelterman's are not; and her CIs are usually more narrow than Zelterman's.

Van der Heijden's et al (in progress) estimators

As outlined in the methods paragraph we first fitted a truncated Poisson model. Initially, variables like SEX (1=male, 0=female), MAR (1=married, 0=not married)), DUT (1=Dutch nationality, 0=otherwise), AGE (in years), INC (1=income from work, 0=otherwise), TOG (1=living together with a partner, 0=otherwise) and SUR (1=of Surinam origin, 0=otherwise) were included in the model. Since not all terms turned out to be significant, a more parsimonious model was obtained by only including, SEX, DUT, AGE, TOG and SUR. In the process of model specification it was also checked whether or not interaction terms should be added to the model and we checked if a quadratic term for AGE had to be included in the model. As it turned out, the simple model without interaction- and quadratic terms fitted well. Under this model (Model II, in Table 3) N was estimated to be 2991 persons. We must note here that this model is inappropriate for these data, since the dispersion parameter is significant. It does, therefore, not come as a surprise that the Truncated Poisson estimator with covariates underestimates N for these data relative to Zelterman's and Chao's estimators.

Table 5.3: **Coefficients and the estimated population size, est(N) of the Homogeneous truncated Poisson model (I), the truncated Poisson regression model (II) and the truncated Negative Binomial regression model (III) (all coefficients significant at p<.05; obs(N)=2029; δ is the dispersion parameter under model III).**

	Model I	**Model II**	**Model II**
Cons	0.16	0.37	-0.14
SEX		0.19	0.22
DUT		0.22	0.26
TOG		0.11	0.14
SUR		0.27	0.32
AGE		-0.02	-0.02
δ		-	0.98
Est(N)	2937	2991	5006

The homogeneous truncated Poisson model without covariates (Model I, in Table 5.3) is even more inappropriate for these data and produced a low est(N) of 2937.

The same set of covariates were fitted in the truncated Negative Binomial regression model (Model III, in Table 3) and the parameter that captured the dispersion, δ, turned out to be significantly different from zero (δ=.98; T=3.98; p=.000). Under this truncated Negative Binomial model with covariates N is estimated to be 5006, which is the highest estimate found so far.

5.4 Discussion

Summary of results

Table 5.4 summarises the point estimates, est(N), and the related estimated 1-year prevalences, est(p), as a percentage of the Rotterdam population, 345675 in the age range 15 - 54 years. We will be using this age-category because 99% of the observed opiate users are included in this category. Because Rotterdam attracts opiate users from outside the city, the estimates are likely to overestimate.

At this point we remind the reader that in the study of Wiessing *et al* (1995) the number of opiate users in Rotterdam in 1994 was estimated to be 2400 - 3500, which is somewhere in the lower range of our own estimates. Still, we observe some variance across the estimates.

Table 5.4: **Estimates of N and the prevalence of opiate use by model (Rotterdam 1994).**

Model	**est(N)**	**est(p) (%)**
Zelterman's (1988) Truncated Poisson estimator	3727	1.08
Zelterman's (1988) pooled estimator	3769	1.09
Chao's (1989) Truncated Poisson estimator	3565	1.03
Chao's (1989) pooled estimator	3640	1.05
Homogeneous Truncated Poisson estimator	2937	0.85
Truncated Poisson estimator + covars	2991	0.86
Truncated Negative Binomial estimator + covars + δ	5006	1.45

The hypotheses revisited

1. We find support for the hypothesis that Zelterman's and Chao's estimators produce about the same results. We obtained 3727 and 3565 respectively. Note also that the 95% CIs show substantial overlap: 3497-3990 and 3348-3818. From this we conclude that Zelterman's and Chao's estimators do indeed produce about the same estimates of N.

2. We also find support for the hypotheses that both Zelterman's

and Chao's estimators produce higher estimates than what will be obtained under the homogenous truncated Poisson model. We obtained 3727 ≈ 3565 > 2937. Since the latter is well below the lower 95% CI limits of Zelterman's and Chao's estimators, we accept this hypothesis. This supports the idea that when we must assume homogeneity, as one is forced to when using the homogenous truncated Poisson estimator, then this leads to an estimate which is too low when, in fact, there is heterogeneity.

3. We find no support neither to accept nor to reject the hypothesis that the pooled estimators of Zelterman and Chao result in higher estimates than the unpooled ones. We found 3769 and 3640 for the pooled estimators, and we found 3727 and 3565 for the unpooled ones. As expected, the pooled estimators of Zelterman and Chao are about the same and both are higher than the respective unpooled estimates. This supports the idea that the pooled estimators somewhat better capture heterogeneity in the population and this, in turn, results in less underestimation. However, we must also point to the fact that the pooled estimates do not fall outside the 95% CIs of the unpooled estimates. So, the difference, if present, is in the expected direction, but for these data we have no statistically significant finding. It is also clear that, for these data, Chao's estimator is lower than Zelterman's, even to the extent that her pooled estimator is lower than his unpooled estimator. So, the support for our expectation is only tentative.

4. In contrast to our expectation we find that the truncated Poisson regression estimator is lower, not higher, than the pooled estimators of Zelterman and Chao: 2991 < 3769 ≈ 3640. This is indicative that this estimator does *not* deal with heterogeneity any better than the pooled estimators of Zelterman and Chao with respect to these data. Here we like to recall what has been said in the methods paragraph: the truncated Poisson regression model can not deal with that part of heterogeneity which has not been modelled through the covariates. The presence of significant dispersion has biased results here.

5. Finally, we expected that the truncated Negative Binomial regression estimator would come up with the highest estimate of N, and this expectation is fully supported. With regard to hypothesis 4 we note again that the dispersion parameter, δ, turned out to be significantly different from zero (δ=.98;

T=3.98; p=.000).

Conclusions and caveats

In the absence of a gold standard it is impossible to say what estimator is the best in terms of its outcome. Having said that, some conclusions can based on the confirmed and rejected hypotheses, while other conclusions can be based on the statistical tests which have been carried out during the model specification process.

- With regard to the data we used, the homogenous Poisson estimator would not be our favourite. It can not account for the observed and unobserved heterogeneity which is present in this data-set. As a consequence it produces an outcome too low relative to the other estimates. This manifest underestimation, in turn, is indicative of the importance of estimators that do better cope with heterogeneity -when, of course, heterogeneity is present.

- If we were to make a choice, we would prefer Zelterman's, or for that matter, Chao's estimator. Both estimators (and their pooled counterparts) produce about the same results and are also in line with the study of Wiessing *et al*. Relative to the truncated Negative Binomial regression model both estimators appear to underestimate somewhat, but this is a well known result. Both estimators have additional advantages. They are easy to calculate -although the 95% CIs of Zelterman require a computer to solve an equation iteratively. More importantly, they are based on readily available data. In principle, only counts of observations of opiate users by a single agency are needed. If covariates are present, then they can be used to stratify the data-set. Further, both estimators are not based on assumptions that are totally unrealistic, but they do assume a constant (re)capture probability of each individual over time. With respect to the interpretation of the f_0 group, we feel that we can safely generalise only to the group of opiate users who have a non-zero latent probability of becoming a client of the methadone programme.

