Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Developing and Implementing Mathematical Models for the Prevalence, Incidence and Initiation into Cocaine Use in Ireland

Ms Anne Snel

Supervisor: Professor Catherine Comiskey

School of Nursing and Midwifery,
Faculty of Health Sciences

A thesis submitted to the
University of Dublin, Trinity College
for the degree of
Doctor of Philosophy

September 30th, 2011

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed

Summary

The National Drug Strategy 2009-2016 under the auspices of the Department of Community, Rural and Gaeltacht Affairs (DCRGA) reported widespread public concern regarding the misuse of cocaine, particularly when combined with other illegal substances, and/or alcohol (DCRGA 2009). In order to ensure best practice in treatment services, it is essential that accurate and up to date prevalence figures of drug use are available, that incidence is monitored as an indicator of trends, that information is available as to what interventions will be most effective in reducing cocaine prevalence and that the dynamics of initiation to cocaine use are better understood.

This study aims to address these identified gaps and its aims and objectives are presented in chapter 1. In chapter 2 the pharmacology of cocaine is described and statistical estimates of cocaine use worldwide, in Europe and in Ireland are provided. Chapter 3 explores the role of statistical and mathematical models as a tool to achieve the aims and objectives as laid out in chapter 1.

In chapter 4, statistical models of prevalence are critically analysed and applied to over 330,00 urinalysis samples obtained from the Drugs and Aids Information System (DAIS) of the Northern Area Health Board (NAHB) in Ireland. For the years 2006-2010 the number of unique individuals in treatment in whose urinalysis samples cocaine was detected were 1092, 1057, 1097, 1027 and 1044 respectively. Truncated Poisson and capture recapture methods are used to produce more robust estimates of the total numbers using cocaine whilst in treatment. These estimates range from 1,306 to 1,860 for 2006, from 1,293 to 1,930 for 2007, from 1,301 to 1,764 for 2008, from 1,339 to 1,457 for 2009 and from 1,354 to 2,527 in 2010. These estimates are then used as benchmark figures to which a multiplier is applied enabling the estimation of cocaine use among the general population in the NAHB. For 2010 this ranges from 7,598

to 10,587. The ratio of those in treatment using cocaine to those in the general population using cocaine lies approximately between 1:6 to 2:9.

In chapter 5, a simple ordinary differential equation model for prevalence and incidence estimation is developed. Last year prevalence rates in Ireland, obtained from two NACD studies (NACD 2006, NACD 2008), are applied to CSO census figures obtained during the same years (CS0 2002, CSO 2006) enabling a model to be built of the dynamics of the change in cocaine use during that period. The model is implemented using two separate quit rates obtained from US research implying average cessation of use after 7.5 and 5.5 years respectively. It investigates, through simulations, the comparative efficacy of pre- and post-initiation interventions aimed at reducing cocaine use. The model predicts that a shortening of the average time of use before cessation by two years would result in a decrease of 48% in the number of users after 35 years. The ordinary differential equation models are developed further within chapter 6 where age dependency and social mixing are introduced via the Who Acquires Infection from Whom (WAIFW) matrix structures. Using the same NACD and CSO data as in chapter 5, these models facilitate exploration of the levels of initiation associated with age related social interaction. The model considered to be best fitting includes movement through age groups and estimates of incidence are produced. In the 15 to 24 age group incidence increased from on average 4,638 in 2002 to 7,359 in 2006. In the 25 to 34 age group the corresponding figure were 2,169 in 2002 and 5,384 in 2006. In the 35 to 44 age group it increased from 530 in 2002 to 1,284 in 2006. The ageing model also suggests that the last year prevalence of cocaine use in the cohort aged 15-44 in 2010 could range between 70,461 and 82,272. Within chapter 7 all results are critically discussed, limitations are identified and recommendations made for policy, practice and future research. In conclusion findings highlight the benefits for service users of greater collaboration between policy makers, treatment planners and mathematical modellers.

Contents

1 Thesis Introduction, Objectives and Overview **1**

1.1 Introduction 1

1.2 Aims and Objectives 4

1.3 Thesis Outline 4

2 Cocaine: Pharmacology and Prevalence both Past and Present **6**

2.1 Cocaine Pharmacology 6

2.1.1 Methods of Administration 8

2.2 History of Cocaine Use 9

2.3 Present Day Use 11

2.3.1 Cocaine Use Worldwide 11

2.3.2 Cocaine Use In Europe 15

2.3.3 Cocaine Use in Ireland 17

2.4 Chapter Summary 21

3 The Mathematical Modelling Process and its Historical Evolution **23**

3.1 Justification of the Use of Mathematical Models 23

3.1.1 The Modelling Process 24

3.1.2 Historical Developments in Epidemiology 31

3.2 Recent Applications of Mathematical Modelling to Drug Use 36

3.2.1 Point Estimation Methods 36

3.2.2 Dynamical Mathematical Modelling Techniques 37

3.3 Chapter Summary 41

4 Point Prevalence Estimation Models **43**

4.1 Introduction to Prevalence Estimation 43

4.2 Data Sources 46

4.2.1 National Drug Prevalence Household Survey 47

4.2.2 Northern Area Health Board Population Figures 47

4.2.3 Pulse Data 48

4.2.4 Drugs and Aids Information System 49

4.3 Methods of Prevalence Estimation 52

4.3.1 Multiplier-Benchmark Method 52

4.3.2 Truncated Poisson 56

4.3.3 Capture-Recapture 59

4.4 Prevalence Estimates 78

4.4.1 Estimates from the NACD Household Survey 79

4.4.2 Truncated Poisson 79

4.4.3 Three Sample Capture-Recapture Method 80

4.4.4 The Multiplier Method 86

4.5 Conclusion and Discussion 87

4.6 Chapter Summary 90

5 Dynamic Models and Intervention Strategies **92**

5.1 Introduction 92

5.2 Literature Review 93

5.3 Mathematical Model 96

5.4 Parameter Estimation 97

5.4.1 Estimating the rate of change in the population μ 98

5.4.2 Rate of cessation of cocaine use 99

5.4.3 Estimating β, the rate of initiation 100

5.5 Model Simulations 101

5.5.1 Model simulation of user numbers after 35 years 101
5.5.2 Simulation of the extent of the reduction in cocaine users resulting from measures aimed at reducing initiation 103
5.5.3 Simulation of the extent of the reduction in cocaine users as a result of effective treatment and intervention 105
5.5.4 Discussion 107
5.6 Chapter Summary 108

6 Dynamic Age Dependent Models and Social Contact Patterns 109
6.1 An Age Differentiated Model 109
6.2 Estimation of the Parameters 114
6.2.1 Population rate of change, μ and Λ 114
6.2.2 Rate of quitting 117
6.2.3 Force of infection parameter 118
6.2.4 Calculating the β values for the four WAIFW matrices 120
6.2.5 Conclusion 122
6.3 A Separate Model for Each WAIFW Structure 122
6.3.1 β values for WAIFW1 123
6.3.2 β values for WAIFW2 124
6.3.3 β values for WAIFW3 125
6.3.4 β values for WAIFW4 126
6.3.5 Conclusion 127
6.4 Improving the Model with an Ageing Parameter 131
6.4.1 Discussion and Comparison of the results 133
6.5 Chapter Summary 137

7 Conclusions and Further Work 139
7.1 Introduction 139
7.2 Strengths and Limitations 140
7.3 Discussion and Further Work 143

A Glossary 145

B Berkeley Madonna 147

C Ethical Approval DAIS 148

List of Figures

2.1 UNODC 2011: Overview of world cocaine prevalence 13

2.2 Trends in cocaine related deaths from 1998 to 2007. 20

2.3 Trends in the number of cocaine seizures reported by Connolly and compiled by the Central Statistics office. 20

2.4 Trends in proceedings of cocaine use reported by Connolly and compiled by the Central Statistics office. 21

3.1 Steps in the development of a mathematical model adapted from Vynnycky & White (2010) 25

3.2 Flowchart SI Model 27

3.3 Flowchart SIS Model 28

3.4 Flowchart SIR Model 29

3.5 Graphic reproduced from Sattenspiel (1990) illustrating the number of publications concerned with mathematical models for contagious processes between 1900 and 1973 which were cited in Bailey (1975) . 38

4.1 The number of DAIS clients who tested cocaine positive between one and ten times in the years 2006 to 2010 52

4.2 Venn diagram showing the numbers testing positive for cocaine in one or more of the three four-month periods in 2006 81

5.1 Census figures and population estimates of 15-34 age group from 2002 to 2020 showing immigration estimates in yellow 98

5.2 Berkeley Madonna Population estimates 2002 to 2020 99
5.3 Model Prediction of cocaine user numbers after 35 years assuming average time of use is 7.5 years 102
5.4 Model Prediction of cocaine user numbers after 35 years assuming average time of use is 5.5 years 103
5.5 Model Prediction of cocaine user numbers after 35 years assuming preventative measures achieve reductions in initiation : quit time 7.5 years 104
5.6 Model Prediction of cocaine user numbers after 35 years assuming preventative measures achieve reductions in initiation : quit time 5.5 years 105
5.7 Model Prediction of cocaine user numbers after 35 years assuming treatment and interventions achieve reductions in user numbers : quit time 7.5 years 106
5.8 Model Prediction of cocaine user numbers after 35 years assuming treatment and interventions achieve reductions in user numbers : quit time 5.5 years 107

6.1 Population of 15-24 year olds extrapolated for the years 2003 - 2006 and 2007 - 2010 from 2002 and 2006 Census Data. The difference seen in 2006 indicates the level of increase due to immigration. . . . 115
6.2 Population of 25-34 year olds extrapolated for the years 2003 - 2006 and 2007 - 2010 from 2002 and 2006 Census Data. The difference seen in 2006 indicates the level of increase due to immigration. . . . 116
6.3 Population of 35-44 year olds extrapolated for the years 2003 - 2006 and 2007 - 2010 from 2002 and 2006 Census Data. The difference seen in 2006 indicates the level of increase due to immigration. . . . 116
6.4 Incidence estimates from 2002 and 2006 for each WAIFW matrix shown per age group 129

6.5 Prevalence trends 2007 and 2010 for each WAIFW matrix shown per age group 130
6.6 Incidence data 2002 and 2006 with an aging component to model . . 135
6.7 Incidence data 2002 and 2006 with an aging component to model . . 136

List of Tables

4.1 Last year prevalence rate with 95% confidence intervals of cocaine use in 2006/2007 (Results for the Northern Area Health Board) 47

4.2 Population by Gender and Age of the Northern Area Health Board . 48

4.3 Urinalysis samples which are cocaine positive in the Drugs and Aids Information System 2006 - 2010 50

4.4 Frequencies of clients in DAIS database who had 1, 2, 3 and so forth cocaine positive urinalyses from 2005 to 2010 51

4.5 Possible sources for benchmark data recommended by Hickman *et al.* (2002) 53

4.6 Contingency table indicating presence or absence in one of three samples 61

4.7 Contingency table indicating presence or absence in one of three samples 61

4.8 Two by two contingency table with marginal totals 62

4.9 Partial Tables created by splitting across levels of Z 65

4.10 Partial Tables created by splitting across levels of Y 66

4.11 Partial Tables created by splitting across levels of X 66

4.12 Partial Tables created by splitting across the levels of one variable . 67

4.13 Three sample contingency table of expected values 71

4.14 Last year prevalence estimates with 95% confidence intervals of cocaine use in 2006 (NACD results for the Northern Area Health Board) 79

4.15 Zelterman and Chao point estimates with confidence intervals of the total number of problem drug users in treatment who are also using cocaine in the Northern Area Health Board 80
4.16 Three sample contingency table showing the numbers present in the samples or their overlaps 82
4.17 Count of the clients present in the first, second or third four month period in each year from 2006 to 2010 84
4.18 Capture-recapture point estimates with confidence intervals of the total number of problem drug users in treatment who are also using cocaine in the Northern Area Health Board 86
4.19 Estimated general population cocaine prevalence based on treatment estimates in the NAHB in 2007 - 2010 87

6.1 Estimated rate of population change, μ and constant of immigration, Λ per age group based on 2002 (NACD 2006) and 2006 prevalence surveys (NACD 2006, NACD 2008) 117
6.2 Force of Infection, λ_i, per age group 120
6.3 β values for each WAIFW structure 121
6.4 Last year prevalence of cocaine users in 2006 predicted by each WAIFW model 122
6.5 Incidence estimates for cocaine use in 2002 - 2006 using WAIFW1 model 124
6.6 Incidence estimates for cocaine use in 2002 - 2006 using WAIFW2 model 125
6.7 Incidence estimates for cocaine use in 2002 - 2006 using WAIFW3 model 126
6.8 Incidence estimates for cocaine use 2002 - 2006 using WAIFW4 model 127
6.9 Prevalence estimates 2007 - 2010 using WAIFW models 128

6.10 Beta and approximate ratio values estimated for each WAIFW model where ageing is included 132
6.11 Incidence estimates 2002 - 2006 using WAIFW models with Ageing 133
6.12 Prevalence estimates 2006 - 2010 using Ageing WAIFW models . . 133

Acknowledgements

The undertaking of this research thesis was enabled by two awards, one by the National University of Ireland Maynooth, of the Pat Hume Scholarship during the first year, for which I am very grateful. In years two, three and four, I was fortunate to be awarded a Ph.D. stipend by the School of Nursing and Midwifery, Trinity College Dublin.

This thesis would not have been possible without the unfailing support, patience, encouragement and professional guidance of my supervisor Professor Catherine Comiskey. I particularly appreciate her aptitude to never criticise but helpfully suggest improvements through the entire four year period, putting up with utmost calm any times of stress and frustration on my part. In her own professional career, she has served as a shining example of what can be achieved through outstanding ability, hard work and unfailing consideration to others.

I would also like to thank all those in the School of Nursing and Midwifery, in particular Professor Cecily Begley for the standards of excellence she has set in the school, Doctor Gabrielle McKee for facilitating my development through the Ph.D. days and other supports for which she was responsible and to both Doctor Elizabeth Curtis and Ms Sylvia Huntley-Moore who gave me the opportunity to improve my SPSS skills by including me in their respective research projects.

A special thanks to Doctor Orla Dempsey for her unfailing help and encouragement throughout the four years and in particular for being my rock in the final few weeks before submission. Thanks also to Dr. Anne-Marie Brady, my study buddy for the diploma in statistics who helped make the commitment two evenings a week, a pleasant and fun experience. Thanks also to Ms Emma Murphy and Doctor Emma White for both their friendship and help throughout the years.

Finally, I could never have completed this research without the support of my children John, Vincent and Evelín and my husband René, who not only coped with

half a wife for four years, kept the household running smoothly and assisted with graphical advise and expertise but in the final days before submission was prepared to show his love for me in a way that must have been excruciating torture for him, by proofreading this thesis. Thank you René.

Chapter 1

Thesis Introduction, Objectives and Overview

1.1 Introduction

According to the United Nations Office on Drugs and Crime between 14.2 and 20.5 million people aged 15-64 used cocaine in 2009 (UNODC 2011). There is a wide range of physical and psychiatric problems associated with cocaine use, many of which are serious and may even be fatal. Cocaine overdoses are unpredictable and combining cocaine with other drugs, particularly alcohol is highly risky (Pennings *et al.* 2002). When cocaine is taken with alcohol it combines in the system to form another drug - cocaethylene - which is more toxic than either drug used alone (Harris *et al.* 2003). At high doses cocaethylene is estimated to be over 25 times more likely to induce death than cocaine by itself (Karan *et al.* 1998). Cocaine dependence is associated with a greater risk of suicide in some cocaine users with certain comorbid conditions (Roy 2009) and this risk is even higher among those who use alcohol and cocaine together in a problematic way (Salloum *et al.* 1996).

In 2007, the National Advisory Commission on Drugs (NACD) commissioned a sec-

ond report on cocaine use in Ireland which concluded that all the indicators pointed to a continued increase in cocaine use, that this cocaine use crossed all social strata and that the impact was very much experienced nationwide. The report stated that cocaine can cause social and economic harm not only to the cocaine user but also to the communities that bear the brunt of the behaviour of the criminal activity associated with the supply of cocaine (NACD 2007). According to two household studies undertaken on behalf of the NACD, lifetime, last year and last month prevalence of cocaine use increased from 2002/2003 to 2006/2007 (NACD 2006, NACD 2008). It is widely recognised however, that household surveys tend to underestimate the prevalence of the use of drugs such as heroin and cocaine because of the hidden nature of their use and the under representation of certain sub populations such as the homeless and those in prison. Hence, Gannon *et al.* (2011) were commissioned by the NACD to research into alternative methods and data sources which could produce more robust estimates.

Estimation of cocaine prevalence is however complicated by the fact that two distinct groups have been identified by both Mayock (2001) and the NACD (2007). The first group consists of so called recreational users who use cocaine as a weekend drug alongside other substances such as alcohol. The other more problematic group tend to use cocaine alongside heroin and other drugs such as benzodiazepines. Cox & Comiskey (2011) in their discussion of the implications of concurrent cocaine use amongst opiate users in treatment found considerable differences in not only drug outcomes but in physical, psychological and crime related outcomes. The first objective of this thesis is to estimate the prevalence of the latter group.

It is now widely acknowledged that initiation into drug use occurs largely through familial and peer social contact (Kandel 1985) and this suggests a similarity with the processes normally associated with the spread of infectious diseases. The application of mathematics to the study of infectious diseases has been traced as far back as 1760,

when Danial Bernouilli used mathematical techniques to evaluate the effectiveness of variolation against smallpox (Anderson & May 1991). The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) have recommended modelling, using mathematical and statistical techniques, as a valuable tool in the process of developing rational policies and interventions in response to drug problems. It is seen as a means to simplify complex processes and structures, to heighten the understanding of key elements, highlighting what data needs to be collected, and a validated model could be used to predict outcomes of different scenarios (EMCDDA 2001). Hence, a second objective of this thesis is to apply infectious disease modelling techniques to cocaine use among the general Irish population and in particular an exploration of intervention strategies. Policy makers must allocate scant resources as effectively as possible in order to reduce user numbers but it is difficult to assess and compare the effectiveness of intervention measures. Behrens & Tragler (2001) developed a model to explore through simulations the most effective intervention strategies but no such model has been developed in the Irish context. This thesis aims to fill this gap.

A third objective is to apply modelling techniques to the problem of initiation to cocaine use. Little is known about age specific initiation though Almeder *et al.* (2001) showed that the introduction of age-specific aspects into drug initiation models gave detailed insight into the underlying processes of drug epidemics and their control. Anderson & May (1991) suggested the use of Who Acquires Infection From Whom (WAIFW) matrices in exploring age specific infection in relation to infectious diseases and such models have been applied to modelling the impact of immunization on the epidemiology of varicella zoster virus (Brisson *et al.* 2000), to the 1957 (Asian) influenza pandemic in the United Kingdom (Vynnycky & Edmunds 2008) and to predicting the impact of measles vaccination in England and Wales (Babbad *et al.* 1995). To date however a model utilising WAIFW matrices has not been applied to initiation to drug use therfore this will be undertaken in the present study.

The aims and objectives as outlined are summarised in the following section.

1.2 Aims and Objectives

In this thesis established statistical and mathematical models are developed and applied to

- Estimate the prevalence of cocaine use among opiate users in treatment
- Investigate by simulation the effectiveness of interventions aimed at prevention of initiation to cocaine use as compared to measures encouraging cessation of cocaine use in the general population
- Investigate age differentiated initiation using WAIFW matrices and in so doing produce age differentiated incidence estimates where incidence is defined to be the number of new cocaine users in a given year

In the following section a breakdown of the thesis by chapter is presented.

1.3 Thesis Outline

In the next chapter the pharmacology of cocaine is described and methods of ingesting the substance and a brief historical overview are presented. Statistical evidence of cocaine use worldwide, in Europe and Ireland is provided. Following that, in chapter 3, reasons are given to justify a mathematical modelling approach to the problem of cocaine use, along with an overview of the types of models commonly available to the modeller. Next, a summary of the most familiar types of models used in epidemiological modelling are described along with the relevant differential equations. Finally, a historical overview shows how the mathematical modelling of infectious diseases has developed since 1760. In chapter 4, standard prevalence estimation methods are described and applied to available treatment data to provide estimates of the number of opiate users in treatment in the Northern Area Health

Board who were concurrently using cocaine for each year from 2006 to 2010. In particular capture-recapture and truncated Poisson methods provide cross validation of the obtained estimates. Results from the 2006/2007 household survey (NACD 2008) contextualise these estimates within the broader population as do figures obtained from the police database for 2006. Chapter 5 develops and validates a simple deterministic compartmental model which models the change in cocaine user numbers in the whole population between 2002 and 2006. This model is then used to simulate the effects of strategies aimed at prevention of cocaine initiation and compares them to effects of treatment and other post-initiation strategies which aim to reduce cocaine user numbers. Simulations allow conclusions to be drawn about which intervention strategies are most effective in reducing cocaine user numbers. Chapter 6 builds on the foundations of the simple model developed in chapter 5 and expands it to develop an age differentiated model suggested by Anderson & May (1991). This model uses the so called WAIFW matrix (Who acquires infection from whom). The model progresses through several stages and its solution provides both incidence estimates for each hypothesised structure along with a model which can be used to extrapolate prevalence into the future. Finally, in chapter 7 all results are critically discussed, limitations are identified and recommendations made for policy, practice and future research.

Chapter 2

Cocaine: Pharmacology and Prevalence both Past and Present

2.1 Cocaine Pharmacology

Cocaine (benzoylmethylecgonine) is an alkaloid extracted from the leaf of the Erythroxylon coca bush, which grows primarily in Columbia, Peru and Bolivia. Like all stimulant drugs, cocaine produces a psychoactive effect by interacting with the central nervous system, stimulating it to perform its ordinary functions more intensely. This system operates through the release of various neurochemical transmitters (from the nerve cells in which they are produced) and their binding to receptor sites on neighboring cells. The constant release and binding of these neurotransmitters forms a pathway of messages that travel throughout the body, sustaining life and making possible the organism's response to environmental stimuli. Cocaine is believed to work by blocking the dopamine reuptake transporter (DAT) and thereby increasing the availability of free dopamine within the brain (Volkow *et al.* 1997). This results in stimulation of the reward system of the brain located in the nucleus

accumbens, amygdala and anterior cingulate area of the brain (Garavan *et al.* 2000, Haney 2008) giving the user a feeling of intense pleasure. However, after periods of prolonged use, the depletion of the stores of dopamine and serotonin can lead to depression, which is known to reach suicidal proportions in some cocaine users with certain comorbid conditions (Roy 2009).

Cocaine also alters the function of a protein known as postsynaptic density-95. Long-term changes in this protein are thought to be involved in the process of learning and memory formation which account for users' strong memories of the drugs pleasurable effects (Koob *et al.* 2004). Cocaine also affects the diencephalon region of the brain responsible for temperature regulation. At the same time it causes constriction of surface blood vessels. The combination of these two effects can lead to hyperthermia (Hall *et al.* 1990).

Cocaine is typically used in conjunction with a range of licit and illicit drugs such as heavier alcohol and ecstasy use (Shearer *et al.* 2007, McCrystal *et al.* 2011). When cocaine is combined with alcohol, a certain amount of the cocaine ($< 10\%$) is biotransformed into cocaethylene (Repetto & Gold 1997). At high doses cocaethylene is estimated to be over 25 times more likely to induce death than cocaine by itself (Karan *et al.* 1998). Some users combine cocaine, either at the same time, or consecutively with heroin in a practice known as "speedballing". This practice in Ireland was found particularly among users who inject cocaine and some users explained they would not use cocaine without heroin to help them "come down" (NACD 2007).

Cocaine is a quick-acting drug but its effects are short lived. When cocaine is injected into the bloodstream it reaches the brain quickly, and users feel its effects within seconds. Inhalation also delivers cocaine quickly to the brain because air passages in the lungs are positioned close to capillary accesses to the bloodstream. When cocaine is sniffed, the onset of effect is slower because the drug must pass through the nasal mucosa before entering the bloodstream. Swallowing cocaine delays delivery to the brain even more because most of the drug is passed through the

gastrointestinal tract before it crosses through cell membranes into the bloodstream. Reinerman & Zimmer (1997) state that whatever the route of administration, within thirty to sixty minutes, the processes of biotransformation and excretion cut in half cocaine's concentration in the blood.

2.1.1 Methods of Administration

There are several ways of using cocaine. Cocaine hydrocloride can be inhaled through the nose, this is commonly referred to as "snorting". The cocaine powder is chopped on a glass plate with a blade to make a finer powder to improve absorption. The powder is arranged in lines, and a line is inhaled through a straw or rolled up bank note. Karch (1996) states that one gram of cocaine might yield 25 to 30 lines. Cocaine hydrochloride powder can also be mixed with water and injected into a vein. The estimates of the time required to reach the brain range from 3 to 30 seconds. Intravenous administration results in 20 times as much cocaine reaching the brain as intranasal use, but the effects last for only about 15 minutes (Strang 1993). Powder cocaine can also be ingested sublingually due to the rich supply of blood in the tissues of the mouth under the tongue. A fourth method, according to Karch (2002), is rectal use, a practice which is popular among male homosexuals as the local anesthetic properties of cocaine allow engagement in otherwise painful sexual practices. Cocaine powder can be mixed with a solvent such as ether and then a base compound such as ammonia to form what is called "freebase". This can be smoked, but although the fumes reach the brain in 7 seconds this method is not popular as the process of creating the "freebase" is dangerous due to the volatility of the chemicals used. A safer product was discovered in the 1980's which was called crack because of the crackling sound it makes when smoked. It is made by mixing cocaine hydrochloride with baking powder and water and then heating it till an oily base forms. Cold water is then added which hardens the cocaine base into white balls. The crack can then be sold in small ready to use quantities fairly cheaply

(Doweiko 2006). In the next section the history of cocaine use to the present day will be examined.

2.2 History of Cocaine Use

The use of cocaine is not a new phenomenon. In South America the leaves of the coca plant are chewed to give energy, ease hunger pangs and to help against altitude sickness. Anecdotally, it has been reported (July 2011) that, a bag of coca leaves can be purchased in Bolivia for five euro cents and travelers in Columbia can pay nine US dollars to visit a coca farm and see how cocaine is made (Youtube 2008). The practice of chewing coca has been scientifically proven to date back to at least 2000 years ago. Using methods originally developed in forensic science, Cartmell *et al.* (1991) tested human hair from mummified bodies to determine the presence of cocaine and benzoylecgonine. The earliest archaeological evidence, however, dates its use as far back as 2500 BC. Small lime containers, thought to be used in coca chewing, were found in southwestern Ecuador, an area inhabited by the Valdivia culture. On the Peruvian coast, coca paraphernalia have been dated to between 2500 and 1800 B.C. Based on the archaeological evidence of coca-leaf tomb offerings, the practice seems to have continued among the various tribes inhabiting the western coast of South America up to the time when they were taken over by the Inca Empire sometime in the 1400s (Plowman 1985).

The ruling Inca family appointed official cocaine collectors, who gathered all the cocaine that was grown and turned it over to the royal family. Since the coca leaf was believed to be of divine origin even prior to the Incas, its use was a privilege reserved for members of the highest classes. This policy gradually eased over the following 200 years. By the time of the Spanish conquistador Francisco Pizarro's advance on Cuzco, which completed his conquest of Peru in 1536, coca had lost much of its earlier significance and was no longer a symbol of exclusive political rank or of social status. The Spanish recognised that the coca habit was important to the health and

motivation of the Andean Indian. Phillip II of Spain in a crown law of 1569 declared the coca habit essential to the well-being of Andean Indians while concurrently urging the missionaries to end the idolatrous use of the plant (Petersen 1977). During the Spanish period coca leaf production experienced a major period of growth. This was due to an apparently more widespread consumption of the leaf in traditional farming communities as well as heavy use among the Indian miners whose numbers increased with the growth of a major mining industry in the southern districts of Potosi in modern day Bolivia. By the end of the sixteenth century, these mines had brought some 100,000 Indian labourers to the region. The coca leaf in this period often served in place of money wages and was the most highly commercialised Indian product in the colonial Andean world, sometimes serving as money in even Spanish commercial exchanges. The trade was so lucrative that Spanish hacienda owners gradually became involved. By the 1950s, 55% of the coca production in the Yungas region of Bolivia was by the haciendas (Klein 1985).

Cocaine leaves were imported to Europe along with gold, silver, tea and tobacco and in the 1860s coca wine became popular, the most popular brand of which was Vin Mariani. This contained 60 gram of cocaine leaves soaked in red wine. Two glasses of this wine would be equivalent to one line of cocaine in the present day. Three events happened which increased the popularity of cocaine (Musto 1989, Karch 1999). The first was the publication in 1884 of Freud's paper "Uber Coca" which extolled the virtues of cocaine. The second was the discovery in the same year by Koller of its anesthetic properties. The third was the discovery of a method to semi-refine cocaine before shipment which increased the amount that could be shipped thus decreasing shipping costs. As a result of these developments, sales of one of the top cocaine producers, Merck, located in Darmstadt, Germany, increased from 3/4 of a pound in 1884 to 158,352 pounds in 1886. In 1886 the well known soft drink Coca-Cola was introduced. It was initially marketed as a patent medicine and contained two main active ingredients, cocaine, extracted from coca leaves and caffeine obtained from kola nuts. It was claimed that the drink cured diseases as varied as morphine

addiction and impotence (Karch 1999). The combination of the ready availability of cocaine and lack of awareness of possible side effects led to waves of what has been described by Doweiko (2006) as epidemics of cocaine abuse in Europe from 1886 to 1891 and in both Europe and the United States between 1894 and 1899.

From around 1910 onwards, articles began to appear in the British Medical Journal and the Lancet about nasal damage due to cocaine snorting (Karch 1999). In the United States the political mood had also changed and was moving towards a prohibition stance both for alcohol and other substances. This culminated in the Harrison Narcotics Act passed in 1914 whereby non-medical cocaine use was prohibited. In the United Kingdom the "Dangerous Drugs Act" was passed in 1920 (Berridge 1978). As a consequence of increased information about possible dangers of cocaine use and general loss in popularity generally, cocaine use diminished in the 20's and 30's and in the US by the 1950s it was no longer considered a problem worthy of law enforcement attention (Musto 1989). Cheaper and more readily available amphetamines became more popular. In the 1960's and 1970's during the "flower power" period LSD and cannabis were popular. Gradually as the dangers of amphetamine use became known among users, their popularity decreased and there was a resurgence in the use of cocaine again in the 1980s. The discovery of a procedure to manufacture "crack" cocaine in the 1980s opened up its use to less affluent sections of the population.

