NACD RESEARCH WORKING PAPER SERIES

Research into methods and data sources for the estimation of prevalence of problematic opiate and cocaine use in Ireland

Maria Gannon, Gordon Hay, Jennifer McKell

School of Social and Political Sciences
University of Glasgow

NACD

The NACD's Working Paper series:

• Explores theoretical, conceptual or methodological issues

• Produces technical or scientific reports

• Publishes original work commissioned by the NACD and not already published elsewhere.

The series will be published only on the NACD website www.nacd.ie

The views in this report are the author's and may not necessarily reflect the views of the NACD.

Contents

List of tables

List of figures

Foreword

The core task of the NACD is to advise Government about the prevalence, prevention, treatment and consequences of problem drugtaking in Ireland. In developing such advice the NACD reviews and evaluates published literature and data but more usually it commissions external research on the range of issues covered by its remit. The results of those studies are published when they have been considered by Government.

In the course of these studies the NACD may obtain information and/or data which might not be included in the final report. This material could for example, explore methodological issues pertaining to a particular topic. It could also derive technical or scientific material which might not be of immediate interest to policy makers but which might be interesting to researchers or workers in the drugs area. The NACD has decided to make these materials available to a wider readership through a Working Paper Series to be published on the NACD website.

This second paper in the series examines some of the methodological issues surrounding the estimation of the level of problematic opiate and cocaine use in Ireland. It arises from the 2009 NACD report *Prevalence of Opiate use in Ireland 2006: a 3-source capture-recapture study* wherein the NACD noted that the results needed to be treated with a considerable degree of caution due to technical complications in the estimation, arising from the very welcome expansion of places on the methadone treatment programme. As a result the NACD committed itself to undertake research during 2010 into other methods of estimating the prevalence of problematic drug use. A project team consisting of Maria Gannon, Gordon Hay and Jennifer McKell from the University of Glasgow was the successful tenderer for research into methods and data sources for the estimation of problematic opiate and cocaine use in Ireland. While their report will be used to inform a NACD study planned for 2012, it was felt that the report would be of interest to others including not only those groups planning to tender for that study, but also those in the Regional and Local Drugs Task Forces involved in local prevalence estimation.

The NACD is grateful to the project team for their detailed evaluation of the methodologies and of the potential data sources in Ireland.

Dr Des Corrigan
Chairperson
November 2011

Authors' acknowledgements

Executive Summary

The National Advisory Committee on Drugs (NACD) commissioned the Centre for Drug Misuse Research at the University of Glasgow to examine the methods and data sources available to estimate the prevalence of problem opiate and cocaine use. This report has three main parts; the first discusses the various existing methodologies for estimating the prevalence of problem drug use and evaluates their usefulness and the feasibility of their application in an Irish context. The second part outlines the data sources available in Ireland and details which data sources lend themselves to which methodologies. A third section provides more detail on the application of the recommended methods.

Following an evaluation of eight different methods of prevalence estimation the study team believe that the application of the following three approaches to Irish data will prove the most successful in estimating the prevalence of opiate use:

- Four sample Capture-recapture (CRC)
- Multiplier methods
- Multiple Indicator method (MIM)

Four sample CRC should be used to estimate prevalence of opiate use for all 26 counties. If this methodology fails to produce a county level estimate then either the MIM or multiplier methods should be used to calculate the estimate. These county estimates can then be summed to give regional estimates and an overall national estimate. It is recommended that the following four data sources should be constructed from the available data for inclusion in a four-sample capture recapture study:

- Central Treatment List
- HIPE
- Garda Síochána (PULSE)
- Probation Service data

We would also advise using the same matching criteria as previous Scottish and English studies. This would mean matching over initials, data of birth and gender. It is critical to the success of a new study to establish a geographical unit of analysis and work should be done to find the most suitable one prior to the beginning of any new prevalence estimation work. This unit of analysis needs to be exacting enough to help with the methodological difficulty of heterogeneity in highly populated areas but also mesh with current administrative geography to give estimates that are of practical use to policy makers and stakeholders. Following a roundtable discussion with prospective data providers, county of residence was identified as the preferred lowest level for estimates to be produced at. These county estimates can then be combined to give estimates at larger geographical areas such as Garda Síochána or HSE regions. This would mean taking into account counties such as Dublin and Tipperary which straddle these larger administrative areas.

In order to estimate the prevalence of cocaine use in Ireland, the study team recommend using a different methodology which combines large household survey data with longitudinal study data. This approach would account for the heterogeneity in the cocaine using population and would produce more than one estimate of cocaine use reflecting the different cocaine using cohorts among the population.

A number of previous studies outlined in the report have suffered from delays when accessing data. The study team recognises the important role these systems have in safeguarding patient data but these delays can prove costly and some form of co-ordination may prove helpful particularly when undertaking a national study. It is recommended that the advisory group and stakeholders work together to streamline the ethics and data access process.

The study team recommend proceeding to a full national study as opposed to commissioning a pilot study based on this methodology. Prevalence estimation studies that follow the approach outlined in this report are typically composed of two parts; the data collection stage and the analysis stage. The first stage which involves accessing and collecting the data consumes the bulk of the time and resources associated with the study. The time and effort related to this stage of the study is not directly related to the amount of data collected therefore commissioning a pilot study focussing on fewer areas would not significantly reduce the cost or effort involved in the study. Another possible reason for commissioning a pilot study would be to investigate possible issues with the data sources involved. These issues can only be identified once the data collection phase has occurred. Therefore a pilot study investigating these issues could take the same time and effort as a full study without the benefit of more comprehensive results.

Introduction

Reliable and consistent information on the prevalence of problem drug use at the local and national level is vital for the planning and provision of services, for the monitoring of policy interventions directed towards drug misuse and for fulfilling Ireland's commitments to the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Problem Drug Use (PDU) is one of the five key indicators of the EMCDDA which commits member states to providing accurate and up-to-date (at least on a three-yearly basis) estimates of the prevalence of problem drug use and drug injecting, and where possible, estimates of the incidence of problem drug use/drug injecting.

Unlike some of the other EMCDDA key indicators, estimating the prevalence of problem drug use is far from a routine task. Specific research studies are required and a specific level of expertise and experience is required to plan and carry out a successful prevalence estimation study. An in-depth knowledge of available data, matched with a firm grounding in the statistical methods is required, specifically when assessing whether the assumptions related to any particular methodological approach are valid.

The EMCDDA has a standard case definition for problem drug use which it seeks to apply across all member states. The case definition encompasses opiate use (including methadone use) and problem stimulant use. The injecting of any drug is included in the PDU case definition and injecting drug use also forms a separate case definition. While it is important for the EMCDDA to adopt a standard case definition, its applicability and relevance may differ across member states or within member states.

There have been previous prevalence estimates for Ireland, including the two most recent studies carried out in 2003 and 2006.

The most recent Irish study concluded that the estimates needed to be treated with caution as the population was not closed. The open/closed population assumption of the capture-recapture method is just one of a series of assumptions and when being statistically rigorous, none of these key assumptions hold completely when studying a covert population like heroin users. Having carried out a large number of studies across Europe, and a series of methodological sub-studies within a large scale UK study (including a sensitivity analysis to examine the impact of the closed population assumption), the research team feel that the methods are relatively robust to violation of the closed population assumption, and that there have may been other more relevant issues that could have affected the prevalence estimates.

Objectives

The objectives of this study are:

- To determine indirect statistical approaches to estimating numbers and rates of problematic opiate and cocaine users in Ireland;
- To identify statistical or practical adaptation that would improve the reliability of the current capture-recapture estimate;
- To identify all data sources in Ireland, which can be used in the estimation of the prevalence of problematic opiate and cocaine use and using a systematic approach, evaluate their potential for use;
- To design pilot studies to test the preferred approaches.

This report begins with a review of the range of methods which can be used to estimate the number of problem opiate and cocaine users and the prevalence of problematic opiate and cocaine use in the population. Three prevalence estimation methods are recommended for future use in Ireland. These are the multiplier method, the capture-recapture method and the multiple indicator method (MIM). The report then goes on to examine the range of data sources in Ireland that can be use with these methods to estimate the prevalence of problematic opiate and/or cocaine use. This includes a matching of the methods to the data sources. In the final section of the report the study team give their recommendations for future prevalence estimation work in Ireland.

Section 1: Prevalence Estimation Methods

Estimating problem drug use amongst national populations requires the consideration of a number of issues. It is important to consider the difficulty of identifying the target population: problem drug users. Due to the illicit nature of their activity, individuals involved in drug use can be reluctant to disclose their involvement, even with regard to support agencies. As a consequence, methods for estimating the numbers of drug users must be able to take account of hidden populations. There are various methods that address the estimation of hidden populations but the choice of a particular method is dependent on factors such as the availability of suitable data as well as the ease of collecting that data. In certain circumstances it maybe necessary or appropriate to use a number of different methods to estimate the overall size of a drug using population. In the following section, we outline a number of methods that have been used to estimate drug using populations including discussion of some of the contexts in which they have been used as well as the strengths and weaknesses of each method.

1.1 Multiplier Methods

Background

The multiplier method is a common approach used to estimate the size of a drug using population. In fact, despite the growing popularity of more complicated statistical methods, the multiplier method remains popular for estimating the size of drug using populations in many European countries. One of the reasons for the popularity of the multiplier is that it is simple to use. The multiplier estimates the size of a drug using population by multiplying the numbers of known drug users recorded in a particular source by a proportion that is thought to reflect a drug user's propensity to be associated with the activity that the source records. One example of this method is the treatment multiplier which uses the number of known drug users attending a treatment service and multiplies it by the proportion of drug users within the population involved in treatment. Other examples of using this method are the police multiplier which refers to arrest data; and the mortality multiplier which employs data on drug related deaths. The proportions used within multiplier methods are typically drawn from other independent research that has measured phenomenon such as involvement with treatment and criminal activity among random samples of the relevant drug using population.

Data requirements

In order to estimate the total population of problem drug users (N) a sample of drug users or benchmark (n) is required and some measure of the probability that an individual from the total population would occur in the sample (p). The estimate is calculated as follows:

$$N = n/p$$

Possible sources of benchmark data are: treatment data, drug related deaths data or arrestee data. The method only requires a count of the number of individuals in contact with the particular data source over a specified period of time; usually one year. The multiplier term is usually gleaned from research carried out in relation to the benchmark data. For example, when using the multiplier method on mortality data the death rate from a longitudinal study could be used.

Example

Multipliers have been employed to estimate drug using populations both at the city and national level. The most prominent study, known as the Drug Indicators Project, was carried out in the 1980s. In that study, researchers used a mortality multiplier to estimate the number opiate dependent users in certain parts of London by employing data from coroners' records and using a combined mortality rate for opiate users attending treatment services, which had been observed in a number of UK studies (Hartnoll, Lewis, Mitcheson and Bryan, 1985). In the Netherlands, Smit and colleagues used a treatment multiplier to estimate the number of problematic opiate users in the Netherlands, in order to provide a comparison of available estimation methods (Smit, Van Laar & Wiessing, 2006). To produce a national estimate the researchers used data on individuals that had participated in out-patient treatment and were recorded in the National Alcohol and Drugs Information System for the year 2001; and multiplied this with the proportion of people attending treatment as obtained from various fieldwork studies conducted in the city of Amsterdam.

Limitations

Strengths of the multiplier method include, as noted, its relative simplicity and its ability to employ existing data and research findings, precluding any expensive and resource intensive data collection and analysis. However, it is also important to highlight the number of weaknesses associated with this method. Firstly the multiplier method is restricted in its ability to estimate drug using populations by the specifics of the data used to create a multiplier. Thus as a treatment multiplier uses data on treatment admissions and a population treatment rate it can only estimate the number of drug users within a wider drug using population that are 'treatment ready'. This therefore excludes from the resulting estimate those drug using individuals that are less likely to present for treatment such as non problematic users or those at an early stage in their drug using careers. Critics of multiplier methods highlight that multipliers also assume a linear relationship between the number of drug users and the proportion of drug users involved in an activity. For example, Korf and colleagues have pointed out that using a mortality multiplier may underestimate levels of drug use due the differences in mortality experienced by different types of drug users (Korf et al, 1994). Higher mortality rates are more associated with injecting drug use than smoking or sniffing as a route of administration; and a temporary increase in the typical purity of drugs used in an area could also inflate mortality rates for a period. Thus despite the apparent initial advantages of using multipliers, their inherently simple approach can often struggle to reflect the complexity of drug use.

1.2 Capture-recapture (CRC) Methods

Background

Capture-recapture methods were first developed to estimate the size of animal populations in the late nineteenth century. They also found prominence in correcting for undercounting in census or disease registers. However in recent decades the use of capture-recapture has become popular in social and health related research, particularly in the estimation of drug use prevalence. Early use of the capture-recapture method involved a two sample procedure, where researchers would estimate the size of a population by ascertaining the overlap between two samples. The literature has typically illustrated this procedure with the example of estimating the total population of fish within a lake

(EMCDDA; Hay et al). This example first describes a researcher capturing a sample of fish, from a lake, marking each individual and then returning them to the water. After a period of time, the researcher then returns to the lake, captures another sample of fish and records how many of the captured fish are marked and were therefore captured within the earlier sample. By then calculating the ratio of captured, to not captured fish it is then possible to work out the overall population of fish within the lake. Estimates produced by this method are considered reliable depending on the non-violation of certain assumptions. These included the equal probability of each fish appearing in both sources; that the population is closed, meaning that between sampling there is no loss or gain of individuals within the overall population; and that individuals captured within samples can be correctly identified from one sample to the next.

As capture-recapture began to be used to investigate the size of human populations, it became apparent that it was more difficult to maintain the assumption that each sample was independent and no individual's appearance in one sample was dependent on their appearance in the other. To combat this issue, researchers began to experiment with three and four sample capture-recapture, otherwise known as multi sample capture-recapture. By using three or four samples instead of two it meant that researchers were able to model the interactions between samples and account for any under or over estimation that could affect the independence of samples.

Data demands

Ideally the capture-recapture method requires three or more data sources to produce robust estimates. A later section of the report will detail the reasons for this. The method relies on identifying the same people across various data sources; therefore a number of identifiers are required in order to ensure that the matching is as accurate as possible. The following data fields would be integral to matching across sources:

- First name initial
- Surname initial
- Date of birth
- Gender.

In order for the 'closed population' assumption to hold, a contact date is also required, if only to exclude those that did not attend the agency during the period in question. Another crucial piece of information is some indication of where the individual is resident. In order to produce meaningful estimates at a local level it is essential that as far as possible individuals are assigned to their area of residence and not the area where they come into contact with the data source in question. For example, if a study uses police data it is important that individuals are assigned to their area of residence and not the area where the offence was committed as there are a high concentration of offences committed in urban centres and if the individual was assigned to that area it would falsely inflate the population of that area and thus affect the estimate.

Example

The capture-recapture method has mainly been used to estimate drug user populations within cities or other large urban areas yet there are some examples of its use at the national level. Most of the

early studies were confided to cities such as those in Glasgow (Frischer, 1992; Frischer et al; 1993), Toulouse (Bello and Chêne, 1997), Dundee (Hay and McKeganey, 1996) and Barcelona (Domingo-Salvany et al, 1995). Brugha et al (1998) used the method in a rural area in the North West of England and Hay et al (unpublished studies) used the methods in mixed urban/rural Health Board areas in Scotland. One of the first attempts to apply the capture-recapture method to estimating the size of a drug user population at the national level was carried out in Wales using data relating to 1994 (Wood et al, 2000). Wood and colleagues provided estimates of serious drug use in Wales by modelling the overlap between multiple data sources including the Welsh Drug Misuse Database, the probation service, police, needle and syringe exchange schemes and three laboratories providing HIV testing. In a journal article published in 2000, the authors admit that the study faced a number of methodological difficulties. Particularly important were the substantial differences in the drug use definitions used by each data source which resulted in the eventual use of a general definition of 'serious drug use'. Since this time, researchers in a number of European countries have used capture-recapture methods to estimate national drug user populations. National prevalence estimates have been produced for Austria, Denmark, Finland, Luxembourg, Sweden, Ireland, Scotland and England. (Kraus et al, 2003). In the UK, Hay and colleagues have used capture-recapture to provide separate and successive national problem drug use prevalence estimates for Scotland and England. Using a four sample method involving data from hospital and treatment admissions, criminal justice reports, police, probation and prison services, the researchers produced local area estimates that were summed to produce a national estimate.

In the context of Irish problem drug use prevalence, the capture-recapture method has been used on three occasions. The first of these involved a study carried out by Comiskey which estimated the prevalence of opiate use in Ireland by estimating a prevalence particular to Dublin, based on the assumption that opiate use was confined to the capital (Comiskey et al, 2001a). Following on from this work, a second and third study investigating opiate use prevalence in the years 2000/2001 and 2006, respectively, was carried out by academics from Trinity College, Dublin (Kelly et al, 2003; 2006). Employing hospital and methadone treatment data as well as police statistics, Kelly and colleagues carried out a three sample capture-recapture process, producing estimates for Dublin, the rest of Ireland and Ireland as a whole.

Limitations

There are a number of advantages to using capture-recapture to produce estimates of the size of a drug user population. A capture-recapture analysis can be carried out through the use of existing data sources which avoids the need to collect data directly and in turn the costs associated with such data collection. Capture-recapture can also be applied systematically over time, allowing for comparison between successive estimates, such as over a period of a few years. Another advantage of capture-recapture is that the technique allows for the calculation of confidence intervals for estimates providing some insight into the reliability of estimates. Capture-recapture depending on the availability of data sources can provide simultaneous estimates for different geographical areas without having to establish certain assumptions for extrapolation to multiple areas.