- We expected that the truncated Poisson estimator with covariates would be an improvement over Zelterman's and Chao's estimators, but the results of this study show that this estimator is inappropriate with respect to the analysis of this particular data-set. Unobserved heterogeneity is present in the data, and under such conditions the truncated Poisson

regression model is simply not the best choice. From this we conclude that it is advisable to use the truncated Negative Binomial regression estimator and test whether or not a dispersion parameter must be included in the model -and not to rely solely on the truncated Poisson regression estimator.

- The truncated Negative Binomial estimator with covariates produced results which were in line with our expectations. This gives strength to the idea that this estimator is a suitable candidate for the analysis of these data. As compared to the estimators of Zelterman and Chao it came up with an higher estimate, which was expected. Its draw back is its computational complexity and the fact that it needs more data than Zelterman's and Chao's estimators.

In conclusion, population size estimators are important because they can play a valuable role in research fields such as unmet needs assessment, health service demand estimation and health service performance evaluation. With these purposes in mind we feel that Zelterman's and Chao's estimators have some appeal, and so has Van der Heijden's *et al* truncated Negative Binomial estimator with covariates. Choosing between these estimators is mainly a matter of data availability and should further be guided by a justifiable concern about heterogeneity. The interpretation of the estimable f_0 group should in all likelihood be confined to a segment of the opiate using population which has a non-zero latent probability of being seen at the methadone programme. How, a possible violation of the constant recapture assumption has affected the results, remains an important but unresolved issue.

More generally, it is not a matter of saying which estimator is good and which one is bad. All estimators are absolutely perfect -that is, just as long as the data do not violate their assumptions. So, the focus should always be on how well an estimator and the data match. This, in turn, urges us to give *both* the assumptions of competing estimators *and* the data-generating process a long and hard look. The population dynamics and the individual careers of opiate users have generated the data in the first place and, naturally, we would prefer estimators which are based on assumptions that are in some way compatible to, or isomorphic with, the data-generating process. Finally, all the estimators which have been discussed, share a single feature: they are all based on a single data-set. At times this can be an advantage, because data from a single source are likely to be more readily available than data from multiple sources. This advantage is also associated with a draw back: we can only generalise to a distinct 'inbound' segment of the opiate using population and not to the whole

population. So, it all comes down to answering three interrelated questions: -what precisely do we want to know? -what kind of data can we get? -and what estimator best helps to find an answer to our research question, given these data?

Acknowledgement

This study has been made possible by the generous financial aid of the *Municipal Health Service of Rotterdam*. We also received support from the *European Monitoring Centre for Drugs and Drug Addiction* (Lisbon), specifically in the form of an invitation to participate in the *EMCDDA Research Group on Local Level Prevalence Estimation* (Glasgow).

References

Brecht, M-L. and Th. D. Wickens Application of multiple capture methods for estimating drug use prevalence. Journal of Drug Issues (1993) 02 229-50.

Chao, A. Estimating animal abundance with capture frequency data. *Journal of Wildlife Management* (1988) 52 295-300.

Chao, A. Estimating population size for sparse data in capture-recapture experiments. *Biometrics*(1989) 45 427-38.

Cormack, R.M. and P.E. Jupp Inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika* (1991) 78 (4) 911-6.

EMCDDA and the Council of Europe *Estimating the Prevalence of Problem Drug Use in Europe. EMCDDA scientific Monograph Series No 1*. Lisbon: EMCDDA, 1997.

Korf, D.J., S.A. Reijneveld & J. Toet Estimating the Number of Heroin Users: A Review of Methods and Empirical Findings from the Netherlands. *The International Journal of the Addictions*, (1994) 29 (11), 1393-1417.

Khorazaty, M.N. el; P.B. Imrey; G.G. Koch and H.B. Wells Estimating the total number of events with data from multiple record systems: a review of methodological strategies. *International Statistical Review*. (1977) 45 129-57.

Pollock, K.H. Modelling capture-recapture and removal statistics for estimation of demographic parameters of fish and wildlife populations: past, present and future. *Journal of the American Statistical Association*. (1991) 86 (413) 225-38.

Seber, G.A.F. *The estimation of animal abundance and related parameters*. London: Charles Griffin, 1982.

Seber, G.A.F. A review of estimating animal abundance. *Biometrics*. (1986) 42 267-92.

Seber, G.A.F. A review of estimating animal abundance II. *International Statistical Review*. (1992) 60 129-66.

Van der Heijden, P.G.M.; D. Zelterman; G.B.M. Engbertsen and J. van der Leun Estimating the number of illegals in the Netherlands with the truncated Poisson regression model. (in progress).

Simeone, R.S.; W.T. Nottingham; L. Holland Estimating the size of a heroine using population: An examination of the use treatment admissions data. *The International Journal of the Addictions*. (1993) 28 (2) 107-28.

Wickens Th. D. Quantitative methods for estimating the size of a drug using population. *Journal of Drug Issues*. (1993) 2 185-216.

Wiessing, L.G.; J. Toet; H. Houweling *et al*. *Prevalentie en risicofactoren van HIV-infectie onder druggebruikers in Rotterdam.* Bilthoven, RIVM / Rotterdam, GGD Rotterdam, 1995

Wilson R.M. and M.F Collins Capture-recapture estimation with samples of size one using frequency data. *Biometrika* (1992) 79 543-53.

Zelterman, D. Robust estimation in truncated discrete distributions with application to capture-recapture experiments. *Journal of Statistical Planning and Inference*. (1988) 18 225-37.

6 Setúbal

ESTIMATING THE NUMBER OF OPIATE USE IN SETÚBAL

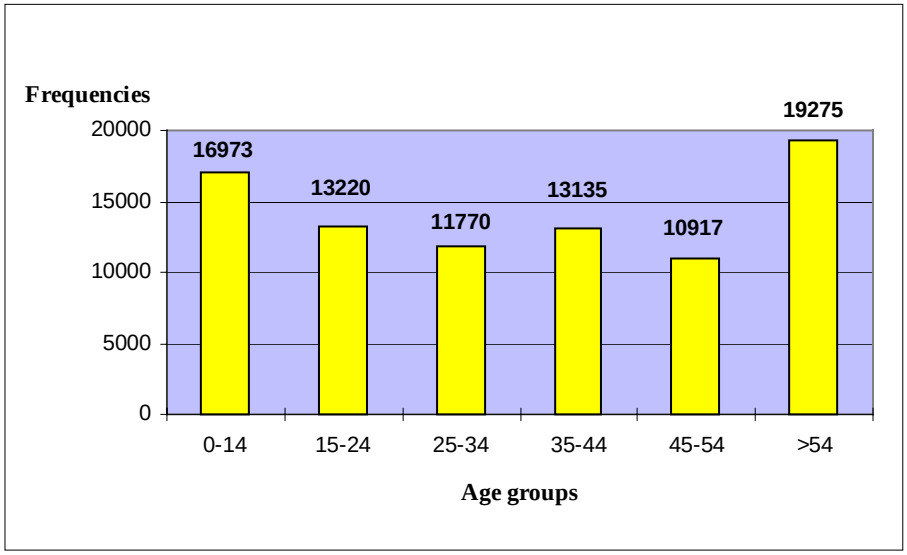
Sofia Freire, Observatório VIDA, Lisbon

Maria Moreira, Observatório VIDA, Lisbon

6.1. Demographic data concerning Setúbal county

As stated in the1991 population census[1], there are 85292 individuals in the Setúbal city. From these, 49043 are 15-54 years old.