In the following section, cocaine use in recent times worldwide, in a European context and in an Irish context will be examined.

2.3 Present Day Use

2.3.1 Cocaine Use Worldwide

Since 1997, the United Nations Office on Drugs and Crime have produced annual reports documenting prevalence and international production and trafficking

of cannabis, opiates, cocaine and amphetamine type stimulants. According to their latest report (UNODC 2011), between 14.2 and 20.5 million people aged 15-64 used cocaine in 2009, representing between 0.3% and 0.5% of the world population. In order to compare prevalence rates worldwide, annual prevalence rates, defined to be the number of users as a percentage of the 15-64 year old population, are reported for the latest available year. Figure 2.1 gives an overview of worldwide cocaine prevalence, and this is followed by a brief summary of prevalence worldwide presented in this report (UNODC 2011). It is important to note however, that these figures are dependent on figures being supplied by national governments, hence countries who have less resources to carry out prevalence estimation are less well represented than more developed and economically better off countries.

North America

The latest figures from the report (UNODC 2011) indicate that North America comprising the United States, Canada and Mexico continues to be the region with the largest cocaine use, accounting for one third of world consumption. Annual prevalence in the region was 1.9% in 2009 representing a decrease of 0.5% on the 2006 figures. The United States continues to have the highest prevalence estimates at 2.4% in 2009 according to the US Department of Health's Substance Abuse and Mental Health Services Administration (SAMSHA) (UNODC 2011). Prevalence in Canada was slightly lower at 1.4%, also showing a decrease from 1.9% in 2008. In Mexico, one of the major supply corridors of cocaine to the US, the annual prevalence of cocaine use relatively low at 0.4%.

South, Central America and the Caribbean

In South and Central America and the Caribbean according to the UNODC report (2011), the estimated number of annual cocaine users ranges between 2.6 and 2.9 million people aged 15-64. In South America annual prevalence rates range between 0.3% and 2.6%, with highest prevalence in Argentina (latest figures from 2006)

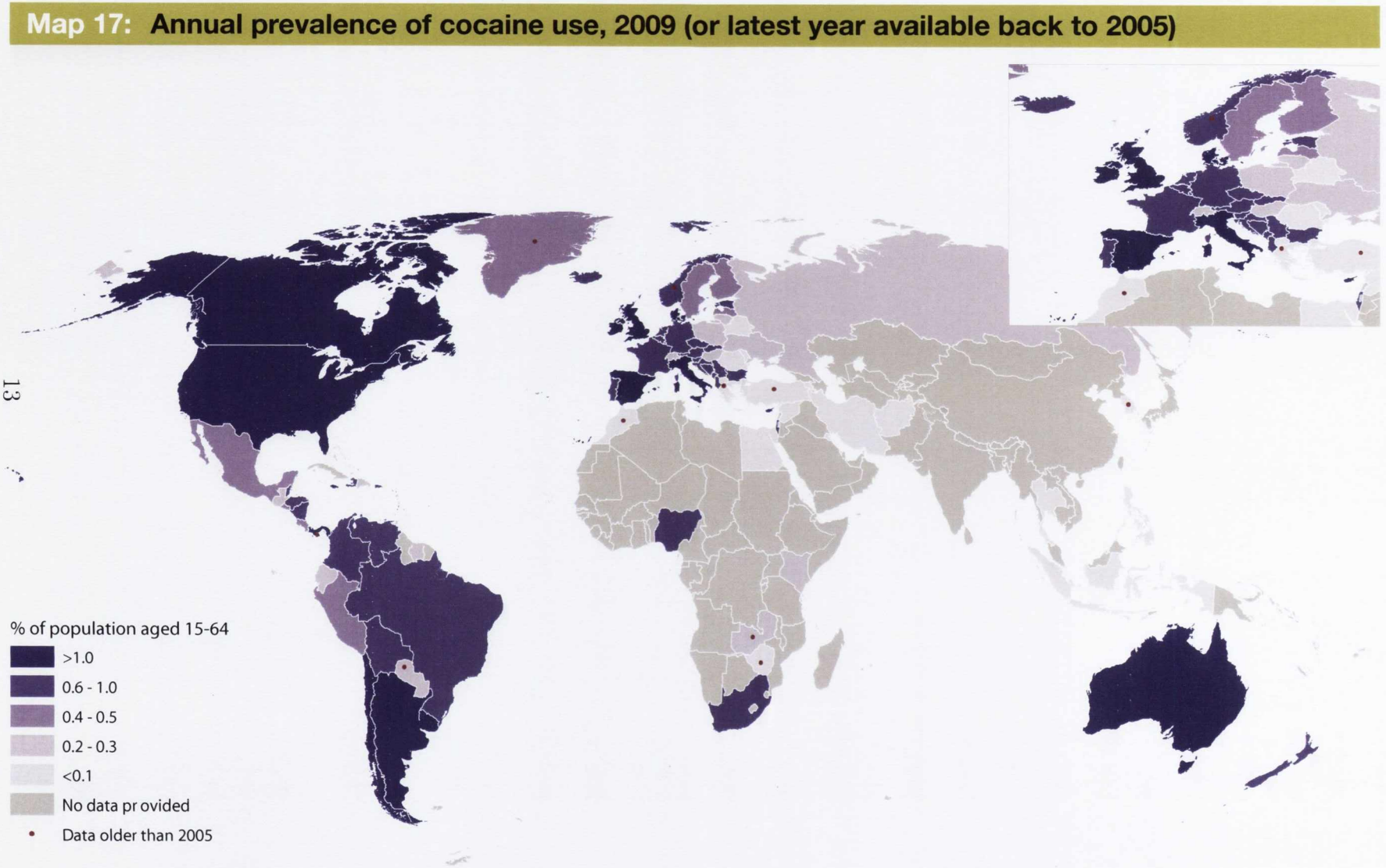

Figure 2.1: UNODC 2011: Overview of world cocaine prevalence

followed by Chile at 2.4% (latest figures from 2008). Although Brazil has a lower prevalence rate of 0.7%, because of its large population, the country has the highest number of cocaine users (900,000) in South America. Prevalence rates in the top cocaine producing countries of Columbia, Bolivia and Peru are at 0.8%, 0.8% and 0.5% respectively. Estimates of rates for Central America and the Caribbean range between 0.2 and 1.7%, though there are no figures available for 12 of the Caribbean islands and all but 2 estimates predate 2006.

Oceania

Only two countries in Oceania have supplied national cocaine prevalence estimates, Australia and New Zealand. Both countries showed increased prevalence between 2003 and 2007 but the most recent trends show a decrease in use. The prevalence rates are 1.9 and 0.6 respectively (UNODC 2011).

Africa

Information on the extent of cocaine use is only available from a limited number of countries in Africa. The annual prevalence of cocaine use is estimated by UNODC (2011) as being between 0.2% and 0.8% of the population aged 15-64, corresponding to between 940,000 and 4.4 million people. This estimate should be viewed with caution due to the lack of data supplied by the national governments. Of the 13 countries in East Africa only one, Kenya, supplied last year prevalence rates of 0.3% based on a household survey. In North Africa, 2 of the 6 submitted estimates, Egypt and Morocco, both less than 0.1%. The highest prevalence rate in Africa of 0.8% was seen in South Africa, though again only 3 of the 11 countries submitted details. Zambia and Zimbabwe reported rates of 0.2% and 0.1% respectively. Of the 25 countries in East and central Africa Nigeria reported rates of 0.7% and Cape Verde of 0.2%.

Asia and the Middle East

Very little is known about cocaine prevalence in these regions but of the 11 countries who did submit rates, 9 had rates of 1% or less and the highest rate reported, Israel, was only 0.6%. This illustrates the reliance of the UNODC on nationally returned figures and the fact that they are not always forthcoming particularly in the less well off nations.

In the following section a more detailed overview of prevalence rates in Europe is given.

2.3.2 Cocaine Use In Europe

Prevalence

The EMCDDA has produced annual reports on "The state of the drug problem in Europe" since 1996. In the latest report, (EMCDDA 2010), prevalence figures were compiled from the latest available national reports, giving the best available estimates of cocaine use in Europe. Lifetime prevalence, defined as the number of the sample who reported ever having used cocaine at the time surveyed, was estimated as being approximately 14 million people or 4.1% of the population. Last year prevalence was reported to be 4 million or 1.9% and last month prevalence 2 million or 0.5% of the population of the region. This prevalence was not evenly distributed, with the countries of Eastern Europe showing very low prevalence rates, whereas four countries were above the average for the region in all three categories, namely the United Kingdom, Spain, Ireland and Italy. When prevalence rates are limited to the 15 to 34 year old age group, the rates for the four highest countries are higher again, ranging from 8.2 to 14.9% for lifetime prevalence, from 3.1% to 6.2% for last year prevalence and from 1.0% to 2.1% for last month prevalence. Interestingly, the Netherlands has one of the lowest prevalence rates at around 1%, considering that it along with Spain is one of the major entry locations of cocaine into Europe. Among 15-34 year olds, the United Kingdom and Spain both have prevalence rates higher

than the US or Canada implying that they have the top two rates for this age group in the world.

Treatment for Cocaine

Treatment figures could be viewed as one indicator of the societal burden of cocaine misuse in Europe. The EMCDDA statistical bulletin reports, for 27 European countries, the percentage of those in treatment who sought treatment for cocaine as their primary drug of misuse in either inpatient or outpatient facilities (EMCDDA 2011a). This figure combines both those starting a treatment episode for the first time in 2008 (or the most recent data to 2008 supplied by the country in question) or those returning to treatment. The average percentage (expressed as a percentage of numbers for which primary drug is known), combined for either type of facility, was 8%. This figure disguises large differences between the countries with low prevalence of cocaine use, mainly in Eastern Europe, and those countries where it has become problematic for a considerable number of individuals. In Spain, where the figure was highest, the percentage of those in treatment who sought treatment for cocaine as their primary drug of misuse was 46 % as compared to nine countries reporting cocaine treatment percentages of less than 1%.

Incidence percentages, defined to be the percentages of new users entering treatment for problem cocaine use for the first time, can be indicative of trends. The average incidence percentage, calculated over the 23 countries who provided such data, increased to 11%. Of the 24 countries, 21 countries reported an increase in the percentage of new clients entering treatment for problematic cocaine use as compared to the percentages of existing clients (EMCDDA 2011a).

Deaths

Deaths caused by cocaine use are another very important indicator of the social burden of its use. However, a causal link between cocaine use and mortality is very difficult to assess. Although there are risks associated with cocaine use such

as increased risk of chest pain, heart attack, stroke and respiratory failure (House 1990), relating cocaine use to mortality remains difficult. Difficulties in establishing any causal links are complicated by the fact that cocaine users tend to use other substances alongside cocaine (Vroegop *et al.* 2009). The reporting of cocaine related deaths to the EMCDDA bears this difficulty out (EMCDDA 2011b). Reporting is incomplete and usually deaths involving cocaine show a combination with other drugs (alcohol, opiates and others). Three countries, namely Germany, Italy and Spain, reported the percentage of drug deaths that were due to cocaine alone and percentages were respectively 12%, 9% and 4%.

2.3.3 Cocaine Use in Ireland

Although Ireland and in particular the Dublin region experienced a heroin epidemic in the 1980's, up to the year 2000 there was little evidence of cocaine use in Ireland. Indeed, in one of the first studies undertaken, Mayock (2001) concluded that, despite indicators of a slight increase in cocaine use during the 1990's as evidenced by treatment data and cocaine seizures, cocaine use remained rare among school-going adolescents. Lifetime use in the general population at that time was around 1%. This situation changed gradually from 2000 onwards.

Prevalence

To date, two population drug surveys have been carried out by the National Advisory Committee on Drugs (NACD) in the Republic of Ireland in conjunction with the Drug and Alcohol Information and Research Unit (DAIRU) in Northern Ireland and the results of a third will be released at the end of 2011 or early 2012. Based on the sample proportions obtained these surveys, conducted in 2002 and 2006 respectively, point estimates and confidence intervals were obtained for lifetime, last year and last month prevalence of cocaine use alongside other drugs (NACD 2006, NACD 2008). The point estimate of Lifetime cocaine prevalence increased from

4.7% to 8.2%, an almost 75% increase. Last year prevalence increased from 2.0% to 3.1% while last month prevalence also was up from 0.7% to 1.0%.

The European School Project Survey on Alcohol and other Drugs (ESPAD) carried out three surveys reporting lifetime use of drugs among 15 to 16 year-old school children including Ireland (ESPAD 2001, ESPAD 2005, ESPAD 2009). For cocaine use the results are available separately for crack and powder cocaine. Among this cohort in Ireland crack cocaine use remained the same in 1999 and 2003 and then increased to 4% in 2007. Powder cocaine use also increased in the same years from 2% to 3% to 4% in this group.

The prevalence surveys undertaken to date indicate that cocaine use in Ireland is increasing and that further study is warranted to estimate both the hidden prevalence and incidence of cocaine use in Ireland.

Treatment data

Treatment data also suggests an increase in problem cocaine use. Bellerose *et al.* (2009) report that one-fifth (10,764) of all cases treated for problem drug use between 2002 and 2007 reported cocaine as a problem substance. The number of cases who reported cocaine as their main problem substance increased by 502%, from 128 in 2002 to 770 in 2007. The number of cases who reported cocaine as an additional problem substance increased by 128%, from 826 in 2002 to 1,885 in 2007. The Research Outcome Study in Ireland (ROSIE) was the first large-scale, prospective, multi-site, drug treatment outcome study in Ireland (Comiskey *et al* 2009). Cox & Comiskey (2007) report that 44% of opiate users in the ROSIE study, presenting for a new treatment episode, reported recent (last 90 days) cocaine use. In a one year follow up study, Cox & Comiskey (2011) found that cocaine users had more co-existing problems. Their analysis revealed that those who used cocaine at intake were more likely to use cocaine at 1-year follow-up, to commit crime, and to be homeless.

This worrying increase in cases presenting for treatment as a result of cocaine use

and the negative impact which cocaine use has on those in treatment suggests that an estimate of the hidden prevalence cocaine use among those in treatment for opiate use would be useful to both service provider and policy maker alike.

Deaths

Many deaths are caused by poisoning (both intentional and unintentional), where the death is directly attributable to the consumption of cocaine (alone or in combination with other substances). Deaths among cocaine users (whether the user is dependent or non-dependent) may very often be indirectly attributed to their drug use. Non-poisoning deaths may be attributed to the medical causes such as cardiotoxic effect of cocaine or to actions taken while under the influence of drugs, such as accidents caused by impaired judgment or to the exacerbation of an already present psychiatric illness placing the individual at a greater risk of suicide (trauma). Over the period, among all drug users the most common causes of death owing to trauma were hanging and road traffic collisions. The most frequent medical causes of death were cardiac events (118, 25.2%), respiratory infections (83, 17.7%) and liver disease (48, 10.2%). The latest figures from the National Drug-Related Deaths Index (2011), where presence of cocaine in the system of the deceased was found (though not necessarily exclusively cocaine), are shown in figure 2.2.

Law Enforcement Data

Routinely gathered data by law enforcement bodies also inform about cocaine use trends. Connolly (2008) shows that proceedings for possession of cocaine undertaken by the Irish police (An Garda Síochána) rose from 1015 in 2003 to 2442 in 2006 (see figure 2.3). Seizures, by both the Police and Customs and Excise, rose steadily up to 2007, after which they declined (Connolly 2010). This is shown in figure 2.4 and it is noteworthy that similar trends were present for seizures of all drugs. It is not known if this is due to a reduction in use of all drugs or due to a change in police activity levels (Connolly 2010).

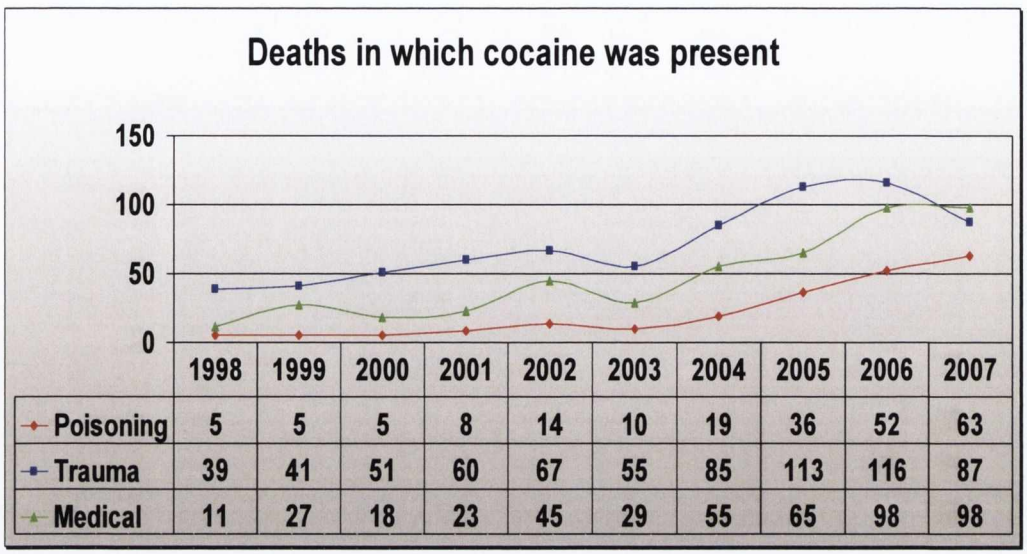

Figure 2.2: Trends in cocaine related deaths from 1998 to 2007.

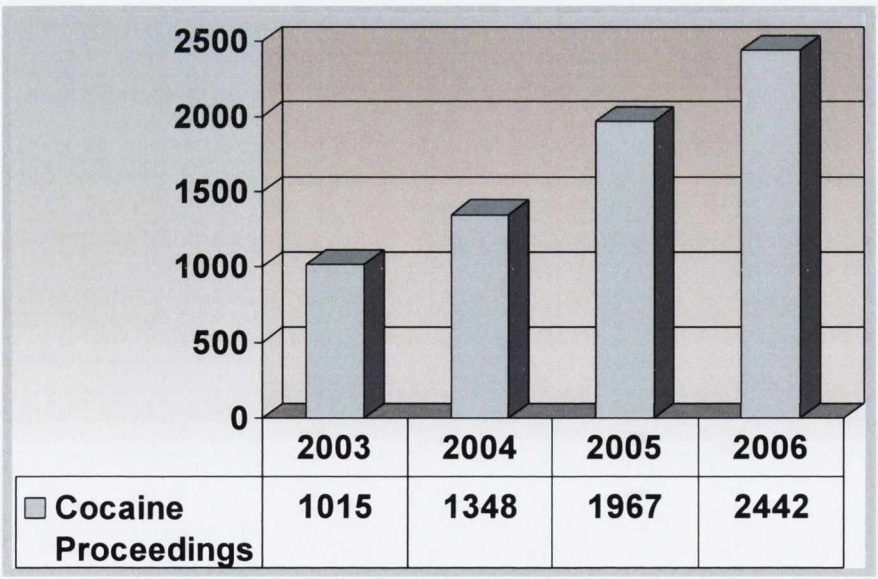

Figure 2.3: Trends in the number of cocaine seizures reported by Connolly and compiled by the Central Statistics office.

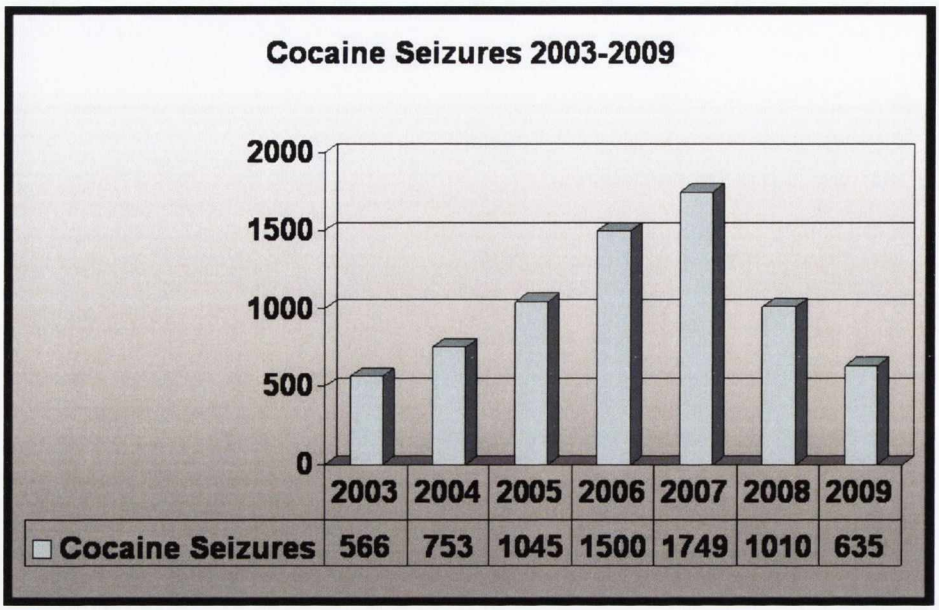

Figure 2.4: Trends in proceedings of cocaine use reported by Connolly and compiled by the Central Statistics office.

2.4 Chapter Summary

In this chapter the pharmacology of cocaine is described in an attempt to explain the motivational factors which encourage people to use the substance and also to explain how the substance can cause problems both physically and mentally to the individuals concerned. Methods of ingesting the substance are detailed as the risks involved in the use of the substance depend on the method of ingestion used. A brief historical overview is given to trace how a substance which has been widely used in South America for over two thousand years and continues to be used to the present day as part of their cultural heritage, became problematic when its active ingredient was isolated in the late 19th century. For this reason, aside from countries in South America most countries have taken a negative stance on the substance, declaring its use illegal and regarding its use as problematic. Statistical evidence of cocaine use is gathered worldwide and an overview of the extent of the problem worldwide, in Europe and Ireland is provided. Having established that all the available routinely gathered indicators point to an increase in the prevalence of cocaine use in Ireland

both in the general population and among those in treatment for opiate use, it is clear that its negative impact on treatment, mortality and crime statistics indicate a pressing need for more precise estimates of its hidden prevalence and in particular of the precise extent of its use among those in treatment. Primary data obtained from a regional health board and spanning the years 2006 to 2010 is available and will be described in detail in chapter 4. Secondary data from a nationally undertaken all-Ireland general population drug prevalence survey is also available for 2002/3 and 2006/7. In order to extract the necessary information from this data as set out in the research aims of this thesis, it will be necessary to use various statistical and mathematical modelling techniques. Hence, in the following chapter the role of statistical and mathematical models in epidemiology and their application to drug use is examined.

Chapter 3

The Mathematical Modelling Process and its Historical Evolution

3.1 Justification of the Use of Mathematical Models

The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) recommend modelling, using mathematical and statistical techniques, as a valuable tool in the process of analysing and make sense of data on drug use and its consequences in the European Union (EMCDDA 2001). Roberts & Heesterbeek (1993) make the following points justifying the use of mathematical models in epidemiology.

- Models provide insight into, and understanding of, the relationships between the mechanisms operating at the level of the individual, and the phenomena that result at the population level.
- Formulating mathematical models requires precision about the underlying assumptions and can reveal potentially useful working hypotheses that might

otherwise go unnoticed. The analysis of mathematical models can lead to the discovery of concepts that turn out to play an important role in the epidemiology of infectious disease.

- An important use of models is in clarifying which parameters have a critical influence on the predicted dynamical behaviour of the population. This may lead to the discovery of key parameters, the numerical value of which may be unknown or, alternatively, the realisation that some parameters that researchers are struggling to measure are irrelevant to the dynamics of the infection.

- Models are valuable for performing thought experiments, for example to evaluate the efficacy of control measures in cases where actual experiments are impossible because of ethical or economic constraints.

Roberts & Heesterbeek (1993) warn however about the utility of using models to predict future trends for the following reasons:

- The most complex models for specific diseases are still (highly) oversimplified.

- Our knowledge of key parameters in the transmission process is often poor.

- Making models more complex rapidly leads to a proliferation of parameters, hardly any of which can be "guesstimated" with accuracy.

Having highlighted both the benefits and limitations of modelling techniques, the next section shows the steps which might be involved in setting up a mathematical model.

3.1.1 The Modelling Process

Vynnycky & White (2010) present a possible cyclic procedure which could be used to develop a mathematical model of the spread of an infectious disease. It is based on a procedure compiled by a group in Rotterdam following its experience of setting

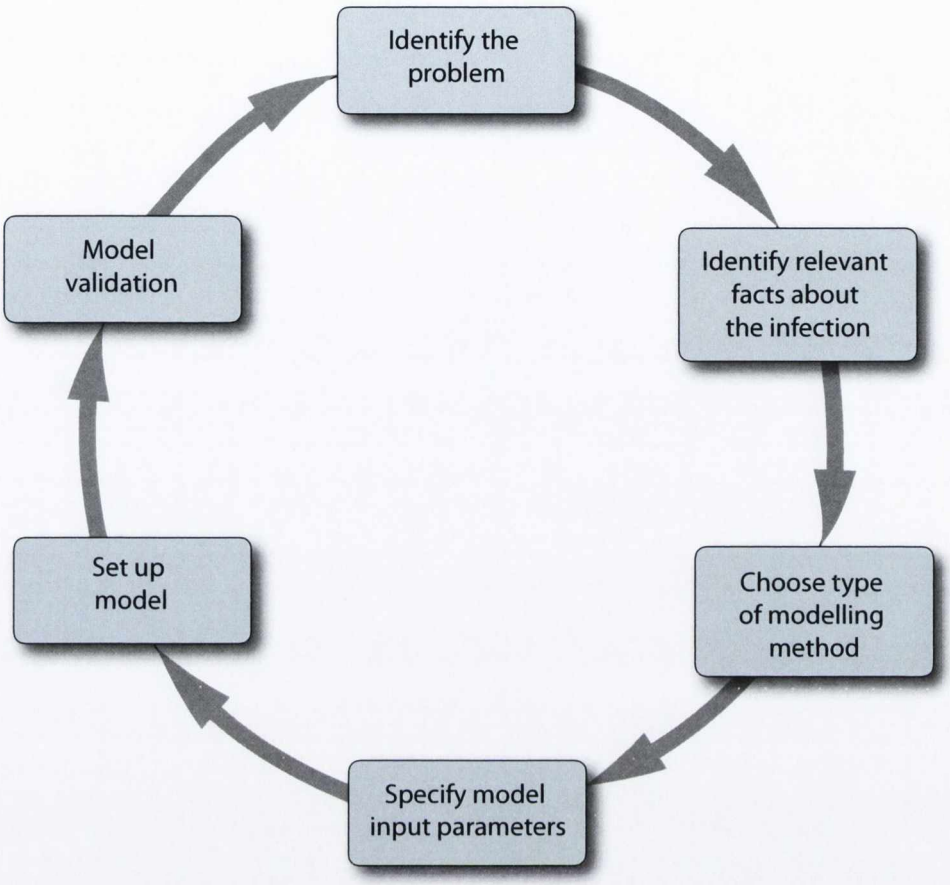

Figure 3.1: Steps in the development of a mathematical model adapted from Vynnycky & White (2010)

up detailed models of the transmission and control of onchocerciasis and schistosomiasis over a number of years. They state that each step in the process may need to be revisited several times before the model is completed. Figure 3.1 illustrates the process diagrammatically. In this chapter, the first three steps are discussed. The remaining steps will be the subject of the remaining chapters.

Step 1: Identify the problem

The problem has been identified as the need to measure the prevalence, incidence and diffusion of cocaine use for the first time in Ireland.

Step 2: Identify the relevant facts about the infection

On first inspection, one may not see a parallel between drug misuse and an infectious disease but on closer scrutiny many parallels can be observed. Firstly, like an infectious disease initiation to drug use occurs in a social context. The NACD (2006) found that the most common way of obtaining cocaine remained through family and friends. The percentage of people obtaining the drug in this manner increased from 33% in 2002/3 to 49% in 2006/7, significantly among women from 24% in 2002/3 to 70% in 2006/7. Secondly, in the same way that a percentage of people will recover from an infectious disease without any intervention, a large proportion of people who use cocaine, cease use of their own accord without any recourse to treatment. Thirdly, just as a person's previous medical history and individual susceptibilities can complicate the progress of an infectious disease, so too can use of other substances alongside cocaine influence a user's path to either problematic use or cessation of use. Finally as with an infectious disease, a prior infection can sometimes increase a person's susceptibility to reinfection, similarly experience with other drugs can often increase a person's likelihood of using a new drug such as cocaine. For the reasons outlined, it seems reasonable that models which have been developed within infectious disease epidemiology could be used within substance using epidemiology. Hence in the following section, we look at three of the modelling structures described by Vynnycky & White (2010) which have been developed to model infectious diseases, the SI model, the SIS model and the SIR model. A more extensive description of infectious disease models can be found in Hethcote (2000).