One of the disadvantages of using multi source capture-recapture is that it is not a simple method but one that requires statistical/epidemiological knowledge and experience, in addition to knowledge and understanding of drug using populations. As highlighted previously, another disadvantage is the

likelihood of violating one of the assumptions required by capture-recapture methodology coupled with common uncertainty as to the extent of violating these assumptions. Although capture-recapture is useful for providing estimates in multiple contexts, this is dependent upon the size of available data sources in those settings. Capture-recapture requires population samples of a reasonable size so as to ensure the production of sensible estimates. For example, it may be difficult to employ capture-recapture methods to produce estimates for a rural area where service provision is patchy and data sources are limited.

1.3 Multivariate Indicator Method (MIM)

Background

The Multivariate Indicator Method (MIM) employs regression techniques to provide prevalence estimates for geographical areas based on the relationship between prevalence estimates and drug use indicators in other areas. To undertake this kind of analysis, a researcher must first collect prevalence estimates for a sample number of areas within a country, a region or a city. These are considered to be the anchor points of the analysis and are typically derived using the capture-recapture method. Then, in the areas in which they aim to estimate prevalence, they must then collect data on a series of variables that reflect levels of drug use. These can include hospital admissions due to drug use, the number of burglaries, rates of hepatitis C prevalence within drug using populations or indeed any readily available, suitably geographically aggregated data that is thought to be correlated with problem drug use. Once this data collection is complete, the researcher can then input the drug use indicators into a regression model and investigate the factors that best explain the prevalence estimates in the anchor point areas. Having established this relationship, the same drug use indicators used in the regression are then collected for the areas that do not have a prevalence estimate and are then used to calculate the corresponding prevalence estimate. A 95 per cent confidence can also be produced for each estimate, and a national estimate can be obtained by summing the local multiple indicator method and/or capture-recapture estimates.

Data demands

In order to produce MIM estimates two important components are required – a number of anchor points and indicator data related to problem drug use. The anchor points are estimates of problem drug for a number of discrete areas. The indicator data is usually publicly available data relating to problem drug use. One advantage of this method is that if there are already estimates available to use as anchor points one only needs to collect indicator data which is aggregate data and does not entail accessing the sensitive identifiers that would be required for a capture-recapture study, for example. This indicator data takes many forms for example: the number of problem drug users in treatment, the number of shop-lifting offences or the number of drug related deaths. However this data is required at the geographical level that estimates are being calculated. If, for example, this was local drug task force level then the study would require the count of drug related deaths, the number of those in treatment and the number of those shop-lifting for each local drug task force area.

Example

This method has been used a number of times to produce estimates of problem drug use in England. One of the earlier demonstrations of the method was by Frischer et al (Frischer 2001). They used four existing estimates from four regions across Britain to produce regional estimates of problem drug at both regional and national level. The indicators they employed included convictions for drug offences, seizures of controlled drugs, treatment data from the Regional Drug Misuse Database, cases of HIV related to injecting drug use and drug related deaths. One feature of this analysis which differs from the straightforward application of the method was the use of 'principal components analysis'. Frischer et al had a relatively small group of anchor points compared to their supply of indicator data. They therefore decided to use principal components analysis to shrink the number of indicators without losing any factors that explain the variation in the indicators; in effect creating 'super-variables'.

Hay et al have used this method along with capture-recapture four times to estimate prevalence in England for a series of Home Office studies. The capture-recapture method was used to produce local level estimates for as many areas as possible and the MIM used to provide estimates for areas where this was not possible. Due to the large number of anchor points available to the team they did not have to use principal components analysis to produce an optimum number of indicators.

Limitations

This method, like others discussed, stands or falls on the quality of data used. Traditionally, researchers have had little choice in the anchor points they employ in the analysis and therefore have had to make do with estimates from previous years or been forced to use a number of estimates concentrated on urban areas. It is unclear what specific effect historic data may have on the estimates but the use of almost exclusively urban anchor points with sparse rural indicator data can make robust model fitting a struggle. Another issue with the application of this method relates to the indicator data. Although most suitable data is publicly available there may be some issues with obtaining the data at the geographical level required. For example, the Office for National Statistics (ONS) in England and Wales when reporting drug related deaths at certain levels censor values under five. This can mean unless unrestricted data is given to researchers there are blank cells for a number of areas which would render that particular indicator useless. Indicator data with the correct geographical data is also essential. For example, if police data is used then it should be area of residence as opposed to the area where the offence was committed. The inclusion of police data with area of offence not residence would inflate the numbers in certain areas and can affect the estimates. This can have an extreme affect for example the City of London has low numbers of residents but a high volume of offences so if offence data with incorrect geography was used there it would result in an over estimate of problem drug users per head of population.

1.4 Nomination Techniques

Background

Estimates of the size of a drug using population can also be ascertained through the use of nomination techniques which involve asking drug users to identify other drug users with which they have been in contact. Such techniques include social network studies and snowball sampling of a

drug using population. Social network studies rely on the assumption that individuals are likely to associate with individuals or groups in which they share something in common. For drug users, this means that they are more likely to spend time with other drug users than individuals who are not involved in drug use (Fraser & Hawkins, 1984 cited in Dunn & Ferri, 1999). It follows that a snowball sampling involves selecting a number of drug using individuals who in turn refer researchers on to other drug users, who then identify their own contacts. The pattern continues until no new referrals can be made.

Data demands

Nomination techniques often result in researchers calculating estimates based on a ratio of 'known' to 'unknown' problem users. This is in effect a form of multiplier method. It requires a sufficient level of contact details for the 'nominated' users but these sensitive data are not actually used to calculate the resultant estimates. In order to collect this level of contact information experienced researchers are needed. Often a study will take advantage of this by getting the researchers to administer a survey or carry out qualitative interviews with participants.

Example

From the literature, it appears that such nomination techniques have been used to estimate drug use prevalence in two ways. The first has amounted to a simple count of drug using individuals. Korf and colleagues describe ethnographic studies conducted in the Netherlands which produced prevalence estimates for both towns and rural areas based on the mapping of drug user networks (Korf et al, 1994). Secondly, nomination techniques have been used to create multipliers which, as described above involves obtaining information on a certain characteristic of a drug using population. This was the approach taken by researchers in Slovenia in a research study that aimed to estimate the national prevalence of heroin use as well as gather more detailed information on drug users (Nolimal, 1996). As a starting point for their sample, the researchers recruited twenty-six people that were receiving methadone maintenance from a treatment centre in the town of Koper. The initial twenty six were then asked to identify other individuals involved in heroin use, and administer a questionnaire, covering a number of different topics, including involvement with treatment. Using the information collected from the questionnaires, the researchers then proceeded to create a treatment multiplier based on a ratio derived from the number of untreated to treated heroin users. Combining this proportion with information on official numbers of heroin users treated with methadone and admitted to hospital between 1991 and 1993, the researchers were able to produce a national prevalence estimate of heroin use. However, despite the production of a national estimate, the researchers highlight the study's large margin for error and stress that the primary merit of the study lay in other information it was able to gain regarding drug users. Problems encountered by the researchers in using this approach included a lack of knowledge regarding how heroin users were chosen by the core sample of individuals; and uncertainty as to whether some individuals had been surveyed on multiple occasions.

Comiskey and colleagues also describe the use of nomination techniques to develop a multiplier in a study that estimated the number of drug users with previous criminal involvement in Ireland (Comiskey et al, 2007). Focusing upon the self-reporting aspect of nomination techniques, Comiskey and colleagues created a multiplier using data relating to criminal offences collected from

participants to the ROSIE study. The Research Outcome Study in Ireland (ROSIE) investigated drug treatment outcomes relating to opiate users across Ireland by surveying a drug using cohort, at entrance to treatment and at later intervals over a period of twelve months. Although the focus of this study was treatment outcomes, the ROSIE survey also asked participants to provide information on other areas of their lives including criminal behaviour. Having ascertained that 3.5% of participants had been arrested for supply offences in the last three months, the researchers multiplied this proportion by the total number of supply offences for Ireland in any quarter of 2003 or 2004, as recorded in official crime statistics.

Limitations

One undoubted advantage of using nomination techniques to estimate drug use prevalence is its simplicity. Such a method does not require an understanding of relatively complicated statistical techniques, especially in studies where the object is simply to count the number of drug users in an area. However, this simplicity is outweighed by the disadvantages of using this method. One disadvantage is the fact that data is directly collected through surveying drug users, which has implications for study costs. Carrying out research in the field can be time consuming and resource intensive. In addition to employing researchers, such a study will likely require dedicated field-workers to actually collect the data. As a consequence, and as indicated by Korf and colleagues, it may be possible to conduct such a study successfully at the local level, but a national study based on these methods is likely to be prohibitively expensive. Nomination techniques also suffer from the fact that they involve the collection of self-report data, which due to the illicit nature of drug use is often viewed as unreliable. Some researchers have attempted to ameliorate this problem by selecting field workers with a particular status that means they are able to elicit reliable responses. This practice has been termed the Privileged Access Interviewer method (Dunn & Ferri, 1999). This method was used in the Slovenian research discussed above but as highlighted this produced other problems for the research study.

1.5 Truncated Poisson Method

Background

The Truncated Poisson Method can be used to produce estimates of problem drug use from a single data source. The method requires information on the number of times an individual has been in contact with a service over a specified period of time. The principal behind the methodology is that the number of people who have attended the service once, twice, three times and so on can be used to estimate the number of people who have been in contact zero times. It is assumed that the pattern of individuals' contacts to the agency in question is 'truncated below one' and therefore follows a Poisson distribution that is 'truncated below one'. This means that the data can be modelled to give an estimate of the 'hidden population' or those who do not attend the agency. Two of the more popular estimators of the hidden population have been developed by Zelterman (Zelterman, 1988) and Chao (Chao, 1987) and the formulae are detailed over.

Zelterman's estimator

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

Chao's estimator

est $(n) = S + f_1^2/(2f_2)$

where

f_1 = the number of people appearing once

f_2 = the number of people appearing twice

S = the sum of all appearances

These estimators give some prominence to the people who occur once and twice in the source. This takes into account the inherent heterogeneity of attendance patterns i.e. those that visit only once or twice are more similar to the hidden population and less like regular attendees. This methodology is subject to similar assumptions as the capture-recapture method. The population is closed, the population is homogeneous and the probability of being observed more than once is constant over time. Each of these assumptions can be violated to a certain extent when applying the method to drug-related data. The closed population assumption is usually satisfied by restricting the study period to a minimum i.e. one year or one month. The homogeneous population assumption can be addressed through stratifying the data into less heterogeneous groups. The final assumption, that of equal probability of observed, is more challenging and can sometimes be helped by keeping the observation period short.

Data demands

This method requires data on the number of times an individual has contacted an agency over a particular time period. For example the number of times an individual has visited a methadone maintenance service over the course of a year. Therefore data providers need to have the ability to identify individuals that are in contact with them and to track their attendances over the course of the study period. The data required could take the simple form of a unique identifier/client code assigned at first contact or if that is not available then some combination of identifiers such as initials, gender and date of birth. Some low threshold agencies such as needle exchanges may not collect the level of data required to distinguish individual users and therefore would not be suitable for use in a truncated Poisson analysis.

Example

When using the truncated Poisson method to estimate the prevalence of problem drug use researchers have based their estimates on various data sources relating to illicit drug use. In Rotterdam the truncated Poisson method was applied to methadone maintenance data to produce estimates of problem drug use (Smit et al, 1997). The researchers felt that the resultant estimates were more accurately reflected the target population for the methadone service than the entire hidden population and therefore the total estimate could be described as an underestimate. This method was also used in New York State (Simeone et al, 1993) where it was again applied to methadone maintenance data. Researchers in Switzerland used police arrestee data in their

truncated Poisson analyses in the early nineties (Knolle, 1997). The truncated Poisson method has also been used by one of the authors of this report to model needle exchange data in order to estimate the number of problem drug users in Riga (Hay, 2003).

Limitations

This estimation technique relies on one data source and therefore the resultant estimates are intrinsically linked to the population in contact with this agency. For example, an estimate of problem drug use derived using needle exchange data would be more reflective of injecting drug users than the wider problem drug using population. Therefore case definition should always be taken into account when discussing the results from a truncated Poisson analysis.

1.6 Back Calculation Method

Background

The back-calculation method has traditionally been used to measure the spread of disease but it can also be applied to drug related data to estimate the prevalence of problem drug use. The general theory underpinning the method relies on knowing the point of infection with the disease, the time of incubation and the rate at which the sample progresses from infection to a particular endpoint. This is then used to 'back-project' to the number of people initially infected. Treating drug misuse as a disease means the back calculation method can be applied to estimate the number of problem drug users. Taking mortality data as an example the point of initiation of injecting is treated as the point of infection, the endpoint is death due to overdose and the rate of progression is the rate at which the sample moves from injecting to death. The method requires a number of years of data in order to successfully model the spread of the 'infection'. The application of this technique assumes consistent reporting and completeness of data also any major fluctuations in the rate of progression to endpoint must be taken into account.

Data demands

This technique requires data from only one source but needs quite detailed information over a number of years. For example the agency involved would have to record when individuals began using drugs. As previously mentioned data completeness is important to the successful application of the method and if over the course of time the rate of progression fluctuates this would have to be taken into account.

Example

Law (Law 2001) et al calculated two sets of estimates of problem drug use in Australia; one produced using methadone maintenance data, the other using overdose data. This gave them the ability to compare estimates and assess their credibility. De Angelis (De Angelis 2004) used mortality data to produce estimates of opiate use in England.

Limitations

This technique requires detailed information from one data source over a period of a number of years. This can mean that if the data source used is not representative of the entire problematic drug using population then the estimates will reflect a subset of this group. It should be borne in mind, as

demonstrated in the study by Law et al, that the choice of data source will determine the case definition of the resultant estimates, with mortality data producing an estimate of injectors and treatment data estimating the wider population of opiate users.

1.7 Waste Water Analysis

Background

In recent years, a new method for investigating levels of local drug consumption has emerged and provided the possibility of further opportunities for estimating drug using populations. This method which has been termed 'sewage epidemiology' has its origins in earlier research which assessed the presence of pharmaceuticals in the environment (van Nuijs et al, 2009; Kasprzyk-Hordern et al, 2009). Research studies that have employed this method have collected water from waste treatment plants and analysed samples for the presence of certain drugs and their metabolites. Having ascertained drug concentration levels, researchers have then combined this information with data on the size of the local population, water flow rates, the extent to which drugs are metabolized within the body and typical drug doses (Kasprzyk-Hordern et al, 2009). Such calculations have resulted in the production of daily consumption rates for the local population based on so many doses per thousand people.

Although some early research in this area was carried out in the US, this method has mainly been developed in Europe, beginning with Zucatto and colleagues' research in Italy which sampled waters from the River Po as well as multiple water treatment plants serving a number of Italian cities (Zucatto et al, 2005). Focusing on levels of cocaine use and its major metabolite benzoylecgonine, the researchers discovered a daily consumption rate of 27 doses per thousand people aged 15 to 34 years.

Data demands

This method doesn't require direct data collection and only involves the collection and analysis of water samples. However, for a study such as this to succeed there must be reliable information on the consumption of the drugs in question. This level of information would generally be garnered from a qualitative study and it is crucial that this data is as timely and representative as possible.

Example

Following the approach taken by the researchers in Italy, Bones and colleagues carried out a similar study in Ireland in 2006 (Bones et al, 2007). The key difference in this research however was that in addition to cocaine, the researchers also investigated the levels of other drugs. Sampling from waste treatment plants and surface waters in and around the greater Dublin area, Bones and colleagues ascertained drug concentration levels for cocaine, benzoylecgonine, morphine, temazepam; and the major metabolite of methadone, EDDP. However, according to the researchers, issues with Irish census data prevented reliable estimation of the daily dose per thousand people. Nevertheless, an important point made concerning this research was that despite the detection of morphine within samples, it was not possible to ascertain to what extent such concentrations referred to illicit heroin use or therapeutic opiates. A Zucatto and Castiglioni study suggests a similar argument in their review of the research into the environmental presence of illicit drugs (Zucatto & Castiglioni, 2009). The authors state that the high concentrations of morphine found in waste water samples collected in Switzerland, Germany and England are likely to be due to greater use of therapeutic morphine in these countries.

The studies previously highlighted in this section have focused upon developing rates of consumption for local areas rather than drug prevalence estimates. However, recent research conducted in Belgium has attempted to use consumption rates to produce a national prevalence estimate for cocaine use (van Nuijs, 2009). In this study, the researchers sampled water from 41 treatment plants across Belgium, on two occasions during 2007/2008. Overall, the treatment plants that were featured, serve more than a third of Belgium's population. For each treatment plant area, the researchers calculated a daily and annual rate of consumption of so many grams for every thousand people. By then summing the total amounts detected in each area for each time period, it was concluded that an average of 1.88 tonnes of cocaine are consumed in Belgium per year. The researchers then divided this amount by the average amount of cocaine consumed by an individual over one year, based on prior assumptions of an average dose (100mg) and weekly usage. This produced the overall number of cocaine users in Belgium, which was then divided by the number of people in Belgium aged between 15 and 64, producing a national prevalence rate of 0.8%. Although this research is acknowledged to have a number of flaws, the researchers also highlighted that the resulting prevalence rate, was similar to a previous national estimate of cocaine use for 2004, produced by the UNODC.