Figure 6.1: **Distribution of group age frequencies.**



The individuals from this county have a low educational level as well as a low income level. Only 25.2% have completed primary school level and 13.6% are illeterate; only 3.2% have a university degree. They work mainly at the industrial production and in manufacturing (20.3%), at security services, house services or similar (16.7%) or are non-qualified workers working in agriculture, industry, trade or other services (16.6%). The unemployment rate is high: 12.2%.

6.2. Data concerning drug misuse in the Setúbal county

The Setúbal county has a significant drug problem. Concerning data from the Ministry of Health[2], the Setúbal county had in 1995 one of the most serious problems related to AIDS cases. From 1983 to 1995, 360 new AIDS cases were reported in this county. This number represents 12.9% of all the reported cases. Moreover, in 1995, 86 people died from AIDS; this number represents 9.2% of all this cases.

[1]Census 91 - Lisboa e Vale do Tejo, Instituto Nacional de Estatística, 1993

[2]Mininstério da Saúde - *Elementos Estatísticos de Saúde: 1995.* Lisboa, Departamento de Estudos e Planeamento da Saúde, 1997

In relation to hepatitis[3], a study made on 379 IV drug users, in 1995 in the Setúbal's specialised public drug treatment centre, shows that 85% of these drug users tested positive for the hepatitis C. The same study shows that 20% of these drug users tested positive for the HIV.

According to data from the Ministry of Justice[4] , there were 464 presumed offenders in 1995. This number represents 7.3% of the overall presumed offenders in Portugal in this year: only 3 counties (Lisbon, Oporto and Faro) presented higher values.

6.3. Description of the sources used in the analysis.

In order to make this study 4 nominal samples were collected. All those samples concerns individuals who live in the Setúbal city:

- drug addicts who were in treatment during 1996 in a specialised public drug treatment centre;
- drug addicts who were in treatment during 1996 in Setúbal General Health Centres';
- drug addicts who were condemned during 1996;
- drug addicts who were receiving in-treatment in Setúbal Hospital.

All the collected names were transformed, so that the drug addicts could remain unrecognised. The names were transformed according to the Group Pompidou Protocol and the final code was also used as the matching criteria.

The code sugested by the Pompidou Group is:

- 3rd letter of the first name
- number of letters of the first name
- 3rd letter of the last name

[3] Godinho et al. (1996) Comportamentos de risco de doenças infecciosas. Avaliação da população rastreada nos CATs de Setúbal e de Almada in *Toxicodependências*, ano 2, nº3

[4] Ministério da Justiça - *Droga: Sumários de Informação Estatística: 1995*. Lisboa, GPCCD, 1996.

- number of letters of the last name

- day of birth

We experienced some difficulties in order to collect all these data. At the end of the study we had only one complete sample - the one concerning the specialised public drug treatment centre, and a partial sample from the health centres.

We weren't able to collect any other sample, mainly because data wasn't easily available, either due to legal reasons (the data concerning condemned drug addicts), or because it wasn't available at all (the data from Setúbal Hospital wasn't ready, because the drug addiction diagnosis was still being codified).

In order to get an estimate of the local prevalence, we worked with the samples from the Health Centres and form the specialised drug treatment centre. As it wasn't possible to get a police sample, the estimate is more likely to refer to problematic opiate users.

Table 6.1: **Description of the data from the three sources.**

Source	Number	% Male	Average Age
CAT Semester 1	191	81.7	28.7
CAT Semester 2	219	85.4	28.3
CAT Total	313	83.4	28.2
Health Centre	40	80.0	29.9
All individuals	339	82.9	28.2

The overall sample consists of 339 heroin users - 313 from the specialised drug treatment centre and 40 from the Health Centres. Most of the individuals were males and aged 15-34 years old. As such 82.9% of all the individuals were males and 83.8% of all the individuals were young people, i.e., aged 15-34 years old.

Table 6.2: **Description of the data from the three sources.**

	Males		Females	
	15-34 years	**35-54 years**	**15-34 years**	**35-54 years**
CAT semester 1	127	29	29	6
CAT semester 2	153	34	26	6
Health Centre	27	5	7	1
Total of contacts	307	68	62	13
Total of individuals	234	47	50	8

6.3.1. The specialised drug treatment centre:

There is a survey on drug addiction[5], made on a national level, that aims at gathering information about demographic data, use patterns and risk behaviours concerning drug addicts who are in treatment in all specialised public treatment centres (CAT).

During 1995, it was collected information concerning 980 individuals. The great majority was male (80.9%) and aged 20 - 24 years old (31.0%) and 25 - 29 years old (28.0%). The average age was 27.4 years old. The great majority consumed heroine as the main drug of abuse (95.7%) and 41.0% were IV users.

The sample from the drug treatment centre (T= 313) consisted of 191 heroin users who looked for help in the first semester (CAT semester 1) and of 219 in the second semester (CAT semester 2). There were an overlap between these two samples: 97 individuals looked for help during first and second semester. The majority of these heorin users were male (81.7% in the first semester and 85.4% in the second semester). The average age was 28.7 years old in the first semester and 28.3 years old in the second one.

6.3.2. The Health Centres:

Only 9 of the 48 GPs working in the Setúbal Health Centres were able to collect the data requested. This was mainly due to the inavailability of the data: data isn't computarizesd or centralized, so it is rather difficult for the GPs to select a sample of the drug addicts they are consulting. Nevertheless, the GPs come from all the Health Centres in the Setúbal city, so all the city had a similar coverage.

The data collected concerns heroin users who looked for help due to health problems. The sample from the Health Centres consisted of 40

[5] Félix da Costa, N., et al. - Tratamento da toxicodependência: estudo sagital de 1995 in *Toxicodependências* , ano 2, nº3, Lisboa, SPTT, 1996.

heroin users, most of who were males (80.0%) and were 29.9 years old (average).

6.4. Results:

As we can see in table 6.3 there was a good degree of overlap between the sources.

Table 6.3: **Overlaps between Helath Centre data and 2 semesters of a specialised treament centre.**

		Semester 1			
		Present		Absent	
		Semester 2			
		Present	Absent	Present	Absent
Health Centres	Present	6	2	6	26
	Absent	91	92	116	-

According to the log-linear analysis, the model that assumes an interaction between semester 1 and semester 2 is the most suitable one, as the deviance is small; 2.14 against 2 degrees of freedom. (Table 6.4)

Table 6.4: **Results of log-linear analysis.**

				Hidden Population		**Total Population**	
Model	χ^2	**df**	**p-value**	**N**	**CI**	**N**	**CI**
S1+S2+H	21.82	3	0.00	155	107-214	494	446-553
S1xS2	2.14	2	0.34	555	281-1084	894	620-1423
S1xH	14.28	2	0.00	135	91-190	474	430-529
S2xH	17.65	2	0.00	138	92-194	477	431-533

The best estimate, according to this model, is 555 (281 is the minimum value and 1084 the maximum one of the 95% Confidence Interval). According to this estimate, and considering that the known population is formed by 339 individuals, the overall population of problematic heroin users, aged 15 to 54 years old, in Setúbal city will be 894 (95% CI 620 - 1423. The estimate represents 1.82% of the overall inhabitants in Setúbal city, aged 15 to 54 years old (Table 6.5).