Compartments used in Epidemiological Models

In order to model the progress of an epidemic in a large population, the population may be divided into so-called compartments. The simplest model is the SI model comprising of two compartments, the susceptibles, S, and the infectious, I. This model is suited to diseases which are spread by human to human contact and from

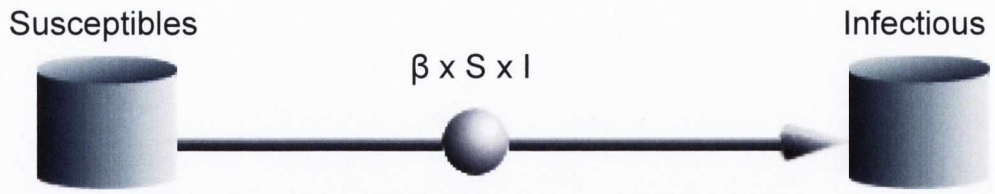

Figure 3.2: Flowchart SI Model

which there is no recovery. It could, for example, be used to model the spread of the HIV virus as HIV infected individuals remain infected and infectious for the rest of their lives. Differential equations are used to model the flow of individuals from one compartment to another over time. Parameters are used to define the rate at which these flows occur. The simplest models assume random mixing in the population and are based on the "mass action" principle introduced by Muench (1959), which assumes that susceptibles become infected at a rate which is proportional to the number of infectious present in the population. A basic SI model which does not include consideration of births or death is then

$$\frac{dS}{dt} = -\beta SI \tag{3.1}$$

$$\frac{dI}{dt} = \beta SI \tag{3.2}$$

where β is the proportion of contacts between susceptible and infectious that result in the transmission of the infection. The equation ensures that the number of susceptibles over time decreases at a rate which is proportional to the number of effective (in the sense of infecting) contacts between susceptible and infectious. This model has been solved analytically by Bailey (1975) and this analysis shows that eventually the entire population will become infected. The flowchart for this model can be seen in figure 3.2

For diseases where the infectious individual can be treated, recover and then reenter the susceptible group, the disease is modelled with an SIS structure. If the recovery rate is α then the flow chart is as shown in figure 3.3 and this is implemented in the

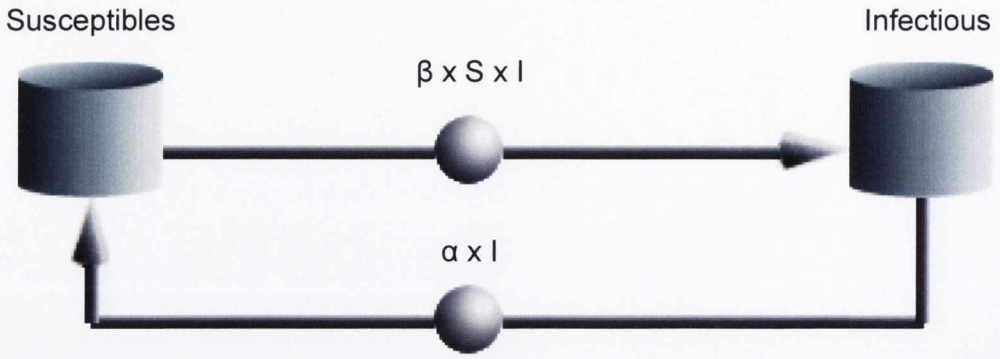

Figure 3.3: Flowchart SIS Model

following equations

$$\frac{dS}{dt} = \alpha I - \beta SI \tag{3.3}$$

$$\frac{dI}{dt} = \beta SI - \alpha I \tag{3.4}$$

The epidemiology of many diseases is such that once a person has been infected they develop immunity to the virus. Examples are measles, smallpox, mumps, varicella among others. Such diseases are modelled using the SIR model. Recovery usually is assumed to occur at a constant rate, γ, proportional to the number of infective individuals in the population. Bailey (1975) has also provided the solution to the simplest variation of this model. Results from the analysis of this model show that there is a critical population size of susceptibles required for an epidemic to occur. If the initial number of susceptibles is below this number, then the disease will immediately die out. If the initial number of susceptibles is above this number, there will be an epidemic of the disease until enough susceptibles have been infected to reduce the susceptible population to a level below the critical size. Following this, the disease will die out. In either case, the disease eventually goes extinct in the population. Figure 3.4 shows the flowchart corresponding to the equations (1.5) to (1.7) below.

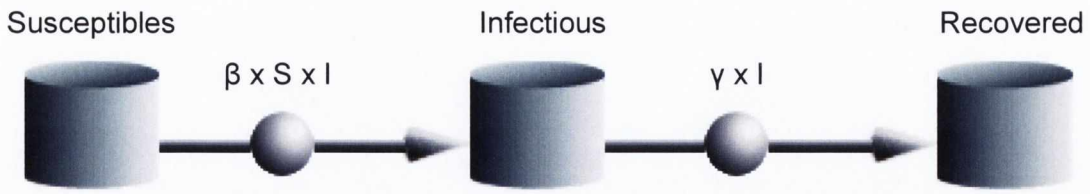

Figure 3.4: Flowchart SIR Model

$$\frac{dS}{dt} = -\beta SI \tag{3.5}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{3.6}$$

$$\frac{dR}{dt} = \gamma I \tag{3.7}$$

The epidemiology of other diseases may require a compartment for those who are infected but not yet infectious due to an incubation period. It is customary to call this compartment E for exposed. Other diseases may require a compartment for maternally conferred immunity, usually called M for maternal.

The models above do not explain the maintenance of diseases at an endemic level or the recurrence of epidemics over time. In order to have either of these situations, there must be continuous influx of new susceptibles to replace those that have become infected. This can be catered for by including births or migration in the model. Deaths can also be included if the modelling period calls for it.

These models, which have been used primarily for infectious diseases, can reasonably be used for initiation to drug use of those susceptible to it by those who are already users. In chapter 5, a simple SIR model will be adapted to model the effects of different intervention strategies on the spread of cocaine use in a population.

Step 3: Choosing the modelling method

In this section some of the most widely used modelling methods in Epidemiology are briefly described. Barnes & Fulford (2002) provide an excellent overview identifying the following modelling approaches:

- **Empirical:** In this approach a curve is fitted through a set of data using techniques such as least squares regression. The best fitting curve can then be used to extrapolate (predict) future data.

- **Deterministic:** The modeller using this approach tries to formulate equations describing the basic fundamental relationships between the variables of the problem. A deterministic system is a system in which no randomness is involved in the development of future states of the system. A deterministic model will thus always produce the same output from a given starting condition or initial state. This approach is widely used in epidemiological modelling and will be used in chapters 5 and 6.

- **Stochastic:** Another approach to modelling is the probabilistic or stochastic approach ("stochastic" comes from the Greek word to guess). Using this method the modeller tries to estimate the probability of certain outcomes based on the available data. This method is typically used for models of small populations when chance variation can play an important role in future states of the system.

- **Simulation:** Once a model has been developed and the key parameters estimated, simulations can be run on the model which test the effects of changes to either the initial conditions or to the parameter estimates. Simulations are assumed to be dynamic, in other words the model changes over time. They can be carried out for both deterministic and stochastic models. In a deterministic model, simulations can be employed to explore "what if" scenarios, they can demonstrate the effects of change in the system which in real life would be difficult to test. Such a simulation will be used in chapter 5 to compare the effects of prevention and treatment in reducing the numbers of cocaine users in a population.

- **Statistical:** Davison (2003) states that the key feature of a statistical model

is that variability is represented using probability distributions, which form the building-blocks from which the model is constructed. Typically it must accommodate both random and systematic variation. The randomness inherent in the probability distribution accounts for apparently haphazard scatter in the data, and systematic pattern is supposed to be generated by structure in the model. The art of modelling lies in finding a balance that enables the questions at hand to be answered or new ones posed. The complexity of the model will depend on the problem at hand and the answer required, so different models and analyses may be appropriate for a single set of data. Statistical models used in the capture-recapture method and the truncated Poisson method will be used in chapter 4.

The remaining steps in the cyclic process will be developed in following chapters. In the next section we give an overview of the development of mathematical modelling in epidemiology.

3.1.2 Historical Developments in Epidemiology

Bailey (1975) provides a comprehensive overview of the history of epidemiology. He points to the work of John Gaunt(1620-1674) and William Petty (1623-1687), whose work studying the London Bills of mortality marked a beginning of the study of vital and medical statistics and the understanding of large scale phenomena connected with disease and mortality. Epidemiologists regard John Snow's work as being a significant early development (Cameron & Jones 1983). His spatial studies on of the locations of cholera outbreaks led to the isolation of the Broad Street pump as being the source of the contamination and ensuing closure of that point responsible in halting the outbreak. Later, in 1873, William Budd established a similar manner of spread for typhoid (Pickles 1948).

Parallel to these detailed investigations were the broader studies of statistical re-

turns made by William Farr (1807-1883), who hoped to discover empirical laws underlying the waxing and waning of epidemic outbreaks. In 1840, Farr effectively fitted a normal curve to smoothed quarterly data on deaths from smallpox, assuming the constancy of "second ratios" of successive pairs of frequencies (Langmuir 1976). Dr John Brownlee, first director of the Statistical Department of the Medical Research Council from 1914 to 1927, fitted various Pearson curves to epidemic data on many diseases occurring a different times and places (Bailey 1975).

Deterministic Models

According to Hethcote (2000), Daniel Bernoulli was one of the first to create a mathematical model in Epidemiology, when in 1760 he formulated and solved a model for smallpox in order to evaluate the effectiveness of variolation of healthy people with the smallpox virus. By the end of the 19th century the general mechanism of epidemic spread, as revealed by bacteriological research, and the long familiarity with epidemiological data, together made possible developments of a new kind. Hamer (1906) in his study of measles epidemics considered that the course of an epidemic must depend on the number of susceptibles and on the number of contacts between them and infectious individuals. The assumptions made be Hamer, have provided the basis of all subsequent deterministic models and seem to have been inspired by the theory of the kinetics of chemical reactions, where it is referred to as the mass-action principle.

Another important early figure was Ronald Ross (1857-1932), the son of a General in the English army who had been stationed in India (Chernin 1988). He commenced the study of malaria in 1892. After two and a half years' failure, Ross succeeded in demonstrating the life-cycle of the parasites of malaria in mosquitoes. In 1902 he was awarded the Nobel prize and for his contribution towards the protection of troops against malaria. Ross developed mathematical models for the study of malaria, initiated in a report on Mauritius in 1908, elaborated in his Prevention

of Malaria in 1911 and further elaborated in a more generalized form in scientific papers published by the Royal Society in 1915 and 1916. These papers represented a profound mathematical interest which was not confined to epidemiology, but led him to make material contributions to both pure and applied mathematics. Although his models involved some use of probability, they were essentially deterministic in nature. Those related to pathometry (the determination of the proportionate number of individuals affected with a certain disease at a given time, and of the conditions leading to an increase or decrease in this number) are best known and to today constitute the basis of much of the epidemiological understanding of insect-borne diseases.

More elaborate mathematical studies of the same general type were undertaken by Kermack & McKendrick (1927) their most outstanding contribution being the Threshold Theorem. This theorem stated that the density of susceptibles must exceed a critical value in order for an epidemic outbreak to occur. Further deterministic work was carried out by Soper (1929) in his work on measles. He looked at difference equations and discovered a damped train of harmonic waves. This damped oscillation was not mirrored in the actual statistics for measles, a fact which Soper was unable to explain.

Of particular interest to the mathematical modeller of drug treatment are the models which deal with interacting species. The Lotka-Volterra equations, also known as the predator-prey equations, are a pair of first order, non-linear, differential equations frequently used to describe the dynamics of biological systems in which two species interact, one a predator and one its prey. They were proposed independently by Alfred J. Lotka in 1925 and Vito Volterra in 1926. According to Simon (1959), Lotka developed much of the methodology used today with regard to solutions of differential equations and analyses of equilibria. Simon states that Lotka in his book, "Elements of Mathematical Biology", first published in 1924, deals with the

general solution of systems of first-order differential equations and of the conditions of stability as well as types of equilibria, stable and unstable.

Stochastic Models

Smaller populations under investigation and the need to include elements of chance in epidemiological mathematical models, led to the development of stochastic models. According to Bailey (1975), McKendrick (1926) was the first to publish a genuinely stochastic treatment of an epidemic process. Whereas in deterministic models one takes the actual *number* of new cases in a short time interval to be proportional to the numbers of both susceptibles and infectious cases, McKendrick assumed that the probability of one new case in a short interval was proportional to the numbers of both susceptibles and infectious cases. This is a "continuous infection" model and entails an individual being himself infectious from the instant he receives infection until the moment he dies, recovers or is isolated. Examples were given of probability distributions for the total number of cases in a household when infection was introduced from outside. This pioneering effort did not attract much attention, however and similar models were only investigated 20 years later according to Bailey (1975).

Bailey describes an alternative probability model which established itself in the 1930's through the work of Greenwood in England and independently in the United States through the work of Lowell J. Reed and Wade Hampton Frost who were using the same kind of ideas in lectures and discussions. This model assumed that the period of infectiousness was comparatively short and that the latent and incubation periods could be regarded as approximately constant. Starting with a single or several cases in a closed group, new cases would then occur in a series of stages or generations. These cases occurring at any stage would, under suitable conditions, have a binomial distribution depending on the numbers of susceptibles and infectious individuals present at the previous stage. There would be a chain of binomial distributions. This theory known as chain-binomial theory was extended by

Greenwood, Bailey, and Abbey in the 1940's and 1950's (Bailey 1975). Some of the most important developments followed from advances made in the early 1940's in the mathematical handling of stochastic processes. Whittle (1955) showed that the probability distribution of the ultimate number of infected individuals in a population could be calculated by solving a certain set of singly recurrent relations and in so doing he provided a stochastic alternative to Kermack & McKendrick's threshold theorem. Bartlett (1957) developed by stochastic models of recurrent epidemics such as measles which overcame the damping problem of earlier Hamer/Soper models and were comparable to empirical results achieved using Monte Carlo simulation methods.

Developments since 1957

Since the appearance of Bailey's first book on mathematical modelling (Bailey 1957), the numbers of mathematical modelling papers have proliferated. The publication in 1991 of Anderson & Mays' *Infectious diseases of humans: dynamics and control* provided a textbook which serves as a starting point for numerous types of models, from static to dynamic models, involving either homogeneous or heterogeneous populations and transmission models ranging direct to indirect transmission. Their work on age-related transmission in chapter 9 of the above text forms the basis of chapter 6 of the present work. The application of this type of deterministic epidemiological modelling to the spread of drug use was initiated in Ireland by White & Comiskey (2007). This present work builds on these foundations.

3.2 Recent Applications of Mathematical Modelling to Drug Use

3.2.1 Point Estimation Methods

In this section an overview is presented of recent applications of mathematical modelling to cocaine use in particular and drug use in general. A review of the literature reveals that the log-linear model as it is implemented in the capture-recapture method has been used extensively. Hay *et al.* (2010) include a comprehensive list of papers which use the method, hence only a brief overview is reported here.

Capture-recapture has been used to estimate the prevalence of opiate or heroin use in Australia (Larson *et al.* 1994, Choi & Comiskey 2003), the United States (Calkins & Aktan 2000), Spain (Domingo-Salvany *et al.* 1998) and the Netherlands (Buster *et al.* 2001) to mention but a few. Capture-recapture has also been used extensively in England and Scotland to estimate problem drug use defined as opiate and/or crack cocaine use by persons aged 15 to 64 (Hay & McKeganey 1996, Hay 2000, Beynon *et al.* 2001, Frischer *et al.* 2006, Hay *et al.* 2010). It has also used to estimate the prevalence of more specifically defined groups, such as injecting drug users or injecting drug users who are HIV positive, in Scotland, Russia and Thailand (Mastro *et al.* 1994, Davies *et al.* 1999, Platt *et al.* 2004). A London study focused on crack cocaine only (Hope *et al.* 2005).

In Ireland the method has been used to estimate opiate use by Comiskey & Barry (2001) and by Kelly *et al.* (2003). Despite the interest in cocaine use in Ireland as evidenced by reports (NACD 2003, Barry & Lawlor 2005, NACD 2007, Connolly *et al.* 2008), to date no study has been published which uses capture-recapture to estimate the hidden use of cocaine in Ireland, though a tender to undertake research into methods and data sources for the estimation of prevalence of problematic

opiate and cocaine use in Ireland was issued by the National Advisory Committee on Drugs (NACD) in 2010. In chapter 4, the method will be used to estimate the prevalence of cocaine use among opiate users in treatment in a local region of Dublin.

The truncated Poisson method is an alternative method which can be used to estimate the prevalence of drug use. McKendrick (1926) used this approach as early as 1926 in India to estimate the number of households unaffected by cholera. The method has since been used by Hser (1993) to estimate the prevalence of illicit drug use in Los Angeles and by Choi & Comiskey (2003) to estimate the number of opiate users in Western Australia. It has been also been applied to the estimation of the prevalence of drug injectors from needle exchange data in Scotland by Hay & Smit (2003) and to estimate the number of drug users in Bangkok (Böhning *et al.* 2004). We will apply this method for the first time in an Irish setting to cross validate the estimates obtained using the capture-recapture method.

3.2.2 Dynamical Mathematical Modelling Techniques

Sattenspiel (1990) writing about mathematical modelling of infectious diseases, jokes that there had been an epidemic associated with the development of such models. Indeed the graph, which is reproduced in figure 3.5 shows how this growth is similar to an epidemic curve. This proliferation of models has continued unabated to the present day showing no signs yet of dying out.

There has unfortunately been no shortage of infectious diseases to model. Alongside the all too familiar infectious diseases of old such as influenza, tuberculosis, measles, rubella, pneumonia, gonorrhea, dengue, yellow fever and so forth, a host of new infectious diseases have emerged. Hethcote (2000) lists a host of newly identified diseases including Lymes disease (1975), Legionnaires disease (1976), toxic-shock syndrome (1978), hepatitis C (1989), hepatitis E (1990), and hantavirus (1993). The human

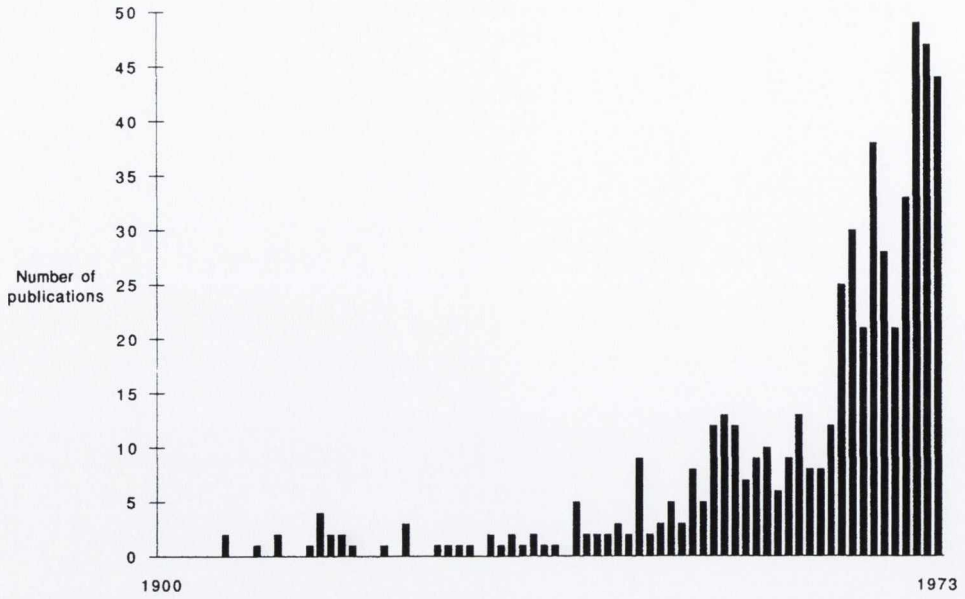

Figure 3.5: Graphic reproduced from Sattenspiel (1990) illustrating the number of publications concerned with mathematical models for contagious processes between 1900 and 1973 which were cited in Bailey (1975)

immunodeficiency virus (HIV), which is the etiological agent for acquired immunodeficiency syndrome (AIDS), emerged in 1981 with devastating effects throughout the world resulting in a large number of mathematical models attempting to throw light on its various aspects.

De Alarcon (1969) was one of the first people identifying the spread of heroin-use from one individual to another as similar to the spread of a contagious disease. His field study among heroin users, newly referred to the psychiatric services, was carried out in 1967 in Crawley, a small town near London. By investigating when and by whom each person was initiated to heroin use he was able to estimate the incidence of heroin use in the town. Mackintosh & Stewart (1979) developed the concept further with their model of a heroin epidemic, the main purpose of which was to assist in the conceptualisation and definition of the primary parameters. In the model they explored intervention strategies and concluded that early intervention was most effective. Rossi's (2004) "mover-stayer" model is a modified version

of an SIR model. The model considers the susceptible population as subdivided into two main groups: the group of stayers who are considered not at risk of "infection", and the group of movers who are. The model is then used to obtain "what if" scenario analyses which are obtained by simulation. Although this model reflects the drug using career very accurately, this very accuracy results in the use of six compartments and a resultant proliferation in the number of parameters the majority of which are then guesstimated.

Whereas European modellers tended to focus on heroin use, models of the drug problem in the United States tended to focus on cocaine use. Behrens & Tragler (2001) examined various models with the motivation of informing policy choices in how scarce resources should be allocated between various cocaine prevention programs. Their conclusion was that control measures, such as treatment and prevention, are most appropriate for specific stages of a drug epidemic and budget allocations across those measures should change over time. They concluded that prevention works best when there are relatively few heavy users at the beginning of an epidemic but that treatment is relatively more efficient at supporting the decline of drug abuse later in the epidemic.

Almeder *et al.* (2001) explore the effects of age differences in the multi-state user/non-user model. In this model the initiation rate involves three factors, namely the basic initiation rate, the reputation of the drug and the effect of prevention programmes. Each factor depends on age. The main factor for the reputation is how much influence a user of age b has on a non-user of age a, and whether the influence is positive or negative. Especially for young non-users, persons who are of the same age or a little older set examples, and their influence is therefore very high but persons who belong to their parents generation have only a small impact. Their model showed that a model with just one state but a heterogeneous age structure can produce cycles of drug use of the type observed historically. Furthermore, such

cycling is not the only possible outcome. It is therefore possible to explore conditions under which cycling occurs and contrast them with conditions under which drug use approaches a constant steady state.

An alternative approach to Almeder *et al.*'s multi state Markov chain age differentiated model is to incorporate age differentiation by means of a WAIFW matrix as suggested by Anderson & May (1991). Brisson *et al.* (2000) used WAIFW matrices to develop and apply a dynamic mathematical model of Varicella Zoster Virus transmission to predict the effect of different vaccination strategies on the age-specific incidence and outcome of infection. Vynnycky & Edmunds (2008) used WAIFW matices to apply an age-structured model of the transmission dynamics of pandemic influenza to data from the United Kingdom from the 1957 (Asian) influenza pandemic. Based on this model they drew conclusions about how school closures implemented at different stages of a pandemic, for different assumptions about contact between individuals, would affect its size and duration in the future. An earlier paper by Babbad *et al.* (1995) used WAIFW matrices in order to predict the impact of measles vaccination in England and Wales. To date this method has not been applied to any drug use.

Sánchez *et al.* (2011) present a model which analyses the evolution of cocaine consumption in Spain and uses the model to predict consumption trends over the next few years. The model is dynamic compartmental and based on epidemiological-type models. The model sensitivity analysis allows examination of intervention strategies. The model, though interesting contains one questionable assumption. The flow assumes that light users move into more regular users into problem users. The only exit from the user compartment is through treatment. This model does not allow for the fact that a large proportion of cocaine users quit without any intervention. Their quit rate is based on treatment success rates only. The model could be applied in the Irish context and we will try to remove what we see as discrepancies in the

model in the process.

In the Irish context the dynamical mathematical modelling of cocaine use has yet to be explored. White & Comiskey (2007) present an ODE compartmental model based on mathematical epidemiology of the drug-using career of opiate users. In this paper the authors identify 3 compartments namely, the susceptible group, the drug user group not in treatment and the drug user group in treatment. Following their model analysis they conclude that efforts to increase prevention are more effective in controlling the spread of habitual drug use than efforts to increase the numbers of individuals assessing treatment.

3.3 Chapter Summary

In this chapter modelling as it relates to infectious diseases and in particular to drug use is described and the various statistical and mathematical approaches are outlined. In estimating prevalence of cocaine use among those in treatment for opiate use, it is decided to apply the statistical method of capture recapture as recommended by both the EMCDDA (2001)and the UNODC (2003) which has been used in numerous peer reviewed papers in numerous countries and for a variety of drugs as outlined in a 3.6.1. The capture recapture estimates will be cross validated by the truncated Poisson method, an alternative method recommended by the EMCDDA (2001). The results of the 2006/2007 all-Ireland general population drug prevalence survey (NACD 2008) will be used both to place these estimates in context and as a benchmark to predict interim general population drug prevalence survey estimates.

Furthermore, infectious disease models from the time of Bernoulli's first model in 1760 are reviewed and their application to the modelling of drug use. It is necessary to choose between stochastic models which are more suited to small populations

where chance variation plays a significant role (Renshaw 1993) and deterministic models which are simpler and more useful on the macro level for large populations (Wiessing *et al.* 2001). It is decided that deterministic models are more appropriate in fulfilling the second objective of this thesis namely to explore intervention strategies which aim to reduce cocaine user numbers among the broader population. By keeping parameters to a minimum these can be estimated and more importantly validated from the known data. Simulations which are applied to validated models are arguably more useful than simulations applied to more complex models which model the complexities of the drug pathways more accurately but include a large proportion of guesstimated parameters. Hence, a simple deterministic model will be developed and validated in chapter 5 and the effects of interventions are simulated through manipulation of the model parameters.

The model in chapter 5 can be seen as the starting point for a more advanced model which uses WAIFW matrices as has been applied to infectious disease modelling and applies it to the problem of age differentiated initiation to cocaine use. This model reflects the iterative cyclic procedure outlined in figure 3.1 as the model development cycles through the stages "set up model" and "model validation" three times before the model is deemed satisfactory. This model further shows how development of a model can cast light on a phenomenon. By developing and validating an age differentiated model, incidence estimates for each age group under various mixing assumptions can be obtained as will be seen in chapter 6.

In the following chapter the first objective of this thesis is tackled, namely to provide an estimate of the hidden prevalence of cocaine use among the opiate using population who have had contact with the treatment services. The statistical methods of capture recapture, truncated Poisson and the multiplier method are explored in detail.

Chapter 4

Point Prevalence Estimation Models

4.1 Introduction to Prevalence Estimation

Within this chapter statistical methods for point prevalence estimation are critically assessed. Models appropriate for the estimation of cocaine use are then discussed and applied to Irish data. Findings on the first estimates of opiate users in treatment who are also using cocaine in a local region in the northern area of the capital city are presented. Results from each of the applied models are discussed in terms of their limitations and general applicability for other international regions.

Measuring the prevalence of an activity which is illicit is not easy. Measuring the prevalence of cocaine use is further complicated by the fact that it, like amphetamines, tends to fall into both the "recreational drug use" (powder cocaine) and "problematic use" (crack and injecting use). Bellerose *et al.* (2009, p. 2) reported two trends in cocaine use in Ireland.

> There appeared to be two patterns of cocaine use among those entering treatment: use of opiates alongside cocaine and use of combinations of

alcohol, cannabis and ecstasy alongside cocaine.

Hence, any efforts at prevalence estimation must take note of this dichotomy. From the statistical evidence presented in chapter 2 it is apparent that cocaine use and drug use in general is not simply an Irish problem. A worldwide underworld business has emerged in the trading of illegal drugs and to countries such as Columbia (the world's largest producer of coca derivatives), and Afghanistan (world's largest producer of opium), their supply has become an important component of their economies (Central Intelligence Agency 2010).

To combat what it saw as a world problem, the United Nations set up the United Nations Office on Drugs and Crime (UNODC) in 1997. One of the three pillars of this organisation is research and analytical work to increase the knowledge and understanding of drugs and crime issues and expand the evidence base for policy and operational decisions. A key component of this pillar is the estimation of the prevalence of the use of illicit drugs.

The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) was established in 1993. The EMCDDA exists to provide the EU and its Member States with factual, objective, reliable and comparable information concerning drugs, drug addiction and their consequences. Both these organisations have recognised the fact that methods of prevalence estimation can be divided into two categories, those which seek to measure the prevalence of more or less stigmatised drug use.

The first of the methods available in assessing cocaine prevalence is the general population drug prevalence survey (also referred to as the household survey in the literature). This method is seen both by the EMCDDA and the UNODC as being a useful tool in assessing the prevalence of cannabis, alcohol and tobacco use but less useful in measuring opiate and cocaine use. The UNODC point to its lack of

usefulness in measuring the use of drugs lacking social acceptability and the under-reporting caused by a lack of representation of groups at particular risk.

> However, for a number of reasons, general population surveys can perform poorly in assessing some types of drug abuse. In particular the more socially unacceptable types of drug abuse, such as heroin or cocaine addiction or drug injection, are often not measured well by household surveys, and underreporting can be a problem. Household surveys are also technically complex and resource intensive undertakings that are simply not practical in many developing countries. (UNODC 2003, p. 3)

In the Irish context, it is important to differentiate between cocaine *addiction* which may be socially unacceptable and *recreational* cocaine use which may not be. Various authors writing about cocaine use in Ireland (Mayock 2001, NACD 2003, Galvin & Cambell, 2010) have indicated that cocaine is seen as a safe drug and they pointed to the lack of stigmatisation of its use in powder form as compared to injecting of cocaine or use of crack cocaine. The reservations expressed by the EMCDDA (1999a, p. 9) are arguably more pertinent in the Irish context.

> There is the methodological problem that general population surveys are less likely to include harder to contact people who may be more likely to be problem drug users, in particular sub-populations such as the homeless or those living in institutions. People may be reluctant to divulge information on matters which are deemed to be socially unacceptable, therefore general population surveys may be more appropriate in assessing the prevalence of recreational drug use, but not for problematic use. In addition very few people use drugs such as heroin therefore detecting the use of such drugs within a general population survey is quite difficult.

Despite these limitations in the 27 EU countries the general population drug prevalence survey continues to be popular with all but two countries, Slovakia and Romania, carrying them out in the last 15 years. In Ireland three household surveys have been carried out to date, the first in 2003/2003 (NACD 2006), the second in 2006/2007 (NACD 2008) and the most recent one in 2010/2011 for which only the first initial results are available (Horgan 2011). In the present study these results are used to contextualise the prevalence of cocaine use among opiate users in the NAHB, the main focus of the present chapter.

In order to estimate the prevalence of cocaine use among those in treatment who, because of their small numbers and more chaotic lifestyles are more likely to be missed in a general population drug prevalence survey, alternative methods of prevalence estimation are required. The UNODC have produced a toolkit outlining indirect methods for estimating the size of the drug problem in a particular country or region (UNODC 2003) which are more appropriate for the estimation of hidden populations. The EMCDDA have also published guidelines to estimate the prevalence of problem drug use (EMCDDA 1999). These methods are most useful in estimating local prevalence and are thus suitable for estimating prevalence in the Health Service Executive (HSE) region of the Northern Area Health Board (NAHB). In order to apply these methods the necessary data sources must be available. In the following section an overview is given of the data sources which form the basis of the present study.

4.2 Data Sources

In this section, the data sources used in this chapter are introduced. Police data, gathered in the course of routine police work is used to enumerate the numbers recorded on the police database with an association with cocaine use. Prevalence rate data, reported in a national general population drug prevalence survey is used

to estimate the number of cocaine users in the general population in the study region. It is expected that this estimate reflects the level of recreational use in the region. Urinalysis test results are analysed using the truncated Poisson and capture-recapture methods to provide an estimate of the hidden prevalence of opiate users in treatment who are also using cocaine.