Perhaps the biggest advantage of using waste water analysis to assess levels of illicit drug use is its ability to provide current information. Whereas other methods must rely on using historic data from previous years, this method can report within a very short time period, including the same year. Another advantage is that although such studies entail the direct collection of data, the collection of water samples is likely to require fewer resources than interviewing individuals or as time consuming as arranging access to existing datasets. However, a number of disadvantages should be acknowledged. Such research studies have relied upon certain assumptions about the amounts of illicit drugs consumed by individuals. Zucatto and colleagues assumed an average dose of 100mg of cocaine in their research, as did the researchers conducting a similar study in Ireland. Van Nuijs and colleagues also adopted an average weekly level of cocaine use based on the findings of two other studies that took place in New York and Amsterdam, respectively, to calculate a national prevalence estimate for Belgium. By basing research on such assumptions it raises questions about how representative resultant findings are. Moreover, waste water analysis can only provide one prevalence estimate for each area in which water is sampled which has negative implications for drug use variability within areas; and the production of estimates by gender, age or ethnic group. It is also a disadvantage that in the case of opiates it appears difficult to distinguish between illicit and therapeutic forms of the drug. This is particularly unfortunate given that, in many European countries, the prevalence of problem opiate use, relative to other drugs, is considered particularly important.

Limitations

The inability of this method to distinguish between opiate drugs used for medical purposes and illicit opiates mean that this technique is not best suited to estimating the prevalence of heroin of other opiate use among the population.

1.8 Preferred Methodologies

It appears that three methods may offer valid estimates of the prevalence of opiate and/or cocaine in Ireland. These three methods are the multiplier method, the four-sample capture-recapture method and the multiple indicator method. These methods are, in some way, related to each other. The two-sample capture-recapture method is quite similar to the multiplier method and the multiple indicator method can be seen as an extension to the simple multiplier method. All three methods can be used with the data in Ireland (either aggregate data that should be readily available or extracted from existing databases) or individual-level data that may be made available subject to data protection/client confidentiality issues being addressed.

The capture-recapture method would be the most data intensive as it uses individual-level data and cross-references across three or preferably four data sources. It is also the most labour intensive approach, not so much in terms of the data processing or analyses, but in terms of negotiating access to data and collecting data from local data holders. Much of the data collated within a capture-recapture study could then be used within a multiplier method analysis or, more importantly, within a multiple indicator method analysis. Other data at the aggregated level could also be used within the multiplier method and multiple indicator method analyses, and such data should be relatively easy to obtain and analyse. It can also be used to assess the validity of the capture-recapture estimates, particularly when assessing prevalence outside of Dublin.

Section 2: Summary of Irish data sources

In this section we describe the available data sources in Ireland that can be used to estimate the prevalence of opiate and/or cocaine use. We are only focussing on data that could be used within capture-recapture, multiplier or multiple indicator method analyses. We begin with the data sources that were used in previous three-sample capture-recapture analyses in Ireland and then go on to examine some other data sources that we feel could be used within a four-sample capture-recapture analysis.

We also examine other data sources that, since they do not collate sufficient identifier information such as initials, may not be useful for a capture-recapture analysis, may be used to generate estimates based on multiplier methods.

2.1 Central Treatment List

Description

The Central Treatment List (CTL) was established in 1993 as an administrative database to monitor the dispensing of methadone treatment. This includes information on people receiving both methadone maintenance and detoxification. In 2008 there were 7,942 continuing cases of methadone treatment in Ireland. Once a person is assessed and prescribed methadone their doctor submits a form to the CTL and the client is added to the database with a unique identification number. Patients prescribed methadone are in receipt of an identification card which they are required to renew every 12 months. This means that the CTL can monitor continuing cases. If a patient stops receiving a methadone prescription an exit form is submitted. The CTL currently only collects data on those receiving methadone, it does not monitor other substitute medication such as suboxone. However, it is anticipated that next summer the CTL will be modified to include other substitute drugs. The database as it stands collects the following information:

■ Name

■ Address

■ DOB

■ Gender

■ District Electoral Division

■ HSE Area

■ Local Health Office Area

■ Task Force Area

■ Date began taking methadone

■ Type of treatment

■ Doctor

■ Clinic.

Data available and possible application

The identifiers held by the CTL mean that its data is suitable for inclusion in a capture-recapture analysis and data from the CTL has been used as part of a number of prevalence estimation studies over the years. The study team would anticipate the inclusion of CTL data in any future prevalence estimation research. There are, however, some issues relating to the CTL that should be borne in mind. The provision of methadone across Ireland is not uniform and there are HSE areas where demand out strips supply. This means that there are a number of people, mostly outside the Dublin area on waiting lists for methadone treatment. The study team would recommend investigating this waiting list data with a view to bolstering the information about opiate users in rural areas and thus possibly aiding the provision of improved estimates of problem drug use for these areas. We would envisage the waiting list data augmenting the CTL data and that this pooled data could be used as a source in a CRC analysis.

2.2 Hospital In-Patient Enquiry System

Description

The Hospital In-Patient Enquiry Scheme (HIPE) is a computer-based system designed to collect demographical, clinical and administrative data on discharges and deaths from acute hospitals nationally across Ireland. Although it is a central database, sufficient identifier information is only held at the source hospital. Much of the time and effort of previous capture-recapture studies in Ireland has been concerned with negotiating access to these local data, including gaining the approval from relevant local ethical committees.

In any future capture-recapture study using HIPE data, it is recommended that alternative approaches to gaining ethical and/or data protection approval are explored by the NACD and/or other relevant national or Government organisations. The reason for this recommendation is that it will be important for the Irish Government to have prevalence estimation undertaken in a cost-effective manner, and there may be more cost-effective ways of gaining approval to use the relevant identifier information. We are not suggesting that the approval of local ethical committees should not be sought, or that local hospital managers should not be involved in the process as both have a key role in protecting patient confidentiality, rather it may be possible to involve such local stake-holders in a more efficient manner. As an example, in a previous Scottish study the approval of a senior medical officer (known as the Caldicott guardian) for each of the 14 local health areas had to be gained prior to receiving patient data. One of the 14 local health areas acted as a lead reviewer for the study and gave approval for the study to be carried out in their area, and provisional approval for the study to be carried out in all other areas. This decision, and more importantly the reasons why approval was given, was relayed to all other areas who were then given the opportunity to make a different decision if they saw fit. This streamlined the process and saved time and resources. In some way, it is like the approach to gaining ethical approval in the United Kingdom where multi-centre research ethics committees can give approval for research studies to be carried out in more than one area, with each local area being able to raise local issues if necessary.

We are suggesting that the NACD should take the lead in exploring this, rather than any researchers carrying out the study, as any negotiations are likely to be high-level between Government and health

organisations. More importantly, this should be something done before a study is put out to tender as a research group would have little option but to submit a tender with sufficient time and associated travel costs for negotiating local access.

It should be noted that the most recent capture-recapture studies in the United Kingdom, such as a study currently being carried out in Scotland by ISD (the Information Services Division of the National Health Service in Scotland), have been advised by Research Ethics Committees that this type of study is an audit using data collected as part of usual care therefore does not require NHS ethical review under the terms of the Governance Arrangement for Research Ethics Committees in the UK[1]. While that advice is subject to interpretation, and any such advice is not readily applicable to Ireland, it is perhaps for the NACD and partners such as the Health Research Board (the Irish equivalent to ISD) to consider whether such an argument may be relevant in Ireland.

Data available and possible application

Notwithstanding the issues raised relating to data access, the data available from HIPE would be suitable for use in a capture-recapture study. The data could also be used to provide multiplier based estimates and there is scope for using the data as an indicator in a multiple indicator analysis.

2.3 Garda data

Description

An Garda Síochána have a database known as PULSE (Police Using Leading Systems Effectively). It holds information on individuals detained or searched under the following categories:

■ Importation of drugs

■ Cultivation or manufacture of drugs

■ Possession of drugs for sale or supply

■ Possession of drugs for personal use

■ Forged or altered prescription offences

■ Obstruction under the Drugs Act.

A capture-recapture study would be particularly interested in people detained or searched for possession of drugs for personal use, or forged/altered prescription offences as such individuals are highly likely to be using drugs. Consideration would have to given to the inclusion of those in possession of drugs for sale or supply, particularly if 'user-dealers' were prominent.

There is a free text field that can be used to identify which drugs the individual had been detained in connection with, and care must be taken to identify all relevant records (i.e. heroin records and diamorphine records). There are various fields that describe the location, although the location where the crime was committed may be more easily extracted than the place where the individual actually lives. In terms of identifier information, initials, dates of birth and sex/gender can easily be extracted.

---

1   http://www.drugmisuse.isdscotland.org/publications/local/prevalence_projectupdate.pdf

Data available and possible application

If it was possible to access data on area of residence then this data could be used as part of a capture-recapture analysis. The data also lends itself to application as an indicator in a multiple indicator analysis it could also be used as a benchmark when deriving estimates using the multiplier method.

2.4 Probation Service Data

Description

The probation service in Ireland works with offenders to reduce criminal behaviour through assessment and supervision orders, community initiatives, community service and young peoples' services. As part of this scoping exercise we have met with members of the service and have been given an overview of their work and their data systems. We have paid particular attention to the type of data they collect relating to offenders substance misuse.

In 2009 the probation service completed 8,273 assessments for the courts and supervised approximately 11,000 offenders in the community. Data is available from four main forms administered over the course of an offender's contact with probation:

■ Form 'A '– Referral form

■ Form 'B' – Decision & Adjournments

■ Form 'C' – Case closure

■ LSI-R – Risk assessment.

Form 'A' is completed when an offender arrives at court and is a general referral into the system. It gives details including the offender's age, gender, probation region and type of offence committed. 'B' forms are completed sometime after referral and are used by judges to gauge offenders' progress and thus make decisions about how to proceed. An offender can have a number of 'B' forms completed on them relating to a particular offence during the course of their contact with probation. The 'C' form is completed when a client ceases contact with probation. It gives details on the date and reason for cessation of contact. The fourth form is the LSI-R (Level of Service Inventory-Revised) or risk assessment. This form is administered along with the 'A' form at the point of referral and is repeated every six months. The LSI-R monitors clients' responses in ten key areas including; criminal history, education/employment, financial, family/marital status and accommodation. Crucially for this study, the LSI-R asks offenders about their drug use. They are asked about their current use and to name their main drug of use. They are also asked if they have committed a drug related offence in the past 12 months.

Data available and possible application

Following discussions with probation service staff the study team believe that data held by the service, in particular the LSI-R data, could be used in a study to estimate the prevalence of opiate use. The identifier data together with the drug information collated by the LSI_R form mean that it should be considered as a data source for a capture-recapture study. Aggregate data could also be used in a multiplier analysis. The study team are cognisant of the fact that the LSI-R completion rate is around

85%; however, as long as the sample provided by the LSI-R data is representative this should not preclude its use in a prevalence estimation study. Another possible issue with this data is the relationship with HSE information. It is probably the case that a large number of offenders in the probation system are already known to treatment agencies. If it were the case that all drug using offenders were known to treatment agencies then the probation service data would be a subset of the treatment data; this would mean that it could not be used as an independent source of data in a CRC study. At the time of writing it is not clear if this is the case and it would require further investigation to reach a conclusion.

Another potentially useful aspect of the probation service data is the information held on those about to leave prison under a supervision order. This means that the probation data could be split into two sources; one a community sample, the other a prison sample. As part of previous prevalence estimation work undertaken by this study team in England probation data has been exploited in this manner and it has been a useful way to maximise the data available. We would recommend that this be explored in the context of the Irish probation data.

2.5 Prison Service Data

Description

The Irish Prison Service (IPS) works with the HSE and other services to provide healthcare facilities for prisoners in Irish jails. At reception into prison all inmates are medically assessed within a specified timeframe and their healthcare needs noted. This could involve continuing care that was initiated in the community, prescribing or referral to specialists if required. These assessments are carried out by a team of 20 doctors employed by the IPS on a full or part time basis. There are also specialist nurse officers working as part of Dedicated Drug Treatment Teams to provided treatment to prisoners with substance misuse issues. The study team would like to explore any data pertaining to substance misuse that may be collated by the doctors at assessment and by the Dedicated Drug Teams. The team would also like to investigate the link between data held by the probation service and the prison data.

Data available and possible application

IPS data could be used to augment other criminal justice data as part of a capture-recapture analysis. This data could also be used as a benchmark when applying the multiplier method. If sufficient data was available at a geographical level it may be possible to use it as an indicator in a multiple indicator study.

2.6 National Drug Treatment Reporting System (NDTRS)

Description

The National Drug Treatment Reporting System is an epidemiological database on treated problem drug use in Ireland. The NDTRS is co-ordinated by staff at the Drug Misuse Research Division (DMRD) of the Health Research Board (HRB). Since 2004, compliance with the NDTRS requires that the total number of cases in treatment with each treatment provider on 1 January each year is returned to the

DMRD and that one form is completed for every case that commences or returns to treatment for problem alcohol or drug use at each treatment centre. Data forms are submitted quarterly. Service providers at drug treatment centres throughout Ireland collect and submit this data. They collect administrative details, demographic characteristics, parameters to measure access to treatment, treatment status (new versus previously treated cases), problem substance use in the month preceding this treatment contact, risk behaviours (associated with injecting drug use) and initial treatment type. The NDTRS does not collate initials therefore may be of limited use within a capture-recapture study. It does, however, collate dates of birth and gender. The drug treatment centres can record the individual's name on their copies of the reporting forms and, if those forms are stored systematically at the local level, they could be used to create an additional source of data within a capture-recapture study. This would require contributing drug treatment centres to allow researchers access to such information.

At the national level staff, at the DMRD of the HRB, compile anonymous, aggregated data. For the purpose of the NDTRS, treatment is broadly defined as 'any activity which aims to ameliorate the psychological, medical or social state of individuals who seek help for their drug problems'. Treatment options include one or more of the following: brief interventions, addiction counselling, medication-free therapy, alternative therapy, psychiatric treatment, medication (detoxification for problem alcohol, opiate or benzodiazepine use, methadone substitution), and social and occupational reintegration. Clients who attend needle-exchange services are not included in this reporting system. Treatment is provided through residential and non-residential services. Drug treatment data are viewed as an indirect indicator of drug misuse as well as a direct indicator of demand for treatment services.

Data available and possible application

In terms of prevalence estimation, the information from the NDTRS could be used to produce different multipliers for use in multiplier method analyses, for example the proportion who are in receipt of methadone. This proportion could possibly be used in conjunction with aggregate data from the Central Treatment List within a multiplier method analysis.

Although matching only on date of birth and gender (within a particularly geographical region) may not be sufficient to ascertain a definite match it may still be possible to use the NDTRS data within some kind of capture-recapture analysis. However, as part of an unpublished sensitivity analyses carried out within a Home Office funded study by Hay et al this type of matching was shown to potentially under-estimate prevalence. Taking into account this reservation the NDTRS data should still be used within the prevalence analyses. This is, in part, due to the high quality of data recorded in that system and other related data in Ireland. Sensitivity analyses can be easily carried out using NDTRS data and/or other data sources to see how many false positive matches would be expected if matching was done on date of birth and gender only. It would then either be possible to adjust the overlap pattern accordingly, or adjust the resultant estimate accordingly. It may, however, be that matching on only date of birth and gender would be sufficient in some areas outside of Dublin where the numbers of drugs users may be expected to be fewer.

2.7 National Psychiatric Inpatient Reporting System (NPIRS)

Description

The National Psychiatric Inpatient Reporting System is a national database which contains information on all individuals admitted to and discharged from psychiatric units and hospitals in Ireland. It is administered by the Mental Health Unit (MHU) based within the Health Research Board and has it's origins in a census of psychiatric institutions that was initiated in the 1960s.

Information recorded within the database includes demographic data such as the patient's address, gender, date of birth and marital status. The database does not contain any details of patients' first or second names but does include a patient number that is allocated by individual hospitals. In the case of most hospitals this patient number is a unique identifier that will be used again if a patient is re-admitted after a previous discharge. A small minority of hospitals do not repeat the use of patient numbers instead using different numbers for different in-patient episodes. The address data collected records the last three lines of an address, excluding the house name or number but including county of residence. Address data refers to patients' accommodation just prior to admission which could relate to another hospital or temporary lodgings such as a hostel. Contributing institutions include all units and hospitals listed on a register of approved centres established under the Mental Health Act of 2001. There are approximately sixty approved centres on this list including eight private facilities. All of these institutions provide inpatient services only. Publicly funded approved centres submit data on a quarterly basis whilst annual data is provided by private facilities and child/adolescent units. Almost all institutions submit data to the MHU electronically. The remainder, provide their data in hard copy which is then entered into the NPIRS database.

Institutions also provide information on primary and secondary diagnoses, at both admission and discharge. This data is recorded with reference to diagnostic groups that are based upon ICD-10 diagnosis codes. Substance use is categorised under the diagnostic group 'other drug disorders'; and employs the codes F11 to F19. In 2009, there were 824 people recorded as having a substance use related primary diagnosis.

A secondary diagnosis is not often reported to the database, though there is uncertainty as to whether this accurately reflects the presence of a secondary diagnosis.

Data available and possible application

The NPIRS is an important source because it can provide aggregate data on individuals with drug use disorders presenting to inpatient psychiatric institutions across Ireland. This dataset could therefore be useful in a prevalence study employing the multiplier method, which requires simple count data related to certain factors associated with drug misuse. For example, the number of people at risk of admission to a psychiatric unit or hospital for a disorder related to opiate dependence could be calculated using the actual number of relevant individuals multiplied by a known proportion of psychiatric inpatients that display opiate dependence. Moreover, as the NPIRS address data is quite detailed, including as it does town and county details as well postal districts for Dublin, it increases the opportunities for producing prevalence estimates for different administrative levels, particularly within a multiple indicator analysis.

The NPIRS would probably be unsuitable for use within study using the capture-recapture method, primarily because it does not contain the first and second names of individual patients. However if the proposed sensitivity analyses looking at the effect of matching in the absence of initials suggests that the NDTRS data could be used in some kind of adapted capture-recapture analysis then the same

might hold for NPIRS data. As with the NDTRS, the identifier information may be available within the institutions contributing to the NPIRS but negotiating access to, and collecting this data from each unit and hospital may introduce delays and higher study costs.