Table 6.5: **Rates per population aged 15 to 54.**

	Hidden Opiate Users	Total Opiate Users	
	Number	Number	Rate (%)
Estimate	555	894	1.82
Lower	281	620	1.26
Upper	1084	1423	2.90

If we consider the stratify analysis, we can see that all confidence intervals seem reasonable and that the obtained estimates seem to reflect the overall known picture of drug addict population in Portugal (Tables. 6.6 and 6.7): the males group aged 15 to 34 years old are the most representative group, as it would be expect according to many indirect indicators, namely, data from courts.

Indeed, data from the Ministery of Justice[6] show that, in 1995, from the 3012 presumed offenders, 77.6% were heroin users, 68.9% were males and that 50.6% aged 21-29 years old and 30.6% aged 30-44 years old.

Table 6.6: **Summary of analysis assuming model with interaction between S1 and S2 for each stratified group (2 df).**

					Hidden Population		**Total Population**	
Sex	**Age**	χ^2	**p**	**known**	**N**	**CI**	**N**	**CI**
Male	15-34	1.83	0.40	234	414	179-951	648	413-1185
Male	35-54	2.78	0.25	47	30	1-145	77	48-192
Male	15-54	1.10	0.58	281	415	194-864	696	475-1145
Female	15-34	2.20	0.33	50	258	36-4500	308	86-4550
Female	35-54	1.53	0.47	8	0	0-1	8	8-9
Female	15-54	2.00	0.37	58	150	28-957	208	86-1015
Both	15-34	2.93	0.23	284	600	281-1301	884	565-1585
Both	35-54	3.79	0.15	55	25	1-109	80	56-164
Both	15-54	2.14	0.34	339	555	281-1084	894	620-1423

[6] Ministério da Justiça - *Droga: sumários de informação estatística (1995).* Lisboa, Gabinete de Planeamento e de Coordenação do Combate à Droga, 1996

Table 6.7: **Rates per stratified population.**

Sex	Age	Estimated Population	Opiate Users		Confidence Interval (%)	
			Number	Rate (%)	Lower	Upper
Male	15-34	12,250	648	5.29	3.37	9.67
Male	35-54	11,750	77	0.66	0.41	1.63
Male	15-54	24,000	696	2.90	1.98	4.77
Female	15-34	12,750	308	2.42	0.67	35.69
Female	35-54	12,250	8	0.07	0.07	0.07
Female	15-54	25,000	208	0.83	0.34	4.06
Both	15-34	25,000	884	3.54	2.26	6.34
Both	35-54	24,000	80	0.33	0.23	0.68
Both	15-54	49,000	894	1.82	1.27	2.90

Although, the confidence intervals are reasonable and the estimate for the overall population seems trustful, there aren't any other studies of this kind in Portugal to which we can compare this figure. Therefore, it is difficult to make statements about the degree in which it reflects the reality.

7 Toulouse

ESTIMATES OF OPIATE USE IN TOULOUSE CONURBATION IN 1995

Dr Pierre-Yves Bello, Observatoire Régional de la Santé en Midi-Pyrénées

Funded by the Observatoire National des Drogues et des Toxicomanies

7.1 Methods

Our area of work was the city of Toulouse which is the capital of the Midi-Pyrénées region (population 2,460,663) and of the Haute Garonne departement (population 925,000). Toulouse has a population of almost four hundred thousand, with nearly three hundred thousand more inhabitants in the suburbs. The population of the Toulouse conurbation is as follows :

Table 7.1: **Population of Toulouse conurbation by sex and age group in 1990.**

	0-14	**15-24**	**25-34**	**35-44**	**45-54**	**>54**
Males	56000	57400	57100	49800	33200	63100
Females	52400	60000	56600	51000	34400	79300
Total	108,400	117,400	113,700	100,800	67700	142,400

We worked on data corresponding to six months of 1995 : May to October.

We contacted all the services taking care of drug users and we asked them if they would agree to participate in the study. Finally we were able to work with ten sources. We would have liked to have carried out a prospective collection of data in all the sources, but it was not possible, so collection was made in part prospectively and in part retrospectively.

We created a questionnaire for this study . Identification variables included were surname initial, name, birthday, sex, and other variables like the residency in the Toulouse conurbation, working, housing, drug use characteristics (main product, main way of use, length of drug use, use of injection), serologic status (HIV, HCV, HBV).

All data were collected on Epi-Info.

Overlaps were identified manually. We sorted our list in three different way :

Surname initial, name birthday

name, surname initial, birthday

birthday, surname initial, name

We created a list of people with presence/absence in each of the ten different sources.

We then selected people who were opiates users and we eliminated people who were not resident in Toulouse conurbation. It is on this selected population that we started the analysis process.

We estimated population size by classical capture-recapture technique for each pair of sources. We also estimated odds-ratio and its confidence interval between sources using Wittes's technical. Thus we had an idea of level of dependency between sources. Then we made fusion of data between the positive dependent sources. We worked with a step by step procedure. Progressively we obtained three groups of sources :

One repressive group including police and jail data

One low threshold group including Intermede data

One socio-sanitary group including the seven other sources

We carried out the log-linear modelling with BMDP 4F program. We used a backward procedure, starting from a saturated model and deleting non informative interaction. When we worked with four groups of sources or with three groups of sources and a variable of stratification (like sex or age group), we preferred to use a forward procedure. We also calculated Akaike criterion.

The BMDP handbook explains that adding 0.5 to each cell when cell frequencies are small permit to compensate inflation in chi square that might occur. We used this option.

7.2 Agencies:

We were able to work with ten different agencies within the Toulouse conurbation, usually working with opiates users. We made a short description of the structures and of the collection of data.

CHS : Centre Hospitalier Spécialisé Gérard Marchant is the psychiatric public hospital of Toulouse. There are four wards on eight which are working more specifically with drug users mainly for maintenance/reduction therapy. Data were collected retrospectively

by mean of the medical information department. Cases were selected using ICD 10 codes.

Intermède (association Clémence Isaure) : is a low threshold structure. It opens each afternoon. Drug users can drink a coffee, wash their clothes, have a shower, exchange syringes, meet a social assistant. During one specific afternoon they can also meet a physician, a nurse or a lawyer. It is an important place for more marginal drug users. It permits drug users to initiate contact with other agencies in Toulouse. Data was collected retrospectively using questionnaire of first encounter with users of Intermède.

Joseph Ducuing Hospital : this is a general hospital which has been developing more specific work with drug users for more than 14 years. This work is linked with other agencies such as Clémence Isaure, Oc Drogue, Association accueil toxicomanie in an association named GRAPHITI. It is also a member of the Toulouse hospital-city net on drug use named Passages. Data was collected prospectively in the emergency ward and in the outpatient ward.

Association Accueil Toxicomanie (AAT) : is an association working with drug users by means of therapeutic work and art workshops (painting, theatre, music, writing). It is a member of the GRAPHITI association. Data was collected prospectively.

Intensive care unit of Rangueil hospital : is a ward of one of the two university hospitals of Toulouse. It receives overdoses. The duration of stay of drug users is often short (24 to 48 hours). Data was collected prospectively.

Clémence Isaure : is a therapeutic association working with drug users. The association comprises different services. One is working on therapy, another is working with accommodation for people in social distress. Intermède is part of the Clémence Isaure association. Clémence Isaure is a member of the GRAPHITI association. Collection of data was made retrospectively using first encounter questionnaires.

The centre Maurice Dide : it is a centre which opened at the end of 1994. It is part of the university hospital of La Grave, and was created by psychiatrists in the psychiatric ward of the university hospital. The centre can prescribe methadone. So some drug users in the Dide centre are using methadone but other are not. The centre's team comprises nurses, psychologist, social assistant and psychiatrists. Collection of data was made prospectively.