4.2.1 National Drug Prevalence Household Survey

The National Advisory Council on Drugs (NACD) carried out a general population drug prevalence survey in the Republic of Ireland during 2006/7. Prevalence percentages for the regional drugs task force areas (RDTFs), including the Northern Area Health Board region were released in 2009 (NACD 2009). The last year prevalence percentages of cocaine use in the Northern Area Health Board can be seen in table 4.1.

Table 4.1: Last year prevalence rate with 95% confidence intervals of cocaine use in 2006/2007 (Results for the Northern Area Health Board)

	Total	Male	Female	15-34	35-64
%	3.3	5.8	0.8	5.6	1.1
CI+	6.2	12.1	2.9	12.2	3.4
CI-	1.5	2.2	0.1	2	0.2

4.2.2 Northern Area Health Board Population Figures

A National Census was undertaken in 2006 and population figures for the local health areas of Dublin North and Dublin North Central can be obtained from the Health Service Executive website (HSE, 2007). These population figures are shown in table 4.2. As the NACD data is chosen from a representative sample of the population, the prevalence percentages from this survey can be applied to the population of the region and the cocaine prevalence among the general population inferred.

Table 4.2: Population by Gender and Age of the Northern Area Health Board

	Total	Male	Female	15-34	35-64
Dublin North Central HO	91,672	46,437	45,235	50,372	41,300
Dublin North HO	154,717	76,215	78,502	72,661	82,056
Total	246,389	122,652	123,737	123,033	123,356

4.2.3 Pulse Data

The PULSE system (Police Using Leading Systems Effectively) is a computer system used by an Garda Síochána, the Irish police force. It was introduced in November 1999 and is run by the PULSE Project teams of the IT section of an Garda Síochána. The area covered by the Northern Area Health Board is covered by two Garda Divisions, the Dublin Metropolitan Northern Region and the Dublin Metropolitan North Central Region. It was originally intended to apply the capture recapture method to this data and other data sources but we were not successful in obtaining the necessary access to initials, gender and date of birth for all the data sources which are prerequisites to matching data sources. Hence this data is merely enumerated to shed some light on the extent of the burden of cocaine use on the legal and justice system. In 2006 data relating to 613 individuals was recorded in the pulse system as a result of an association with cocaine. The precise level and nature of this association is however not specified. Of these, 533 were male (87%) and 80 were female (13%). There were 504 in the 15 to 34 age group, 105 in the 35 to 64 age group and 4 were in neither.

To put these figures in perspective, nationally there were 3231 cocaine associated records. Of these 89.1% (n =2880) were male and 10.9% (n = 351) were female. Age was recorded for 99.4% (n = 3210) of the cases with the highest percentage of 49.2% (n = 1589) recorded for the 12 to 25 age group followed by the 26 to 34 age group accounting for 36.6% (n = 1183) followed in turn by the 35 to 44 age group accounting for 11% (n = 69). Those aged over 45 accounted for 2.5% (n = 82).

4.2.4 Drugs and Aids Information System

Drugs and Aids Information System (DAIS) is a shared client care record database which is used by drug treatment workers in the Northern Area of the HSE who are treating opiate users. This system was introduced into to some clinics in the HSE East Coast area in 2005 and subsequently rolled out to the Northern Area Health Board. To date, over 76 staff use the database which contains data on over 3,000 active clients at thirteen sites and mobile needle exchange services. The ethical approval to use this data was obtained from Trinity College Dublin and a copy of this approval can be found in Appendix A.

Urinalysis data spanning the period January 2006 to December 2010 is available as part of this system. Overall there were 330,802 different urinalyses tests performed during the period on 4,229 unique individuals. Unique individuals were identified by a unique client ID. These tests were performed weekly. In all cases, client samples were tested for the presence of opiates, cocaine, EDDP (a methadone metabolite detectable in urine as confirmation of methadone consumption/compliance) and creatinine. High creatinine levels indicate a pure test while low amounts of creatinine in the urine indicate a manipulated test, either through the addition of water in the sample or by drinking excessive amounts of water. On some occasions, four additional tests were performed for alcohol, benzodiazepines, cannabis and amphetamines. Approximately 12% of the samples tested positive for cocaine (n = 37,886).

Over the five year period, the numbers of samples testing positive and the number of individuals involved increased up to 2008 when it peaked and then started to decline. This is illustrated in table 4.3.

At present, urinalysis remains the most reliable tool for identification of the presence of most substances (Wolff & Farrell 1999). The presence of cocaine can be detected in a urine sample for 6 to 8 hours but cocaine is rapidly converted to benzoylecgonine

Table 4.3: Urinalysis samples which are cocaine positive in the Drugs and Aids Information System 2006 - 2010

Year	Number of cocaine positive samples	Number of cocaine positive individuals
2006	8,008	1,084
2007	7,056	1,037
2008	8,247	1,096
2009	7,349	1,023
2010	7,175	1,043

in alkaline urine at room temperature. This cocaine metabolite can be detected for two to three days. Clients are tested weekly but it is likely that in some cases clients have used cocaine outside of the detection time frame. Also episodes of care can be interrupted and clients' usage during periods of absence from care would then naturally go undetected.

It is quite likely, therefore that there is a hidden population of clients in treatment who are in fact also using cocaine who have gone undetected. It was possible to count the number of times clients tested positive and so create a frequency distribution of counts. The Truncated Poisson method, to be outlined in the next section, can be used when this distribution of counts is available to estimate the number of clients who were undetected, in other words those seen zero times in the Poisson distribution. In 2006, 1,084 clients tested positive for cocaine. Of this number 317 (29% of the sample) were found to be cocaine positive once, 168 clients tested cocaine positive in two tests, 91 in three tests and 60 in four tests. Two clients tested positive in fifty tests. To estimate the population of cocaine users, N, for 2006, 2007, 2008 2009 and 2010 using the Zelterman (1988) and Chao (1987) estimators, equations (4.3) and (4.6) respectively, the numbers testing once, twice and three times are required. The frequencies f_1, f_2 and f_3 up to f_{10} along with the highest found frequency for each year are shown in table 4.4. The mean of the frequency distributions in the years 2006, 2007, 2008, 2009 and 2010 are 7.4, 8.8, 9.6, 9.2 and

8.9 respectively and these are also shown in table 4.4.

Table 4.4: Frequencies of clients in DAIS database who had 1, 2, 3 and so forth cocaine positive urinalyses from 2005 to 2010

Frequency	2006	2007	2008	2009	2010
f_1	317	316	315	300	353
f_2	168	151	174	150	115
f_3	91	100	91	100	105
f_4	60	77	70	69	64
f_5	48	52	46	49	36
f_6	53	42	37	37	43
f_7	29	31	35	28	35
f_8	39	25	26	21	35
f_9	20	20	20	32	24
f_{10}	20	24	19	25	20
..	..	..	..	..	..
f_{47}	2	0	1	1	1
..	..	..	..	..	..
f_{50}	2	0	0	0	0
Average	7.4	8.8	9.6	9.2	8.9

The scatter plot in figure 4.1 shows the number of clients who tested positive between one and ten times in the years 2006 to 2010. This graph of the frequencies suggests that the data is approximately Poisson.

The method of capture-recapture, which is critiqued in detail in the next section, is also applied to this urinalysis data. In order to apply the capture recapture method it is necessary to have good reliable data, preferably from three or more sources. In the present study we were unfortunately not successful in obtaining data from three independent data sources which we could match using a unique identifier. Instead, the approach of splitting the urinalysis data from each year into three four month partitions was used in order to apply the 3-sample capture-recapture method per year. Central to the method is the identification of the numbers present in two or more data sets. This is possible in DAIS due to the presence of a unique identifier.

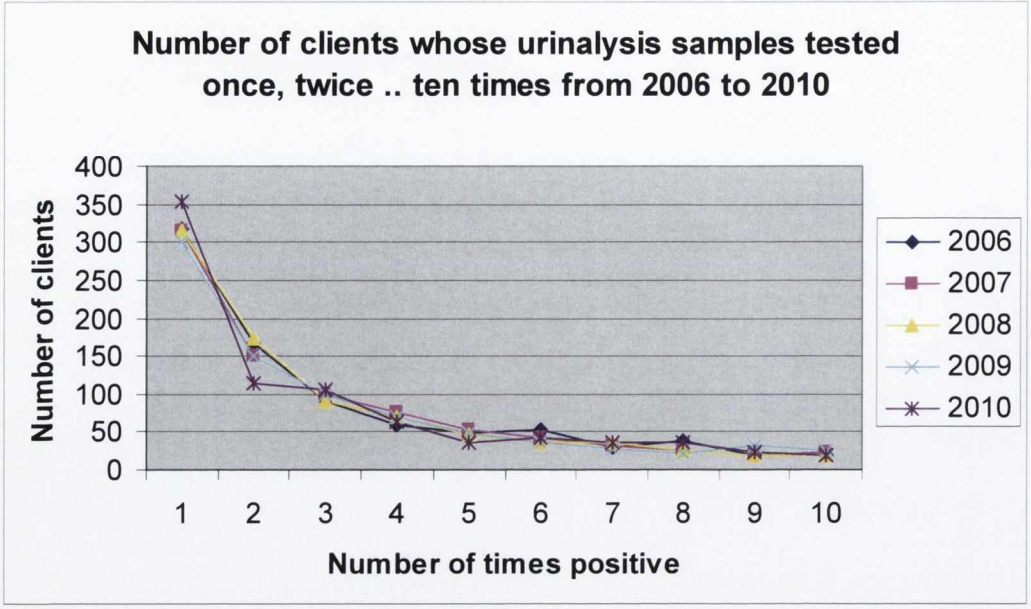

Figure 4.1: The number of DAIS clients who tested cocaine positive between one and ten times in the years 2006 to 2010

4.3 Methods of Prevalence Estimation

4.3.1 Multiplier-Benchmark Method

In multiplier-benchmark studies, the research makes use of existing routinely gathered data (the benchmark) along with a multiplier to produce an estimate of the target population. Examples of routinely gathered data in Ireland are data from the Police Using Leading Systems Effectively (Pulse) system, drug treatment data reported quarterly to the Health Research Board (HRB) and drug related deaths, also reported to the HRB. The idea behind the multiplier is that on average a proportion of the target population will appear in any of these datasets and a representative sample can be surveyed to assess what this proportion actually is. The success of this method is based on the quality and level of representation of the benchmark data. For example in the context of using this method to estimate the prevalence of cocaine use, account has to be taken of the heterogeneity of the population. Are crack cocaine users or problem drug users for example more likely to appear on the Pulse

system than powder cocaine users? Estimating the associated multiplier requires, usually, a small, separate sub-study. Hickman *et al.* (2002) have recommended the use of a number of possible sources for benchmark data and these sources are listed in table 4.5.

Table 4.5: Possible sources for benchmark data recommended by Hickman *et al.* (2002)

Data Source	Example
Specialist drug treatment	Drug users on methadone, attending treatment agencies, or in residential care
Low threshold drug agencies	Drug users attending drop-in sites or contacted by outreach workers
Needle exchange programmes	Drug users registered at needle exchange programmes
Casualty ward	Drug users attending casualty ward because of an overdose
Laboratory	Drug users tested for HIV, HCV or hepatitis B virus
Police/prisons	Drug users arrested or imprisoned for drug offences or for other crimes
Probation	Drug users on probation
Social services-assessments	Drug users assessed by local social services
Hostels for drug users	Drug users living in hostels
Addict registers	Drug users reported to a central register
Surveys of problem drug users	Community surveys of drug users
Overdose deaths	Number of deaths due to opiate overdose

An early paper by Hartnoll *et al.* (1985) illustrates the application of this technique, using deaths amongst drug users. The Global Assessment Program on Drug Abuse toolkit II (UNODC 2003) describes the method used by Hartnoll using a hypothetical benchmark figure of 3,000 and a hypothetical multiplier of one in 50 as follows. To apply the multiplier procedure to estimate the number of drug users in a given year, Hartnoll uses two things:

- The number of deaths to drug users in that year, say 3,000. This acts as the

fixed benchmark in the calculation.

- The death rate amongst drug users in that year, say 2 per cent, or 1 in 50 dying in the year. This provides the multiplier in the calculation.

The calculation of the number of drug users is straightforward once the multiplier is known. If 1 in 50 die, then the overall population must have been 3,000 x 50 = 150,000. The required multiplier may be obtained by a separate study among, for example, drug users known to treatment services. Often the initial study group who agree to participate in the study are encouraged to recruit their friends and acquaintances to the study in a procedure known as snowball sampling. Using this method allows the researcher to expand the number of recruits which may help to improve the accuracy of the study. However, as sample members are not selected from a sampling frame, snowball samples are subject to numerous biases. For example, people who have many friends are more likely to be recruited into the sample.

Sometimes it is possible to use figures from already published data, if they are appropriate. Archibald *et al.* (2001) used the number of HIV tests made on injecting drug users in Toronto in 1996 recorded as 4,050 as their benchmark figure for the number of injecting drug users. The estimate of the multiplier, which indicated what proportion of injecting drug users who had tests in the same period, was taken from a survey carried out in a different city. The assumption was that the proportion in the study would be comparable to the proportion in Toronto at that time. In the study, the proportion of users tested for HIV was known from the other study to be 25 per cent, or 1 in 4. The total number of injectors was then calculated to be 4 x 4,050, or 16,200 people.

The UNODC has recommended the multiplier method for developing countries in particular who may not have a sufficient availability of data or the resources to carry out expensive research (UNODC 2003). In an unpublished Serbian report (Comiskey *et al.* 2011) the multiplier method was used extensively to estimate the size of the

hidden injecting drug user population.

Assumptions of the Model

The assumptions on which the multiplier method is based should be carefully considered. First, it must be assumed that the benchmark data are accurate. Unfortunately, routine data sources can be notoriously inaccurate, because of underreporting or incomplete data collection. For example, in the Toronto case study, the authors raised the possibility that the laboratories may undercount the number of HIV tests carried out and that clinicians ordering tests do not always specify that someone was an injector. Therefore, the benchmark total may need to be adjusted to take account of that type of underreporting.

The method assumes that a correctly defined multiplier is available. When calculating the multiplier, the key question is really the following: Is the person recorded in the benchmark figure? If so, then the multiplier matches the benchmark successfully. The method also assumes that an unbiased estimate of the multiplier is available. Ideally, that estimate will be obtained from a representative sample of problem drug users and collected over the specific time period and for the specific place corresponding to the benchmark to be used. That rarely happens. Truly random representative samples of problem drug users do not exist and the best available option is to recruit subjects in a way that limits any potential bias.

One of the key requirements is, of course, that the multiplier is a fair representation of the connection between the benchmark count and the overall target population. There may be marked geographical heterogeneity in the true value of the multiplier, for example, treatment rates are very different in urban and rural areas. In this case, putting those areas together in a single multiplier can be misleading. The use of stratification of the population can be used if data are available separately for each stratum to overcome that heterogeneity problem.

Violation of one or all of those assumptions is clearly possible, providing ample opportunity for the study to give an inaccurate prevalence estimate. However, it is

clearly preferable assuming that sufficient data sources are available to use of several multipliers allowing cross validation of the resulting estimates.

4.3.2 Truncated Poisson

The truncated Poisson method is used when a single data source of counts is available for a specific time frame. Counting the repeated entries in the list leads to a frequency distribution of counts and, if an appropriate model for this count distribution can be found, an estimate of the (unobserved) frequency of zero entries in the list can be constructed. Count models differ in the way the count distribution is specified. The simplest model is the Poisson one. The Poisson distribution is used to mathematically model random behaviour in space or time. The truncated Poisson method of calculating prevalence is based on the premise that a data source consisting of frequency counts of a random variable follows this distribution. For example, in Ireland, the police use a system known as the Pulse system (Police Using Leading Systems Effectively) to record crime data. As each individual has a unique identifier, a data set of counts can be compiled consisting of the number of entries in the system relating to cocaine use an individual had with the police in a given time period. A frequency list can then be created consisting of f_1, f_2, f_3 and so forth, the counts of the numbers of people who have appeared once, twice three or more times on the list. Clearly, the count of persons who have committed a crime but have not been caught by the police represents a count of the hidden population f_0 with zero frequency. If we define p_x to be the probability that an arbitrarily chosen user has x contacts with the police in the time frame under study, then p_0 is the probability that an individual has zero contacts with the police. If p_0 is unknown, a conventional approach assumes that the frequencies arise from a Poisson distribution where $p_0 = e^{-\lambda}$, $p_1 = e^{-\lambda}\frac{\lambda}{1!}$, $p_2 = e^{-\lambda}\frac{\lambda^2}{2!}$ .....etc. The average number of contacts λ is the unknown parameter in the distribution and can be estimated by an iterative maximum likelihood procedure. This procedure, described in detail by

Böehning *et al.* (2004) can be summarised as follows. An initial estimate for the population of cocaine users, $\hat{N}$, is assigned (say $\hat{N} = n$), where n is the total of the observed frequencies, and the first estimate for λ is

$$\hat{\lambda} = \frac{c}{\hat{N}} \tag{4.1}$$

where c is the total number of contacts observed for the n users. A new estimate for N can now be calculated using the Horvitz-Thompson estimator (Horvitz & Thompson 1952)

$$\hat{N} = \frac{n}{1 - p_0} \tag{4.2}$$

where $p_0 = e^{-\lambda}$. From here, a new estimate of λ is built according to 4.1, then again a new estimate of $\hat{N}$ is constructed according to 4.2 with the current value of $\hat{N}$. This algorithmic procedure will cycle back and forth between steps 4.1 and 4.2 until convergence. The resulting estimate is the maximum likelihood estimate of $\hat{N}$. Various authors have proposed estimators of the population N which can be used instead of the maximum likelihood procedure. Wilson & Collins (1992) reviewed 14 different truncated Poisson methods estimators and recommended the use of Chao's (1987) bias adjusted estimator for relatively small data and the use of Darroch & Ratcliffs (1980) estimator otherwise. To estimate the prevalence of opiate users in Western Australia, Choi & Comiskey (2003) used the Zelterman (1988) estimator and the aforementioned Chao (1987) estimator. Del Rio Vilas & Böhning (2008) used the same estimators to calculate the total population of sheep holdings in Great Britain which were infected with scrapie, a fatal, degenerative disease that affects the nervous systems of sheep and goats. Zelterman (1988) in his estimate replaces maximum likelihood estimation by a moment estimator. His observation was that $\lambda = \frac{ip}{p_{i-1}}$ so that a natural estimator of the form $\lambda = \frac{if_i}{f_{i-1}}$ arises naturally. Typically, i will be two or three. The Zelterman estimator of the population size

(denoted with subscript Z) is

$$\hat{N_Z} = \frac{n}{1 - e^{(\frac{-2f_2}{f_1})}} \tag{4.3}$$

Zelterman (1988) claimed that his estimator produces reasonable estimates of the population size even if the homogenous Poisson model is violated. Del Rio Vilas & Böhning (2008) provided an associated estimator for the variance of the Zelterman estimator as seen in equation (4.3).

$$Var(\hat{N_Z}) = nG(\hat{\lambda})\left[1 + nG(\hat{\lambda})\hat{\lambda}^2\left(\frac{1}{f_1} + \frac{1}{f_2}\right)\right] \tag{4.4}$$

where

$$G(\hat{\lambda}) = \frac{e^{-\hat{\lambda}}}{(1 - e^{-\hat{\lambda}})^2} \tag{4.5}$$

However, according to Del Rio Vilas & Böhning (2008) one criticism of the Zelterman estimator is the large associated variance. He states that Chao's estimate tends to be lower but it has the advantage of having a smaller variance. Denoted with a subscript C it is given by

$$\hat{N_C} = n + \frac{{f_1}^2}{2f_2} \tag{4.6}$$

$$Var(\hat{N_C}) = \frac{1}{2}\frac{{f_1}^2}{f_2}\left(1 - \frac{{f_1}^2}{2f_2 n + {f_1}^2}\right) + \frac{{f_1}^3}{{f_2}^2}\left(1 + \frac{1}{4}\frac{f_1}{f_2}\left(1 - \frac{f_2}{n}\right)\right) \tag{4.7}$$

It should be noted that both estimators (4.3) and (4.6) are primarily based on the lower frequencies (f_1 and f_2). This emphasis on the lower frequency classes makes sense. People seen rarely (only once or twice) are likely to bear a greater resemblance to people never seen, than people seen very often. In addition, the emphasis on the lower frequency classes makes the estimators robust in the presence of heterogeneity, e.g. persons seen very often may form a different subgroup as compared to persons rarely seen. The influence of the persons often seen is weighted down in this estimator and therefore heterogeneity, if present, is likely to exercise a relatively small influence. Finally, emphasis on the lower frequency classes results

in a further advantage in that the estimator is known to perform rather well even with sparse data.

Assumptions of the Model

The assumptions of the truncated Poisson model are as follows:

- The frequencies follow a Poisson distribution.
- The population is closed, in other words the true population size, N, is unaffected by migration, birth and death during the period under review. This is addressed by limiting the study period to one year.
- The individual probabilities of being observed and re-observed are constant during the study period
- The population itself is homogeneous. For clients in treatment, detected heroine use can have consequences in that they will not be given "take-away" methadone and this acts as a form of deterrent. However, as cocaine use is not the focus of treatment, its presence in a sample though noted carries no adverse consequences. It can be assumed therefore that detection depends entirely on the random event of cocaine use.

4.3.3 Capture-Recapture

According to Seber (1982) the capture-recapture was first used in 1896 by Petersen to estimate the size of the plaice population and was subsequently independently developed by Lincoln in 1930 to estimate populations of waterfowl. The technique involved the capture of a sample of the species, and the marking of the caught samples in some way so they would be identifiable if captured again. The fish/animals were then released and a second sample was taken. The recaptured animals/fish were then identified and counted. The calculation of N, the total population estimate, was based on the assumption that the number captured in the first capture would

be proportional to the proportion recaptured in the second sample. The methodology has recently been extended to the count of hidden human populations and in particular to the estimation of prevalence. The use of the CRM (capture-recapture method) is recommended by both the EMCDDA (1999a) and the UNODC (2003). A proportion of drug users inadvertently come into contact with official agencies each year and records are kept providing a count equivalent to a capture for the agency in question. Examples of such agencies are the police, the courts, the prisons, drug treatment agencies, social welfare agencies and accident and emergency units, to name but a few. The count of users as reported by a specific agency can be seen as a capture. Providing the individuals can be uniquely identified, a count of the number of individuals caught once, twice, three or more times by the agencies will be available. If these counts are from a single agency then this is a single sample capture-recapture method.

Methodological Background

The following description of the theoretical background to the capture-recapture method is a synopsis of the method as described by Bishop *et al.* (1975) with theoretical backup from Agresti (1996). The theory is illustrated with either two or three samples as the method can be expanded to include more than three samples if necessary. Let us say we have three counts of users from three different sources and the individuals in each source are identifiable in such a way that we can count the number of people who appear in two or three of the sources as well as in only one. The data can be represented in a contingency table which shows a count of those present or absent in a particular sample as follows.

In Table 4.6, the variable a represents the number present in all three samples. The variable b is the number present in both samples 1 and 3 but absent in sample 2. The total number present in sample 1 is a + b + c + d, the total present in sample 2 is a + e + c + g and the total present in sample 3 is a + b + e + f. Another useful notation is to denote the elements of the table by x_{ijk} as shown in table 4.7.

Table 4.6: Contingency table indicating presence or absence in one of three samples

		Sample 3			
		present		absent	
	Sample 2	present	absent	present	absent
Sample 1	present	a	b	c	d
	absent	e	f	g	h

Table 4.7: Contingency table indicating presence or absence in one of three samples

		Sample 3			
		present		absent	
	Sample 2	present	absent	present	absent
Sample 1	present	x_{111}	x_{121}	x_{112}	x_{122}
	absent	x_{211}	x_{221}	x_{212}	x_{222}

Provided the layout of the table 4.7 is adhered to, a suffix of 1 denotes presence in a specific sample and 2 denotes absence in that sample. Hence for example, x_{221} has a 2 in the first position so it is absent in sample 1. Similarly it is absent in sample 2 but present in sample 3. The variable x_{222}, the count of those who were not present in any of the three samples, is the unknown count that the methodology seeks to estimate. In the following, the notation used by Bishop *et al.* (1975) is adhered to by letting

N = the total population and

n = the total observed population (a + b + c + d + e + f + g).

Independence

A key assumption of the model is independence. In contingency tables there are different ways in which the assumption of independence is implemented. One of these involves probabilities and is the method used by Pearson when he formulated the Chi Squared test. This method is based on probability and is easily demonstrated

by means of a 2 x 2 table.

Table 4.8: Two by two contingency table with marginal totals

Sample 2		present	absent	Total
Sample 1	present	$a(x_{11})$	b(x_{12})	a+b
	absent	$c(x_{21})$	$d(x_{22})$	c+d

In probability theory: If A and B are two independent events then

$$P(A \text{ and } B) = P(A).P(B)$$

So if sample 1 and sample 2 are independent then we would expect the count labeled a in cell x_{11} to be

$N \times$ (*probability of being present in sample* 1) $\times$ (*probability of being present in sample* 2)

In Table 4.8 this becomes

$$E(x_{11}) = N\left(\frac{a+b}{N}\right)\left(\frac{a+c}{N}\right) \tag{4.8}$$

The notation $E(x_{11})$ here means the expected value of the entry in the cell in row one and column one. One often sees equation (4.8) also written as

$$\frac{(row\ total)(column\ total)}{N}$$

This is the approach that Pearson adopted when he developed the χ^2 test.
An alternative approach to independence is as follows. If sample 1 and sample 2 are independent then the proportion present in sample 1 should be independent of presence or absence in sample 2. Using the entries from Table 4.8 above this means

that

$$\begin{aligned} \frac{a}{a+c} &= \frac{b}{b+d} \\ a(b+d) &= b(a+c) \\ \Rightarrow ab+ad &= ba+bc \\ \Rightarrow ad &= bc \end{aligned} \tag{4.9}$$

Testing Independence with the Odds ratio

The result in equation (4.9) can be approached from another angle, namely the odds ratio. In many applications which are represented in a 2 x 2 table, one random variable is regarded as the response variable while the other random variable is regarded as the explanatory variable. We will name them X (represented by sample 1 in table 4.8) and Y (represented by sample 2 in table 4.8) respectively. There are two possible outcomes for Y of which one may be regarded as success and the other as failure. Each level of X can be seen as a distribution of Y, the so called conditional distribution. If we let π_i be the probability of success at the i^{th} level of X, the probabilities $(\pi_i, 1-\pi_i)$ form the conditional distribution of Y in row i. Then odds_1 is defined to be

$$\frac{\pi_1}{1-\pi_1}$$

and odds_2 is defined to be

$$\frac{\pi_2}{1-\pi_2}$$

where odds_1 for example is the probability of success divided by the probability of failure. The odds ratio θ (also called the cross-product ratio) is then the ratio of these two odds given by

$$\theta = \frac{odds_1}{odds_2}$$

The odds ratio can be expressed using the cell counts as follows

$$\begin{aligned}\theta &= \left(\frac{x_{11}}{x_{11}+x_{12}} \div \frac{x_{12}}{x_{11}+x_{12}}\right) \Big/ \left(\frac{x_{21}}{x_{21}+x_{22}} \div \frac{x_{22}}{x_{21}+x_{22}}\right) \\ &= \frac{x_{11} \div x_{12}}{x_{21} \div x_{22}} \\ &= \frac{x_{11} \times x_{22}}{x_{12} \times x_{21}} \\ &= \frac{ad}{bc} \end{aligned} \tag{4.10}$$

In the special case where X and Y are independent this ratio is equal to 1. In other words

$$\begin{aligned} &\frac{ad}{bc} = 1 \\ \Rightarrow \quad &ad = bc \end{aligned} \tag{4.11}$$

and we again have equation (4.9) above.

In the capture-recapture method we are concerned with estimating the missing cell x_{22} which is the count of those absent in each sample. From the assumption of independence

$$\begin{aligned} \frac{x_{11} \times x_{22}}{x_{12} \times x_{21}} &= 1 \\ \textit{and hence} & \\ x_{22} &= \frac{x_{12} \times x_{21}}{x_{11}} \end{aligned} \tag{4.12}$$

This is the standard formula used to estimate the missing cell in the capture-recapture methodology with two samples, hence the assumption of independence is crucial in the use the method when there are only two samples available.

Next we consider the odds ratio in the 2 x 2 x 2 case. In a capture-recapture

setting, this is the case where we have three datasets and the contingency table shows the number of cases present or absent in any of the three samples. Let us call the third variable Z. We can calculate the odds ratio in each 2 x 2 table at each level of Z and the odds ratios give us a measure of the association between X and Y at each level of Z. If the odds ratio is the same at each level we would say that the association between X and Y is independent of Z. Z may have more than 2 levels but if the X-Y association, as measured by the odds ratios, is the same at each level of Z this is called homogeneous association. We will demonstrate, using table 4.6 above, that for the 2 x 2 x 2 case, if there is homogeneous association between X and Y at each level of Z, then there will also be homogeneous association between Y and Z at each level of X and also between X and Z at each level of Y. We will first examine the so called partial association in the partial tables created by considering each level of Z as a separate 2 x 2 table as illustrated in table 4.9.

Table 4.9: Partial Tables created by splitting across levels of Z

Present in Sample 3 (Z)				**Absent in Sample 3 (Z)**			
		Sample 2 (Y)				*Sample 2 (Y)*	
		pres	abs			pres	abs
Sample 1	pres	a	b	*Sample 1*	pres	c	d
(X)	abs	e	f	*(X)*	abs	g	h

If the association between X and Y is independent of the level of Z the odds ratio of X-Y at each level of Z will be the same and therefore

$$\frac{af}{be} = \frac{ch}{dg}$$

This implies that

$$adfg = bceh \tag{4.13}$$

Next we look at the association between X and Z at each level of Y shown in table 4.10.