2.8 National Drug-related Deaths Index (NDRDI)

Description

The National Drug-related Deaths Index is a census of drug-related deaths (such as those due to accidental or intentional overdose and deaths among drug users (such as those due to hepatitis C and HIV) in Ireland. The data is collected from a number of sources, including the Coroner Service, the General Mortality Register, the Central Treatment List and HIPE.

Data available and possible application

It can perhaps be used as a distinct data source in a capture-recapture analysis (since full identifiers can be extracted) however care must be taken in selecting a relevant time period which will not violate some of the key methodological assumptions. One approach would be to use drug-related death data for 2010 and use data pertaining to 2009 for all other data sources. These data can also be used within a mortality multiplier analysis. While such an analysis may not be as statistically robust as a capture-recapture analysis, it may be of particular use in comparing prevalence across different areas of Ireland or assessing the validity of estimates derived for areas outside of Dublin.

One of the more interesting potential uses of the NDRDI takes advantage of the fact that it already cross-references between contributing sources, some of which are directly applicable for using within a capture-recapture analysis such as HIPE or the Central Treatment list. It may be possible to either use the NDRDI within various different multiplier method analyses, or even directly within a capture-recapture analysis.

2.9 Applicable data sources

Table 1 below summarises the information available from each of the data sources detailed in this section. If also indicates the suitability of the data with regard to its use in a particular analysis. From the table it is clear that a number of the data sources could be used as part of all three preferred methodologies. If researchers were to apply the capture-recapture method then we would recommend using four large, representative data sources;

■ Central Treatment List

■ Hospital In-patient Enquiry System

■ Garda data

■ Probation service data combined with relevant prison data.

Should a capture-recapture analysis not produce robust estimates for more rural areas then multipliers could be used to provide estimates for these areas. An alternative way to provide estimates for areas where capture-recapture may fail is to apply the multiple indicator method. The estimates from the capture-recapture analysis would be used as anchor points and the other applicable data could be input as indicators. This would however rely on the aggregate data being available at a suitable geographical level.

Table 1: Comparison of possible data sources for inclusion in an Irish study to estimate the prevalence of problematic opiate and/or cocaine use

Source	Information available	Suitable for Multiplier analysis	Suitable for CRC analysis	Suitable for MIM analysis	Suitable for downloading to an Excel File
CTL	■ Initials ■ DOB ■ Gender ■ Geographical data ■ Contact date	✔	✔	✔	✔
HIPE	■ Initials ■ DOB ■ Gender ■ Geographical data ■ Contact date ■ Drug	✔	✔	✔	✔
PULSE (Garda data)	■ Initials ■ DOB ■ Gender ■ Geographical data[1] ■ Contact date ■ Drug	✔	✔	✔	✔
Probation Service data	■ Initials ■ DOB ■ Gender ■ Geographical data ■ Contact date ■ Drug[2]	✔	✔	✔	✔
Prison Service data	?	?	?	?	?
NDTRS	■ Initials[3] ■ DOB ■ Gender ■ Geographical data ■ Contact date ■ Drug	✔	?	✔	✔
NPIRS	■ DOB ■ Gender ■ Geographical data ■ Contact date ■ ICD 10	✔	✘	✔	✔
NDRDI	■ Initials ■ DOB ■ Gender ■ Geographical data ■ Contact date ■ Drug	✔	✔	✔	✔

1 There may be issues with the geographical data as it is easier to note where the offence was committed not necessarily where the offender is resident
2 The drug data would be collated from the LSRI forms and linked to the identifiers from the referral form
3 Contributing treatment agencies would have to be approached to provide clients initials

Section 3: Further information on favoured methodologies

This section focuses on the three methods which would be most applicable to the data sets available in Ireland. More detailed examples are given of all three methods and some of the procedural issues that can arise during their application are discussed.

3.1 Multiplier Methods

As outlined previously in section 1, the multiplier method is a simple way of estimating unknown populations such as prevalence of drug misuse. The method uses the available information on the population in question as a benchmark, e.g. number of drug related deaths, and applies a multiplier that is related to the population and has normally been derived from a small scale study.

An example of one such multiplier is shown below in order to demonstrate how to construct a multiplier and how to then apply it to some benchmark data.

Construction of the mortality multiplier

In this example we will use the results of a qualitative study to construct a mortality multiplier. The Drug Outcome Research In Scotland (DORIS) study is a longitudinal study that recorded the number of drug-related deaths within its cohort of drug users over a certain period of time. Since the total number of drug users within the cohort is known the percentage of drug users who died in a year can be calculated:

0.746% of DORIS participants died in 2002.

Therefore we can say that for every one death there are roughly 134 drug users (100/0.746). We can now use this mortality multiplier to produce an estimate for the number of drug users in Scotland. From the Drug Misuse Statistics in Scotland report (2007), there were 248 heroin/morphine related deaths recorded in Scotland in 2002. Using the multiplier, it is estimated that there are 33,244 drug users in Scotland (248* 134.048) in 2002.

This example illustrates how simple it is to produce estimates using this method however it should be borne in mind that the benchmark data must match the multiplier in order for the estimate to be accurate. As previously discussed, the choice of benchmark data will effect the case definition. In the case of the Irish data there are a number of possible sources of benchmark data and it would be wise to utilise as much of this data as possible to produce a number of estimates that can be assessed and quality checked. Producing more than one estimate and looking at more than one data source allows us to assess the credibility of the estimates and to put them into context.

3.2 Multivariate Indicator Method

The Multiple Indicator Method (MIM) provides prevalence estimates for all areas by using known estimates of prevalence for a few areas, called anchor points, in combination with indicator data relating to problem drug use. Essentially this method uses indicator data to model prevalence using regression techniques. This relationship can be expressed as an equation which takes the form:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + c$$

Where y is the dependent variable and a1x1 etc are the explanatory variables and c is a constant term.

A very simple example of this would be the relationship between prevalence and treatment. This can be expressed as:

$$y = ax$$

Where y is the prevalence rate and x is the treatment rate. Figure 1 illustrates this relationship using data from the English prevalence study carried out by Hay et al. As part of this study estimates were produced for all 33 Drug Action Teams (DATs) in the London Government Office Region. After applying capture-recapture techniques the study team produced 27 estimates for the region. The six remaining DAT estimates were produced using MIM. Figure 1 is a plot of prevalence versus treatment; with prevalence on the y-axis and the treatment rate on the x-axis. All 27 CRC estimates are marked on the plot and the line of 'best fit' is calculated using a statistical package such as SPSS.

Figure 1: Prevalence rate versus treatment rate for London Drug Action Teams
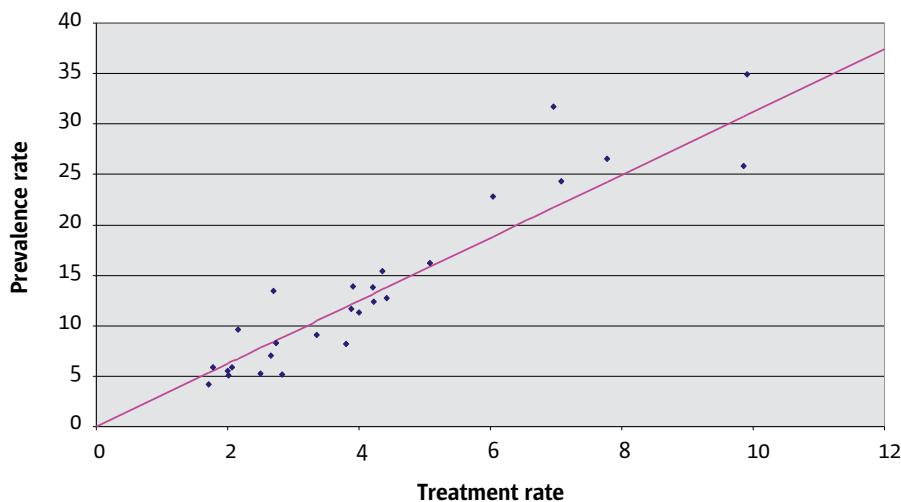
This 'line of best fit' or model can be expressed as:

prevalence rate = -1.227 + 3.500 x treatment rate

Since we know the treatment rate for the six remaining areas without estimates we can now use this equation to calculate the prevalence rate for these places. For example, the prevalence rate for Westminster is 6.56 per thousand population. Using this information in the equation gives:

prevalence rate = -1.227 + 3.500 x 6.56

prevalence rate = 21.73

This is the simplest model and is essentially a treatment multiplier. The actual model used in the final analysis included four indicators; treatment data, population density, prison data and police data on drug offences.

All available indicator data can be used to model prevalence using this method. However it must be at the correct geographical level and converted into rates, this must also be done with the anchor points. It is then a matter of carrying out the regression according to your preferred method. This regression can be carried out forwards, backwards or stepwise. Forward selection considers all available indicators but only add the most correlated to the model. Correlated indicators are only added to the model if they improve the fit. Backwards elimination begins with all indicators in the model and excludes those that are least correlated. The exclusion of indicators continues until the best fitting model is arrived at. Stepwise regression begins in a similar manner to forward selection but after an indicator is added to the model, checks are done to see if any other indicator should then be excluded. This means that in order to apply this method an inclusion and exclusion criteria must be decided upon. Due to these extra checks stepwise regression would be the recommended approach to modeling the data.

3.3 Capture-recapture Methods

This section gives more details of the application of the Capture-recapture method. A Capture-recapture study can be divided into three main stages; data processing (which includes access, collection and cleaning), analysis and finally model selection. In this part of the report we will discuss each stage in turn outlining the processes involved and issues that have occurred in previous analyses that those applying the method should be cognisant of.

To begin with we will examine the five assumptions that we must be aware of when applying the capture-recapture methodology:

■ The population should be closed i.e. no movement in or out of the population during the study period

■ Perfect matching i.e. the same individual should be identifiable across sources

■ Data sources should be representative

■ Everyone has an equal chance of appearing in any individual data source

■ Presence in one source does not influence presence in another.

These assumptions impact on the choice of data sources that the method can be applied to. Some assumptions have more of an effect than others; for instance the closed population one can be dealt with by keeping the study period sufficiently short. We know from the literature that individuals can use

opiates for a number of years therefore a 12 month period is normally a short enough time span to ensure that the population is static. There are other forms of prevalence estimation that attempt to model open populations (Jolly 1965) however the 12 month period has proved successful in a number of studies over the years and can simplify the analysis. This assumption places little demand on data sources other than one should be able to identify when an individual had contact with that source during the study period and it is clear that all data sources mentioned in table 1 could fulfil this criteria.

The perfect matching assumption rests on our ability to identify an individual to a satisfactory degree from the data fields we have permission to access. As discussed previously with the advisory group a unique identifier across all sources would be an easy solution to this issue. However the nature of data systems in Ireland and other countries across Europe has meant that this is not possible and researchers have had to develop other ways of identifying individuals across sources while protecting privacy and keeping in mind data protection issues. The following list of identifiers has been sufficient for identifying individuals across data sources at a local geographical level in previous capture-recapture studies;

- Forename initial
- Surname initial
- Date of birth
- Gender

For previous Scottish studies a portion of the postcode of residence was the geographical data unit of analysis. For the English study it was Drug Action Team area of residence. These geographical units of analysis were settled upon because they ensured a small enough sample that would not contain too many individuals with the same identifiers and yet would be large enough to make it difficult to work out a person's identity from the data fields involved. There are issues with using these types of identifiers. For example, Mary O'Leary born on the 16th May 1978 would appear in a CRC data source as MO1605782, with the 2 indicating her gender. Certain initials both forename and surname are more common than others for example surnames beginning with M, O or S. However initials in combination with the other information it should be enough to identify individuals across sources if there is sufficient geographical data. Another potential pitfall is an individual with different identifiers in different data sources. For example, if Mary presented for treatment at the beginning of the year as MO1605782 then married and took her partners' name and appeared in the HIPE data as MM1605782. This can also occur when people give their full name to one source and a nickname to another. The capture-recapture method has proven to be robust to a certain amount of this variability but this can cause difficulty when fitting models, therefore it is important for researchers to have a good working knowledge of the data sources involved so that they would be able to spot these issues.

The third assumption that data sources should be representative can mean that certain data is excluded from the outset for example drug treatment agencies that specialise in women only or young people only. If unrepresentative data sources are used in an analysis this would make model fitting difficult and cast doubt on the credibility of the resultant estimates.

The fourth assumption can prove difficult to uphold. It relates to equal probability of capture. If service provision is uniform across an area then every individual has an equal chance of appearing in the data however if service provision is patchy then this assumption can be violated to some extent. Again this means that researchers should make every effort to become well acquainted with the prospective data sources.

The last assumption, that of presence in one source impacting on presence in another can be relaxed using log-linear modelling. This modelling is an attempt to account for the relationships that we know exist between data sources and therefore can result in a final model that will accurately reflect the flow of data between sources and in turn yield more robust estimates.

Outline of procedures

Data access, collection and processing

Once the prospective data sources have been decided upon there are a number of ethical and procedural permissions that must be received prior to the commencement of data collection. These issues can be time consuming and have hampered previous studies. It is vital that the study team and in particular the advisory group work closely with prospective data providers to ensure that data access issues are dealt with in a timely fashion. This may require the setting up of new procedures.

Once formal permissions have been received the data collection process can proceed. It is vital that researchers plan for not only the data extraction but also its safe passage from provider to researcher whether that transfer is electronic or person to person. This important step can be overlooked but it is potentially where the data is most at risk. Although the data described above is non-identifiable it is preferable that it is encrypted; this can be done on site when the data is being collected.

After collection data must be cleaned within source to remove irrelevant and invalid records. Researchers must then decide on their approach to incomplete data and be consistent in how they process it or remove it. Once this has been done any duplicates within source must be removed.

Matching

Matching across data sources in order to identify individuals that appear in more than one source is crucial to the success of the study. Once the matching criteria are decided upon the process is straight forward and can be accomplished with little trouble by a number of different software packages. Researchers can match across all of the identifiers provided; this is referred to as 'hard matching' or can decide to only match over a select number of identifiers 'soft matching'. It may be the case that data quality issues with one source forces researchers to use soft matching; if for example the forename initials are largely incomplete for one data source. Hay et al have looked at the issues of hard and soft matching as part of the Home Office study. They concluded that using surname initial instead of both initials did not greatly affect the resultant estimates but using neither initial did tend to result in underestimates. Once the overlaps between sources have been calculated a Venn diagram similar to the one below can be produced.

Figure 2: Venn diagram of three-sample capture-recapture data (Dundee 1994)
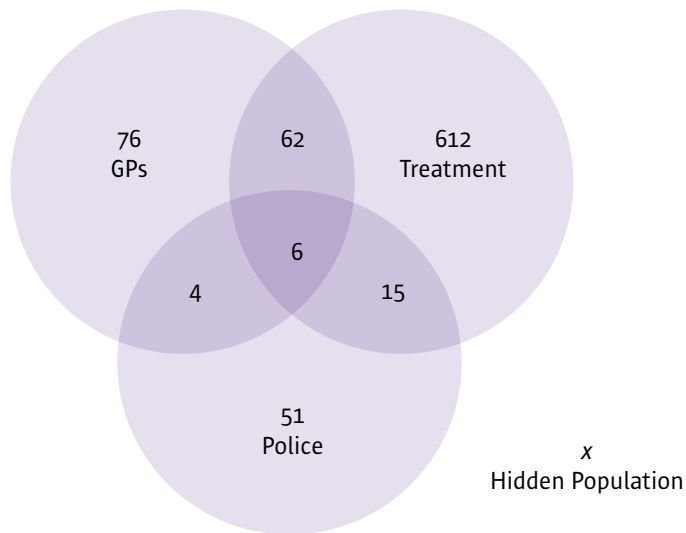
Figure 2 gives the overlap pattern for data from Dundee in 1994. From this we see that researchers found six individuals in all three sources, 10 in GP and police and 612 in treatment only. This data can also be represented in the form of a contingency table, with the numbers of individuals present or absent noted in each cell as in table 2 below.

Table 2: Contingency table summarising data from a Capture-recapture study (Dundee 1994)

		Treatment			
		Present		Absent	
		GPs			
		Present	Absent	Present	Absent
Police	**Present**	6	15	4	51
	Absent	62	612	76	-

From this table we see that 21 are present in both police and treatment and 51 are absent in both treatment and GP but present in police data. The final arrangement to the data in table 3 is the format it takes for input into the statistical package for analysis.

Table 3: Data table summarising data from a capture-recapture study (Dundee 1994)

Treatment	GPs	Police	Count
1	1	1	6
1	1	0	62
1	0	1	15
1	0	0	612
0	1	1	4
0	1	0	76
0	0	1	51
0	0	0	-

In table 3 above the data is in one column to the right, the sources are represented by each column to the left and presence or absence is noted by a 1 or a 0. The final cell in the data column is left blank, this represents the number in none of the sources and this is what we will use the overlap pattern to estimate.

For a three-sample capture-recapture analysis there are 8 possible models that can be fitted, each one describes the different relationships between the sources. Although we have anecdotal information on how the data sources interact with one another we cannot be sure this is the underlying truth of the situation. Instead of fitting just the models that describe what we believe is the situation we should be data led and fit all of the models. This way we can use a combination of statistical measures of fit and credibility checks to select the best model. All eight models can be expressed by the following formulae:

$$\text{Constant} + p1 + p2 + p3$$

$$\text{Constant} + p1 + p2 + p3 + p1*p2$$

$$\text{Constant} + p1 + p2 + p3 + p1*p3$$

$$\text{Constant} + p1 + p2 + p3 + p2*p3$$

$$\text{Constant} + p1 + p2 + p3 + p1*p2 + p1*p3$$

$$\text{Constant} + p1 + p2 + p3 + p1*p2 + p2*p3$$

$$\text{Constant} + p1 + p2 + p3 + p1*p3 + p2*p3$$

$$\text{Constant} + p1 + p2 + p3 + p1*p2 + p1*p3 + p2*p3$$

Where p1, p2 and p3 are the sources and * denotes an interaction between two sources. Note that the first equation has no * term this is the simplest model which assumes independence. The final equation is known as the saturated model and assumes there are interactions between all sources.