Antenne toxicomanie : this is a structure for drug users located in the Saint Michel jail. It prepares drug users for their release from the jail. It is not mandatory, so the population studied is not the whole population of drug users present in the jail. Collection of data was made prospectively.

Passages : is one of the two methadone centre of Toulouse. It opened at the end of 1994. Referral to the centre can only be made by one of the four members of the Toulouse hospital-city net on drug use (AAT, Josep Ducuing Hospital, Clémence Isaure, general practioner's net). Collection of data was made retrospectively.

Police : the total data is a mix of data from the drug use unit of urban police (prospectively collected) and of the data from the central bureau in Paris (collected retrospectively). Questions about viral status were not included in the police questionnaire.

7.3 Description of data collected by the ten sources

Sources were asked to collect data on each drug user seen in the study period. We selected people who where identified as drug users. We eliminated people who were not declaring use of any opiates and also those who reported living outside of the Toulouse conurbation. For some sources there was a high percentage of people who had not answered this question. The number of drug users by sources was ranging from 18 to 260. Two sources were under 50 people (Passages and Intensive care). Mean age was between 28 years and 10 months and 31 years and 11 months. Mean beginning age was between 17 years and 6 months. Sex-ratio was ranging from 1.5 (Clémence Isaure) to 7.6 (Jail). Five sources had a sex-ratio equal or less than two and a half, all were belonging to socio-medical sources. Five sources had a sex-ratio more than two and a half, two were belonging to repressive sources. Level of homelessness was varying from 9% to 38% and working from 8% to 44%.

Table 7.2: **Description of opiates users living in Toulouse, in ten sources : from May 1995 to October 1995.**

	Total	Opiate users	Mean age	1st use	75% before	Male: Female	Homeless	Working
CHS	233	139	28.1	19.6	23	3.1	37%	14%
Intermède	316	260	30.1	18.1	23	3.2	38%	14%
J Ducuing	184	167	30.1	19.7	24	2.4	12%	44%
AAT	97	91	28.1	18.6	21	1.9	20%	21%
Inten. care	20	18	29.0	17.6	-	1.3	35%	33%
C Isaure	121	115	29.6	18.4	23	1.5	20%	8%
M Dide	61	56	30.4	18.0	22	2	9%	34%
Jail	95	77	28.7	20.8	26	7.6	18%	24%
Passages	33	33	31.1	17.9	20	3.7	18%	33%
Police	116	107	29.8	21.1	25	2.7	14%	17%
All	1106	799	29.8	20		2.7	23%	25%

Heroin was the main product used in all but one source. Current injection was widely used by our population of opiates users (50% to 89%). The number of people who answered and who knew their viral status varied a lot by sources and level of prevalence was calculated only for people whom viral (HIV, HCV or HBV) status was known. HIV prevalence was ranging from 9% to 33%, HCV prevalence ranged from 41% to 88% and HBV prevalence from 32% to 63%.

Table 7.3: **Description of opiates users living in Toulouse, in ten sources : from May 1995 to October 1995**

	N	Heroin*	Injector (Main)	Injector (All)	HIV	HCV	HBV
CHS	139	?	70%	?	15%	53%	32%
Intermède	260	84%	72%	89%	17%	55%	33%
J Ducuing	167	62%	54%	59%	18%	59%	44%
AAT	91	70%	58%	63%	16%	32%	32%
Intensive care	18	83%	83%	83%	25%	86%	75%
C Isaure	115	78%	75%	80%	17%	67%	49%
M dide	56	54%	38%	74%	27%	54%	63%
Jail	77	88%	50%	50%	9%	41%	62%
Passages	33	49%	52%	55%	33%	88%	42%
Police	107	97%	69%	72%	-	-	-
Global	799	72%			16%	54%	37%

* heroin used as main product

7.4 Making less sources

As in our work of 1994, we produced estimates of the size of the opiate-using population using 2 sources and estimates of odds-ratio using three sources (two sources studied and the others). These

estimates allowed us to identify a couple of sources which were positively dependent (OR >1 and $p<0.05$). Thus we joined the most dependent sources. This process permitted us to lower our number of sources from 10 to 4 or 3. The 2 sample capture-recapture estimates and the corresponding odds ratios are presented as appendix 7.1.

The first step was fusion of Joseph Ducuing Hospital with Clémence Isaure (OR = 5.03). Second step was fusion of Police with jail. They had an OR of 3.75 and no significant link with any other sources. Third step was fusion of AAT with CHS and Dide centre. Forth step was fusion of Passages with the intensive care unit. Fifth step was the fusion of the Passages-Intensive care unit with the Clémence Isaure-Ducuing hospital group. We obtained four groups of sources. There was only one significant OR which was between the jail-police group and the AAT-CHS-Dide group. We made a medico-social group joining AAT-CHS-Dide group with Clemence Isaure-Ducuing-Passages-Intensive care unit group. Thus we had a "repressive" group a "care" group and an "Intermède" group. We decided to work our log-linear models with those three sources.

7.5 Models and first estimates

7.5.1 Global population

For the global estimates we were rather comfortable because all the strategies of model choice gave us the same result. Forward, backward technique and use of the Akaike criterion indicated that the R, LS model was the best. We estimated the empty cell to 1379 people and the total estimate was 2178 with a confidence interval from 1780 to 2734. This model implies an independence of the repressive source from the two other groups of sources. The use of classical capture-recapture technique gave an estimate of 2110 people (CI 95% 1570-2649) between S and R and of 2218 (CI 95% 1340-3095) between R and L. The LS interaction is quite logical because Intermede has established relations with various services.

Our estimate is higher than in 1994 (n=1150). We believe that the inclusion in 1995 of the police data and of Intermede data modified in an important way the representativeness of our sample.

Table 7.4: **Contingency table for three sources on opiates users in Toulouse, May-October 1995.**

		Low threshold (L)			
		yes		no	
		Repressive (R)		**Repressive**	
		yes	no	yes	no
Sanitary (S)	yes	6	79	34	389
	no	13	162	116	0

Table 7.5: **Log-linear models, population estimates, 95% confidence interval for opiates users in Toulouse, using three sources**

Model	**df**	**G^2**	**p-value**	**AIC***	**N222**	**Total**	**95%CI**
RL, RS, LS	0	0.03	1.00	0	1190	1990	703-3277
RS, LS	1	0.03	0.86	-1.97	1402	2201	1340-3060
R, LS	2	0.06	0.97	-3.94	1350	2149	1700-2598
					1379	2178	1780-2734
R,L,S	3	6.22	0.10	0.22	-	-	-

* Akaiké criterion.

7.4.2 Stratification on sex

First we carried out separate analyses for males and females and then we created a global model including sex as a variable. This model gave an estimate for males and an estimate for females.

For males the best model (table 7.7) was the same as for the global population and gave an estimate of 1820 male opiates users. For females the best model (table 7.8) was the simplest without interaction : R,L,S. It gave an estimate of 390 female opiates users. The sum of both estimates gave 2206 opiates users which was close to the global estimate.

The global model (table 7.9) gave a model identical as for the global population (R,LS) with an extra interaction between sex and sanitary sources. It gave a global estimate of 2175, near from the separate estimates but increased the number of females and lowered the number of males.