Table 4.10: Partial Tables created by splitting across levels of Y

Present in Sample 2 (Y)				**Absent in Sample 2 (Y)**			
		Sample 3 (Z)				*Sample 3 (Z)*	
		pres	abs			pres	abs
Sample 1	pres	a	c	*Sample 1*	pres	b	d
(X)	abs	e	g	*(X)*	abs	f	h

If the odds ratio at each level of X-Z is the same at each level of Y then

$$\frac{ag}{ce} = \frac{bh}{df}$$

This again implies that

$$adfg = bceh$$

Finally we look at the association between Y and Z at each level of X shown in table 4.11.

Table 4.11: Partial Tables created by splitting across levels of X

Present in Sample 1 (X)				**Absent in Sample 1 (X)**			
		Sample 3 (Z)				*Sample 3 (Z)*	
		pres	abs			pres	abs
Sample 2	pres	a	c	*Sample 2*	pres	e	g
(Y)	abs	b	d	*(Y)*	abs	f	h

Here again if the odds ratio is the same for Y-Z at each level of X then

$$\frac{ad}{bc} = \frac{eh}{fg}$$

Again we see that

$$adfg = bceh$$

This is again the same as equation (4.13) and also must be true under the assumption of homogeneous association.

We have shown that if the X-Y conditional odds ratios are the same at each level

of Z then the same property holds for the X-Z and the Y-Z conditional odds ratios (Agresti 1996, p 59). When this is the case we say there is *no three-way interaction between the variables.* We note that for each pair-wise odds ratio shown above that the numerator in the case of equal odds ratio is the product of the boldface cells in table 4.12. This v-shaped diagonal arrangement of the numerator is true no matter which sample we control for.

Table 4.12: Partial Tables created by splitting across the levels of one variable

$\mathbf{x_{111}}$	x_{121}	x_{112}	$\mathbf{x_{122}}$
x_{211}	$\mathbf{x_{221}}$	$\mathbf{x_{212}}$	x_{222}

Definition of three-way independence

The preceding discussion allows us now to give a definition of what we mean by three-way independence. Under this assumption of independence we are then able to find an estimate for the missing cell.

Definition: *In a three-way table we say there is no three-factor interaction if the odds ratio between any two of the three variables remains the same across all the levels of the third variable.*

In the 2 x 2 x 2 model this implies that

$$\frac{x_{111} \times x_{221}}{x_{121} \times x_{211}} = \frac{x_{112} \times x_{222}}{x_{122} \times x_{212}} \tag{4.14}$$

In a capture-recapture model the cell count x_{222} which denotes the count of those absent from all three samples is unknown and the aim of the method is to estimate this count. An estimate of the unknown cell follows immediately from equation (4.14) and is

$$x_{222} = \frac{x_{111} \times x_{221} \times x_{212} \times x_{122}}{x_{211} \times x_{121} \times x_{112}} \tag{4.15}$$

The Loglinear Model

We now look at the loglinear model approach to estimating the value of the missing cell and use the independence criteria given in equation (4.8). We first focus on the two-way contingency table.

Let π_{i+} = the probability of an outcome being in row i of sample 1, the sample X, where i = 1 .. 2

and

let π_{+J} = the probability of being in column j of sample 2, the sample Y, where j = 1 .. 2. Then we can formally express the expected value in cell ij as

$$E(x_{ij}) = N(\pi_{i+})(\pi_{+j}) \tag{4.16}$$

If we take the log of both sides we get

$$\ln E(x_{ij}) = \ln N + \ln \pi_{i+} + \ln \pi_{+j}. \tag{4.17}$$

If we follow the notation used by Bishop *et al.* (1975) this gives us a linear model for the log of the expected value in a specific cell under the assumption of independence of the two samples as follows:

$$l_{ij} = \mu + \mu_{1(i)} + \mu_{2(j)} \tag{4.18}$$

where l_{ij} is the natural log of the entry in the ith row and jth column, $\mu_{1(i)}$ is the cell specific effect of an entry being in row i of the first sample and the i and j in the brackets are a reminder of which level is describing the so called effect. Hence the log of the expected cell value consists of a general population effect and separate row and column effects.

This loglinear model is called the independence model as there is no term indicating an interaction between the two variables. If the model requires that there is interac-

tion between the two terms, then a parameter can be introduced into the equation to model this effect. Equation (4.18) is then adapted to include an interaction term $\mu_{12(ij)}$

$$l_{ij} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{12(ij)} \tag{4.19}$$

Here the term $\mu_{12(ij)}$ is a parameter which adds an interaction effect between X and Y to the estimated cell value. Clearly this effect could be negative as well as positive. As effects at each level of each sample are modelled with a separate parameter, redundancy is introduced into the models. For example, in the independence model (4.18), one $\mu_{1(i)}$ and one $\mu_{2(j)}$ are redundant. Agresti (1996) states that there is no unique way to calculate the parameters of a given model. To eliminate the redundancy one can set the first or last level of each factor equal to zero or one can let the sum of the parameters equal to zero. These models have been implemented in various statistical software packages. According to Agresti (1996) software such as Generalized Linear Interactive Modelling (GLIM) (Aitkin 1989) and Statistical Package for the Social Sciences (SPSS) (Nie *et al.* 1970) take the former approach while Statistical Analysis System (SAS 1966) takes the latter. The model for the three-way contingency table is based on the same principle but clearly there are more interactions possible. There are three possible two-way interactions, XY, YZ and XZ but also a three-way interaction XYZ. A model is said to be fully saturated if the number of parameters is equal to the number of cells being modelled. A fully saturated model for a three-way table is as follows:

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \mu_{23(jk)} + \mu_{13(ik)} + \mu_{123(ijk)} \tag{4.20}$$

However other models are possible involving less interactions. These models are:

1. The three samples are independent and the model is given by:

   $l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)}$

2. One sample is independent of the other 2 and the model is given by:

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} \qquad \textit{(sample 3 is independent of 1 and 2)}$$

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{13(ik)} \qquad \textit{(sample 2 is independent of 1 and 3)}$$

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{23(jk)} \qquad \textit{(sample 1 is independent of 1 and 2)}$$

3. Two pairs of samples are related and the model is given by:

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \mu_{23(jk)}$$

(samples 1 and 3 are not related)

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \mu_{13(ik)} \qquad \textit{(samples 2 and 3 are not related)}$$

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{13(ik)} + \mu_{23(jk)} \qquad \textit{(samples 1 and 2 are not related)}$$

4. All pairwise interactions are present:

$$l_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{13(ik)} + \mu_{23(jk)} + \mu_{13(ik)}$$

In the next section, it is shown how Bishop *et al.* (1975) calculate the value of the missing cell under the assumption of three-way independence for these four models.

Estimation of cells in a three-way table under the various model assumptions

It is interesting to note that, for the four models described above, the expected cell values for each model are estimated directly using probabilistic methods. The models are used merely to demonstrate the pattern of interactions. On page 58, Bishop *et al.* (1975) state that "*once the sufficient statistics are available, then the*

expected cell counts can be calculated without the intermediate step of computing the model parameters". They show that, in the case of the response variable having either a Poisson or Multinomial sampling distribution, the sufficient statistics are the same for each distribution, though they vary per model. In the following description the notation presented in table 4.7 above is used to refer to the cell entries and the Bishop *et al.* (1975) notation is used to refer to the expected values $\hat{m}_{ijk}$ of the cells as shown in table 4.13.

Table 4.13: Three sample contingency table of expected values

		Sample 3			
		present		absent	
Sample 2		present	absent	present	absent
Sample 1	present	$\hat{m}_{111}$	$\hat{m}_{121}$	$\hat{m}_{112}$	$\hat{m}_{122}$
	absent	$\hat{m}_{211}$	$\hat{m}_{221}$	$\hat{m}_{212}$	$\hat{m}_{222}$

They also use a subscript + to denote summing the cells over the level replaced by a +. For example x_{++1} refers to the sum of the counts in all cells present in sample 3, in other words

$$x_{++1} = x_{111} + x_{121} + x_{211} + x_{221}$$

x_{1+1} refers to the sum of those present in sample 1 and sample 3 in other words

$$x_{1+1} = x_{111} + x_{121}$$

x^*_{++1} refers to the sum of counts in the cells in sample 3 but excluding cell x_{221} in other words

$$x^*_{++1} = x_{111} + x_{121} + x_{221}$$

1. The three samples are independent:

   Bishop *et al.* (1975) tell us that the maximum likelihood equations for the independent samples model cannot be solved directly and they suggest using the Deming-Stephen iterative proportional fitting procedure to estimate the

cells. The algorithm, given by Deming and Stephan in 1940, matches marginal totals by using iterative scaling of a table whose entries are initially all equal to unity.

2. One sample is independent of the other 2. Let us assume that the third sample is independent of sample 1 and sample 2. This option is the most complex as logic and probability is used to build up an estimate for each cell in the table. We let the total population N = the sum of the 8 cells in the table and $n = Nx_{222}$ Each cell is estimated using probability methods and the assumption of three independent samples.

   We start with the estimated value for the cell x_{221}. This is an estimate of those only present in sample 3 but absent in the other 2. We are assuming sample 3 is independent of the other two samples. Our best estimate would have been $\frac{x_{221}}{(x_{221} + x_{222})}$ but as x_{222} is not known we only have one possible estimate for $\hat{m}_{221}$, and that is the observed value of x_{221}.

   As we fill in the remaining cells we need to take account of the fact that this cell is fixed and that n must therefore be reduced by the value of this cell. We let $n^* = n - x_{221}$ as this will now be used to calculate the probability of presence or absence in the remaining cells. The * is to indicate exclusion of x_{221}.

   We will show in detail how $\hat{m}_{111}$ is calculated and then report the results for the remaining cells. The probability of a case being in sample 1 and sample 2 is

$$\frac{x_{111} + x_{112}}{n^*}$$

   which can be written more simply as

$$\frac{x_{11+}}{n^*}$$

   where the + indicates summation over both values of sample 3. The proba-

bility of a case being in sample 3 is

$$\frac{x_{111} + x_{121} + x_{211}}{n^*}$$

which we can write more simply as

$$\frac{x_{++1}}{n^*}$$

We are assuming that presence in sample 1 AND sample 2 is independent of presence in sample 3 therefore the probability of a case being in all three samples is the product of these two probabilities. The expected value of x_{111} then becomes

$$\begin{aligned}
\hat{m}_{111} &= (n^*)(\text{probability in sample 1 AND 2})(\text{probability in sample 3}) \\
&= \frac{(n^*)(x_{11+})(x^*_{++1})}{(n^*p)(n^*)} \\
&= \frac{(x_{11+})(x^*_{++1})}{(n^*)}
\end{aligned}$$

The other cells are calculated in a similar fashion.

$$\begin{aligned}
\hat{m}_{121} &= (n^*)(\text{probability in sample 1 AND NOT in 2})(\text{probability in sample 3}) \\
&= \frac{(n^*)(x_{12+})(x^*_{++1})}{(n^*)(n^*} \\
&= \frac{(x_{12+})(x^*_{++1})}{(n^*)}
\end{aligned}$$

$$\begin{aligned}
\hat{m}_{211} &= (n^*)(\text{probability NOT in sample 1 AND in 2})(\text{probability in sample 3}) \\
&= \frac{(x_{21+})(x^*_{++1})}{(n^*)}
\end{aligned}$$

$$\begin{aligned}\hat{m}_{112} &= (n^*)(\text{probability in sample 1 AND 2})(\text{probability NOT in sample 3}) \\ &= \frac{(x_{21+})(x^*_{++2})}{(n^*)}\end{aligned}$$

$$\begin{aligned}\hat{m}_{122} &= (n^*)(\text{probability in sample 1 AND NOT in 2})(\text{probability NOT in sample 3}) \\ &= \frac{(x_{12+})(x^*_{++2})}{(n^*)}\end{aligned}$$

$$\begin{aligned}\hat{m}_{212} &= (n^*)(\text{probability NOT in sample 1 AND 2})(\text{probability NOT in sample 3}) \\ &= \frac{(x_{21+})(x^*_{++2})}{(n^*)}\end{aligned}$$

We can now estimate the missing cell $\hat{m}_{222}$ by filling the values we have just estimated into the equation

$$\hat{m}_{222} = \frac{\hat{m}_{111} \times \hat{m}_{221} \times \hat{m}_{212} \times \hat{m}_{122}}{\hat{m}_{211} \times x_{121} \times \hat{m}_{112}}$$

$$\begin{aligned}\hat{m}_{222} &= \frac{\frac{(x_{11+})(x^*_{++1})}{(n^*)} \times \frac{x_{221}}{1} \times \frac{(x_{12+})(x^*_{++2})}{(n^*)} \times \frac{(x_{21+})(x^*_{++2})}{(n^*p)}}{\frac{(x_{12+})(x^*_{++1})}{(n^*)} \times \frac{(x_{21+})(x^*_{++1})}{(n^*)} \times \frac{(x_{11+})(x^*_{++2})}{(n^*)}} \\ &= \frac{(x_{221})(x_{++2})}{(x^*_{++1})} \qquad (4.21)\end{aligned}$$

It follows that the estimate of the total population $\hat{N}$ is n + $\hat{m}_{222}$

3. Two pairs of samples are related. Let us assume the first and the second and the second and the third. If the second is not present there will be no relationship possible between the first and the third. Therefore the expected counts for the cells where the second sample is absent must equal the observed counts.

$$\hat{m}_{121} = x_{121} \quad \hat{m}_{122} = x_{122} \quad \hat{m}_{221} = x_{221}$$

$$
\begin{aligned}
\hat{m}_{111} &= \text{(total present in sample 2)} \\
&\quad \text{(probability in sample 1 AND 2)} \\
&\quad \text{(probability in sample 2 AND 3)} \\
&= x_{+1+} \times \frac{x_{11+}}{x_{+1+}} \times \frac{x_{+11}}{x_{+1+}} \\
&= \frac{(x_{11+})(x_{+11})}{(x_{+1+})} \qquad (4.22) \\
\hat{m}_{112} &= \text{(total present in sample 2)} \\
&\quad \text{(probability in sample 1 AND 2)} \\
&\quad \text{(probability in sample 2 AND NOT in 3)} \\
&= \frac{(x_{11+})(x_{+12})}{(x_{+1+})} \qquad (4.23) \\
\hat{m}_{211} &= \text{(total present in sample 2)} \\
&\quad \text{(probability NOT in sample 1 AND in 2)} \\
&\quad \text{(probability in sample 2 AND in 3)} \\
&= \frac{(x_{21+})(x_{+11})}{(x_{+1+})} \\
& \qquad (4.24)
\end{aligned}
$$

$$
\begin{aligned}
\hat{m}_{212} &= \text{(total present in sample 2)} \\
&\quad \text{(probability NOT in sample 1 AND in 2)} \\
&\quad \text{(probability in sample 2 AND NOT in 3)} \\
&= \frac{(x_{21+})(x_{+12})}{(x_{+1+})} \\
\hat{m}_{222} &= \frac{\dfrac{(x_{11+})(x_{+11})}{(x_{+1+})} \times \dfrac{x_{221}}{1} \times \dfrac{(x_{122})}{1} \times \dfrac{(x_{21+})(x_{+12})}{(x_{+1+})}}{\dfrac{(x_{12+})}{(1)} \times \dfrac{(x_{21+})(x_{+11})}{(x_{+1+})} \times \dfrac{(x_{11+})(x_{+12})}{(x_{+1+})}} \\
&= \frac{(x_{221})(x_{122})}{(x_{121})} \qquad (4.25)
\end{aligned}
$$

It follows that the estimate of the total population $\hat{N}$ is n + $\hat{m}_{222}$

The variance is as follows:

$$Variance = (\hat{m}_{222})^2 \left(\frac{1}{x_{221}} + \frac{1}{x_{122}} + \frac{1}{x_{121}} + \frac{x_{121}}{x_{221}x_{122}} \right)$$

4. All pairwise relationships are present.

   For this model, as the number of cells is equal to the number of therms in the model, there are zero degrees of freedom. Hence the Maximum Likelihood Estimates (MLEs) of the expected cell counts are the observed values (Bishop *et al.* 1975, p. 241). Thus

$$\hat{m}_{222} = \frac{x_{111}x_{221}x_{212}x_{122}}{x_{211}x_{121}x_{112}}$$

   The estimate of the asymptotic variance is

$$Variance = (\hat{m}_{222})^2 \left(\frac{1}{x_{111}} + \frac{1}{x_{121}} + \frac{1}{x_{211}} + \frac{1}{x_{221}} + \frac{1}{x_{112}} + \frac{1}{x_{122}} + \frac{1}{x_{212}} + \frac{1}{\hat{m}_{222}} \right)$$

   In this section Bishop *et al.*'s probabilistic approach to estimating the missing cell and corresponding variance has been presented under various assumptions. This approach will be used as well as the standard SPSS general linear modelling procedure to estimate the missing cell in the capture-recapture section in order to allow a comparison of the two methods.

Assessing goodness of fit

The goodness of fit of the models can be assessed by using the Pearson chi square statistic χ^2, defined as

$$\chi^2 = \sum_{all\ cells} \frac{(O-E)^2}{E}$$

Where O stands for the observed count and E for the maximum likelihood estimate in each cell. Alternatively, the goodness if fit of the models can be assessed by using

the G^2 statistic defined as

$$G^2 = -2\sum O \times log\left(\frac{O}{E}\right)$$

which is also distributed, under the null hypothesis, as a central χ^2 distribution with appropriate degrees of freedom. For model selection, Hook & Regal (1995) recommend the use of three criteria, the Akaike Information Criterion (AIC) (Akaike 1974) and two forms of the Bayesian Information Criterion (BIC), one suggested by Schwarz (1978) and the other by Draper (1995). The formulae for these criteria are as follows:

$$\begin{aligned} AIC &= G^2 - 2(df) \\ SIC &= G^2 - (\ln N_{obs})(df) \\ DIC &= G^2 - (\ln \frac{N_{obs}}{2\pi})(df) \end{aligned}$$

where G^2 is the deviance and df is number of degrees of freedom associated with the model, N is the known population of drug users and ln denotes the natural logarithmic function. Once the best model has been selected, various methods for producing a confidence interval can be used. A method favoured by Cormack (1992) and by Hook & Regal (1995) which recognises that the estimate for the hidden population is derived from an asymmetric distribution has commonly been used.

Assumptions of the Model

The assumptions of the capture-recapture model are as follows:

- the samples/lists must be representative of the population under study. Individuals not seen (those that we wish to estimate) have to have similar characteristics to those who are seen.
- the samples must come from a closed population (closed refers to the assumption that individuals do not enter or leave the population during the study

period). This assumption is only reasonable for short time periods (e.g. up to about one year). During longer periods there will be some individuals who join the group of interest (i.e. initiate regular drug use) and others who leave it (i.e. achieve abstinence or die). If prevalence is to be estimated over a long period of time, other methods considering open populations need to be applied.

- each list must be homogeneous, i.e. the probability of selection into a sample/list must be constant for all individuals.
- the lists must be mutually independent, i.e. the individual probability of selection into a list must not be influenced by the presence or absence of the person in another list.

4.4 Prevalence Estimates

In this section the prevalence estimation methods described in section 4.3 are applied to the data sources presented in section 4.2 in order to obtain prevalence estimates of cocaine use in the Northern Area Health Board. We first obtain an estimate of the prevalence of cocaine use in the general population by applying the NACD reported rates for the region (NACD 2009) to the population census figures for the region (HSE 2007). We then use both the truncated Poisson methods and the capture-recapture methods together with the DAIS urinalysis sample results to obtain an estimate of the true number of opiate users in contact with the treatment services who are also using cocaine. We expect this estimate to be a subset of the general population estimate as DAIS clients are drawn from those with an address in the area and exclude the homeless and those in prisons who are treated by different services. Finally we estimate a multiplier based on the general and treatment estimates and apply this to extrapolated population figures for the years 2007 to 2010.

4.4.1 Estimates from the NACD Household Survey

In this section we apply the rates presented in table 4.1 to the population figures presented in table 4.2 to obtain estimates for the total population of cocaine users in the Northern Area Health Board in 2006. These figures together with the breakdown per gender and per age group are presented in table 4.14.

Table 4.14: Last year prevalence estimates with 95% confidence intervals of cocaine use in 2006 (NACD results for the Northern Area Health Board)

	Total	Male	Female	15-34	35-64
	8,131	7,114	990	6,890	1,357
CI+	15,276	14,841	3,588	15,010	4,194
CI-	3,696	2,698	124	2,461	247

4.4.2 Truncated Poisson

We used the frequencies shown in table 4.4 to estimate the total population of problem drug users in treatment who are also using cocaine in the Northern Area Health Board. These frequencies are obtained from urinalyses carried out in the period 2006 to 2010 and represent the frequencies of those who tested cocaine positive once, twice and three times. Although drug users in the system are tested regularly while they are in continuous treatment, positive samples may go undetected as they are outside the detection time frame and will definitely be undetected when the clients are not in treatment due to a break in an episode of care. The aim of this analysis is to estimate the total number of problem drug users in treatment who are also using cocaine for each year from 2006 to 2010. The results, as presented in table 4.15, show the Zelterman point estimate with corresponding confidence intervals obtained using equations (4.3) and (4.4) and the Chao point estimate with corresponding confidence intervals obtained using equations (4.6) and (4.7).

Table 4.15: Zelterman and Chao point estimates with confidence intervals of the total number of problem drug users in treatment who are also using cocaine in the Northern Area Health Board

	Zelterman	Chao
Total Population 2006 (95% CI)	1,671 (1,482 - 1,860)	1,391 (1,306 - 1,476)
Total Population 2007 (95% CI)	1,717 (1,504 - 1,930)	1,388 (1,293 - 1,483)
Total Population 2008 (95% CI)	1,640 (1,461 - 1,819)	1,382 (1,301 - 1,463)
Total Population 2009 (95% CI)	1,625 (1,426 - 1,823)	1,327 (1,239 - 1,415)
Total Population 2010 (95% CI)	2,181 (1,834 - 2,527)	1,586 (1,431 - 1,740)

4.4.3 Three Sample Capture-Recapture Method

Capture-Recapture has been used in numerous studies to estimate local prevalence. Hay *et al.* (2010) listed 22 studies which used the method. The majority of these studies, however, use 3 or more data sources. In the present study we were not successful in obtaining three suitable data sources. However an alternative approach to capture-recapture with three datasets is to partition one data set into three samples to which the methodology is applied. This approach is described by Domingo-Salvany (1996) in an EMCDDA monograph. The method has been applied by Hser (1993) to obtain population estimates of illicit drug users in Los Angeles County and by Brugal & Domingo-Salvany (2004) to obtain prevalence estimates of problematic cocaine consumption in Barcelona, Spain. We have followed this approach and split our data for each year from 2006 to 2010 into three 4-month samples. In 2006, for example, Sample 1 contains the client-IDs for all cocaine positive clients from January to April, sample 2 for those from May to August and sample 3 contains the client-IDs for all cocaine positive clients from September to December. The numbers present or absent in the samples and their overlaps in 2006 can be seen in table 4.16. The italicised figure in brackets replicates the subscript notation used in table 4.13 in section 4.3.3, where 1 denotes presence in a sample and 2 denotes absence in a sample. It is also convenient to illustrate this data by means of a Venn diagram (see figure 4.2).

Using the methodology of Bishop *et al.* (1975) as outlined in section 4.3.3 we first

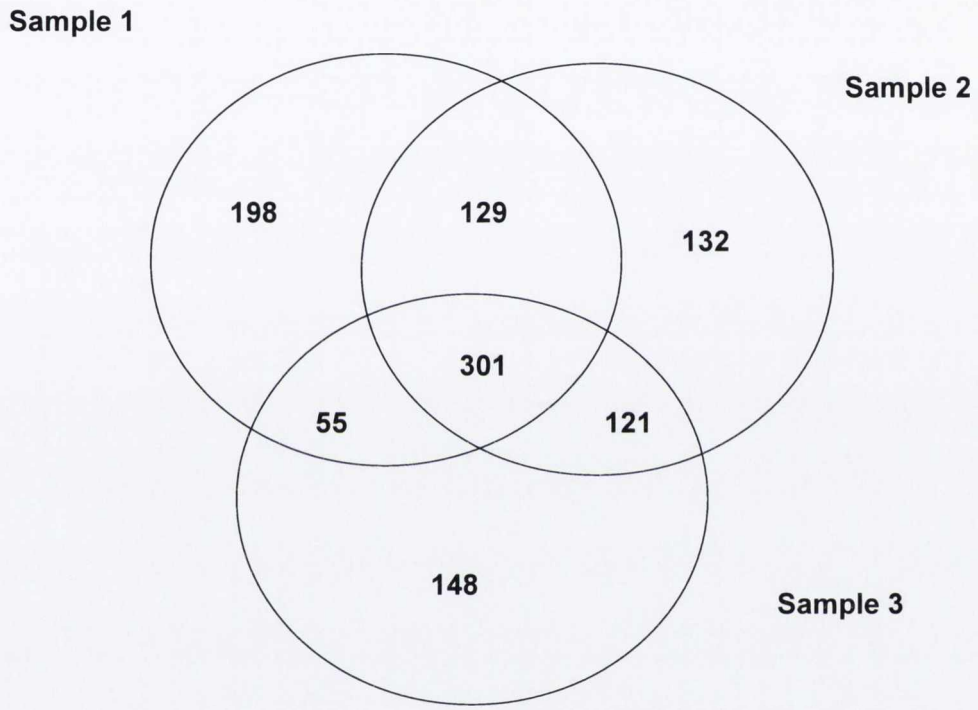

Figure 4.2: Venn diagram showing the numbers testing positive for cocaine in one or more of the three four-month periods in 2006

Table 4.16: Three sample contingency table showing the numbers present in the samples or their overlaps

			Sample 3		
		present		absent	
Sample 2		present	absent	present	absent
Sample 1	present	301*(111)*	55*(121)*	129*(112)*	198*(122)*
	absent	121*(211)*	148*(221)*	132*(212)*	

assume that all pairwise relationships are present (option 4). This is a reasonable assumption in the context of treatment numbers as many clients attend the clinics for a number of years, though not necessarily continually. Under this assumption, the estimate for the missing cell is as follows:

$$\hat{m}_{222} = \frac{x_{111}x_{221}x_{212}x_{122}}{x_{211}x_{121}x_{112}}$$

Filling in the figures from table 4.13 gives:

$$\hat{m}_{222} = \frac{301 \times 148 \times 132 \times 198}{121 \times 55 \times 129}$$

$$= 1,356$$

This is the estimate of the hidden number of cocaine users. The point estimate for the population of opiate users who are in treatment and who are also using cocaine under the assumption of all pairwise relationships present is therefore 2,440. Following Bishop *et al.* (1975) the asymptotic variance of the total estimate is

$$Variance = (\hat{m}_{222})^2 \left(\frac{1}{x_{111}} + \frac{1}{x_{121}} + \frac{1}{x_{211}} + \frac{1}{x_{221}} + \frac{1}{x_{112}} + \frac{1}{x_{122}} + \frac{1}{x_{212}} + \frac{1}{\hat{m}_{222}} \right)$$

Filling in the data gives

$$Variance = (1356)^2 \left(\frac{1}{301} + \frac{1}{55} + \frac{1}{121} + \frac{1}{148} + \frac{1}{129} + \frac{1}{198} + \frac{1}{132} + \frac{1}{1356} \right)$$

$$= 106,020$$

The standard deviation is then

$$= 326$$

The 95% confidence interval for the population of opiate users who are in treatment and who are also using cocaine under the assumption of all pairwise relationships present is [1802,3078]. This estimate seems very high and would imply that between 25% and 43% of the sample surveyed in the NACD general population drug prevalence survey who used cocaine in the 2006 were in treatment. It is also higher than the Zelterman and Chao estimates for 2006 reported in table 4.15.

We now examine the option of two pairs of samples being related for which the point estimate was given by equation (4.25). The most likely is sample 1 and 2 and sample 2 and sample 3. The estimate for $\hat{m}_{222}$ using this model is

$$\hat{m}_{222} = \frac{x_{221}x_{122}}{x_{121}}$$

Filling in the figures from table 4.13 gives:

$$\hat{m}_{222} = \frac{(148)(198)}{55}$$

$$= 533$$

This is the estimate of the hidden number of cocaine users in opiate treatment. Hence, the point estimate for the population of opiate users who are in treatment and who are also using cocaine under the assumption that sample 1 and sample 2 are related and sample 2 and sample 3 are related is 1,617.

$$Variance = (\hat{m}_{222})^2 \left(\frac{1}{x_{221}} + \frac{1}{x_{122}} + \frac{1}{x_{121}} + \frac{x_{121}}{x_{221}x_{122}} \right)$$

Filling in the figures from table 4.13 gives:

$$Variance = (533)^2 \left(\frac{1}{148} + \frac{1}{198} + \frac{1}{55} + \frac{55}{(148)(198)} \right)$$

$$= 9,046$$

The standard deviation is then

$$= 95$$

The 95 % confidence for this model is [1,430,1803]

The Bishop *et al.* (1975) method was applied to the years 2007, 2008, 2009 and 2010 also and the results are reported in table 4.18 alongside the SPSS estimates for cross validation purposes. In table 4.17 the numbers in each sample and in the sample overlaps are shown according to the coding used in table 4.13. For example the code 121 means that for the year in question, the reported number of clients tested cocaine positive in the months January to April and in the months September to December.

The capture-recapture method (CRM) implemented according to the Bishop *et al.*

Table 4.17: Count of the clients present in the first, second or third four month period in each year from 2006 to 2010

Code	2006	2007	2008	2009	2010
221	148	152	166	120	191
212	132	159	135	125	133
211	121	118	132	106	117
122	198	153	166	202	157
121	55	61	58	77	64
112	129	106	125	111	111
111	301	288	314	282	270
222					

(1975) approach has also been implemented by the statistical software program Statistical Package for the Social Sciences (SPSS). Since the release of SPSS version 15,

the Advanced Models module of SPSS can be used to carry out a capture-recapture analysis of data from multiple sources. The present analysis is obtained by using the Generalized Linear Models analysis commands that are contained within the Analyze menu of SPSS Version 16 for Microsoft Windows. The analysis can also be performed by using Loglinear analysis commands. The procedure on how to input the data into GLIM and SPSS (loglinear analysis only) has been explained in detail by Hay *et al.* (EMCDDA 1999a).