Table 4 lists the estimates that resulted from the Dundee 3-sample analysis. As you can see there is a lot of variation with the lowest estimate 716 and the highest 3598.

Table 4: All eight prevalence estimates and models for Dundee with corresponding 95% confidence intervals, deviance and degrees of freedom

Model	Estimate	Lower	Upper	Deviance	df
constant+p1+p2+p3	921	699	1214	13.78	3
constant+p1+p2+p3+p1*p2	1530	943	2482	6.91	2
constant+p1+p2+p3+p1*p3	716	514	996	6.52	2
constant+p1+p2+p3+p2*p3	966	726	1286	11.72	2
constant+p1+p2+p3+p1*p2+p1*p3	969	342	2748	6.12	1
constant+p1+p2+p3+p1*p2+p2*p3	2081	1164	3721	0.85	1
constant+p1+p2+p3+p1*p3+p2*p3	750	531	1059	5.39	1
constant+p1+p2+p3+p1*p2+p1*p3+p2*p3	3598	912	14201	0.00	0

Figure 3 gives the estimates with their associated 95% confidence intervals. Again there is much variation in the intervals as well as estimates. The five estimates that come from models that are thought to adequately fit the overlap pattern (and thus provide statistically sensible estimates) are shown in the lighter (green) shading. Model 8 (which is the saturated model that has all possible interactions included) is by far the biggest, and as it uses all available information, the model has a deviance value of zero. Figure 3 perhaps demonstrates one of the main difficulties of relying on the three-sample analysis; there are different estimates that come from models that appear to fit the data and these estimates can differ quite considerably. Different measures of how well the statistical model fits the available overlap pattern (such as the AIC or lowest deviance) would suggest that different estimates were the best. In short, it is a bit 'hit or miss' as to whether any of those estimates is the best estimate. A further issue is demonstrated by looking at the confidence intervals. The confidence interval for estimate 8 (the saturated model) is extremely wide. If, therefore, that model/estimate was deemed to be the best (or indeed the only model to fit the data) then the size of the confidence interval may make the usefulness of the estimate questionable.

Figure 3: All eight estimates with corresponding 95% confidence intervals
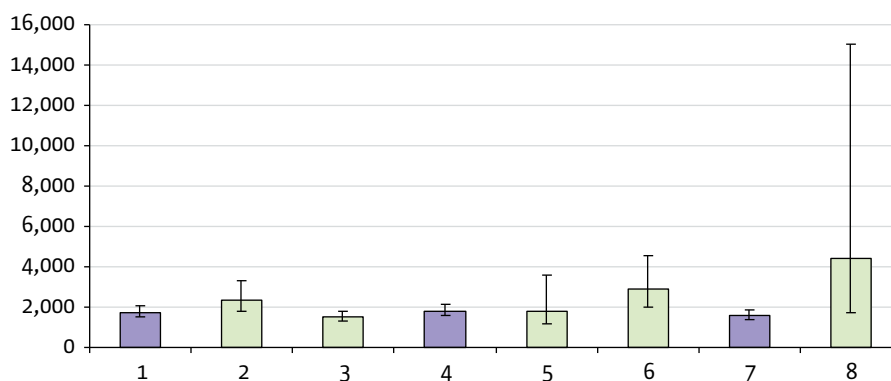
Although this issue can and sometimes does happen when there are four sources available, Figures 4 and 5 provide some of the results from a typical four-sample analysis (derived from a study using data for 2008/09 in Barking & Dagenham, United Kingdom, extracted from Hay et al 2010). In Figures 4 and 5 only the simplest 22 models are presented. All of these models provide a good enough fit to the data. In contrast to the three-sample analyses described above, the estimates seem more similar and the additional information gained from having at least 22 models (or at most 114 models) to consider would give a study more confidence in selecting any individual model/estimate as the best estimate.

Figure 4: Various estimates from a four-sample capture-recapture analysis.
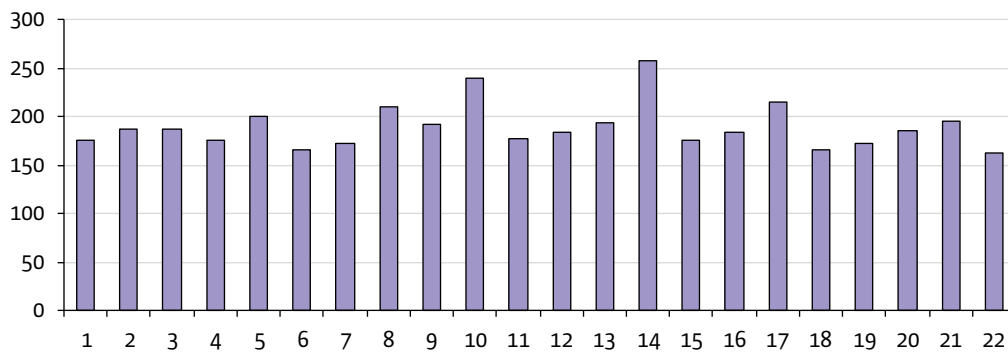
Figure 5: Various estimates and associated 95% confidence intervals from a four-sample capture-recapture analysis.
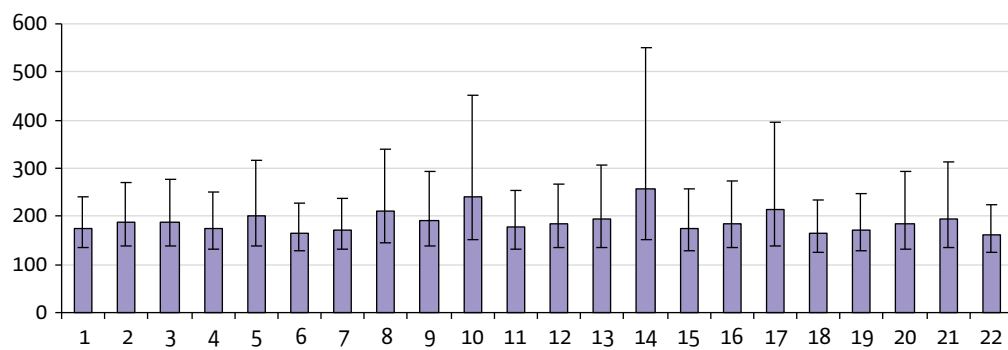
In the three sample analysis, the saturated model is often the best fitting model and this model can offer estimates that are just not credible. It is rare that the saturated model would be seen as the best estimate within a four-sample capture-recapture analysis. This is an additional benefit of using four sources within the analyses.

It is important for a fourth data source to be found to use within a capture-recapture analysis. If a suitable fourth data sources cannot be found, then consideration should be given to splitting a single

data source into two data sources that cover two distinct time periods. While this is a bit of a compromise, it has been successfully used in other areas, in particular Barcelona (Domingo-Salvaney et al, 1998)

Whether we have decided to use three or four sources there are a number of statistical criteria that can assist us in identifying a good fitting model. One way of identifying a good fitting model is examining the deviance value compared with the degrees of freedom. The lower the deviance compared to the chi-squared value for those degrees of freedom the better the fit. Two of the most popular measures of goodness of fit are Akaike's Information Criterion (AIC) and Schwarz's Information Criterion (SIC) and both are based on some manipulation of the deviance value and the degrees of freedom. The AIC can be expressed as follows:

$$AIC = G^2 - 2(df)$$

Where G^2 is the deviance and df the degrees of freedom associated with the model. An AIC value less than 5 or ideally less than 0 indicates a good fitting model.

Schwarz's Information Criterion is calculated using the following equation:

$$SIC = G^2 - ((\ln(n)) * (df))$$

Where $\ln(n)$ is the natural log of n the known data. Again a low SIC value indicates a well fitting estimate. Another way of judging the relative merit of an estimate is by comparing it to a weighted estimate constructed using the SIC and other estimates, this can be done using the following formula:

$$\sum_{\text{mod}\,el=1}^{n} (\exp(-1(SIC_n/2) * est_n) \Big/ \sum_{\text{mod}\,el=1}^{n} \exp(-1(SIC_n/2))$$

A good fitting model will be close to this weighted estimate.

Stratification

When conducting a capture-recapture analysis it may not be possible to fit a model to the data. In the case of some rural areas there may be a lack of data but there can also be issues with model fitting in areas with an abundance of data. From the literature we know that drug users can turn up in different sources depending on their gender or age. For example, younger users are more likely to turn up in criminal justice sources. This introduces heterogeneity into the sample and makes it difficult to describe it using a mathematical model. This can be dealt with by splitting or stratifying the data into more homogenous groups either by gender or age group or both gender and age group. When estimating the prevalence of problem drug use in Scotland Hay et al stratified the data in a number of areas by gender and by the following EMCDDA chosen age groups, 15-25, 25-34, 35-64. One disadvantage of stratifying the data is the added complication of combining confidence intervals.

Model selection is not an exact science and the statistical measures of how well the model fits the data must be tempered by knowledge of the local drug problem so that the credibility of the estimate can be assessed.

3.4 Specific data collection issues

The study team have recommended applying a mixture of methods to produce problem drug use estimates in Ireland; CRC, MIM and multiplier methods. As outlined earlier in this section these methodologies require slightly different types of data. In this section specific examples of the types of data required are shown and the possible demands on prospective data providers are discussed.

Example of multiplier method data

The data required for the multiplier method is a count of individuals in a particular geographical area. This data, unlike CRC data does not include identifiers. Table 2 below gives is an example of this type of data; it is mortality data for Scotland by Health Board area.

Table 5: Drug-related deaths by heroin/morphine by NHS Board Area

NHS Board	2007
Ayrshire & Arran	12
Borders	1
Dumfries & Galloway	6
Fife	19
Forth Valley	18
Grampian	34
Greater Glasgow & Clyde	109
Highland	12
Lanarkshire	32
Lothian	20
Orkney	-
Shetland	2
Tayside	16
Western Isles	-
Scotland	**289**

This data can be provided in a simple, widely used format such as excel or a 'csv' file if preferred. The study team believe that it would not take an inordinate amount of time for this data to be extracted once the geography was decided upon. Once the data was extracted arrangements could be made for its secure transfer to the research team.

Example of capture-recapture data

The data fields required for a CRC study are as follows:

- Contact date
- Forename initial
- Surname initial
- Date of birth
- Gender
- Area of residence
- Drugs used

Initials, data of birth and gender are the fields that are essential for matching. Contact date is needed to ensure the individual was in contact with the data source for the time period under investigation. Some indication of the area of residence, as outlined in section 3.3, helps to give robust estimates at a sub-national level and an indication of the type of drug used means that we will only include the type of problem drug user we are interested in e.g. opiate users. All four data sources that we have recommended using hold this data in some format however it may require some manipulation before it is in the format needed for analysis. For example, many data sources hold full names therefore the initials would have to be extracted. Table 3 shows some simulated data in a format suitable for CRC analysis.

Table 6: Data in format for use in CRC study

Contact Date	Forename	Surname	Sex	DOB	County of Residence
19/01/1994	R	M	1	11/01/1957	Dublin
02/02/1994	F	P	2	14/05/1970	Dublin
04/02/1994	M	M	1	22/03/1962	Kildare
04/02/1994	J	R	1	06/07/1949	Meath
16/02/1994	K	D	1	10/12/1978	Dublin
16/02/1994	R	H	1	12/04/1984	Dublin

Three of the four recommended data sources have provided this type of data for prior prevalence studies using CRC so we believe they would be able to provide it again possibly within a similar time frame. However the probation data has not been used in this type of study before. The data required for CRC is from two separate probation forms; the LSRI form (drug data) and the referral form (all other data) therefore researchers would require data from both and would have to link them. This could mean that probation data may require more time for data collection. Another point relevant to the probation data relates to completeness. The LSRI data is 85% complete however we do not know how complete the specific drug data field is. A CRC study is not a census and therefore does not need this field to be 100% complete however we would pay close attention to the coverage of the data.

Due to the sensitive nature of the data required for a CRC study and the manipulation it may need prior to its collection, the study team should be in close contact with data providers during the data access/collection process. In the past this research team has been willing to assist data providers in any manipulation of the data should they find that helpful however due to data protection concerns this would mean providing a space on site for the researcher to work in. Other researchers may have a different approach to this issue. For previous studies we have encrypted the data prior to receiving it, therefore no potentially identifiable data leaves the data collection site. Again, this is a standard part of our research procedures but may not be the case for other researchers. Our team can work with the data providers so that they can encrypt the data themselves or a researcher can visit the site and encrypt the data for them if this is preferable. Traditionally we have worked with data provided in excel format but if data providers prefer other formats any professional research unit should be able to accommodate their request. Ultimately researchers should strive to make the data collection process as low impact as possible for data providers.

Section 4: Estimating the prevalence of cocaine use

Estimating the prevalence of cocaine use has always presented a challenge for researchers. The nature of the drug means that it is difficult to ascertain the severity of an individuals' use from the type of data normally used to estimate prevalence. For example, when estimating the prevalence of opiate use researchers can be fairly certain that an opiate user appearing in treatment data has the same drug using profile as one present in police data. This is not so for cocaine use as someone using the drug 'recreationally' who does not believe they have a problem with the drug could turn up in a police source and a more problematic user could appear in the treatment source. Due to the level of data collated in these sources it is not possible to distinguish between problematic and recreational cocaine users in the police data. When applying the Capture recapture (CRC) method to this kind of data on cocaine use it is difficult to fit a model due to the heterogeneity that exists in the data. One possible way of dealing with this would be to isolate out any crack users since they are readily identifiable as problematic users and would therefore be a more homogeneous group across data sources. This theory can run into trouble again when using police data as most of the drug information is prepared in forensic laboratories that test the seized materials and only test for the presence of cocaine and do not mention in all cases whether the cocaine took the form of powder or rocks of crack. For areas with significant crack using populations CRC estimation is possible as Hay et al have demonstrated in previous studies. It should be borne in mind that the case definition used was those using crack cocaine and/or opiates, recognising the large number of problematic users that take both drugs. However the nature and extent of Irish cocaine use as highlighted in the report by the NACD 'An Overview of Cocaine Use in Ireland' (2007) would suggest that there are insufficient numbers of crack users to use the CRC method to estimate crack use and that some measure of powder cocaine use would be more informative for Irish policymakers and service providers.

With this in mind there are other approaches to estimating the number of cocaine users which take advantage of existing survey data. Hay et al have used these methods as part of a previous study to size the Scottish drug market and estimate the associated social and economic costs. When faced with the issue of estimating the size of the drugs market researchers had to deal with the entire market not just problematic users. This meant that when estimating the numbers of cocaine users they examined two separate groups, recreational and problematic. Information on recreational use was sourced from two large general population surveys, the Scottish Crime and Victimisation Survey and the Scottish School Adolescent Lifestyle Substance Misuse Survey. The Scottish Crime and Victimisation Survey (SCVS) now known as the Scottish Crime and Justice Survey (SCJS) is a large scale general population survey which monitors adults' experiences and perceptions of crime in Scotland. This survey is the Scottish equivalent of the British Crime Survey and has a specific section covering respondents' drug use. Survey respondents are 16 and over and the last sweep had 16,000 participants. The Scottish School Adolescent Lifestyle Substance Use Survey (SALSUS) looks at levels of smoking, drinking and drug use among young people in Scotland. The survey is carried out among a sample of Scottish schools focussing on pupils aged 13 to 15 years old. Previous surveys in this series have had sample sizes of around 34,000. Researchers used powder cocaine use data from these surveys to produce national estimates of recreational cocaine use. For estimating the number of problematic cocaine users the research team employed the results from the latest round of interviews from the Drug Outcome Research in Scotland (DORIS) study.

There were issues that needed to be addressed, including taking last month, last year or lifetime prevalence in the general population surveys, consistency of definitions across surveys and representativeness of the drug treatment cohort.

The same approach could be carried out in Ireland using the relevant general population surveys and the best available information on problem drug users. The Scottish study included young people as the main aim of the study was to estimate the total size of the Scottish cocaine market. It may be more appropriate to limit the analysis in Ireland to the adult population. It is difficult to recommend what the best source of information on problem drug users use of powder or crack cocaine would be, however it would be relatively straightforward to derive the relevant information from a range of sources such as any data held by HRB or data from any individual treatment service. Data from a survey been carried out in Irish Prisons may also be relevant, or a study being carried out by the Irish Probation Service. Once the information from a range of sources is available then a decision, based also on sample size, can be taken as to what is the best one. In Scotland the Scottish Drug Misuse Database (the treatment demand data) was also considered but it was thought to be affected by under-reporting of secondary or tertiary drug use.

Section 5: Conclusions and recommendations

There are five main recommendations for estimating the prevalence of opiate/cocaine use in Ireland:

1.   A combination of estimation methods should be used calculate the prevalence of opiate use in Ireland; with 4-sample capture-recapture the main methodology employed

The study team recommends using a mixture of three methods to estimate prevalence of opiate use in Ireland. A 4-sample capture-recapture analysis should be carried out for all 26 counties; where it is not possible to calculate an estimate using capture-recapture either MIM or multiplier methods should be used. We favour a 4-sample CRC analysis as this produces more robust estimates than a 3-sample analysis. The four samples we suggest using are:

- Central Treatment List
- HIPE
- Garda Síochána (PULSE)
- Probation Service data.