Table 7.6: **Contingency table of opiates users in Toulouse conurbation between may and October 1995**

				Low Threshold Structure			
				Yes		No	
				Repressive		**Repressive**	
				Yes	No	Yes	No
Sex	Male	**Sanitary**	Yes	5	55	19	266
			No	11	127	100	-
	Female		Yes	1	24	15	122
			No	2	35	16	-

Table 7.7: **Log-linear models and estimates for males opiates users in Toulouse conurbation in 1995**

Model	**DF**	**G^2**	**p-value**	**AIC**	**N222**	**Total**	**95% CI**
RL,RS,LS	0	0.05	1.00	0.05	1315	1898	
RL,LS	1	0.03	0.86	-1.97	1373	1956	
R,LS	2	0.50	0.78	-3.50	1280	1863	1442-1501
R,L,S	3	8.96	0.03	2.96			

Table 7.8: **Log-linear models and estimates for females opiates users in Toulouse conurbation in 1995.**

Model	**DF**	**G2**	**p-value**	**N222**	**Total**	**95% CI**
RS,RL,SL	0	0.00	1.00	115	330	
RL,LS	1	0.02	0.90	130	345	
RL,S	2	0.55	0.76	161	376	
R,L,S	3	1.51	0.69	175	390	323-487

Table 7.9: **Log-linear models and estimates for opiates users in Toulouse conurbation in 1995 including sex in the model**

Model	**df**	**G^2**	**p-value**	**AIC**	**N222 Males**	**Total Males**	**N222 Females**	**Total Females**
RM,LSM*	4	1.86	0.76	-6.14				
R,LSM	5	2.71	0.74	-7.29				
R,LS,SM	7	5.55	0.59	-8.45	1126	1709	251	466
					795-1587	1378-2170	158-379	373-594
R,L,SM	8	11.5	0.18	-4.53	823	1406	183	398
					646-1040	1229-1623	124-256	339-471
R,L,S,M	9	29.5	0.0005	11.5				

*M: Males (yes/no)

7.5.3 Stratification by age group

We carried out separate analyses for people from 15 to 34 years old and for people from 35 to 54 years old. For the youngest group, we found the usual model (R,LS) and a total estimate of 1709 people (table 7.11). For the people from 35 to 54 years old we found the same model and a total estimate of 426 people (table 7.12). Adding both estimates gave us a total of 2135 people which is in the range of previous estimates.

Table 7.10: **Contingency table of opiates users in Toulouse conurbation between May and October 1995.**

				Low Threshold Structure			
				Yes		No	
				Repressive		**Repressive**	
				Yes	No	Yes	No
Age	15-34	**Sanitary**	Yes	6	58	31	312
			No	11	129	102	-
	35-54		Yes	0	21	2	60
			No	2	32	14	-

Table 7.11**: Log-linear models and estimates for 15-34 years opiates users in Toulouse conurbation in 1995**

Model	**DF**	**G^2**	**p-value**	**AIC**	**N222**	**Total**	**95% CI**
RS,RL,SL	0	0.02	1.00	0.02	1076	1725	550-2420
RS,LS	1	0.01	0.93	-1.98	-	-	
R,LS	2	0.21	0.90	-3.79	1060	1709	1340-2080
R,L,S	3	3.74	0.29	-2.26	836	1485	1270-1700

Table 7.12: **Log-linear models and estimates for 35-54 years opiates users in Toulouse conurbation in 1995**

Model	**DF**	**G^2**	**p-value**	**AIC**	**N222**	**Total**	**95% CI**
RS,RL,SL	0	0.22	1.00	0.22	34	165	-
RL,RS	1	0.50	0.48	-1.52	-	-	-
R,LS	2	1.99	0.37	-2.01	395	426	80-970
L,RS	2	2.56	0.28	-1.44	-	-	-
R,L,S	3	7.49	0.06	1.49	139	270	190-350

7.5.4 Opiates use prevalence estimates

The global prevalence estimate for the 15-55 years old population of the conurbation is always around 5.4 per thousand people. 35-54 years old people have a lower prevalence (2.4‰) than 15-34 years old people (7.4‰). Females have a significantly lower prevalence of 2.3 for per thousand compared to 8.5 for males.

Table 7.13: **Best model, opiates users population size estimate, and prevalence estimates for the 15-55 years old population in the Toulouse conurbation**

	Best Model	Estimate	95% CI	Population	Prevalence	95%CI
Global	**R,LS**	2178	1780-2734	404,100	5.4‰	4.4‰-6.8‰
Sex	**R,LS,SM**	2175	1751-2764	404,100	5.4‰	4.3‰-6.8‰
Males	**R,LS**	1709	1378-2170	200,900	8.5‰	6.9‰-10.8‰
Females	**R,L,S**	466	373-594	203,200	2.3‰	1.8‰-2.9‰
Age		2135	-	404,100	5.3‰	-
15-34	**R,LS**	1709	1340-2080	231,100	7.4‰	5.8‰-9.0‰
35-54	**R,LS**	426	80-970	173,000	2.4‰	0.5‰-5.6‰

7.6 Discussion and conclusion

Nearly all (excepted for females) the best models showed an interaction between the low threshold structure and socio-sanitary structures. All models were showing an independence of the repressive group with others. Estimates were very coherent between each others, all were giving an estimate of the population of opiates users between 2100 and 2200 people. Clearly, some questions remain.

What sources for a city? Clearly a city has to be very careful in what sources it will use to do such an estimate. Comparing our studies of 1994 and 1995, we observed that our estimate changed from 1150 to 2178. In work we did not present here we showed that the inclusion of police and low threshold structure were mainly responsible for that change. So we do think that inclusion of repressive and low threshold structures are important to ensure a more correct estimate of the size of the opiates users population.

If this kind of observation is true between two years in the same city, it will be also true between two different cities. So to be able to compare prevalence estimates between cities imply to be able to define what are the sources that will be used or at least to have a good

description of the sources.

We want to emphasise on the gain of exhaustively and quality obtained by the use of a partly prospective collection of data. We do think that a prospective collection of data have to be evaluated in the preparation of a capture-recapture study. If it is possible to implement it, it will permit an improvement of the accuracy of the final estimates. At least in our French context.

How to choose a model? We do not have a very define strategy. We can start with an independent model and add interaction step by step. We can start wit a saturated model and delete interaction step by step. We can use Akaike's index. When the three methods do agree we can be very confident. When there are discrepancies between methods, you have to choose also on the basis of your knowledge of the relation between sources in the conurbation you are studying.

As an ending and open point, we believe that estimates are not so important. What is important is what uses will be made or won't be made of these estimates. To produce information is only a starting point in a public health process.