With three samples there are 8 possible loglinear models which can be fitted. These are the independence model which includes a parameter for each sample (the main effects), 3 models which add in one interaction term as well as the main effects, 3 models which add in a two-way interaction term and finally the saturated model which includes all three two-way interaction terms. The Akaike information criterion (AIC) (Akaike 1974) and the Bayesian information criterion (BIC) (Schwarz 1978) are reported in the SPSS generalized linear model output and the models with the lowest AIC and BIC values excepting the saturated model are chosen. The best fitting model was found for each year to be that containing the two-way interactions between sample 1 and sample 2 and between sample 2 and sample 3 plus the main effects of sample 1, sample 2 and sample 3. SPSS reports the natural log point estimate.

The point estimates and confidence intervals of the population of opiate users who are in treatment and who are also using cocaine for the years 2006 to 2010 are shown in table 4.18 for the best fitting model along with the remaining estimates obtained using the Bishop *et al.* method. The important result that emerges from the comparison of the prevalence estimates obtained in SPSS and those obtained using the more probability based Bishop *et al.* approach is that the point prevalence estimates are identical. This has implications for prevalence estimation perhaps in countries where a more formula driven approach may be preferred particularly in situations where neither SPSS nor GLIM are available.

Table 4.18: Capture-recapture point estimates with confidence intervals of the total number of problem drug users in treatment who are also using cocaine in the Northern Area Health Board

	Bishop *et al.*	SPSS
Total Population 2006 (95% CI)	1,617 (1,430 - 1,803)	1,618 (1,465 - 1,837)
Total Population 2007 (95% CI)	1,418 (1,284 - 1,552)	1,418 (1,310 - 1,573)
Total Population 2008 (95% CI)	1,571 (1,406 - 1,737)	1,571 (1,437 - 1,764)
Total Population 2009 (95% CI)	1,338 (1,232 - 1,444)	1,338 (1,253 - 1,457)
Total Population 2010 (95% CI)	1,512 (1,354 - 1,669)	1,512 (1,383 - 1,694)

4.4.4 The Multiplier Method

Within this section the multiplier method is used to provide an estimate of the in treatment multiplier based on the 2006 data. This multiplier is then applied to the treatment data estimates obtained using the truncated Poisson and CRM estimates for the years 2007 to 2010 to obtain a general estimate of cocaine prevalence in the region studied. The multiplier method requires two pieces of data in order to estimate last year prevalence. From the previous sections we now have three point estimates for the number of opiate users in treatment in 2006 who are also using cocaine. We will express these estimates as a fraction of the NACD point estimate for the NAHB area in the same year to create a multiplier rate. We will then use this multiplier to calculate estimates of the total number of cocaine users in the NAHB area in the years 2007 to 2010. We first express the Zelterman estimate for the number of opiate users in treatment who are also using cocaine as a fraction of the point NACD point estimate.

The Zelterman rate, $rate_1$ becomes

$$\begin{aligned} rate_1 &= \frac{1671}{8131} \\ &= 0.206 \end{aligned} \tag{4.26}$$

The Chao rate, $rate_2$ becomes

$$\begin{aligned} rate_2 &= \frac{1391}{8131} \\ &= 0.171 \end{aligned} \tag{4.27}$$

The CRM rate, $rate_3$ becomes

$$\begin{aligned} rate_3 &= \frac{1617}{8131} \\ &= 0.199 \end{aligned} \tag{4.28}$$

$Rate_1$ and $rate_2$ are then applied to the Zelterman and Chao estimates reported in table 4.15 and the CRM estimate reported in table 4.18 to produce estimates for the prevalence of last year cocaine use in the NAHB for the years 2007, 2008 and 2009. These estimates can be found in table 4.19.

Table 4.19: Estimated general population cocaine prevalence based on treatment estimates in the NAHB in 2007 - 2010

Year	Zelterman Based Estimate	Chao Based Estimate	Capture-Recapture Based Estimate
2007	8,335	8,117	7,126
2008	7,961	8,082	7,894
2009	7,888	7,760	6,723
2010	10,587	9,275	7,598

4.5 Conclusion and Discussion

During the consultation meetings arranged prior to the introduction of the national drug strategy 2009 - 2016, frontline staff voiced their observation of increased cocaine use among their clients (DCRGA 2009). The ethnographic study by Saris & O'Reilly (2010) reiterated this observation. These are indicators of a changing

reality in relation to drug misuse which present challenges to a service which historically developed in response to the heroin problem of the 1980s (Butler & Mayock 2005). In order to facilitate the response of local service providers an attempt is made here to quantify the extent of the problem using standard statistical methods recommended by the EMCDDA and UNODC. As three distinct data sources were not available, the single source capture-recapture approach was used as an alternative.

Routinely gathered urinalysis samples, obtained from a region in Dublin north central and the Dublin northern suburban area, are used. Although the weekly collection of urinalysis samples has attracted some criticism in recent times (Ó Súilliobháin 2009, Department of Health 2005) it is hoped that the present study adds value to these weekly urinalysis results.

Standard methods, truncated Poisson and single source capture-recapture, were used to estimate the total prevalence of cocaine use including the unobserved cases. Cases could be unobserved because clients have used cocaine outside the detection frame or because they have (temporarily) ceased attending treatment. The use of the different statistical techniques served to cross validate the results. The point estimates of those in treatment in the NAHB who were also using cocaine were in the range 1,300 to 1,800 for the five years with the Zelterman estimator and CRM giving higher estimates than the Chao estimator in general. If we take these numbers as representing the local health office area of Dublin North Central and Dublin North, this represents between 0.52% and 0.73% of the population.

These estimates give no indication however of the intensity of use. To measure this, the percentage of positive tests for each individual and the average percent positive for the cohort for each successive year was calculated. The average percent of positive tests for the years 2006 to 2010 were 28%, 28%, 27%, 26% and 27%

respectively. These percentages could be a low estimate as 20% of tests were on a Thursday and 20% were on a Friday which means that if the pattern of use is for cocaine to be a "weekend drug" it is possible that 40% of the tests were unlikely to be positive in the first instance. The association between day of week and cocaine use may be worth further investigation to test the "weekend drug" hypothesis. For comparison purposes the average percent of positive opioid tests for the cohort for the same period was estimated. For the years 2006 to 2010 these were 56%, 58%, 58%, 60%, and 56% respectively. The percentage of heroin positive tests is high considering that its presence can lead to a client having to present to the clinic more often to obtain their methadone prescription which one would be expect to act as a deterrent. However no such repercussions exist for cocaine use. The levels of cocaine use identified in the present study are therefore likely to cause concern among service providers as it has been found that poly-drug use was a major factor associated with fatal drug overdose Byrne (2001). In particular, the fact that there has been no noticeable reduction, rather a slight increase in 2010, is likely to be of particular interest and possibly unexpected due to the continuing recession in Ireland since 2008. The implication for service providers is that further training may be necessary to deal with this polydrug use. A further point worth noting is that new cocaine-like substance such as Mephedrone have appeared in recent year in Ireland (RTE 2010) and it is not known if such substances could show a false positive for cocaine use (personal communication with Claymon Laboratories September 2011). Further research into this issue would be useful. Furthermore, the average percentages of positive tests could mask differences between those in short and long term contact with the treatment services. It would be useful to examine the possible effect of time in treatment in a further study.

Finally, in table 4.14 a point estimate of 8,131 last year cocaine users was obtained by applying the percentages of last year cocaine users reported in the all-Ireland general population drug prevalence survey (NACD 2007) to the known population

figures in the region. It is useful to compare this point estimate with the estimates obtained for 2006 of cocaine use among those in treatment. These ranged from 1,306 (the Chao 2006 lower Confidence interval estimate in table 4.15) to 1,837 (the capture-recapture SPSS upper confidence interval estimate in table 4.18). The ratio of those in treatment using cocaine to those in the general population using cocaine ranges between 1:6 to 2:9.

4.6 Chapter Summary

In this chapter cocaine prevalence estimation techniques were critiqued. The NACD all-Ireland general population drug prevalence survey reported the percentage of those using cocaine among the general population. We calculated the numbers using cocaine in the region based on the CSO census information for 2006. Urinalysis samples were used to obtain estimates of the total number of opiate users who have contact with the treatment services and who use cocaine. Both Zelterman and Chao Truncated Poisson methods were used, the Chao estimator in general produced lower estimates. The single source capture-recapture prevalence method was implemented using the Bishop *et al.* method and the statistical software program SPSS. The point estimate of treatment clients using cocaine was similar with either method, confidence intervals differed slightly.

The estimates for 2006 ranged from 1,306 (Chao) to 1,860 (Zelterman) compared to the observed number of 1,084. In 2007 these estimates ranged between 1,284 (capture-recapture) and 1,930 (Zelterman) as opposed to the observed 1,037. The estimates for 2008 ranged from 1,301 (Chao) to 1,819 (Zelterman) compared to the observed number of 1,096. In 2009 these estimates ranged between 1,232 (capture-recapture) and 1,823 (Zelterman) as opposed to the observed 1,023. Finally, in 2010 the estimates ranged between 1,354 (capture-recapture) and 2,527 (Zelterman) as opposed to the observed 1,043. The point estimates for 2006 were combined with the numbers using cocaine to obtain benchmark rates. The multiplier-benchmark

technique was applied to produce estimates of the numbers of last year cocaine users in the general population in 2007 [7,126 - 8,335], 2008 [7,894 - 8,082], 2009 [6,723 - 7,888] and 2010 [7,598 - 10,587].

While we hope that the treatment estimates provided are of use to treatment service providers in informing their practice, policy makers need to know what measures are most effective in dealing with the cocaine problem. There is a need to provide an evidence base to assist in decision making process involved in deciding whether resources should be more efficiently directed towards prevention of initiation or towards treatment of problematic use. In chapter 5, dynamic models are developed which aim to cast some light on this question.

Chapter 5

Dynamic Models and Intervention Strategies

5.1 Introduction

In chapter 3 a general introduction to mathematical modelling was presented and in chapter 2 some background information on cocaine in general and its use in Ireland was outlined. In this chapter a simple model is developed which aims to compare the effects of intervention strategies which aim to prevent initiation to drug use (pre-initiation strategies) with those intervention strategies which aim to reduce the numbers of existing users (post-initiation strategies). The model is limited to population of 15 to 34 year olds, identified by the Health Research Board (HRB 2006) as the most at risk population, but could easily be extended to a wider population. The model is verified by means of figures obtained from the two National Advisory Committee on Drugs (NACD 2006, NACD 2008) commissioned household surveys and then simulations are run on a hypothetical population showing the effects of pre-initiation strategies versus post-initiation strategies on the development of the "epidemic". First, the literature on intervention strategies is reviewed.

5.2 Literature Review

Drug policy in Ireland is best characterised by the legislation that has been enacted and by the National Drug Strategy (Pike 2008). Irish legislation defines as criminal offenses the importation, manufacture, trade in and possession of most psychoactive substances, including cocaine. The principal criminal legislative framework is laid out in the Misuse of Drugs Acts (MDA) 1977 and 1984, and the Misuse of Drugs Regulations 1988. The offences of drug possession and possession for the purpose of supply are the principal forms of criminal charge used in the prosecution of drug offences in Ireland. The National Drug Strategy (DCRGA 2009) has an overall strategic objective: "To continue to tackle the harm caused to individuals and society by the misuse of drugs through a concerted focus on the five pillars of supply reduction, prevention, treatment, rehabilitation and research".

Pre initiation drug use prevention strategies

Supply reduction and prevention could be classified as pre-initiation drug use prevention measures. The former, carried out by the Garda National Drugs Team, seeks to reduce demand for drugs by disrupting the supply at source. However, the effectiveness of the "War on Drugs" drug policy approach in reducing the demand for drugs has been disputed by various authors among others Bayer (1991) and World Health Organisation (WHO 2008). At local level, Garda units police the drugs situation, and the Garda Diversion Programme is aimed at preventing young people from becoming involved in drug misuse (Irish Youth Justice Service 2011).

Prevention efforts target secondary school students through the social, personal and health education (SPHE) program, which places a large emphasis on alcohol, drug and solvent use. "Walk Tall" is the awareness programme run in primary schools (Morgan 2003). Most Local Drugs Task forces, Health Promotion Units of the Health Service Executive, the Gardai, and community or voluntary groups throughout the

country operate drugs awareness programmes as a prevention measure. The Drugs Awareness Programme (DAP) is a charity and a division of Crosscare (Crosscare 2011), the social care agency of the Dublin Diocese. This programme trains and facilitates people with leadership skills to develop drug-related initiatives in their own area, for example, needs assessment, support groups, peer education, school programmes and service development. Community Awareness of Drugs (CAD) runs drugs education programmes for parents/carers and offers training to community workers from both the voluntary and statutory sector (HSE 2011). The difficulty in assessing the effectiveness of such programs is clear. As Ronald Brogan of the US Drug Abuse resistance program (D.A.R.E) has remarked "How do you prove that drug use would have occurred if it were not prevented from happening?" (Brogan 2008).

Post initiation of drug use interventions

In Ireland, treatment for problem drug use is provided in outpatient, inpatient, low-threshold and general practice settings. According to Reynolds & Fanagan (2008) the numbers seeking treatment for cocaine use has increased from 81 in 2001 to 552 in 2006 of which 43 were new cases in 2001 and 342 were new cases in 2006. Hence the ratio of new cases to existing cases increased from 5:10 to 6:10.

In the legal context, efforts have been made to implement an arrest referral type scheme as according to Connolly (2005) this has been shown to be successful in the UK. The Probation Service is the appointed statutory body to which the Courts refer offenders who are given the opportunity to attend a drug awareness programme and thereby avoid a custodial sentence. According to Giaquinto (2007) this scheme as yet has not been properly implemented with programmes being made available on a fairly ad hoc basis. Nevertheless according to the Health Research Board (2010) the Drug Treatment Court would continue to operate as an alternative to imprisonment up to 2011 is to be continued for a further eighteen months.

The question naturally arises as to which of these approaches is most effective, pre- or post-initiation measures. The question has become even more pressing in this climate of continuing economic downturn. Ireland's drug-related budget for 2010 showed a decrease of approximately 4.5% in 2009, down from 277 million to 264 million (Health Research Board 2010).

Within this chapter a mathematical model for cocaine use is introduced and simulations are run on the model to assess the impact of both pre-initiation and post-initiation measures. White & Comiskey (2007) in their model of the opiate using career identified three compartments, susceptibles, users not in treatment (including those who were in treatment and relapsed) and those in treatment. Their focus was on treatment pathways. In the present study we will also include a susceptible compartment but it is decided that in order to compare the effects of measures aimed at preventing initiation and those which assist users to quit use it is not necessary to complicate the model with two user compartments. This decision is also driven by the fact that the rate at which users quit is only known as an overall population rate. Quitting can occur naturally. Best *et al.* (2006) argued forcibly that the majority of cocaine users will "mature" naturally out of using cocaine. It can also occur because of an intervention and we are interested in measuring the relative effects of such interventions compared to interventions aimed at preventing initiation. Whether cocaine users quit of their own accord or with assistance from some form of treatment, the model needs to include a compartment for those who were cocaine users and have ceased use. To sum up, three compartments will be used in this model for susceptibles, users and recovereds. The model parameters are described in detail in the section on parameter estimation.

5.3 Mathematical Model

The model proposed is a simple SIR model consisting of three populations, susceptibles (S), users (U) and recovered/removed (R). It is as follows:

$$\frac{dS}{dt} = \mu S + \beta SU \tag{5.1}$$

$$\frac{dU}{dt} = \beta SU - qU - pU \tag{5.2}$$

$$\frac{dR}{dt} = qU + pU \tag{5.3}$$

where

- $S(t)$ = the number of people in the population at time t not currently using and thus deemed susceptible to cocaine use
- $U(t)$ = the number of people in the population who are using cocaine at time t
- $R(t)$ = the number of people who were users but who have ceased using either of their own accord or through treatment at time t
- μ is the rate of population change, the combined effect of births and deaths
- β is the rate at which suscepibles are initiated to cocaine use by other users
- q is the rate at which users cease using cocaine and is assumed to be constant
- p is the rate at which users cease using cocaine due to an intervention

The following assumptions are made:

- The susceptible population is changing at a rate μ which is proportional to its size.

- The rate of initiation of one susceptible into cocaine use, β, is proportional to the number of current users hence the initiation of all susceptibles is proportional to the number of current users and current susceptibles. Changes in this parameter will be used to model intervention strategies such as education, advertising campaigns or any other strategy which has a possible effect on reducing initiation into cocaine usage.

- As data is not currently available about average time of cocaine use before cessation, it is assumed that data from the United States can be applied in the Irish context. Two studies provide estimates, the first by Everingham & Rydell (1994) and the second by Caulkins *et al.* (2004).

- When users recover they are no longer susceptible, relapse is not included in the model.

- q is the average rate at which users cease using cocaine.

- p is the rate at which users cease using cocaine due to some intervention.

- $N = S + U + R$, the total population is assumed to be constant.

The assumption of the model is that the rate at which susceptibles are initiated into cocaine use is proportional to the rate of mixing between users and susceptibles according to the so called "mass action" principle. This was first introduced by Muench (1959) as we saw in chapter 3.

5.4 Parameter Estimation

Three parameters need to be estimated in the model, μ, the rate of change of the susceptible population, β, the rate of initiation to cocaine use and q, the rate of cessation of cocaine use. The parameters μ and β are estimated heuristically using the program Berkeley Madonna based on known initial and final estimates of both

susceptible and user populations. (See Appendix A for information on Berkeley Madonna).

5.4.1 Estimating the rate of change in the population μ

The rate of change in the population, μ, is estimated in Berkeley Madonna based on the population census figures obtained from the Central Statistics Office (CSO 2002, CSO 2006). Based on the census figures of 2002 and 2006 predictions were made of the size of the population in the age group 15 to 34 years up to 2016 and 2020 respectively assuming no decrease due to death or emigration. It emerged that the population had grown in the years 2002 to 2006 due to immigration. These predictions can be seen in figure 5.1. The immigration between 2002 and 2006 averaged over the four years is highlighted in yellow. From the diagram there appears

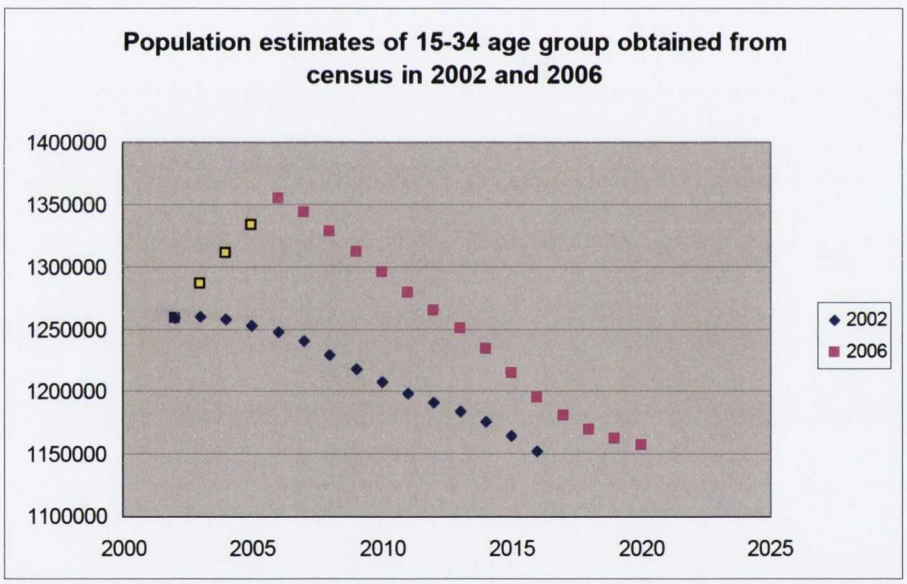

Figure 5.1: Census figures and population estimates of 15-34 age group from 2002 to 2020 showing immigration estimates in yellow

to be three distinct rates of population change, an increase between 2002 and 2006, a decrease between 2006 and 2016 followed by a lesser rate of decrease from 2016 to 2020. Hence, three distinct rates are estimated in Berkeley Madonna heuristically

based on the 2006 census figures using a simplified model shown in equation (5.4).

$$\frac{dS}{dt} = \mu S \tag{5.4}$$

The values of the rate of change in the population μ so obtained are

- from 2002 to 2006 μ = 0.0195849
- from 2007 to 2015 μ = -0.0126625
- from 2016 to 2020 μ = -0.0046109

The Berkeley Madonna generated graph of the size of the susceptible population is shown in figure 5.2.

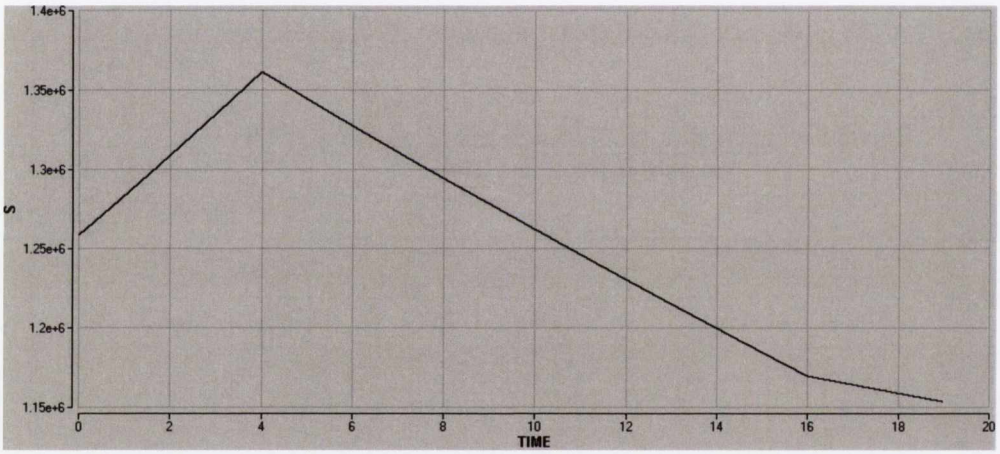

Figure 5.2: Berkeley Madonna Population estimates 2002 to 2020

5.4.2 Rate of cessation of cocaine use

There are no Irish data on rates of ceasing cocaine use at this moment in time. However very good data are available from two studies in the United States. The Rand corporation, a research and development non-profit institution published a study "Modelling the demand for cocaine" in 1994 (Everingham & Rydell 1994). The main focus of the paper was on so called light and heavy users and the role each played in the demand for cocaine. They developed a two-state Markovian model but

information on cohort retention was also produced. The average time to cessation according to the paper was 7.55 years. The estimate of the rate of ceasing use is therefore 1 divided by 7.55, in other words 0.132471.

A second paper written by Caulkins *et al.* (2004) reexamined both the Everingham & Rydell (1994) study and another study of cocaine prevalence authored by Abt for the National Drug Control Policy. By merging the two studies Caulkins *et al.* (2004) produced a model of prevalence estimates which spanned the years from 1972 to 2000. The average rate of cessation in this study was 0.177, which is equivalent to an average time to cessation of 5.64 years.

As Ireland and the United States are at different stages of their "epidemic" it is difficult to assess with any certainty which of the two estimates are preferable - Everingham & Rydell's estimate (1994) may be more pertinent to the Irish context mirroring as it does the early years of the cocaine "epidemic" in the States. But the paper by Caulkins *et al.* (2004) has an impressive span and is more current. The decision was taken to produce models using each one and assess the sensitivity of the model to the different cessation rates. It is reasonable to assume that average cessation rates change over time. In the early phase when a drug has become popular and before the dangers become apparent, one might expect recreational users to use the drug for a longer time. However as problems associated with the drug emerge over time, one might expect the quit rate to increase.

5.4.3 Estimating β, the rate of initiation

The final parameter to be estimated is β, the rate of initiation into cocaine use which is caused by users mixing with susceptibles. Two estimates of this parameter were found in Berkeley Madonna by running the model described in equations (5.1) to (5.3) and finding the value of β which allowed the NACD point estimate of the number of users in 2002 (2% of the total population in 2002) to grow to the NACD point estimate of the number of users in 2006 (3.1% of the total population in 2006).

The population of 15 to 34 year olds in 2002 is 1,258,891. The population of 15 to 34 year olds in 2006 is 1,355,171. The initial values for the model are therefore

$$S(0) = 1233713 \quad U(0) = 25178 \quad R(0) = 0 \quad t(0) = 2002 \tag{5.5}$$

After four years we expect the number of users to be

$$U(4) = 42010 \quad t(4) = 2006 \tag{5.6}$$

Using the Everingham & Rydell (1994) cessation rate of 0.132471 the value of β is found to be 2.2243 x 10^{-7}. Using the Caulkins *et al.* (2004) cessation rate of 0.177 the value of β is found to be 2.61067 x 10^{-7}.

5.5 Model Simulations

5.5.1 Model simulation of user numbers after 35 years

The model described in equations (5.1) to (5.3) is run twice over a time period of 35 years using the initial conditions shown in equation (5.6). Estimates for each compartment susceptible, user and recovered are produced for each year. In the first run the Everingham & Rydell (1994) rate for q and the derived rate for β are used. The Berkeley Madonna graphical display for this run is shown in figure 5.3. It should be noted that in Berkeley Madonna the scale used on the diagram is different on the left (showing number of susceptibles) and right (showing the numbers of users). Also the number of recovered is not shown in order to aid interpretation of the graphic. Berkeley Madonna also presents the output in tabular format which allows one to read off the precise numbers in any compartment (for example users) or any parameter at any time within the model time frame.

According to the model, if the cessation rate is that reported in the Everingham & Rydell (1994) study, implying average time before quitting cocaine use is seven and

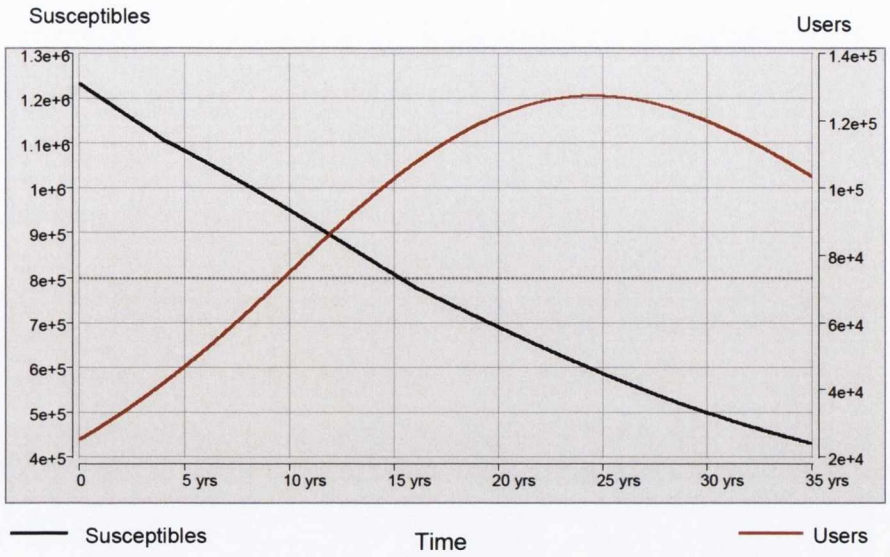

Figure 5.3: Model Prediction of cocaine user numbers after 35 years assuming average time of use is 7.5 years

a half years, then the number of users will have grown to 103,480 after 35 years. This is over 10% of the total population predicted by the model. It would have peaked after 24.5 years at 127,483 users and then started to decline.

The Caulkin's cessation rate, which implies an average time of use before cessation of five and a half years, estimates 59,038 users after 35 years. According to the model this figure represents just over 5% of the population at that time. This model predicts that the number of users would peak after 19.8 years at around 97,174. The Berkeley Madonna graphical display for this run is shown in figure 5.4.

Hence, the models predict that a shortening of the average time of use before cessation by two years would result in a decrease of 48% in the number of users after 35 years. Based on these simulations, it is concluded that changes in the average time it takes to quit a drug, which may be expected as the result of an increase in awareness of the dangers associated with the drug, will have a major impact on the number of users and the span of the epidemic.

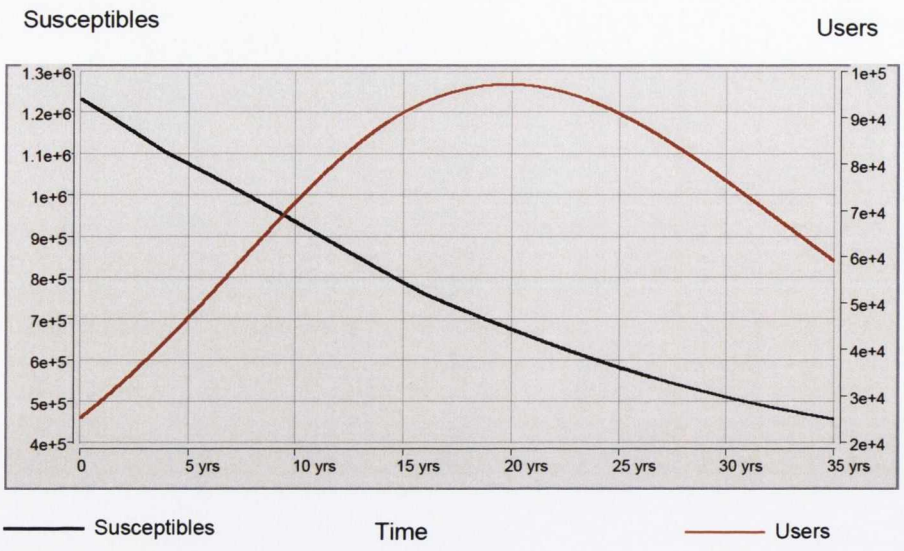

Figure 5.4: Model Prediction of cocaine user numbers after 35 years assuming average time of use is 5.5 years

5.5.2 Simulation of the extent of the reduction in cocaine users resulting from measures aimed at reducing initiation

In this section, simulations will explore the effects of pre-initiation and post-initiation preventative measures on the number of cocaine users after a period of 35 years using both the quit rate with its corresponding β rate and the Caulkins *et al.* (2004) quit rate with its corresponding β rate. The effectiveness of strategies aimed at preventing initiation to cocaine use can be simulated through manipulation of the parameter β in equation (5.2). The effects of a 1%, 2%, 3%, 4% and 5% reduction in the numbers of people starting to use cocaine on total cocaine user numbers over the 35 year time period is simulated. This simulation is easily performed in Berkeley Madonna as it provides an option called *Batch runs*. A parameter of choice can be modified in equal increments as many times as desired. The β parameter was modified five times in increments of 1% during two runs using the different quit rates. The results of these runs can be seen in Figures 5.5 and 5.6 for the Everingham & Rydell (1994) quit rate and the Caulkins *et al.* (2004) quit rate respectively.