It is our understanding that, following preliminary enquiries, all four data sources recommended (including the newly proposed Probation Service data) can provide data in a format suitable for a CRC analysis. This approach would mean that each county would have a local estimate and estimates could be produced for other geographical areas of interest to stakeholders such as HSE region or probation area.

2.   Carrying out a pilot study using the recommended approach would not be an effective use of time and resources

Studies using capture-recapture methodologies to estimate opiate prevalence are generally composed of two phases; the data access and collection phase and the analysis phase. The first phase requires the most time and resources. This time and effort is not directly related to the amount of data being collected therefore carrying out a pilot study by collecting data for only a few areas would not significantly reduce the time or cost involved. Another reason to commission a pilot study would be to investigate specific issues relating to the data. Any possible issues with data can only be identified and dealt with once the data has been accessed and the analysis begins therefore a pilot study focussing on these issues would still require the same amount of time and effort as a full study but without the benefit of more comprehensive results. It is for these reason that we believe a full study would be the best use of resources.

3.   A geographical unit of analysis needs to be identified in order to carry out a successful study

In order for the CRC method to work successfully analyses need to be carried out at a sufficiently low level also stakeholders and in particular data providers would find local level estimates useful for planning purposes. It is also critical that this data refers to area of residence and not area of contact. For example, in Garda data this would be the area of residence of the offender not the area that the offence was committed in. Following roundtable discussions with a group of prospective data providers, county level was identified as the lowest geographical area of choice for reporting estimates. These county estimates could then be combined to give estimates at the level of HSE region or Garda region. The county estimates would also be summed to give an overall national

estimate. However it is clear from the administrative geography of Ireland that some counties for example Dublin and Tipperary are split between regions. This means that a sub-county analysis would have to be conducted for these areas.

4.   **In order for the study to be successful, timely and cost effective issues surrounding research governance need to be improved**

The data required to successfully match across sources in a CRC analysis are:

- Initials
- Date of Birth
- Gender.

Encryption can be used to transform this sensitive data into an unrecognisable code prior to transfer from data provider to researcher. However as with any research of this nature ethical approval is required. The different data providers will also need information on data security and the procedures for safe disposal of the data once the study is completed. Previous CRC studies carried out in Ireland have required multiple ethics applications to access HIPE data from a number of hospitals around that country. This can lead to significant delays and is a waste of time and resources for both the researchers and those assessing the study. The study team would recommend that the NACD together with data providers investigate the issues surrounding streamlining research governance. Prior to any future prevalence work there should be a concerted effort at awareness raising among stakeholders as to the nature and extent of prevalence estimation, the data collection it entails and the benefits of the resultant estimates.

5.   **Use a different method to estimate cocaine prevalence**

Due to the nature of powder cocaine use and therefore the way cocaine users can be identified through current Irish data sets we would recommend using a different method of prevalence estimation than that suggested for opiate use. For example, a heroin user that appears in the police data is in the same cohort as a heroin user appearing in the hospital episodes data however the same cannot be said of a powder cocaine user. This means that if CRC were employed to the cocaine data it would not be able to produce a valid estimate. Therefore we would recommend using similar methods to the ones outlined in section 4 of the report; using a combination of large household survey data with data from a longitudinal study of substance misuse to construct models that would better reflect the different levels of powder cocaine use among the Irish population.

References

Bello, P.-Y., and Chêne, G. (1997) A capture-recapture study to estimate the size of the addict population in Toulouse, France In EMCDDA Scientific Monograph Series, Number 1, *Estimating the prevalence of problem drug use in Europe*. EMCDDA, Lisbon

Bones, J., Thomas, K.V. & Paull, B (2007) 'Using environmental analytical data to estimate levels of community consumption of illicit drugs and abused pharmaceuticals' *Journal of Environmental Monitoring* 9: 701-707

Brugha, R. F., Swan, A. V., Hayhurst, G. K., et al. (1998) A drug misuser prevalence study in a rural English district. *European Journal of Public Health*, 8: 34-36.

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

Comiskey, C.M. (2001b) *'Methods for estimating prevalence of opiate use as an aid to policy and planning' Substance use and misuse* 36(1&2): 131-150

Comiskey, C.M., & Barry, J.M. (2001a) 'A capture-re-capture-study of the prevalence and implications of opiate use in Dublin' *European Journal of Public Health* 11(2): 198-200

De Angelis, D., Hickman, M. & Yang, S. (2004) 'Estimating Long term Trends in the Incidence and Prevalence of Opiate use/Injecting drug use and the Number of Former users: Back-calculation Methods and Opiate Overdose Deaths' *American Journal of Epidemiology* 160(10): 994-1004

Comiskey, C., Saris, J. & Pugh, S. (2007) 'Estimating the prevalence of opiate use in Ireland and the implications for the criminal justice system' Probation Journal 54(1): 22-35

Domingo-Salvany, A., Hartnoll, R. L., Maguire, A., et al. (1995) Use of Capture-Recapture to Estimate the Prevalence of Opiate Addiction in Barcelona, Spain, 1989. *American Journal of Epidemiology*, 141: 567-574.

Domingo-Salvany A, Hartnoll, R, Maguire, A, Brugal, T, Albertin, P, Caylà, J, Casabone, J and Suelves, J. (1998) Analytical considertations in the use of capture-recapture to estimate prevalence: Case studies of the estimation of opiate use in the metropolitan area of Barcelona. *American Journal of Epidemiology*, 148: 732-740.

Dunn, J. and C. P. Ferri (1999). "Epidemiological methods for research with drug misusers: review of methods for studying prevalence and morbidity." *Revista De Saude Publica* 33(2): 206-215.

European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). *'Methodological Guidelines to Estimate the Prevalence of Problem Drug Use on the Local Level'* Lisbon: EMCDDA, December 1998. http://www.emcdda.europa.eu/html.cfm/index58061EN.html

Frischer, M. (1992) Estimated prevalence of injecting drug-use in Glasgow. *British Journal of Addiction*, 87: 235-243.

Frischer, M., Leyland, A., Cormack, R., 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.

Frischer, M., Hickman, M., Kraus, L., Mariani, F & Wiessing, L (2001) A comparison of different methods for estimating the prevalence of problematic drug misuse in Great Britain. *Addiction*, 96: 1465-1476.

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, London: Drug Indicators Project.

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: 469-472

Hay, G. (1997) The selection from multiple data sources in epidemiological capture- recapture studies. *Statistician*, 46: 515-520.

Hay, G. (2003) 'Truncated Poisson Analyses of data from the Riga needle exchange.' The Synergy Project, Washington DC.2003.

http://www.fhi.org/en/CountryProfiles/BalticSeaRegion/res_needle.htm

Hay, G. & Gannon, M. (2006) 'Capture-Recapture estimates of the local and national prevalence of problem drug use in Scotland' *International Journal of Drug Policy* 17: 203-210

Hay, G., Gannon, M., MacDougall, J., Millar, T., Williams, K., Eastwood, C. & McKeganey, N. (2008) *'National and regional estimates of the prevalence of opiate use and/or crack cocaine use 2006/2007: a summary of key findings'* Research Report 9 London: Home Office

Hay, G., Gannon, M., Casey, J. and Millar T. (2010) *Estimates of the prevalence of opiate and/or crack cocaine use, 2008/09. Sweep 5 reports.* London, National Treatment Agency for Substance Use. Available at: http://www.nta.nhs.uk/uploads/glasgowprevalencestudysweep5-technicalreport2008-09[0].pdf

Jolly, G.M. (1965) Explicit estimates from capture-recapture data with both death and immigration – Stochastic model. *Biometrika* 52: 225-247

Kaspyrzk-Hordern, B., Dinsdale, R.M. & Guwy, A. (2009) 'Illicit drugs and Pharmaceuticals in the Environment – Forensic applications of environmental data. Part 1: Estimation of the usage of drugs in local communities' *Environmental Pollution* 157: 1753-1777

Kelly, A., Carvalho, M., & Teljeur, C. (2003) *'Prevalence of opiate use in Ireland 2000-2001: A 3 Source Capture-Recapture Study'* National Advisory Committee on Drugs, Stationary Office, Dublin

Kelly, A., Teljeur, C. & Carvalho, M. (2009) *'Prevalence of opiate use in Ireland 2006: A 3 Source Capture-Recapture Study'* National Advisory Committee on Drugs, Stationary Office, Dublin

Knolle, H. (1997) 'Incidence and Prevalence of Illegal Drug Use in Switzerland in the 1980s and 1990s: An Analytical Study' *Substance Use and Misuse* 32(10): 1349-1368

Korf, D.J., Reijneveld, S.A. & Toet, J. (1994) 'Estimating the number of heroin users: A review of methods and empirical findings from the Netherlands' *The International Journal of the Addictions* 29(11): 1393-1417

Kraus, L., Augustin, R., Frischer, M., Kummler, P., Uhl, A. & Wiessing, L. (2003) 'Estimating the prevalence of problem drug use at national level in countries of the European Union and Norway' Addiction 98(4): 471-485

Law, M. G., Lynskey, M., Ross, J., & Hall, W. (2001) 'Back projection estimates of the number of dependent heroin users in Australia' Addiction 96: 433-443

National Advisory Committee on Drugs (2007) *'An Overview of Cocaine Use in Ireland: II'* Dublin: NACD http://www.nacd.ie/publications/Overview_Cocaine_Use_2.pdf

Nolimal, D. (1996) 'Estimating the extent of the heroin problem in Slovenia: application of the key informant approach and the nomination technique where there are no other reliable sources of information' *Bulletin on Narcotics* 48(1-2): 121-134

Simeone, R. S., Nottingham, W.T., & Holland, L. (1993) 'Estimating the size of a heroin using population: A examination of the use of treatment admissions data' *International Journal of the Addictions* 28(2): 107-128

Smit, F., Toet, J. & van der Heijden, P. (1997) 'Estimating the number of opiate users in Rotterdam using statistical models for incomplete count data.' In European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) *Methodological pilot of local level prevalence estimates*. Lisbon: EMCDDA.

Smit, F., Van Laar, M. & Wiessing, L. (2006) 'Estimating problem drug use prevalence at national level: Comparison of three methods' *Drugs: Education, Prevention and Policy* 13(2) 109 -120

van Nuijs, A.L.N., Pecceu, B., Theunis, L., Dubois, N., Charlier, C., Jorens, P.G., Bervoets, L., Blust, R., Meulemans, H., Neels, H. & Covaci, A. (2009) 'Can cocaine use be evaluated through analysis of waste water? A nation-wide approach conducted in Belgium' *Addiction* 104: 734 -741

Wood, F., Bloor, M. & Palmer, S. (2000) 'Indirect prevalence estimates of a national drug using population: the use of contact-recontact methods in Wales' *Health, Risk and Society* 2(1) 47-58

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

Zucatto, E., Chiabrando, C., Castiglioni, S., Calamari, D., Bagnati, R., Schiarea, S. & Fanelli, R. (2005) 'Cocaine in surface waters: a new evidence based tool to monitor community drug abuse' *Environmental Health: A Global Access Science Source* 4: 14

Zucatto, E. & Castiglioni, S. (2009) 'Illicit drugs in the environment' Philosophical Transactions of the Royal Society A – Mathematical, Physical and Engineering Sciences 367 (1904): 3965-3978

Appendix 1: Three sample analyses using SPSS

This section describes how the SPSS statistical package can be used within the analysis of data from three sources within a capture-recapture analysis. SPSS is a powerful software package for data management and analysis. Various versions will be in use on different computer systems, however the analyses described in this report requires the equivalent of Version 6.1 for Microsoft Windows or a more recent version. The examples presented within this report were obtained by using Version 13.0 for Microsoft windows using the Loglinear analysis commands that are contained within the Analyse menu. Further information on the SPSS package can be found at http://www.spss.com

Log-linear modelling – the independent example

We first look at fitting a log-linear model in SPSS for the simplest case where all sources are assumed to be independent.

The first step is to enter the data into SPSS. The data should be entered in the following format (note in the data table you completed yesterday absence from a particular source was represented by a 0, however SPSS requires the presence and absence to be denoted by 1s and 2s, with 1s denoting presence:

p1	p2	p3	w	r
1	1	1	1	6
2	1	1	1	4
1	2	1	1	15
2	2	1	1	51
1	1	2	1	62
2	1	2	1	76
1	2	2	1	612
2	2	2	0	0

The column w is used as weights signifying that the 8th value in the series of counts is missing and that we want to weight it out of the analysis.

It may be convenient to define each variable p1, p2, p3, w and r as type 'Numeric' with 0 decimal places.

The variable which stores the seven counts, r, needs to be identified as such. This can be done by using the

    Data
        Weight Cases...
            Weight cases by

command, where the 'Frequency Variable' is selected to be r.

To find the estimate of the hidden population size, the parameter estimates for that model must be obtained. This is done by using the

Analyze
> **Loglinear**
>> **General**

command. Here again p1, p2 and p3 are entered into the **Factor(s)** box and w is entered into the **Cell Structure** box. The **Distribution of Cell Counts** should be set to be Poisson.

The **Model**... window should be used to enter the specific model we want to fit, in this case the independence model:

Specify 'custom' model.

Select 'main effects' from the drop-down menu found under 'Build Terms', then transfer p1, p2 and p3 over to 'terms in model' box and click continue.

Use the **Options**... window to specify that the **Estimates** are required. The plots that SPSS suggests are not required.

The complete output from such an analysis can be found at the end of this section. However the tables we are most interested in are shown below:

Parameter Estimates[b, c]

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	6.825	.141	48.402	.000	6.549	7.102
[p1 = 1]	-.414	.136	-3.050	.002	-.681	-.148
[p1 = 2]	0[a]	.	.	.	.	.
[p2 = 1]	-2.380	.120	-19.877	.000	-2.614	-2.145
[p2 = 2]	0[a]	.	.	.	.	.
[p3 = 1]	-3.090	.142	-21.787	.000	-3.368	-2.812
[p3 = 2]	0[a]	.	.	.	.	.

a This parameter is set to zero because it is redundant.
b Model: Poisson
c Design: Constant + p1 + p2 + p3

In SPSS the estimate of the hidden population size can be calculated as the exponential of the value of the constant parameter, which in this case is exp(6.825) = 921. A 95% confidence interval for the size of the hidden population can be obtained by calculating the exponentials of the vales SPSS gives as a confidence interval for the constant parameter (in this case the confidence interval would be 699 to 1214).

We now have an estimate for the hidden population of 921 with 95% confidence interval (699, 1214). It is necessary, however to get an idea of how well this model fits the data. To do this we look at the likelihood ratio:

Goodness-of-Fit Tests[a, b]

	Value	df	Sig.
Likelihood Ratio	13.784	3	.003
Pearson Chi-Square	14.002	3	.003

a Model: Poisson
b Design: Constant + p1 + p2 + p3

The terms 'likelihood ratio' is used in SPSS, whereas some other statistical packages call that term the 'deviance' or 'scaled deviance'. It is similar to (although not exactly the same as) a value from the χ^2 distribution with the given number of degrees of freedom (df) and the χ^2 distribution is commonly used to help in deciding whether this independence model does actually fit the data (and thus describe or account for any potential relationships between the three data sources). The 95% value for a χ^2 distribution with 3 degrees of freedom is 7.81 therefore, since the likelihood ratio (or deviance) attached to this model is far greater than 7.81, it may be sensible to conclude that the independence model does not adequately describe any potential relationships between data sources.

Syntax to carry out the same analysis

Instead of using the menu-driven commands, the following SPSS syntax can be used, after the data has been entered in the above format.

```
Weight by r.
Genlog
        p1 p2 p3
        /cstructure = w
        /model = Poisson
        /print estim
        /plot none
        /design p1 p2 p3. (or whatever the model you are trying to fit is)
```

Typical SPSS session that provides an estimate of the size of a hidden population using data from 3 sources

General Loglinear

Data Information

		N
Cases	Valid	8
	Missing	0
	Weighted Valid	826
Cells	Defined Cells	8
	Structural Zeros	1
	Sampling Zeros	0
Categories	p1	2
	p2	2
	p3	2

Convergence Information[a,b]

Maximum Number of Iterations	20
Converge Tolerance	.00100
Final Maximum Absolute Difference	.00059[c]
Final Maximum Relative Difference	.00106
Number of Iterations	6

a.   Model: Poisson

b.   Design: Constant + p1 + p2 + p3

c.   The iteration converged because the maximum absolute changes of parameter estimates is less than the specified convergence criterion.

Goodness-of-Fit Tests[a,b]

	Value	df	Sig.
Likelihood Ratio	13.784	3	.003
Pearson Chi-Square	14.002	3	.003

a.   Model: Poisson

b.   Design: Constant + p1 + p2 + p3

Cell Counts and Residuals[a,b]

p1	p2	p3	Observed Count	Observed %	Expected Count	Expected %	Residual	Standardized Residual	Adjusted Residual	Deviance
1	1	1	6	.7%	2.562	.3%	3.438	2.148	2.279	1.826
		2	62	7.5%	56.323	6.8%	5.677	.756	1.455	.744
	2	1	15	1.8%	27.676	3.4%	-12.676	-2.410	-3.486	-2.641
		2	612	74.1%	608.439	73.7%	3.561	.144	1.533	.144
2	1	1	4	.5%	3.877	.5%	.123	.062	.065	.062
		2	76	9.2%	85.238	10.3%	-9.238	-1.001	-2.657	-1.020
	2	1	51	6.2%	41.885	5.1%	9.115	1.408	2.605	1.362
		2	0	.0%	.000	.0%	.	.	.	.

a. Model: Poisson

b. Design: Constant + p1 + p2 + p3

Parameter Estimates[b,c]

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Constant	6.825	.141	48.402	.000	6.549	7.102
[p1 = 1]	-.414	.136	-3.050	.002	-.681	-.148
[p1 = 2]	0[a]	.	.	.	.	.
[p2 = 1]	-2.380	.120	-19.877	.000	-2.614	-2.145
[p2 = 2]	0[a]	.	.	.	.	.
[p3 = 1]	-3.090	.142	-21.787	.000	-3.368	-2.812
[p3 = 2]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

b. Model: Poisson

c. Design: Constant + p1 + p2 + p3

Correlations of Parameter Estimates[a,b,c]

	Constant	[p1 = 1]	[p2 = 1]	[p3 = 1]
Constant	1	-.958	-.715	-.578
[p1 = 1]	-.958	1	.650	.525
[p2 = 1]	-.715	.650	1	.392
[p3 = 1]	-.578	.525	.392	1

a. Model: Poisson

b. Design: Constant + p1 + p2 + p3

c. Redundant parameters are not displayed.