Appendix 7.1: **Estimates of drug users population size in Toulouse by capture-recapture with two sources and estimates of odds-ratios with 95% confidence intervals.**

		Police	AAT	Intensive care	CHS	C Isaure	Ducuing	Intermède	M Dide	Jail
A	N (95% CI)	1241(450-2032)								
	OR (95% CI)	1.46 (0.34-5.35)								
BRRg	N (95% CI)	-	-							
	OR (95% CI)	-	-							
CHS	N (95% CI)	1679 (662-2696)	1072 (531-1614)	531(138-924)						
	OR (95% CI)	1.3 (0.43-3.65)	**2.69 (1.01-6.89)**	1.77 (0.0-15.34)						
C-Is	N (95% CI)	1788 (546-3032)	1778 (438-3118)	1101	1475 (679-2272)					
	OR (95% CI)	0.87 0.21-3.12)	1.03 (0.24-3.72)	-	0.44 (0.07-1.94)					
D	N (95% CI)	2015 (788-3242)	812 (506-1119)	1595	1567 (851-2283)	811 (543-1079)				
	OR (95% CI)	1.47 (0.57-3.62)	**4.2 (1.68-10.3)**	-	1.14 (0.45-2.78)	**5.03 (2.45-10.3)**				
In	N (95% CI)	2562 (1157-3966)	2400 (1022-3778)	1239 (168-2310)	1404 (942-1867)	916 (669-1164)	1289			
	OR (95% CI)	0.56 (0.16-1.71)	1.2 (0.46-2.99)	0.79 (0.00-6.54)	1.6 (0.73-3.44)	**4.89 (2.35-10.1)**	**3.12 (1.65-5.84)**			
M	N (95% CI)	683 (290-1076)	873 (233-1513)	-	1139 (368-1910)	1652 (177-3187)	1367 (438-2296)	2478 (624-4334)		
	OR (95% CI)	3.08 (0.0-15.7)	**6.9 (2.0-22.4)**	-	3.18 (0.84-11.0)	2.23 (0.49-8.59)	**3.89 (1.26-11.6)**	1.12 (0.25-4.33)		
jail	N (95% CI)	526 (313-738)	-	740	1212 (491-1934)	3015 (0-6318)	1871 (576-3166)	1696 (833-2558)	2222	
	OR (95% CI)	**3.75 (0.81-14.9)**	-	-	1.2 (0.28-4.37)	0.7 (0.11-3.17)	**0.87 (1.20-3.14)**	1.18 (0.38-3.43)	-	
Passages	N (95% CI)	1835 (0-4881)	446 (162-729)	214 (0-428)	1189 (109-2267)	985 (93-1877)	570 (282-859)	633 (585-681)	-	-
	OR (95% CI)	0.85 (0.00-6.32)	**4.86 (1.29-16.7)**	8.93 (0.00-50.6)	1.52 (0.35-5.68)	1.85 (0.42-6.94)	**3.95 (1.3-11.6)**	**4.88 (1.69-14.0)**	-	-

* N : number of people estimated by mixing the two sources
* 95% IC : 95% confidence interval
* OR : odds-ratio (significant ones in bold)

8 Vienna

ESTIMATING THE NUMBER OF OPIATE USERS IN VIENNA

Dan Seidler, Emergency Dept., General Hospital, Vienna

Alfred Uhl, Ludwig Boltzmann-Institute For Addiction Research, Vienna

8.1 Introduction

Vienna is the capital city of Austria and has a population of over 1.5 million, with about 900,000 aged between 15 and 54. The age profile is described in table 8.1

Table 8.1: **Age profile of the Vienna population**.

Age (years)	0 - 14	15 - 24	25 - 34	35 - 44	45 - 54	> 54
Males	122,286	88,708	152,123	115,049	112,765	159,173
Females	115,407	90,234	148,227	113,410	117,311	257,903
Total	237,693	178,942	300,350	228,459	230,076	417,076

The largest proportion of persons dependent on 'hard drugs' in Vienna are multi-drug users; besides opiates, cocaine and cannabis, addicts in Vienna also consume tranquillisers, stimulants and designer drugs. The consumption of heroin and cocaine has increased over the last few years and the injecting of heroin together with cocaine now constitutes a major problem. Although the number of emergency drug related case has decreased the number of contacts to low threshold services has increased, as has the sales of syringes to combat the spread of HIV; substiture prescribing with methadone and other drugs is similarly increasing.

8.2 Data Sources Used

Four data sources supplied data for this research project, these being;

- P - Police data on any incident involving opiates.
- D - Cases of Drug Related Deaths related to opiate use.
- A - Emergency Ambulance Transports where the emergency doctor diagnosed "acute opiate intoxication.
- H - Hospital Admissions of persons diagnosed with "opiate intoxication".

Drug related deaths have a zero catchability after dying - in the average in 50% of the reference period. This results in an almost 50% reduction of the probability to be recaptured in the other samples. (A 50% reduction in re-catchability naturally doesn't corresponds with a 50% reduction in the number of recaptures, since there are multiple recaptures too, but under most circumstances the reduction will not be much less then 50%).

Because of this peculiar situation, in order to avoid a dramatic overestimation of the hidden population, we chose drug related deaths from the period following the reference period of the other 3 samples. The samples P, A and H represent incidents from 1/9/1993 through 31/12/1993; the sample D

represents incidents from 1/1/1994 through 31/12/1994. The samples P and D consist of all incidences from the reference period, the samples A and H represent a more or less random sample of all incidences all over Vienna in the reference period.

Key identifiers of all four samples are date of birth and sex. In the samples D, A and H the initials are known additionally and in case of P a unique identification code is supplied, but unfortunately not the initials. This allowed us to subtract double counts from within all 4 samples based on sufficient information and to identify matches between the D, A and H samples based on sufficient information but we couldn't determine matches between P and the other 3 samples based on sufficient information, since sex and date of birth are not specific enough. Having no alternative, but to match P and the other 3 samples on sex and date of birth only, we proceed as follows. If the was one or more persons with identical age and sex in sample P and in one of the other samples, we counted one match.

To summarise, the identification of double counting and the matching of persons who definitely were not in sample P constituted no problem, but we have to expect, that some of the matches between P and the other samples are false matches due to insufficient identifiers. We have to expect a certain overestimation of matches and therefore an underestimation the hidden population. It should be possible to correct the influence of too many matches caused by insufficient identifiers based on probability calculations, but we didn't do so in the following calculations. Correcting data before identifying double counts and matches was done by the following procedure. We used a simple computer algorithm to identify near matches within and between the samples and than we corrected the data manually based on plausibility considerations.

In the following table the numbers of subjects for all 4 samples are given. The last column gives you a chance to estimate the frequency of wrong matches based on sex and age only.

Table 8.2: **Number of cases identified from four samples in Vienna.**

	Including double counting	Excluding double counting	Erroneous classifications*
Hospital (H)	84	78	0
Ambulance (A)	302	229	2
Police (P)	899	727	17
Deaths (D)	143	143	2
Total	1,428	1,177	

* Different cases erroneously classified identical if only date of birth and sex are used as identifiers.

In the third column we can see, that 17 out of the 727 relevant cases in sample "P" would erroneously be classified identical, if we used "date of birth" and

"sex" only and no additional identifier. This is 2.3% of the cases. Considering that we identified 69 matches between sample "P" and the other samples, we may crudely expect something between 1 through 3 false matches. To create no wrong impression: We neither developed an exact way to estimate the influence of insufficient identifiers based on the distribution of the existing identifiers yet, nor did we do anything to correct our calculations for this source of bias. Since we overestimated the number of matches due to insufficient identifiers, and since we didn't yet correct this bias based on probability calculations, our estimation of the hidden population must be treated as an underestimation.

Table 8.3: **Description of the 4 samples by age and sex.**

	n	**Mean Age**	**% Male**
Deaths	143	26.6	87
Police	727	24.9	81
Ambulance	229	23.1	66
Hospital	78	22.8	68

The oldest sample were the Drug Deaths, the next oldest sample the Police Data, the next oldest sample the emergency ambulance transports and the youngest sample the patients treated in hospital. The percentage of males was significantly higher in the older samples than in the younger samples.