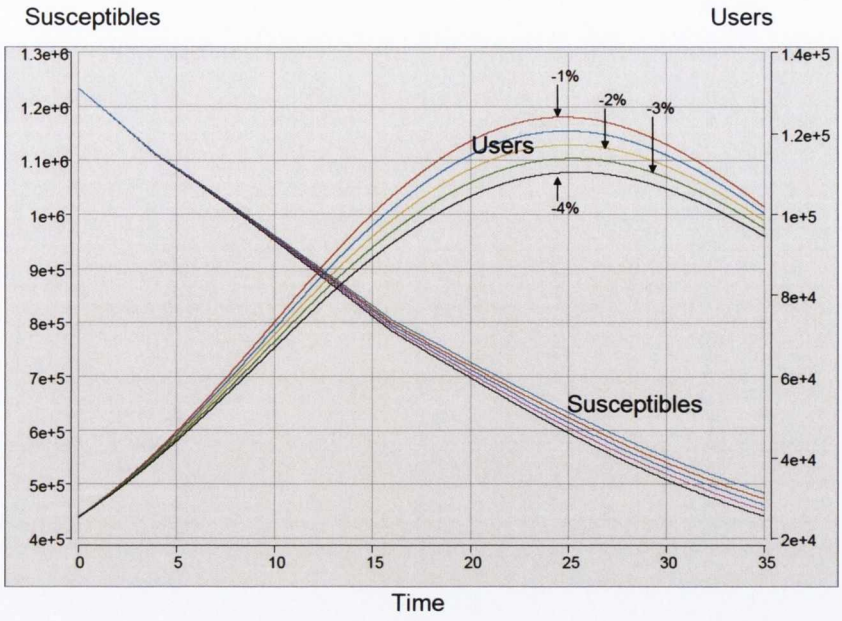

Figure 5.5: Model Prediction of cocaine user numbers after 35 years assuming preventative measures achieve reductions in initiation : quit time 7.5 years

The first model estimates that decreasing the rate of initiation successively by 1% will lead to some reduction in user numbers after 35 years. The numbers, according to the model would be 101,906, 100,228, 98,453, 96,585 and 94,630. The ratio of the reduction in initiation rates to reduction in user numbers after 35 years is in the region of a ratio of 1:2 in percentage terms. In other words a 1% reduction in the rate only results in a 2% reduction in user numbers after 35 years. According to this model, user numbers peak between 24 and 25.5 years at 124,049, 120,623, 117,214, 113,824 and 110,457.

The second model with an average time of using of 5.5 years has much lower user numbers after 35 years and also much lower maximum numbers. The numbers are 58,258, 57,397, 56,457, 55,440, and 54,350 after 35 years and the maximum numbers reached around 20 years for each model are 94,181, 91,218, 88,290, 885,399 and 82,551 respectively. The ratio in the reduction of initiation to final number reduction after 35 years is only of the order 4:3 with the reduction in maximum numbers being better at a ratio of 3:1. In comparing the two models the most

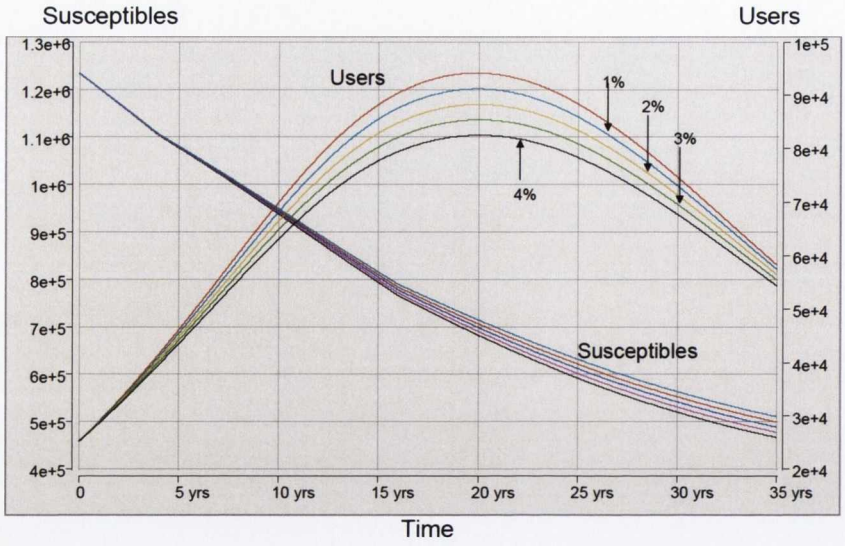

Figure 5.6: Model Prediction of cocaine user numbers after 35 years assuming preventative measures achieve reductions in initiation : quit time 5.5 years

interesting idea to emerge is that a reduction in average time to cessation, which seems to occur naturally as a drug falls out of favour with the population, creates the most striking reduction in User numbers compared to the hypothesised reductions achieved through measure aimed at reducing initiation.

5.5.3 Simulation of the extent of the reduction in cocaine users as a result of effective treatment and intervention

The effectiveness of treatment and other intervention strategies, if successful, in reducing the overall cocaine use in the population can be simulated through manipulation of the parameter p in equation (5.2). This parameter has had the value 0 in the precious simulations. The effects of a 1%, 2%, 3%, 4% and 5% reduction in the numbers of people using cocaine due to these interventions is explored in the model. These parameter changes are applied successively to the model using the Everingham & Rydell (1994) quit rate and the Caulkins *et al.* (2004) quit rate. The results of the simulations are presented in figures 5.7 and 5.8 respectively.

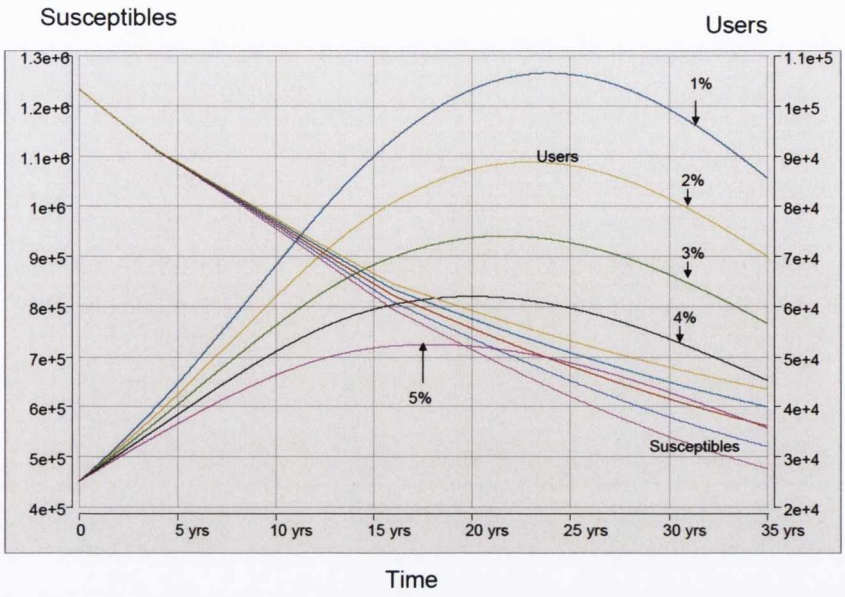

Figure 5.7: Model Prediction of cocaine user numbers after 35 years assuming treatment and interventions achieve reductions in user numbers : quit time 7.5 years

The model, using the Everingham & Rydell (1994) quit rate suggests that reductions of users numbers of 1%, 2%, 3%, 4%, 5% as a result of interventions will result in user numbers of 85,594, 70,033, 56,647, 45,285 and 35,785 respectively after 35 years. These figures represent reductions of 17%, 32%, 45%, 56% and 65% respectively in overall user numbers when compared to the model predictions with no interventions present. The time required to reach the maximum number of users in the model decreases as the effects of policies aimed at reducing user numbers takes effect. It reduces from 24 years to 18 years and the numbers reached are 106,513, 88,791, 74,053, 62,031 and 52,457.

The model using the Caulkins *et al.* (2004) quit rate echoes these results. After 35 years reductions are noted of 16%, 30%, 42%, 53% and 62% in user numbers compared to numbers expected with no such interventions. This translates into user numbers of 83,312, 71,522, 61,648, 53,510, 46,803 respectively as a result of the 1%, 2%, 3%, 4% and 5% reduction in user numbers.

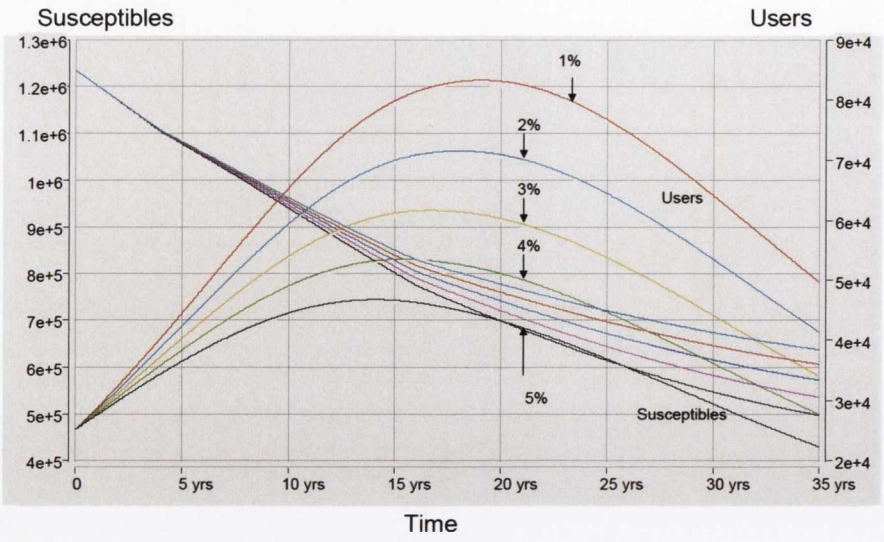

Figure 5.8: Model Prediction of cocaine user numbers after 35 years assuming treatment and interventions achieve reductions in user numbers : quit time 5.5 years

5.5.4 Discussion

The model, which has been developed, enables the exploration of how prevention or education (pre-initiation efforts) as compared to treatment (post-initiation efforts) impact on the number of users over time. The results show that prevention which results in a decrease in the numbers starting use of the drug is not as effective as treatment which succeeds in removing users from the user group. Based on this simple model, the largest reduction in the numbers however, is due to an increase in the rate of ceasing use. This seems to occur naturally with time as dangers related to the drug's use becomes known. For policy makers this would indicate that the most effective approach to dealing with the drug problem is by providing effective treatment and rehabilitation and by keeping the public informed about the dangers as they become known. Efforts at prevention on the other hand seem to produce rather lower returns as defined by a reduction in user numbers. The model is however simple and these results should be accepted with caution. For example, it is known that drug users in treatment relapse usually more than once before they

are successful in ceasing use after treatment. The model does not include relapse and this could be included in an improved adaptation in the future. This could perhaps produce more robust estimates of the effect of various drug prevention strategies.

5.6 Chapter Summary

This chapter demonstrates how mathematical models can be used to simulate the effects of different policy strategies on the average behaviour of individuals. It uses last year prevalence rates obtained from two NACD studies (NACD 2006, NACD 2008) and applies these rates to CSO census figures obtained during the same years (CS0 2002, CSO 2006) enabling a model to be built of the dynamics of the change in cocaine use during that period. A key parameter in the model is the rate at which users cease use, the quit rate. Two rates based on the work of Everingham & Rydell (1994) and Caulkins *et al.* (2004) respectively were used. The models predict that a shortening of the average time of use before cessation by two years would result in a decrease of 48% in the number of users after 35 years. It is difficult to quantify the effect of prevention strategies but this model indicates that prevention is less effective in reducing the numbers using cocaine than strategies such as treatment which reduce the numbers of users. However there are many limitations to the model simulated in particular the lack of an Irish quit rate and the simplicity of the model. As the model presented within this chapter is a very simple one, a more detailed age structured model is introduced and simulated within the next chapter with a view to examining age specific initiation in more detail.

Chapter 6

Dynamic Age Dependent Models and Social Contact Patterns

6.1 An Age Differentiated Model

In the previous chapter, a simple intervention model was developed to explore the effects of prevention (pre-initiation measures) as compared to post-initiation measures in the control of cocaine use. In this chapter a new model is developed to accommodate three different age groupings, 15-24, 25-34 and 35-44 years with the aim of exploring the effects of age dependent initiation into cocaine usage through social contacts on overall incidence and prevalence. This choice is motivated by the fact that the household surveys jointly commissioned by the National Advisory Committee on Drugs (NACD) in the Republic of Ireland and the Drug and Alcohol Information and Research Unit (DAIRU) in Northern Ireland supplied the data in this format (NACD 2006, NACD 2008). In addition, the EMCDDA have indicated that the 15-35 year old group is the most at risk of illicit drug use. Anderson & May (1991) suggested the use of Who Acquires Infection from Whom (WAIFW) matrices

to explore how mixing structures between the various age groups might affect the spread of infectious diseases. The methodology they suggest will be explored and applied to the spread of cocaine use in an Irish population. However, the developed models could be applied in any setting or for any drug and the lessons learnt from the process of applying the model will be universal. The new model, taking into account the three age classes and using subscripts 1, 2 and 3 to identify the age groups 15-24, 25-34 and 35-44 respectively, gives the following three groups of equations:

$$\frac{dS_1}{dt} = \Lambda_1 + \mu_1 S_1 - \lambda_1 S_1 \tag{6.1}$$

$$\frac{dU_1}{dt} = \lambda_1 S_1 + \mu_1 U_1 - qU_1 \tag{6.2}$$

$$\frac{dR_1}{dt} = qU_1 + \mu_1 R_1 \tag{6.3}$$

$$\frac{dS_2}{dt} = \Lambda_2 + \mu_2 S_2 - \lambda_2 S_2 \tag{6.4}$$

$$\frac{dU_2}{dt} = \lambda_2 S_2 + \mu_2 U_2 - qU_2 \tag{6.5}$$

$$\frac{dR_2}{dt} = qU_2 + \mu_2 R_2 \tag{6.6}$$

$$\frac{dS_3}{dt} = \Lambda_3 + \mu_3 S_3 - \lambda_3 S_3 \tag{6.7}$$

$$\frac{dU_3}{dt} = \lambda_3 S_3 + \mu_3 U_3 - qU_3 \tag{6.8}$$

$$\frac{dR_3}{dt} = qU_3 + \mu_3 R_3 \tag{6.9}$$

where

- S_i = the number of susceptibles in age group i
- U_i = the number of cocaine users in age group i
- R_i = the number of users in age group i who have ceased using
- Λ_i = age specific rate of population increase caused by immigration

- μ_i the age specific rate of population change caused by migration or death and will be negative if the population is decreasing
- λ_i = the age specific force of infection which is defined as the per capita rate at which susceptibles are infected
- q = the average rate at which users cease using

The assumption of the model is that the rate at which susceptibles are initiated into cocaine use is proportional to the rate of mixing between users and susceptibles according to the so called "mass action" principle introduced by Muench (1959). Hence the age specific force of infection can be described for $i = 1$ to 3 as

$$\lambda_i = \beta_i * U_i \tag{6.10}$$

where β_i is the rate of initiation to cocaine use caused by this mass action within age group i. However, in the WAIFW matrix theory proposed by Anderson & May (1991), the force of infection is further refined by defining the rate at which each group initiates each other group into cocaine use. We have,

$$\lambda_i = \sum_{i=1}^{3} \beta_{ij} * U_j \tag{6.11}$$

where β_{ij} is the rate at which susceptibles in group i are initiated into cocaine use by users in group j, for example β_{12} is the rate at which 15-24 year olds are initiated into cocaine use by their contacts in the age group 25-34 years old. In other words

$$\lambda_1 = \beta_{11} * U_1 + \beta_{12} * U_2 + \beta_{13} * U_3 \tag{6.12}$$

$$\lambda_2 = \beta_{21} * U_1 + \beta_{22} * U_2 + \beta_{23} * U_3 \tag{6.13}$$

$$\lambda_3 = \beta_{31} * U_1 + \beta_{32} * U_2 + \beta_{33} * U_3 \tag{6.14}$$

This can also be written in matrix form as

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}$$

The $n \times n$ matrix containing the age specific transmission rates β_{ij}

$$\begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix}$$

is called the WAIFW matrix.

Prevalence figures for U_1, U_2, U_3 can be obtained from the NACD Household survey estimates for 2002 and these are used as initial values for U_1, U_2 and U_3. In an epidemiological context the age specific force of infection can be calculated from case records or seriological data. However in the case of initiation to drugs no such data is available. Hence, starting with the NACD point estimates of the number of cocaine users in 2002 (NACD 2006), a heuristic methodology which was previously used in chapter 5 is employed to calculate the rates λ_1, λ_2 and λ_3 which result in the 2006 NACD point estimates being attained according to the model.

Once λ_1, λ_2 and λ_3 are known equations (6.12) to (6.14) must be solved. However, equations (6.12) to (6.14) involve nine unknowns which can therefore only be solved mathematically by putting constraints on the β's. In an infectious disease model the most common assumption is to assume the WAIFW matrix is symmetric. This means that, for example, β_{12} is the same as β_{21}. In other words the rate at which 15-24 year olds are initiated into cocaine use by their contacts in the age

group 25-34 years old is the same as the rate at which 25-34 year olds are initiated into cocaine use by their contacts in the age group 15-24 years old. However in the context of initiation to drug use this does not make epidemiological sense. It would seem reasonable that a 15-24 year old is less likely to initiate a 35-44 year old into drug use than vise versa. The aim in this chapter is therefore to develop a model which uses WAIFW matrices to simulate age specific initiation and discard those which the model proves to be impossible or highly unlikely. In this way, it is expected that insight will be gained into the complex social structures of drug use initiation.

Four different matrix structures are proposed called WAIFW1, WAIFW2, WAIFW3 and WAIFW4 respectively. These structures are exploratory in nature rather than epidemiologically appropriate, as to date no studies were found which had examined age related initiation rates. The rate at which users in a particular age group initiate their peers is on the diagonal. Thus, for example, users in the 15-24 years age group (referred to henceforth as the initiators where that is the function they are performing) initiate their peers (referred to henceforth as the initiated where that is the function they are performing) at a rate of β_1. Users in the 25-34 years age group initiate their peers at a rate of β_2 while users in the 35-44 years age group initiate their peers at a rate of β_3. WAIFW1 assumes that each age group is initiated solely by their own age group and WAIFW2 assumes that each age group is initiated at its own initiator rate by all groups. They have the following form

$$\text{WAIFW1} \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \beta_2 & 0 \\ 0 & 0 & \beta_3 \end{pmatrix} \qquad \text{WAIFW2} \begin{pmatrix} \beta_1 & \beta_1 & \beta_1 \\ \beta_2 & \beta_2 & \beta_2 \\ \beta_3 & \beta_3 & \beta_3 \end{pmatrix}$$

WAIFW3 assumes that each group initiates its own age group and the group next in age. The non-peer group are initiated at the non-peer initiator rates. The WAIFW4 model is the most complex. The 15-24 age group initiates the 25-34 at a rate of β_2,

which is assumed to be a lower rate than the peer rate. They do not initiate users in age group 35-44 at all. The 25-34 age group initiates the 15-24 at the presumed lowest rate of β_3 and does not initiate the 35-44 age group at all. Finally, the 35-44 age group do not initiate the youngest group at all and initiates their peers and the group 25-34 at the same rate of β_3.

$$\text{WAIFW3} \begin{pmatrix} \beta_1 & \beta_2 & 0 \\ 0 & \beta_2 & \beta_3 \\ 0 & 0 & \beta_3 \end{pmatrix} \qquad \text{WAIFW4} \begin{pmatrix} \beta_1 & \beta_2 & 0 \\ \beta_3 & \beta_2 & 0 \\ 0 & \beta_3 & \beta_3 \end{pmatrix}$$

6.2 Estimation of the Parameters

6.2.1 Population rate of change, μ and Λ

The NACD (2002) general population drug prevalence survey and the NACD (2006) general population drug prevalence survey both coincided with censuses hence precise population data for each age group was available from the Central Statistics Office (CSO) for those years. The 2002 population data was imported into Microsoft Excel and a projection of the population in 2006 was calculated as described in chapter 5. It was known in this period that there had been immigration into Ireland this procedure was undertaken in order to estimate just how significant the immigration effect was. Scatter plots of the population data for our three age groups based on this estimation are presented below (see figures 6.1, 6.2 and 6.3). From these plots it is clear that for each age group, an increase in population occurred in 2006 over and above the figures calculated from the 2002 data. Ireland during this period attracted a lot of migrants, most markedly in the 25-34 age group. It is also clear that the population is starting to decline from 2008 in the 15-24 age group with or without migration, the same is true in the 25-34 age group, though the decline is less dramatic, and the only group showing an increase in population during the period is the 35-44 age group. It is interesting to note the differing patterns in the

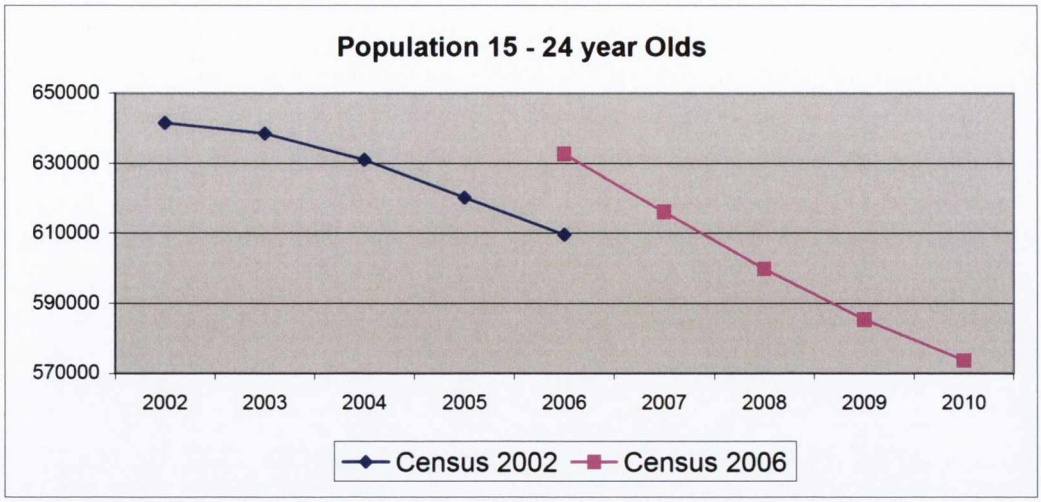

Figure 6.1: Population of 15-24 year olds extrapolated for the years 2003 - 2006 and 2007 - 2010 from 2002 and 2006 Census Data. The difference seen in 2006 indicates the level of increase due to immigration.

age groups strengthening the case in favour of an age differentiated model.

It was decided to calculate μ_1, μ_2, μ_3 as the proportional change in the population from 2002 to 2006 using the estimated population figures and add an immigration constant which would bring the figures up to the observed population for each group. There are two important advantages to incorporating this concept into the model. Firstly simulations which compared this method of implementing the population change produced very different population estimates long term as compared to the model which had no immigration parameter. Secondly the immigration parameter could be turned off readily with a simple if statement in the program Berkeley Madonna. Hence the rate of change of the population μ_i ($i = 1$ to 3) which results in the estimated population in 2006 is calculated for each group.

Berkeley Madonna requires an initial estimate for the parameters in order to run the program. The populations in 2002 and 2006 were known from the census data. An initial estimate for μ_i was calculated by solving for μ_i in the standard compound interest formula seen in equation (6.15).

$$P_{2002}(1+\mu)^4 = P_{2006} \tag{6.15}$$

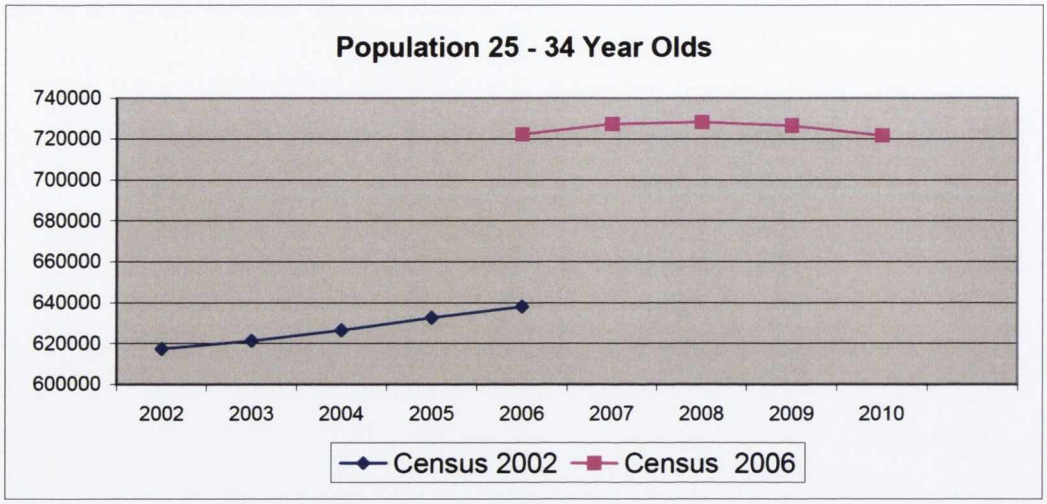

Figure 6.2: Population of 25-34 year olds extrapolated for the years 2003 - 2006 and 2007 - 2010 from 2002 and 2006 Census Data. The difference seen in 2006 indicates the level of increase due to immigration.

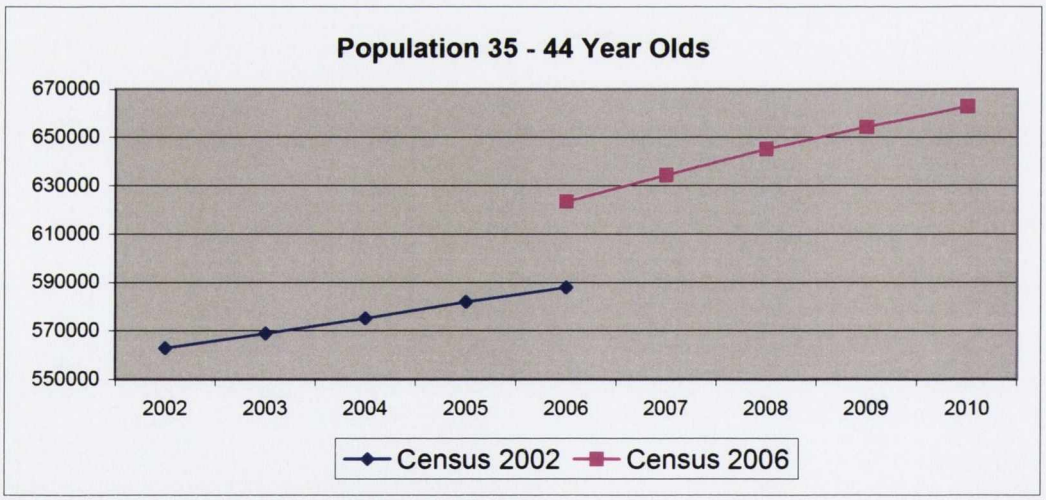

Figure 6.3: Population of 35-44 year olds extrapolated for the years 2003 - 2006 and 2007 - 2010 from 2002 and 2006 Census Data. The difference seen in 2006 indicates the level of increase due to immigration.

where P_{2002} is the population in 2002 and P_{2006} is the population in 2006. This estimate is then fine-tuned in Berkeley Madonna using the well known exponential population growth model shown in equation (6.16).

$$\frac{dP_{2002}}{dt} = \mu P_{2006} \tag{6.16}$$

Next, an immigration constant is calculated for each age group using equation (6.17). An initial estimate was estimated as being a quarter of the difference in population between the estimated population in 2006 and the census population in 2002 Census and then the parameter estimate was fine-tuned in Berkeley Madonna. Hence the change in the population is now described by equation (6.17) below.

$$\frac{dP_{2002}}{dt} = \Lambda + \mu P_{2006} \tag{6.17}$$

This procedure was followed for all three age groups and the parameter estimates per age group are shown in table 6.1.

Table 6.1: Estimated rate of population change, μ and constant of immigration, Λ per age group based on 2002 (NACD 2006) and 2006 prevalence surveys (NACD 2006, NACD 2008)

Age Group	$\boldsymbol{\mu}$	$\boldsymbol{\Lambda}$
15-24	-0.0042795	7068
25-34	0.00692757	25265
35-44	-0.006646	10908

6.2.2 Rate of quitting

In chapter 5 two American estimates for the parameter q were presented, one from the Everingham & Rydell (1994) study and the second from the more recent Caulkins *et al.* (2004) study. As this study incorporated the results from both the Everingham & Rydell (1994) study and another study by Abt and spanned the years from 1972 to 2000, we decided to use the latter estimate for this chapter. The average quit

rate used is therefore 0.177, which is equivalent to an average span of cocaine use of 5.5 years.

6.2.3 Force of infection parameter

The force of infection usually denoted λ is the per capita rate of acquisition of infection. This is a key parameter in any model which hopes to replicate the dynamics of the spread of infection. This model was first proposed by Muench (1959) and was described as catalytic because of its similarities to equations commonly employed in the study of chemical reactions. The simplest model of transmission is

$$\frac{dS}{dt} = -\lambda S \tag{6.18}$$

$$\frac{dI}{dt} = \lambda S \tag{6.19}$$

The rate λ at which susceptibles become infected at time t is proportional to the number of susceptibles at time t. With respect to infectious diseases two main methods are used to estimate λ:

- Case records
- Serological data

In the case of infectious diseases many infections are notifiable and in many countries case records have been kept for various diseases since the start of the 20th century. This implies the availability of a time series of prevalence or incidence data being available which can then be used to estimate λ. A further epidemiological statistic that is available is presence or absence of antibodies specific to a particular organism obtained from immunological assays. The only possibly comparable routinely gathered data available in a drug use context are urinalysis test results which are available in the Irish situation only in a treatment context.