Covariances of Parameter Estimates[a,b,c]

	Constant	[p1 = 1]	[p2 = 1]	[p3 = 1]
Constant	.020	-.018	-.012	-.012
[p1 = 1]	-.018	.018	.011	.010
[p2 = 1]	-.012	.011	.014	.007
[p3 = 1]	-.012	.010	.007	.020

a. Model: Poisson

b. Design: Constant + p1 + p2 + p3

c. Redundant parameters are not displayed.

Appendix 2:
Sample ethics & data access forms

NHS REC Form   Reference: IRAS Version 3.1

Welcome to the Integrated Research Application System

IRAS Project Filter

The integrated dataset required for your project will be created from the answers you give to the following questions. The system will generate only those questions and sections which (a) apply to your study type and (b) are required by the bodies reviewing your study. Please ensure you answer all the questions before proceeding with your applications.

Please enter a short title for this project (maximum 70 characters)
Estimating the Prevalence of Problem Drug Use

1. Is your project research?

◉ Yes   ○ No

2. Select one category from the list below:

○ Clinical trial of an investigational medicinal product

○ Clinical investigation or other study of a medical device

○ Combined trial of an investigational medicinal product and an investigational medical device

○ Other clinical trial or clinical investigation

○ Study administering questionnaires/interviews for quantitative analysis, or using mixed quantitative/qualitative methodology

○ Study involving qualitative methods only

◉ Study limited to working with human tissue samples, other human biological samples and/or data *(specific project only)*

○ Research tissue bank

○ Research database

If your work does not fit any of these categories, select the option below:

○ Other study

2a. Please answer the following question(s):

a) Will you be taking new samples primarily for research purposes (i.e. not surplus or existing stored samples)?   ○ Yes   ◉ No

b) Will you be using surplus tissue or existing stored samples identifiable to the researcher?   ○ Yes   ◉ No

c) Will you be using only surplus tissue or existing stored samples not identifiable to the researcher?   ○ Yes   ◉ No

d) Will you be processing identifiable data at any stage of the research (including in the identification of participants)?   ○ Yes   ◉ No

e) Please confirm that you will be processing only anonymised or effectively pseudonymised data:

◉ Yes, only anonymised or pseudonmyised data   ○ No

3. In which countries of the UK will the research sites be located?*(Tick all that apply)*

☐ England
☑ Scotland

Date: 1

NHS REC Form   Reference: IRAS Version 3.1

☐ Wales
☐ Northern Ireland

3a. In which country of the UK will the lead NHS R&D office be located:

○ England
● Scotland
○ Wales
○ Northern Ireland
○ This study does not involve the NHS

4. Which review bodies are you applying to?

☑ NHS/HSC Research and Development offices
☑ Research Ethics Committee
☐ National Information Governance Board for Health and Social Care (NIGB)
☐ Ministry of Justice (MoJ)
☐ National Offender Management Service (NOMS) (Prisons & Probation)

5. Will any research sites in this study be NHS organisations?

● Yes   ○ No

6. Do you plan to include any participants who are children?

○ Yes   ● No

7. Do you plan at any stage of the project to undertake intrusive research involving adults lacking capacity to consent for themselves?

○ Yes   ● No

Answer Yes if you plan to recruit participants aged 16 or over who lack capacity, or to retain them in the study following loss of capacity. Intrusive research means any research requiring consent in law. This includes use of identifiable tissue samples or personal information, except where application is being made to the NIGB Ethics and Confidentiality Committee to set aside the common law duty of confidentiality in England and Wales. Please consult the guidance notes for further information on the legal frameworks for research involving adults lacking capacity in the UK.

8. Do you plan to include any participants who are prisoners or young offenders in the custody of HM Prison Service or who are offenders supervised by the probation service in England or Wales?

○ Yes   ● No

9. Is the study, or any part of the study, being undertaken as an educational project?

○ Yes   ● No

10. Will this research be financially supported by the United States Department of Health and Human Services or any of its divisions, agencies or programs?

○ Yes   ● No

Date: 2

NHS REC Form   Reference: IRAS Version 3.1

Integrated Research Application System
Application Form for Research limited to working with human tissue samples and/or data

NHS
National Patient Safety Agency
National Research Ethics Service

Application to NHS/HSC Research Ethics Committee

The Chief Investigator should complete this form. Guidance on the questions is available wherever you see this symbol displayed. We recommend reading the guidance first. The complete guidance and a glossary are available by selecting Help.

Please define any terms or acronyms that might not be familar to lay reviewers of the application.

Short title and version number: (maximum 70 characters - this will be inserted as header on all forms)
Estimating the Prevalence of Problem Drug Use

Please complete these details after you have booked the REC application for review.

REC Name:

REC Reference Number: **Submission date:**

PART A: Core study information

1. ADMINISTRATIVE DETAILS

A1. Full title of the research:

A3-1. Chief Investigator:

Title Forename/Initials Surname

Post
Qualifications
Employer
Work Address

Post Code
Work E-mail
* Personal E-mail
Work Telephone
* Personal Telephone/Mobile

Date: 3

NHS REC Form	Reference:	IRAS Version 3.1

Fax

This information is optional. It will not be placed in the public domain or disclosed to any other third party without prior consent.
A copy of a current CV (maximum 2 pages of A4) for the Chief Investigator must be submitted with the application.

A3-2. National coordinating investigator (for a multicentre trial) **or principal investigator** (for a single centre trial)

○ National coordinating investigator

○ Principal investigator

Given name
Family name
Qualification
(MD...)
Institution name
Institution department name
Street address
Town/city
Post Code

Country

Work E-mail
* Personal E-mail
Work Telephone
* Personal
Telephone/Mobile
Fax

This information is optional. It will not be placed in the public domain or disclosed to any other third party without prior consent.
A copy of a current CV (maximum 2 pages of A4) for the Chief Investigator must be submitted with the application.

A4. Who is the contact on behalf of the sponsor for all correspondence relating to applications for this project?
This contact will receive copies of all correspondence from REC and R&D reviewers that is sent to the CI.

Title Forename/Initials Surname

Address

Post Code
E-mail
Telephone
Fax

A5-1. Research reference numbers. *Please give any relevant references for your study:*

Applicant's/organisation's own reference number, e.g. R & D (if available):
Sponsor's/protocol number:
Protocol Version:

Date: 4

| NHS REC Form | Reference: | IRAS Version 3.1 |

Protocol Date:
Funder's reference number:
Project website:

A5-2. Is this application linked to a previous study or another current application?

◯ Yes   ◯ No

Please give brief details and reference numbers.

A5-3. US DHHS grant application.

PHS grant application number:
Name of Program Director:

2. OVERVIEW OF THE RESEARCH

To provide all the information required by review bodies and research information systems, we ask a number of specific questions. This section invites you to give an overview using language comprehensible to lay reviewers and members of the public. Please read the guidance notes for advice on this section.

A6-1. Summary of the study. *Please provide a brief summary of the research (maximum 300 words) using language easily understood by lay reviewers and members of the public. Where the research is reviewed by a REC within the UK Health Departments Research Ethics Service, this summary will be published on the website of the National Research Ethics Service following the ethical review.*

A6-2. Summary of main issues. *Please summarise the main ethical and design issues arising from the study and say how you have addressed them.*

A10. What is the principal research question/objective? *Please put this in language comprehensible to a lay person.*

A11. What are the secondary research questions/objectives if applicable? *Please put this in language comprehensible to a lay person.*

A12. What is the scientific justification for the research? *Please put this in language comprehensible to a lay person.*

A13. Please summarise your design and methodology. *It should be clear exactly what will happen to the research participant, how many times and in what order. Please complete this section in language comprehensible to the lay person. Do not simply reproduce or refer to the protocol. Further guidance is available in the guidance notes.*

4. RISKS AND ETHICAL ISSUES

RESEARCH PARTICIPANTS

A17-1. Please list the principal inclusion criteria (list the most important, max 5000 characters).

Date: 5

NHS REC Form   Reference: IRAS Version 3.1

A17-2. Please list the principal exclusion criteria (list the most important, max 5000 characters).

RECRUITMENT AND INFORMED CONSENT

In this section we ask you to describe the recruitment procedures for the study. Please give separate details for different study groups where appropriate.

A27-1. How will potential participants, records or samples be identified? Who will carry this out and what resources will be used? *For example, identification may involve a disease register, computerised search of GP records, or review of medical records. Indicate whether this will be done by the direct healthcare team or by researchers acting under arrangements with the responsible care organisation(s).*

INCENTIVES AND PAYMENTS

A47. Will individual researchers receive any personal payment over and above normal salary, or any other benefits or incentives, for taking part in this research?

○ Yes   ○ No

A48. Does the Chief Investigator or any other investigator/collaborator have any direct personal involvement (e.g. financial, share holding, personal relationship etc.) in the organisations sponsoring or funding the research that may give rise to a possible conflict of interest?

○ Yes   ○ No

NOTIFICATION OF OTHER PROFESSIONALS

PUBLICATION AND DISSEMINATION

A50. Will the research be registered on a public database?

○ Yes   ○ No

Please give details, or justify if not registering the research.

A51. How do you intend to report and disseminate the results of the study? *Tick as appropriate:*

☐ Peer reviewed scientific journals
☐ Internal report
☐ Conference presentation
☐ Publication on website
☐ Other publication
☐ Submission to regulatory authorities
☐ Access to raw data and right to publish freely by all investigators in study or by Independent Steering Committee on behalf of all investigators
☐ No plans to report or disseminate the results
☐ Other (please specify)

Date: 6

NHS REC Form	Reference:	IRAS Version 3.1

5. Scientific and Statistical Review

A56. How have the statistical aspects of the research been reviewed? *Tick as appropriate:*

☐ Review by independent statistician commissioned by funder or sponsor
☐ Other review by independent statistician
☐ Review by company statistician
☐ Review by a statistician within the Chief Investigator's institution
☐ Review by a statistician within the research team or multi−centre group
☐ Review by educational supervisor
☐ Other review by individual with relevant statistical expertise

In all cases please give details below of the individual responsible for reviewing the statistical aspects. If advice has been provided in confidence, give details of the department and institution concerned.

Title Forename/Initials Surname

Department
Institution
Work Address

Post Code
Telephone
Fax
Mobile
E-mail

Please enclose a copy of any available comments or reports from a statistician.

A57. What is the primary outcome measure for the study?

A58. What are the secondary outcome measures? *(if any)*

A59. What is the sample size for the research? *How many participants/samples/data records do you plan to study in total? If there is more than one group, please give further details below.*

Total UK sample size:
Total international sample size (including UK):
Total in European Economic Area:

Further details:

A60. How was the sample size decided upon? *If a formal sample size calculation was used, indicate how this was done, giving sufficient information to justify and reproduce the calculation.*

A61. Will participants be allocated to groups at random?

NHS REC Form   Reference: IRAS Version 3.1

◯ Yes   ◯ No

A62. Please describe the methods of analysis (statistical or other appropriate methods, e.g. for qualitative research) by which the data will be evaluated to meet the study objectives.

6. MANAGEMENT OF THE RESEARCH

A63. Other key investigators/collaborators. *Please include all grant co‑applicants, protocol co‑authors and other key members of the Chief Investigator's team, including non-doctoral student researchers.*

A64. Details of research sponsor(s)

A64-1. Sponsor

A64-2. Please explain how the responsibilities of sponsorship will be assigned between the co-sponsors listed in A64-1

A65. Has external funding for the research been secured?

☐ Funding secured from one or more funders

☐ External funding application to one or more funders in progress

☐ No application for external funding will be made

A67. Has this or a similar application been previously rejected by a Research Ethics Committee in the UK or another country?

◯ Yes   ◯ No

Please provide a copy of the unfavourable opinion letter(s). You should explain in your answer to question A6-2 how the reasons for the unfavourable opinion have been addressed in this application.

A68. Give details of the lead NHS R&D contact for this research:

Title Forename/Initials Surname

Organisation
Address

Post Code
Work Email
Telephone
Fax
Mobile

Date: 8

NHS REC Form	Reference:	IRAS Version 3.1

Details can be obtained from the NHS R&D Forum website: http://www.rdforum.nhs.uk

A69-1. How long do you expect the study to last in the UK?

Planned start date:

Planned end date:

Total duration:

 Years: Months: Days:

A69-2. How long do you expect the study to last in all countries?

Planned start date:

Planned end date:

Total duration:

 Years: Months: Days:

A70. Definition of the end of trial, and justification in the case where it is not the last visit of the last subject undergoing the trial [1]

A71-1. Is this study?

◯ Single centre

◯ Multicentre

A71-2. Where will the research take place? *(Tick as appropriate)*

☐ England

☐ Scotland

☐ Wales

☐ Northern Ireland

☐ Other countries in European Economic Area

Does this trial involve countries outside the EU?
◯ Yes   ◯ No

A72. What host organisations (NHS or other) in the UK will be responsible for the research sites? *Please indicate the type of organisation by ticking the box and give approximate numbers of planned research sites:*

☐ NHS organisations in England

☐ NHS organisations in Wales

☐ NHS organisations in Scotland

☐ HSC organisations in Northern Ireland

☐ GP practices in England

☐ GP practices in Wales

☐ GP practices in Scotland

☐ GP practices in Northern Ireland

☐ Social care organisations

Date: 9

NHS REC Form   Reference: IRAS Version 3.1

☐ Phase 1 trial units
☐ Prison establishments
☐ Probation areas
☐ Independent hospitals
☐ Educational establishments
☐ Independent research units
☐ Other (give details)

Total UK sites in study:

A definition of the end of the trial should be incorporated in the protocol.

Date: 10

PART C: Overview of research sites

Please enter details of the host organisations (Local Authority, NHS or other) in the UK that will be responsible for the research sites. *For NHS sites, the host organisation is the Trust or Health Board. Where the research site is a primary care site, e.g. GP practice, please insert the host organisation (PCT or Health Board) in the Institution row and insert the research site (e.g. GP practice) in the Department row.*

PART D: Declarations

D1. Declaration by Chief Investigator

1. The information in this form is accurate to the best of my knowledge and belief and I take full responsibility for it.

2. I undertake to abide by the ethical principles underlying the Declaration of Helsinki and good practice guidelines on the proper conduct of research.

3. If the research is approved I undertake to adhere to the study protocol, the terms of the full application as approved and any conditions set out by review bodies in giving approval.

4. I undertake to notify review bodies of substantial amendments to the protocol or the terms of the approved application, and to seek a favourable opinion from the main REC before implementing the amendment.

5. I undertake to submit annual progress reports setting out the progress of the research, as required by review bodies.

6. I am aware of my responsibility to be up to date and comply with the requirements of the law and relevant guidelines relating to security and confidentiality of patient or other personal data, including the need to register when necessary with the appropriate Data Protection Officer. I understand that I am not permitted to disclose identifiable data to third parties unless the disclosure has the consent of the data subject or, in the case of patient data in England and Wales, the disclosure is covered by the terms of an approval under Section 251 of the NHS Act 2006.

7. I understand that research records/data may be subject to inspection by review bodies for audit purposes if required.

8. I understand that any personal data in this application will be held by review bodies and their operational managers and that this will be managed according to the principles established in the Data Protection Act 1998.

9. I understand that the information contained in this application, any supporting documentation and all correspondence with review bodies or their operational managers relating to the application:

   - Will be held by the main REC or the GTAC (as applicable) until at least 3 years after the end of the study; and by NHS R&D offices (where the research requires NHS management permission) in accordance with the NHS Code of Practice on Records Management.
   - May be disclosed to the operational managers of review bodies, or the appointing authority for the main REC, in order to check that the application has been processed correctly or to investigate any complaint.
   - May be seen by auditors appointed to undertake accreditation of RECs.
   - Will be subject to the provisions of the Freedom of Information Acts and may be disclosed in response to requests made under the Acts except where statutory exemptions apply.

10. I understand that information relating to this research, including the contact details on this application, may be held on national research information systems, and that this will be managed according to the principles established in the Data Protection Act 1998.

11. I understand that the summary of this study will be published on the website of the National Research Ethics Service (NRES), together with the contact point for enquiries named below. Publication will take place no earlier than 3 months after issue of the ethics committee's final opinion or the withdrawal of the application.

Contact point for publication(Not applicable for R&D Forms)

NRES would like to include a contact point with the published summary of the study for those wishing to seek further information. We would be grateful if you would indicate one of the contact points below.

☐ Chief Investigator

☐ Sponsor

☐ Study co-ordinator

Date: 12

NHS REC Form   Reference: IRAS Version 3.1

☐ Student
☐ Other – please give details
☐ None

Access to application for training purposes *(Not applicable for R&D Forms)*
Optional – please tick as appropriate:

☐ I would be content for members of other RECs to have access to the information in the application in confidence for training purposes. All personal identifiers and references to sponsors, funders and research units would be removed.