Obviously the samples are not randomly taken from the same population - The catchability of different cohorts is different for different samples and this implies that the samples and the population are quite heterogeneous.

To examine the calculations based on 2 samples only, we calculated all possible 2×2 combinations out of 4 samples (6 calculation). We also collapsed the highly correlated samples H and A to HA (2 additional calculations). The 8 estimations ranged from 303 through 5198 and these are presented as table 8.4.

Table 8.4: **Results of all 2-sample capture-recapture.**

Samples	N	Overlap	Minimum	Estimate	Maximum
P	727				
D	143	20	3114	5198	7282
P	727				
A	229	50	2543	3330	4117
P	727				
H	78	19	1833	2985	4137
P	727				
A+H	248	54	2581	3339	4097
D	143				
A	229	15	1173	2183	3193
D	143				
H	78	4	165	2789	5413
D	143				
A+H	248	16	1227	2217	3207
A	229				
H	78	59	270	303	336

8.3. Three-sample capture-recapture analyses.

When the data is restricted to the 15 to 54 year old age group, table 8.5 describes the overlap between the different samples. As the ambulance source and the hospital source are similar they have been combined into a single source AH. Tables 8.6-8.12 present the same data stratified by age and sex.

Table 8.5: **Overlaps between each source for all age groups.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	5	13	11	113
	Absent	49	658	179	-

Table 8.6: **Overlaps between each source for males, aged 15-54.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	4	11	9	100
	Absent	29	542	124	-

Table 8.6: **Overlaps between each source for males aged 15-24.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	3	7	4	39
	Absent	18	274	70	-

Table 8.7: **Overlaps between each source for males, aged 25-34.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	1	4	5	40
	Absent	11	221	39	-

Table 8.8: **Overlaps between each source for males, aged 34 - 54.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	0	0	0	21
	Absent	0	47	15	-

Table 8.9: **Overlaps between each source for females, aged 15-54.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	1	2	2	13
	Absent	20	116	55	-

Table 8.10: **Overlaps between each source for females, aged 15-24.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	1	2	2	5
	Absent	16	70	38	-

Table 8.11: **Overlaps between each source for females, aged 25-34.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	0	0	0	5
	Absent	4	40	15	-

Table 8.12: **Overlaps between each source for females, aged 34-54.**

		Police			
		Present		Absent	
		Amb/Hosp			
		Present	Absent	Present	Absent
Deaths	Present	0	0	0	3
	Absent	0	6	2	-

Using the statistical package GLIM, the following estimates and confidence intervals were obtained.

Table 8.13: **Analysis for total sample.**

Model	**G^2**	**d.f.**	**p-value**	**N**	**95% CI**	
Independence	14.20	3	0.003	3708	3105-4497	
+AHxD	9.20	2	0.010	4004	3289-4972	
+DxP	7.29	2	0.026	3187	2636-3935	
+AHxP	11.63	2	0.003	4480	3328-6318	
+AHxD+AHxP	**0.76**	**1**	**0.3818**	**6747**	**4332-11668**	←
+AHxD+DxP	4.43	1	0.353	3432	2765-4379	
+AHxP+DxP	6.99	1	0.082	2867	2023-4635	
Saturated	0.00	0	1.000	10525	3479-32933	

From table 8.13, the preferred model included interactions between the ambulance / hospital source and both of the other sources. The estimated total number of opiate users would therefore be 6747, 95% CI: 4332-11668.

The stratified data was also analysed using the three-sample capture-recapture method. Tables 8.14 to 8.19 present some of these analyses. It was not possible to obtain a sensible estimate over all the stratifications and it was also not possible to fit every model within each group.

Table 8.13: **Males aged 15-54 years.**

Model	G^2	d.f.	p-value	N	
Independence	13.17	3	0.004	3278	
+AHxD	6.16	2	0.046	3689	
+DxP	8.26	2	0.016	2747	
+AHxP	12.19	2	0.002	3715	
+AHxD+AHxP	**0.95**	**1**	**0.330**	**5746**	←
+AHxD+DxP	3.70	1	0.055	3137	
+AHxP+DxP	7.23	1	0.007	2197	
Saturated	0.00	0	1.000	10297	

Table 8.14: **Males aged 15-24 years.**

Model	G^2	d.f.	p-value	N	
Independence	**5.76**	**3**	**0.124**	**1390**	←
+AHxD	2.59	2	0.273	1521	
+DxP	5.37	2	0.068	1311	
+AHxP	5.74	2	0.057	1423	
+AHxD+AHxP	1.64	1	0.201	1942	
+AHxD+DxP	2.54	1	0.111	1481	
+AHxP+DxP					
Saturated	0.00	0	1.000	4700	

Table 8.15: **Males aged 25-34 years.**

Model	G^2	d.f.	p-value	N	
Independence	8.56	3	0.036	1245	
+AHxD	**3.03**	**2**	**0.220**	**1481**	←
+DxP	3.51	2	0.173	929	
+AHxP	8.15	2	0.017	1405	
+AHxD+AHxP					
+AHxD+DxP	0.29	1	0.588	1105	
+AHxP+DxP					
Saturated	0.00	0	1.000	2188	

Table 8.16: **Females aged 15-54 years.**

Model	G^2	d.f.	p-value	N	
Independence	**1.56**	**3**	**0.668**	**554**	←
+AHxD	1.45	2	0.483	562	
+DxP	0.67	2	0.716	522	
+AHxP					
+AHxD+AHxP	0.06	1	0.803	963	
+AHxD+DxP	0.62	1	0.431	528	
+AHxP+DxP	0.63	1	0.426	566	
Saturated	0.00	0	1.000	1231	

Table 8.17: **Females aged 15-24 years.**

Model	G^2	d.f.	p-value	N	
Independence	**0.69**	**3**	**0.875**	**289**	←
+AHxD	0.02	2	0.989	300	
+DxP	0.69	2	0.708	288	
+AHxP	0.54	2	0.763	258	
+AHxD+AHxP	0.02	1	0.893	309	
+AHxD+DxP	0.02	1	0.881	300	
+AHxP+DxP	0.35	1	0.552	229	
Saturated	0.00	0	1.000	351	

Table 8.18: **Females aged 25-34 years**

Model	G^2	d.f.	p-value	N	
Independence	2.80	3	0.424	284	
+AHxD	2.07	2	0.355	264	
+DxP	**0.90**	**2**	**0.637**	**233**	←
+AHxP					
+AHxD+AHxP					
+AHxD+DxP					
+AHxP+DxP					
Saturated	0.00	0	1.00	1599	

8.4 Discussion

The number of opiate addicts for Vienna has been estimated between 5,000 and a maximum of 10,000 for many years by now. Because of this the estimate 6,747 with a CI (4332, 11668) fits in nicely into the expectations. The problems are that only 1,028 persons out of the estimated 6,747 are known and that the estimation based on the 4-sample-capture-recapture yields much lower results even though the dependency between the samples H and A should be taken care using the interaction terms in the best model.

To contrast these figures:

- 3,817 persons have been registered by the police from 1982 through 1996 because of heroin related offences and

- Dan Seidler estimated 4,500 based on opiate related Drug deaths under the assumption that the annual mortality rate is 3%. A smaller mortality rate would result in a higher estimate.

When the results from this study are placed within the size of the population aged 15 to 54, the estimate of 6,467 is 0.69 of the population within that age group.