Signature: ....

Print Name:

Date: *(dd/mm/yyyy)*

Date: 13

D2. Declaration by the sponsor's representative

If there is more than one sponsor, this declaration should be signed on behalf of the co−sponsors by a representative of the lead sponsor named at A64-1.

I confirm that:

1. This research proposal has been discussed with the Chief Investigator and agreement in principle to sponsor the research is in place.

2. An appropriate process of scientific critique has demonstrated that this research proposal is worthwhile and of high scientific quality.

3. Any necessary indemnity or insurance arrangements, as described in question A76, will be in place before this research starts. Insurance or indemnity policies will be renewed for the duration of the study where necessary.

4. Arrangements will be in place before the study starts for the research team to access resources and support to deliver the research as proposed.

5. Arrangements to allocate responsibilities for the management, monitoring and reporting of the research will be in place before the research starts.

6. The duties of sponsors set out in the Research Governance Framework for Health and Social Care will be undertaken in relation to this research.

7. Where the research is reviewed by a REC within the UK Health Departments Research Ethics Service, I understand that the summary of this study will be published on the website of the National Research Ethics Service (NRES), together with the contact point for enquiries named in this application. Publication will take place no earlier than 3 months after issue of the ethics committee's final opinion or the withdrawal of the application.

Signature: ....

Print Name:

Post:

Organisation:

Date: *(dd/mm/yyyy)*

HES Data Request Service Application Form

**Episode records,
Episode records including sensitive data and
Patient-identifiable episode records**

**Part 3A: Inpatient data
(including maternity and adult critical care)**

Introduction

This document is Part 3A of the HES Data Request Service Application Form. You can find additional copies of this document, and other documents relating to the Data Request Service, on HESonline [www.hesonline.nhs.uk].

This document consists of the following sections:
1   Data set
2   Filters
3   Fields
4   Other details

1   Data set

a) Data year				
☐ 1989-90	☐ 1990-91	☐ 1991-92	☐ 1992-93	☐ 1993-94
☐ 1994-95	☐ 1995-96	☐ 1996-97	☐ 1997-98	☐ 1998-99
☐ 1999-00	☐ 2000-01	☐ 2001-02	☐ 2002-03	☐ 2003-04
☐ 2004-05	☐ 2005-06	☐ 2006-07	☐ 2007-08	☐ 2008-09
☐ 2009-10				
☐ Provisional 2010-11 (the data provided will be the latest cumulative monthly data for 2010-11. Further details regarding monthly data are available on HESonline [www.hesonline.nhs.uk].				

b) Additional Data Linkage	
If you have a particular cohort of patients that you want to identify in this data set (using NHS number or other patient identifiers) then that would most likely require our Trusted Data Linkage Service (TDLS). If you are not asking us to identify a cohort of patients or link patients across time or datasets then it is likely that the request will fall within the standard HES extract service.	
The data I am requesting requires additional data linkage	☐

2   Filters

a) Filter details

Filter details need to be completed within this section (otherwise this will result in full and complete data years being provided), please ensure filters are included as appropriate to your approvals, for example where only specific patient records have been approved, this needs to be stated here.

HOSPITAL EPISODE STATISTICS: HES Data Request Service Application Form, Part 3A

3 Fields

a) Fields to be extracted (please select (☒).Fields shown in **bold** below are **sensitive/ identifiable** and will need DAAG/ECC approval. Details of the DAAG and ECC approval processes can be found on the Approval process page on HESonline [http://172.20.1.134/Ease/servlet/ContentServer?siteID=1937&categoryID=846]. For help with the fields in this section, please refer to the field descriptions in the HES Data Dictionary, which is available on the Data Dictionaries page of HESonline [http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=289]).

Patient

Note: some related fields can be found under maternity.

- ☐ Administrative & legal status of patient (category) [1]
- ☐ Administrative category (admincat) [2]
- ☐ Age at end of episode (endage)
- ☐ Age at start of episode (startage)
- ☐ Baby's age in days (neodur)
- ☐ **Date of birth - patient (dob)**
- ☐ Month and year of birth (mydob)
- ☐ Date of birth check flag - patient (dob_cfl)
- ☐ Ethnic category (ethnos) [3]
- ☐ **Legal category of patient (leglcat)** [2]
- ☐ **Legal group of patient (legalgpa)** [4]
- ☐ **Local patient identifier (lopatid)** [5]
- ☐ **NHS number (newnhsno)** [5]
- ☐ Patient identifier - HES generated (extract_hesid) [5]
- ☐ Postcode district of patient's residence (postdist)
- ☐ **Postcode of patient (homeadd)**
- ☐ Sex of patient (sex) [6]

Admissions

- ☐ Date of admission (admidate)
- ☐ Admission date check flag (adm_cfl)
- ☐ Date of decision to admit (elecdate)
- ☐ Date of decision to admit check flag (elec_cfl)
- ☐ Method of admission (admimeth)
- ☐ Source of admission (admisorc)
- ☐ First regular day or night admission (firstreg)
- ☐ Waiting time (elecdur)

Discharges

- ☐ Date of discharge (disdate)
- ☐ Discharge date check flag (dis_cfl)
- ☐ Destination on discharge (disdest)
- ☐ Method of discharge (dismeth)

Episodes and Spells

- ☐ Bed days within the year (bedyear)
- ☐ Beginning of spell (spelbgin)
- ☐ Date episode ended (epiend)
- ☐ Date episode started (epistart)
- ☐ Duration of spell (speldur)
- ☐ End of spell (spelend)
- ☐ Episode duration (epidur)
- ☐ Episode order (epiorder)
- ☐ Episode end date check flag (epie_cfl)
- ☐ Episode start date check flag (epis_cfl)
- ☐ Episode status (epistat)
- ☐ Episode type (epitype)
- ☐ Hospital provider spell number (provspno) [5]
- ☐ Ward type at start of episode (wardstrt) [7]
- ☐ Discharge ready date (disreadydate)

Clinical

- ☐ All Diagnosis codes (diag_nn)
- ☐ Primary diagnosis - 3 characters (diag3)
- ☐ Primary diagnosis - 4 characters (diag4)
- ☐ External cause of injury or poisoning (cause)
- ☐ External cause code - 4 characters (cause4)
- ☐ External cause code - 3 characters (cause3)
- ☐ All Operative procedure codes (oper_nn) [8]
- ☐ Main operative procedure - 3 characters (oper3) [8]
- ☐ Date of operation (opdte_nn) [8]
- ☐ Operation status code (operstat) [8]
- ☐ Post-operative duration (posopdur) [8]

HOSPITAL EPISODE STATISTICS: HES Data Request Service Application Form, Part 3A

☐ Pre-operative duration (preopdur) [8]

☐ Patient classification (classpat)

☐ Intended management (intmanig) [5]

☐ Main specialty (mainspef)

☐ Treatment specialty (tretspef)

Healthcare Resource Groups [9]

☐ Dominant procedure (domproc)

☐ Healthcare resource group: version 3.1 (hrg_3.1) [10]

☐ Healthcare resource group: version 3.5 (hrg_3.5) [11]

☐ NHS-generated HRG code (hrgnhs)

☐ NHS-generated HRG code version number (hrgnhsvn)

Organisation

☐ Commissioner code (purcode) [12]

☐ Commissioner code status (purval) [3]

☐ Commissioner's Regional Office (purro)

☐ Commissioner's Strategic Health Authority (purstha)

☐ Commissioning serial number (csnum) [2]

☐ Health Authority area where patient's GP was registered (gppracha) [2]

☐ Primary care group (pcgcode) [13]

☐ Primary care trust of responsibility - historic (pctcode02) [14]

☐ Primary care trust of responsibility - current (pctcode06) [15]

☐ Primary Care Trust area where patient's GP was registered (gpprpct) [16]

☐ Provider code - 5 character (procode) [17]

☐ Provider code - 3 character (procode3/procodet) [17]

☐ Site code of treatment (sitetret) [5]

☐ Provider type (protype) [17, 11]

☐ Regional Office area where patient's GP was registered (gppracro) [2]

☐ Strategic Health Authority area where patient's GP was registered (gpprstha) [16]

Geographical

☐ **Census Output Area, 2001 (oacode)** [11]

☐ Census Output Area, 2001 (6 character) (oaward) [11, 18]

☐ County of residence (rescty)

☐ Current electoral ward (currward & resladst) [18]

☐ Electoral ward in 1981 (ward81)

☐ Electoral ward in 1991 (ward91) [19, 15]

☐ Government office region of residence (resgor)

☐ Government office region of treatment (gortreat)

☐ Health Authority of residence (resha)

☐ Health Authority of treatment (hatreat)

☐ Local authority district (resladst)

☐ **Ordnance Survey grid reference (gridlink)**

☐ Patient's Health Authority of residence, provided by NHS (pctnhs)

☐ Patient's Primary Care Trust of residence - historic (respct02) [20]

☐ Patient's Primary Care Trust of residence – current (respct06) [15]

☐ Patient's Strategic Health Authority of residence - historic (resstha02) [20]

☐ Patient's Strategic Health Authority of residence - current (resstha06) [15]

☐ Primary Care Trust area of treatment (pcttreat) [16]

☐ Region of treatment (rotreat)

☐ Regional Office of residence (resro)

☐ Strategic Health Authority area of treatment (sthatret) [16]

Socio-economic

☐ Lower Super Output Area (soal)

☐ Middle Super Output Area (soam)

☐ Rural/Urban Indicator (rururb_ind)

☐ IMD Crime Domain (imd04c)

☐ IMD Education Training and Skills Domain (imd04ed)

☐ IMD Health and Disability Domain (imd04hd)

☐ IMD Barriers to Housing and Service Domain (imd04hs)

☐ IMD Income Domain (imd04i)

HOSPITAL EPISODE STATISTICS: HES Data Request Service Application Form, Part 3A

☐ IMD Income affecting Adults Domain (imd04ia)

☐ IMD Income affecting Children Domain (imd04ic)

☐ IMD Living Environment Domain (imd04le)

☐ IMD Overall Rank (imd04rk)

Practitioner

☐ Code of GP practice (gpprac) [5]

☐ **Consultant code (consult)** [5]

☐ **Code of patient's registered or referring general medical practitioner (reggmp)** [5]

☐ **Person referring patient (referrer)** [5]

☐ Pseudonymised consultant team code (pconsult)

☐ Pseudonymised code of patient's registered or referring general medical practitioner (preggmp)

☐ Pseudonymised referrer code (preferer)

☐ Referring organisation code (referorg)

Augmented/critical care period [21] **(up to 9 per record)**

☐ Augmented care location (acploc)

☐ Augmented care period data quality indicator (acpdqind) [4]

☐ Augmented care period disposal (acpdisp)

☐ Augmented care period end date (acpend)

☐ Augmented care period period indicator (acpplan)

☐ **Augmented care period local ID (acplcid)**

☐ Augmented care period number (acpn)

☐ Augmented care period outcome indicator (acpout)

☐ Augmented care period source (acpsour)

☐ Augmented care period speciality function code (acpspef)

☐ Augmented care period start date (acpstar)

☐ High-dependency care level (depdays)

☐ Intensive care level days (intdays)

☐ Number of augmented care periods within episode (numacp)

☐ Number of organ systems supported (orgsup)

Maternity **(up to 9 per record)**
Note: Some related fields can be found under 'Patient'

☐ Anaesthetic given during labour or delivery (delprean)

☐ Anaesthetic given post-labour or delivery (delposan)

☐ Antenatal days of stay (antedur)

☐ **Birth date (baby) (dobbaby)**

☐ Birth order (birordr)

☐ Birth weight (birweit)

☐ Change of delivery place (delchang)

☐ Delivery method (delmeth)

☐ Delivery place (actual) (delplac)

☐ Delivery place (intended) (delinten)

☐ First antenatal assessment date (anasdate)

☐ Gestation period in weeks at first antenatal assessment (anagest)

☐ Length of gestation (gestat)

☐ Live or still birth (birstat)

☐ Method to induce labour (delonset)

☐ Mother's age at delivery (matage)

☐ **Mother's date of birth (motdob)**

☐ Number of babies (numbaby)

☐ Number of previous pregnancies (numpreg)

☐ Postnatal stay (postdur)

☐ Resuscitation method (biresus)

☐ Sex of baby (sexbaby)

☐ Status of person conducting delivery (delstat)

☐ Neonatel level of care (neocare) [19]

☐ Well baby flag (wellbaby) [22]

Psychiatric

☐ Age at psychiatric census date (censage)

☐ Carer support indicator (carersi) [5]

☐ Date detention commenced (detndate)

☐ Date detention commenced check flag (det_cfl)

☐ **Detention category (detncat)** [2]

☐ Duration of care to psychiatric census date (cendur)

HOSPITAL EPISODE STATISTICS: HES Data Request Service Application Form, Part 3A

☐ Duration of detention (detdur)

☐ **Legal group of patient (psychiatric) (legalgpc)** [4]

☐ **Legal status classification (leglstat)**

☐ Marital status (psychiatric) (marstat)

☐ Mental category (mentcat)

☐ Psychiatric patient status (admistat) [19]

☐ Status of patient included in psychiatric census (censtat)

☐ V code indicator (vind) [23]

☐ Ward type at psychiatric census date (cenward)

Adult Critical Care

☐ Critical Care Start Date (ccstartdate)

☐ Critical Care Start Time (ccstarttime)

☐ Critical Care Unit Function (ccunitfun)

☐ Critical Care Unit Bed Configuration (unitbedconfig)

☐ Critical Care Admission Type (ccadmitype)

☐ Critical Care Admission Source (ccadmisorc)

☐ Critical Care Source Location (ccsorcloc)

☐ Advanced Respiratory Support Days (aressupdays)

☐ Basic Respiratory Support Days (bressupdays)

☐ Advanced Cardiovascular Support Days (acardsupdays)

☐ Basic Cardiovascular Support Days (bcardsupdays)

☐ Renal Support Days (rensupdays)

☐ Neurological Support Days (neurosupdays)

☐ Gastro-intestinal Support Days (gisupdays)

☐ Dermatological Support Days (dermsupdays)

☐ Liver Support Days (liversupdays)

☐ Organ Support Maximum (orgsupmax)

☐ Critical Care Level 2 Days (cclev2days)

☐ Critical Care Level 3 Days (cclev3days)

☐ Critical Care Discharge Status (ccdisstat)

☐ Critical Care Discharge Destination (ccdisdest)

☐ Critical Care Discharge Location (ccdisloc)

☐ Critical Care Discharge Date (ccdisdate)

☐ Critical Care Discharge Time (ccdistime)

☐ Critical Care Discharge Ready Date (ccdisrdydate)

☐ Critical Care Discharge Ready Time (ccdisrdytime)

☐ Best Match Flag (bestmatch)

☐ Critical Care APC Relationship (ccapcrel)

System

☐ Combined grossing factor (gross_b) [24]

☐ Coverage grossing factor (gross_a) [24]

☐ Date data received by NHS wide clearing service (subdate)

☐ NHS number indicator (nhsnoind) [5]

☐ OPCS 4.3 used flag (opcs43) [25]

☐ Origin of primary care group (pcgorig) [13]

☐ Origin of primary care trust of responsibility - historic (pctorig02) [14]

☐ Origin of primary care trust of responsibility - current (pctorig06) [15]

☐ Record identifier (epikey)

HOSPITAL EPISODE STATISTICS APPLICATION FORM

4 Other details

a) Other details

Footnotes

[1] Up to and including 2001-02.

[2] From 2000-01 onwards.

[3] From 1995-96 onwards.

[4] From 2002-03 onwards.

[5] From 1997-98.

[6] The field SEX will be sensitive when combined with DOB and Homeadd.

[7] Up to and including 2000-01.

[8] These fields reflect all procedures and interventions recorded through OPCS 4.4. OPCS 4.4 was introduced in 2007-08.

[9] Fields providing relative costs of treatment and costs per day (Treat, Hotel and totcost), that were used for economic modelling, have now been removed from the available list. An issue occurred when processing HRG4 fields in 2009-10 HES data. A new date for the release of this information will be announced as soon as possible.

[10] Available from 1995-96 to 2005-06.

[11] From 2003-04 onwards.

[12] Please refer to 4.2 of the HES User Guide on www.hesonline.nhs.uk

[13] Pcgcode field retained for years 1999-2000 to 2001-02. Not available from 2002-03 onwards.

[14] Field historically derived from 1997-98 to 2001-02 on the same basis as for 2002-03.

[15] Field currently only available from 2006-07.

[16] Field historically derived from 1999-2000 to 2001-02 on the same basis as for 2002-03.

[17] Procode3 (previously known as PRODMUT) identifies an individual hospital provider by using the first three characters of PROCODE. Up to 2003-03 this field was called Procode3. From 2003-04, it became Procodet. Procodet contains the 3-character code except where five characters are needed to identify a distinct organisation and PROTYPE contains a summarised description of the organisation type. You will need to determine whether 3- or 5-character Procode is needed.

[18] A request for multiple ward codes might need to be treated as sensitive, due to the small areas identified by differencing. Please seek advice from the HES team on your requirement.

[19] From 1996-97 onwards.

[20] Field historically derived from 1996-97 to 2001-02 on the same basis for 2002-03. Derived from 2006-07 on the same basis as for 2002-03.

[21] From 1997-98 to 2005-06.

[22] The well baby flag indicates which episodes relate to healthy, live infants.

[23] Up to 1995-96, then replaced by ADMISTAT and NEOCARE.

[24] Not yet available for 2002-03 or 2003-04.

[25] Available from 2005-06.

National Advisory Committee on Drugs
Hawkins House
Dublin 2

www.nacd.ie
nacd@nacd.ie
01 6354283

NACD

National Advisory
Committee on Drugs