In the present study we use the results from two studies undertaken in 2002/2003

and 2006/2007 to obtain a point estimate of the number of users. As we have noted earlier, the General Population Drug Prevalence Survey in Ireland and Northern Ireland were conducted by MORI MRC on behalf of the National Advisory Committee on Drugs (NACD) in the Republic of Ireland and the Drug and Alcohol Information Research Unit (DAIRU) at the Department of Health, Social Services and Public Safety (DHSSPS) in Northern Ireland. The core objective of the research was to provide robust data regarding the prevalence of (licit and illicit) drug use amongst the general population. The provision of estimates for both jurisdictions and on an all island basis was the priority of the research based on the guidelines produced by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). The study provides figures for last month prevalence, last year prevalence and lifetime prevalence for twelve drugs in the Republic of Ireland and in Northern Ireland.

Our focus here is on the last year prevalence figures of cocaine use (both powder and crack) for the age-groups 15-44 in the Republic of Ireland and these are reproduced in chapter 4 in table 4.1. We use the point estimates in 2002 and 2006 to help estimate the rate of transmission for the three age groups the age specific force of infection λ_i. Our approach here, as in chapter 5, is heuristic the model described in equations (6.1) to (6.9) to calculate the parameter. There is an interaction in the model between the rate of initiation into cocaine use and the rate of quitting, which is solved in Berkeley Madonna using the Runga Kutta numerical method. For example for the 15-24 age group we use the point estimate to obtain the initial number of users (2.7% of the population) and estimate the value of λ which in combination with the quit parameter results in a number of users equivalent to 3.8% of the overall population. This estimate is produced by solving equations (6.1) to (6.9) in Berkeley Madonna for each age group and the derived estimates can be seen in table 6.2

Table 6.2: Force of Infection, λ_i, per age group

Age Group	Force of Infection
15-24	0.0093741
25-34	0.0069357
35-44	0.002779

6.2.4 Calculating the β values for the four WAIFW matrices

Anderson & May (1991) suggested that once the force of infection is known that the β values can be calculated algebraically for the required matrix structures. For WAIFW1 equations (6.12) to (6.14) are solved giving,

$$\lambda_1 = \beta_1 U_1 \tag{6.20}$$

$$\lambda_2 = \beta_2 U_2 \tag{6.21}$$

$$\lambda_3 = \beta_3 U_3 \tag{6.22}$$

For WAIFW2 equations (6.12) to (6.14) are solved giving,

$$\lambda_1 = \beta_1 (U_1 + U_2 + U_3) \tag{6.23}$$

$$\lambda_2 = \beta_2 (U_1 + U_2 + U_3) \tag{6.24}$$

$$\lambda_3 = \beta_3 (U_1 + U_2 + U_3) \tag{6.25}$$

For WAIFW3 equations (6.12) to (6.14) are solved giving,

$$\lambda_1 = \beta_1 U_1 + \beta_2 U_2 \tag{6.26}$$

$$\lambda_2 = \beta_2 U_2 + \beta_3 U_3 \tag{6.27}$$

$$\lambda_3 = \beta_3 U_3 \tag{6.28}$$

For WAIFW4 equations (6.12) to (6.14) are solved giving,

$$\lambda_1 = \beta_1 U_1 + \beta_2 U_2 \tag{6.29}$$

$$\lambda_2 = \beta_3 U_1 + \beta_2 U_2 \tag{6.30}$$

$$\lambda_3 = \beta_3 U_2 + \beta_3 U_3 \tag{6.31}$$

It should be noted also at this point that structures which appear to be reasonable from a theoretical point of view have no solution mathematically. For example the following matrix which assumes that each age group initiates the other groups at their own age specific rate cannot be solved.

$$\begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \beta_1 & \beta_2 & \beta_3 \end{pmatrix}$$

The calculated β values, for the initial values of the U_i are presented in table 6.3. These β estimates were then used in each WAIFW matrix structure model and the models were run for four years from 2002 to 2006 to assess the accuracy of the model user estimates as compared to the NACD estimates. These models are described in detail in the next section. Here, we simply present the resulting estimates for the number of users in each age group in table 6.4.

Table 6.3: β values for each WAIFW structure

	WAIFW1	**WAIFW2**	**WAIFW3**	**WAIFW4**
$\boldsymbol{\beta_1}$	5.41199E-07	3.32863E-07	3.01218E-07	3.97119E-07
$\boldsymbol{\beta_2}$	8.64154E-07	2.46279E-07	5.17904E-07	3.1094E-07
$\boldsymbol{\beta_3}$	9.87211E-07	9.86791E-08	9.87211E-07	2.56342E-07

Table 6.4: Last year prevalence of cocaine users in 2006 predicted by each WAIFW model

	WAIFW1	WAIFW2	WAIFW3	WAIFW4	Expected*
15-24	30108	36033	40464	34063	24044
25-34	38026	27544	40241	26660	17338
35-44	15578	10275	15613	12445	6234
Total	83713	73851	96318	73168	47616

**Note: The Expected number column shows the number estimated from the 2006 NACD point estimate for each age group.*

6.2.5 Conclusion

The method suggested by Anderson & May (1991) of estimating the λ_i's and then using these estimates to solve equations (6.20) to (6.30) was followed. Comparing the estimates obtained for the four matrix structures with the known point estimates shows that the process suggested by Anderson & May does not work well and provides an over estimate. However the concept remains a useful one. In the following section, models which replicate the WAIFW structures but which replace the force of infection λ_i by $\beta_i U_i$ are presented and these models are now used in Berkeley Madonna to estimate alternative WAIFW structure dependent β_i. These models, given the start prevalence estimates in 2002 for the age group in question, will produce the 2006 prevalence estimates for the age group in question.

6.3 A Separate Model for Each WAIFW Structure

In this section a separate model is developed for each WAIFW structure and these models allow new model dependent estimates to be calculated for the β_i parameters. These β_i parameters are found by forcing the model to arrive at the expected number of users after four years given the known initial number of users in 2002. Once these parameters are found the term involving β (the $\beta_{ij} U_i S_i$ term) gives a continuous estimate of the incidence, and yearly incidence can be found in the table output

provided by Berkeley Madonna. Per WAIFW model the incidence is reported for each year from 2002 to 2006.

6.3.1 β values for WAIFW1

The model outlined in equations (6.1) to (6.9) can be modified by replacing λ with βU as appropriate for the WAIFW matrix structure as described in section 6.1 and the β values can then be estimated in Berkeley Madonna. We have,

$$\frac{dS_1}{dt} = \Lambda_1 + \mu_1 S_1 - \beta_1 S_1 U_1 \tag{6.32}$$

$$\frac{dU_1}{dt} = \beta_1 S_1 U_1 + \mu_1 U_1 - q U_1 \tag{6.33}$$

$$\frac{dR_1}{dt} = q U_1 + \mu_1 R_1 \tag{6.34}$$

$$\frac{dS_2}{dt} = \Lambda_2 + \mu_2 S_2 - \beta_2 S_2 U_2 \tag{6.35}$$

$$\frac{dU_2}{dt} = \beta_2 S_2 U_2 + \mu_2 U_2 - q U_2 \tag{6.36}$$

$$\frac{dR_2}{dt} = q U_2 + \mu_2 R_2 \tag{6.37}$$

$$\frac{dS_3}{dt} = \Lambda_3 + \mu_3 S_3 - \beta_3 S_3 U_3 \tag{6.38}$$

$$\frac{dU_3}{dt} = \beta_3 S_3 U_3 + \mu_3 U_3 - q U_3 \tag{6.39}$$

$$\frac{dR_3}{dt} = q U_3 + \mu_3 R_3 \tag{6.40}$$

The assumption of this model is that initiation for 15 to 24 year olds is due to contact with 15 to 24 year olds. Similarly it is assumed that initiation for 25 to 34 year olds and for 35 to 44 year olds is due to contact with their respective age groups.

The estimates for β_1, β_2 and β_3 obtained in Berkeley Madonna are 4.4606×10^{-7}, 5.5281×10^{-7} and 6.2161×10^{-7} respectively. In other words the rate of initiation in the oldest group is the highest of the three being approximately 40% higher than for the 15 to 24 year olds and 12 % higher than for the 25-34 year olds. The

ratio of β_1 to β_2 to β_3 according to this model is approximately 5:6:7. This seems epidemiologically counter intuitive. Nevertheless, the incidence figures estimated in the model from 2002 to 2006 are presented in table 6.5.

Table 6.5: Incidence estimates for cocaine use in 2002 - 2006 using WAIFW1 model

	2002	**2003**	**2004**	**2005**	**2006**
15-24	4823	5201	5589	5981	6375
25-34	2703	3339	4166	5250	6676
35-44	980	1209	1504	1885	2380

6.3.2 β values for WAIFW2

Next the model outlined in equations (6.1) to (6.9) is modified in order to calculate β values for the WAIFW2 model of mixing. We have,

$$\frac{dS_1}{dt} = \Lambda_1 + \mu_1 S_1 - \beta_1(U_1 + U_2 + U_3)S_1 \tag{6.41}$$

$$\frac{dU_1}{dt} = \beta_1(U_1 + U_2 + U_3)S_1 + \mu_1 U_1 - qU_1 \tag{6.42}$$

$$\frac{dR_1}{dt} = qU_1 + \mu_1 R_1 \tag{6.43}$$

$$\frac{dS_2}{dt} = \Lambda_2 + \mu_2 S_2 - \beta_2(U_1 + U_2 + U_3)S_2 \tag{6.44}$$

$$\frac{dU_2}{dt} = \beta_2(U_1 + U_2 + U_3)S_2 + \mu_2 U_2 - qU_2 \tag{6.45}$$

$$\frac{dR_2}{dt} = qU_2 + \mu_2 R_2 \tag{6.46}$$

$$\frac{dS_3}{dt} = \Lambda_3 + \mu_3 S_3 - \beta_3(U_1 + U_2 + U_3)S_3 \tag{6.47}$$

$$\frac{dU_3}{dt} = \beta_3(U_1 + U_2 + U_3)S_3 + \mu_3 U_3 - qU_3 \tag{6.48}$$

$$\frac{dR_3}{dt} = qU_3 + \mu_3 R_3 \tag{6.49}$$

The assumption in WAIFW2 is that each group mixes with other users irrespective of age but at a rate which pertains to their own age group. The initiation parameter

β for each group so obtained is very different in structure to that obtained for WAIFW1.

The estimates for β_1, β_2 and β_3 obtained in Berkeley Madonna for the age groups 15-24, 25-34, 35-44 are 2.4596×10^{-7}, 1.8204×10^{-7} and 7.289×10^{-8} respectively. The ratio of β_1 to β_2 to β_3 according to this model is approximately 7:5:2. This ratio seems more intuitively plausible. The incidence figures estimated in the model from 2002 to 2006 can be seen in table 6.6.

Table 6.6: Incidence estimates for cocaine use in 2002 - 2006 using WAIFW2 model

	2002	**2003**	**2004**	**2005**	**2006**
15-24	4323	4851	5461	6164	6972
25-34	3124	3673	4329	5110	6041
35-44	1150	1335	1557	1820	2133

6.3.3 β values for WAIFW3

Equations (6.1) to (6.9) for the WAIFW3 model become

$$\frac{dS_1}{dt} = \Lambda_1 + \mu_1 S_1 - (\beta_1 U_1 + \beta_2 U_2) S_1 \tag{6.50}$$

$$\frac{dU_1}{dt} = (\beta_1 U_1 + \beta_2 U_2) S_1 + \mu_1 U_1 - q U_1 \tag{6.51}$$

$$\frac{dR_1}{dt} = q U_1 + \mu_1 R_1 \tag{6.52}$$

$$\frac{dS_2}{dt} = \Lambda_2 + \mu_2 S_2 - (\beta_2 U_2 + \beta_3 U_3) S_2 \tag{6.53}$$

$$\frac{dU_2}{dt} = (\beta_2 U_2 + \beta_3 U_3) S_2 + \mu_2 U_2 - q U_2 \tag{6.54}$$

$$\frac{dR_2}{dt} = q U_2 + \mu_2 R_2 \tag{6.55}$$

$$\frac{dS_3}{dt} = \Lambda_3 + \mu_3 S_3 - \beta_3 U_3 S_3 \tag{6.56}$$

$$\frac{dU_3}{dt} = \beta_3 U_3 S_3 + \mu_3 U_3 - q U_3 \tag{6.57}$$

$$\frac{dR_3}{dt} = qU_3 + \mu_3 R_3 \tag{6.58}$$

The assumption of this model is that each age group will initiate their own age group and the age group immediately younger than them. So, for example, users in the age group 15-25 will not initiate those who are 25-34 or 35-44 and those in the oldest age group will not initiate those in the 15-24 age group.

The estimates for β_1, β_2 and β_3 are 2.5297×10^{-7} , 3.3212×10^{-7} and 6.2161×10^{-7} respectively. The ratio of the βs to each other is approximately 10:13:25. The incidence figures emerging in this model are presented in table 6.7 but the initiation rate for the 35-44 seems to be an overestimate.

Table 6.7: Incidence estimates for cocaine use in 2002 - 2006 using WAIFW3 model

	2002	**2003**	**2004**	**2005**	**2006**
15-24	4398	4854	5425	6144	7051
25-34	2691	3337	4179	5279	6721
35-44	980	1209	1504	1885	2380

6.3.4 β values for WAIFW4

Finally the model is modified in order to calculate the β values for the WAIFW4 model of mixing. We have,

$$\frac{dS_1}{dt} = \Lambda_1 + \mu_1 S_1 - (\beta_1 U_1 + \beta_2 U_2) S_1 \tag{6.59}$$

$$\frac{dU_1}{dt} = (\beta_1 U_1 + \beta_2 U_2) S_1 + \mu_1 U_1 - qU_1 \tag{6.60}$$

$$\frac{dR_1}{dt} = qU_1 + \mu_1 R_1 \tag{6.61}$$

$$\frac{dS_2}{dt} = \Lambda_2 + \mu_2 S_2 - (\beta_3 U_1 + \beta_2 U_2) S_2 \tag{6.62}$$

$$\frac{dU_2}{dt} = (\beta_3 U_1 + \beta_2 U_2) S_2 + \mu_2 U_2 - qU_2 \tag{6.63}$$

$$\frac{dR_2}{dt} = qU_2 + \mu_2 R_2 \tag{6.64}$$

$$\frac{dS_3}{dt} = \Lambda_3 + \mu_3 S_3 - (\beta_3 U_2 + \beta_3 U_3) S_3 \tag{6.65}$$

$$\frac{dU_3}{dt} = (\beta_3 U_2 + \beta_3 U_3) S_3 + \mu_3 U_3 - q U_3 \tag{6.66}$$

$$\frac{dR_3}{dt} = q U_3 + \mu_3 R_3 \tag{6.67}$$

This model assumes that the 15-25 age group are initiated into cocaine use by their own age group at their own initiator rate and by the older 25-34 at the 25-34 initiator rate. The 25-34 age group are initiated by the 15-24 at the 35-44 initiator rate and by their own group at their own initiator rate. The oldest age group are initiated by mixing with their own group and the 25-34 at the 35-44 initiator rate. The estimates for β_1, β_2 and β_3 are 2.7819×10^{-7}, 2.818×10^{-7} and 1.605×10^{-7} respectively. The ratio of the βs is approximately 2:2:1. The incidence from 2002 to 2006 for this model can be found in table 6.8.

Table 6.8: Incidence estimates for cocaine use 2002 - 2006 using WAIFW4 model

	2002	**2003**	**2004**	**2005**	**2006**
15-24	4419	4908	5475	6132	6891
25-34	3072	3640	4319	5129	6098
35-44	974	1221	1521	1885	2328

6.3.5 Conclusion

The four models were verified by ensuring that, for each age group, the point prevalence figure was arrived at in 2006: this figure is called the expected value in table 6.4. The rate of initiation in all but the WAIFW2 model is counter intuitive. One would not expect the 35-44 year old age group to have the highest rate of initiation. It is difficult to predict whether the 15-24 age group or the 25-34 would have the highest initiation rate. Although the younger age group are the most likely to experiment, cocaine is an expensive drug and the older age group may be more likely to be able to afford it. The conclusion has to be reached that the model is flawed.

In particular the WAIFW3 model, which intuitively seems plausible, produces very unlikely initiation rates. The most likely flaw is the lack of movement in the model from one age group to the next due to ageing. This flaw is tackled in the following section. It is interesting to compare the incidence estimates produced by the four WAIFW models in a line plot for each age group provided in figure 6.4. It is clear from this plot that for 15-24 year olds, the incidence predictions for all but the WAIFW1 model are very similar. For the 25-34 year olds, WAIFW1 and WAIFW3 are producing similar estimates as are WAIFW2 and WAIFW4. For the 35-44 age group the incidence rates estimates for WAIFW1, WAIFW3 and WAIFW4 are very similar with WAIFW3 starting from a higher base in 2002 but producing a lower estimate in 2006. Even though the four WAIFW models have flaws we run the models and obtain the model predictions of user number for the years 2007, 2008, 2009 and 2010. These model predictions are shown in table 6.9. Graphical displays are presented in figure 6.5 illustrating the prevalence trends for the three age groups. Graphical displays are presented in figure 6.5 which illustrate the prevalence trends

Table 6.9: Prevalence estimates 2007 - 2010 using WAIFW models

		2007	**2008**	**2009**	**2010**
15-24	**WAIFW1**	25874	27741	29590	31357
	WAIFW2	26659	29726	33232	37124
	WAIFW3	26825	30303	34595	39771
	WAIFW4	26563	29501	32853	36569
25-34	**WAIFW1**	21635	27267	34469	43383
	WAIFW2	20752	24707	29232	34255
	WAIFW3	21725	27397	34627	43576
	WAIFW4	20828	24901	29593	34830
35-44	**WAIFW1**	7763	9743	12284	15507
	WAIFW2	7416	8793	10380	12182
	WAIFW3	7763	9743	12284	15507
	WAIFW4	7670	9438	11572	14088

for the three age groups.

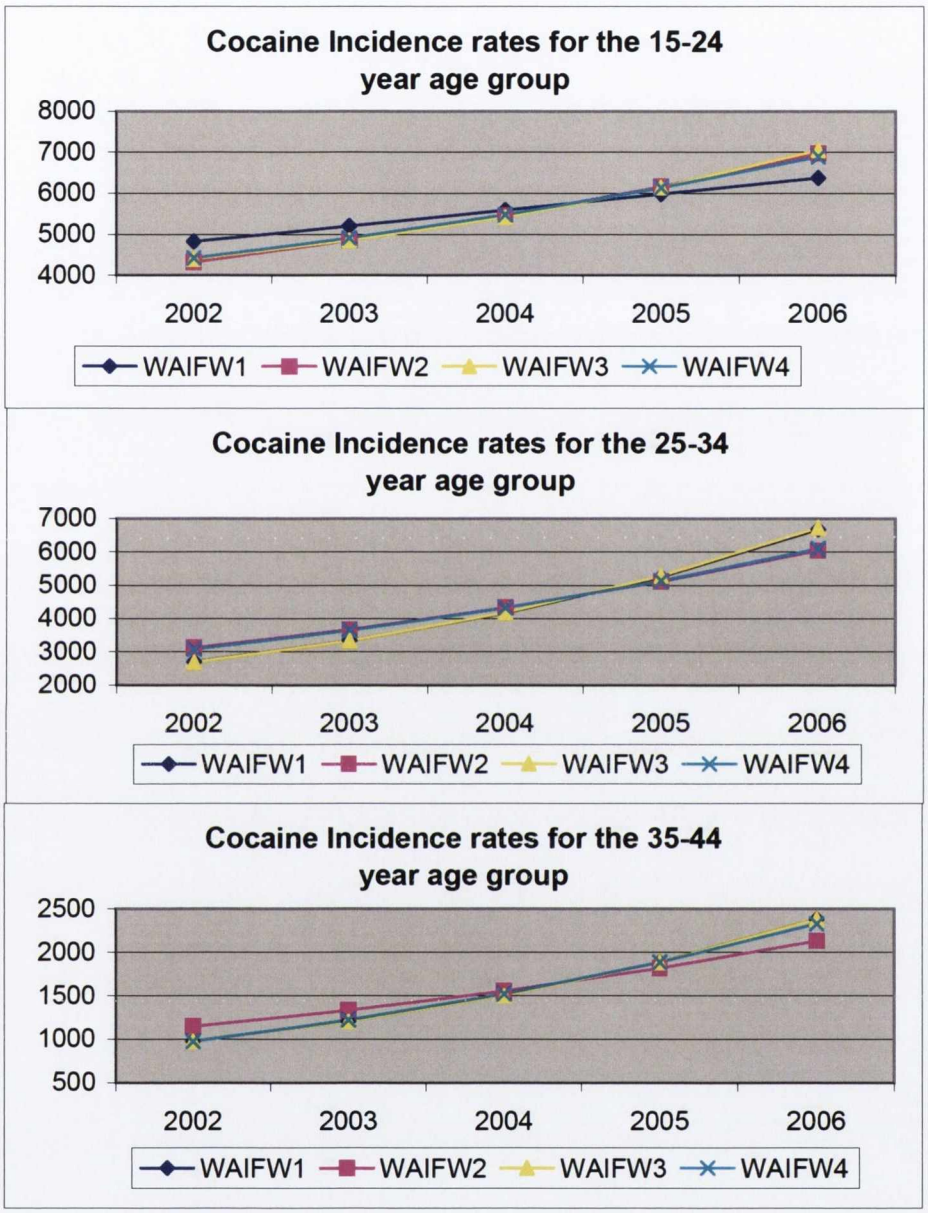

Figure 6.4: Incidence estimates from 2002 and 2006 for each WAIFW matrix shown per age group

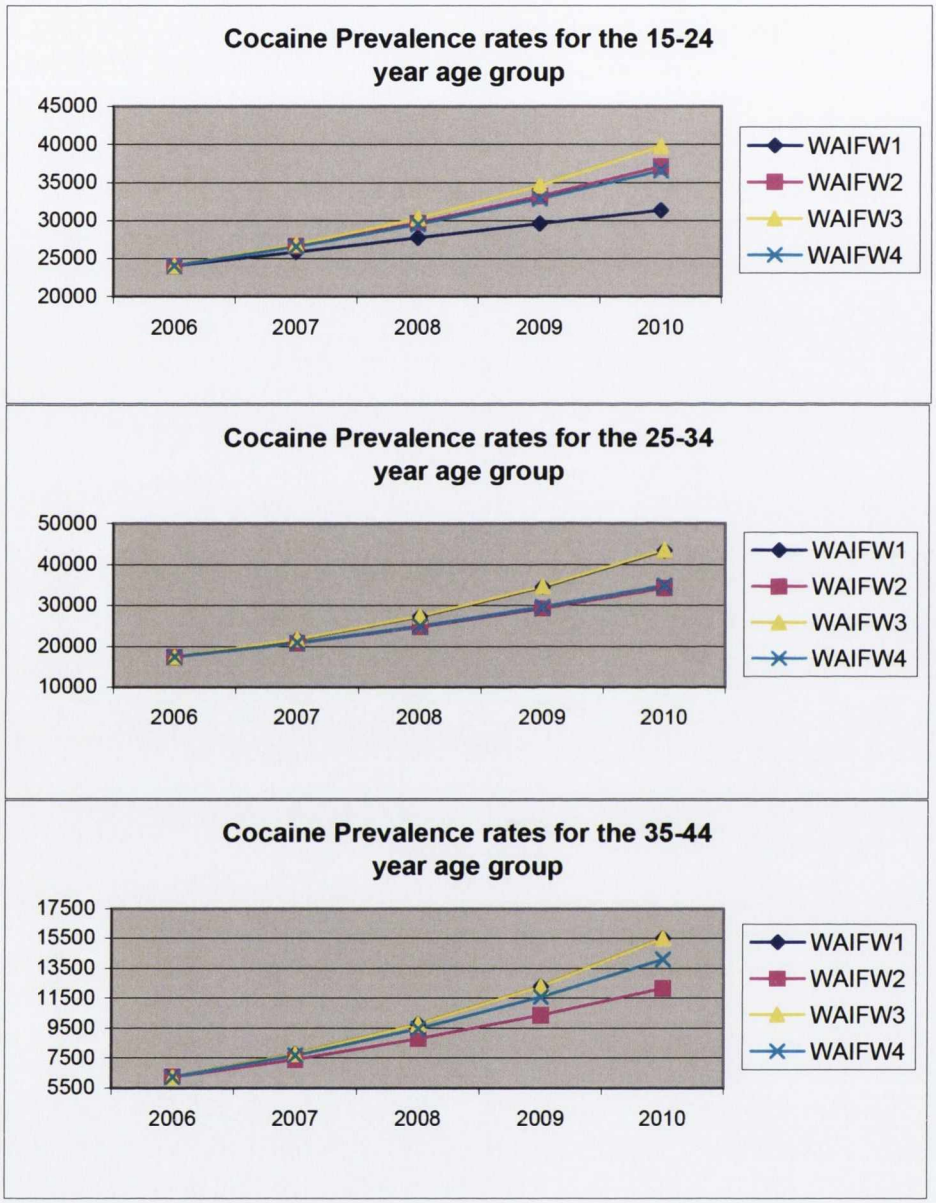

Figure 6.5: Prevalence trends 2007 and 2010 for each WAIFW matrix shown per age group

6.4 Improving the Model with an Ageing Parameter

The WAIFW models presented in section 6.3 are limited as they do not include the natural ageing process. In other words there is no movement of any group (susceptibles, users or recovereds) from 15 to 24 year olds to the next group of 25 to 34 year olds or from 25 to 34 year olds to the next group of 35 to 44 year olds. The problem with this simplified approach is that although it is easier to calculate the parameters, there is a danger that the rate of initiation would be over estimated or underestimated as the user numbers in the two older age groups do not receive the input of users who through the natural ageing process should have entered their age group.

We define a new parameter δ with the value of 0.1, which allows an average of one tenth of the age cohort to move to the next age group each year. Introduction of this parameter into the model means however, that the parameter Λ, the rate of immigration and μ, the rate of population change has to be reevaluated for each age group. There is also a problem created for the 15-24 year old group as there is no compartment from which they receive an influx of new population. The solution chosen, which has been used in many models, is to introduce a constant input to susceptibles and users in the age group. Clearly there was no need for such input to the recovereds. These 2 constant terms called C_S and C_U respectively are based on the 2002 CSO figures of 14-15 year olds. The constant terms are found by dividing this number according to the 2002 NACD point prevalence rates into C_S and C_U. The improved model is shown only for the WAIFW1 matrix structure as the changes are the same for the other three models.

$$\frac{dS_1}{dt} = C_S + \Lambda_1 + \mu_1 S_1 - \beta_1 S_1 U_1 - \delta S_1 \tag{6.68}$$

$$\frac{dU_1}{dt} = C_U + \beta_1 S_1 U_1 + \mu_1 U_1 - qU_1 - \delta U_1 \tag{6.69}$$

$$\frac{dR_1}{dt} = qU_1 + \mu_1 R_1 - \delta R_1 \tag{6.70}$$

$$\frac{dS_2}{dt} = \delta S_1 + \Lambda_2 + \mu_2 S_2 - \beta_2 S_2 U_2 - \delta S_2 \tag{6.71}$$

$$\frac{dU_2}{dt} = \delta U_1 + \beta_2 S_2 U_2 + \mu_2 U_2 - qU_2 - \delta U_2 \tag{6.72}$$

$$\frac{dR_2}{dt} = \delta R_1 + qU_2 + \mu_1 R_2 - \delta R_2 \tag{6.73}$$

$$\frac{dS_3}{dt} = \delta S_2 + \Lambda_3 + \mu_3 S_3 - \beta_3 S_3 U_3 - \delta S_3 \tag{6.74}$$

$$\frac{dU_3}{dt} = \delta U_2 + \beta_3 S_3 U_3 + \mu_3 U_3 - qU_3 - \delta U_3 \tag{6.75}$$

$$\frac{dR_3}{dt} = \delta R_2 + qU_3 + \mu_3 R_3 - \delta R_3 \tag{6.76}$$

The parameters are calculated for each WAIFW matrix structure as in sections 6.3.1 up to 6.3.4. The model change appears to produce more credible β values which are presented, along with ratio estimates, in table 6.10. Incidence rates for each WAIFW structure are to be found in table 6.11, prevalence rates in table 6.12. The results are also shown graphically in figures 6.6 and 6.7.

Table 6.10: Beta and approximate ratio values estimated for each WAIFW model where ageing is included

	β_1	β_2	β_3
WAIFW1	4.73E-07	4.39E-07	3.35E-07
Ratio	12:11:8		
WAIFW2	2.59E-07	1.49E-07	3.86E-08
Ratio	5:3:1		
WAIFW3	2.78E-07	3.26E-07	3.36E-07
Ratio	9:11:11		
WAIFW4	2.89E-07	3.03-07	8.49E-08
Ratio	29:30:8		

Table 6.11: Incidence estimates 2002 - 2006 using WAIFW models with Ageing

	2002	2003	2004	2005	2006
WAIFW1					
15-24	5115	5568	5999	6405	6786
25-34	2147	2760	3483	4326	5296
35-44	528	655	818	1025	1281
WAIFW2					
15-24	4551	5166	5836	6566	7357
25-34	2559	3065	3629	4252	4938
35-44	608	713	834	971	1126
WAIFW3					
15-24	4638	5196	5826	6543	7359
25-34	2169	2776	3501	4363	5384
35-44	530	657	821	1027	1284
WAIFW4					
15-24	4646	5223	5853	6539	7287
25-34	2381	2951	3588	4296	5080
35-44	516	655	818	1008	1228

Table 6.12: Prevalence estimates 2006 - 2010 using Ageing WAIFW models

		2006	2007	2008	2009	2010
15-24	**WAIFW1**	24044	25570	27017	28338	29490
	WAIFW2	24044	26286	28572	31383	34093
	WAIFW3	24044	26319	28905	31781	34883
	WAIFW4	24044	26205	28571	31087	33667
25-34	**WAIFW1**	17338	20544	24144	28007	31936
	WAIFW2	17338	20106	23112	26259	29424
	WAIFW3	17388	20696	24622	29050	33849
	WAIFW4	17388	20277	23510	26929	30382
35-44	**WAIFW1**	6234	7601	9232	11136	13287
	WAIFW2	6234	7386	8665	10062	11547
	WAIFW3	6234	7611	9274	11251	13540
	WAIFW4	6234	7517	8987	10634	12418

6.4.1 Discussion and Comparison of the results

Two groups of models have been developed which examine the effects of different social mixing structures on the incidence and prevalence of cocaine use, and will be

referred to as the simple and ageing models. Both model groups estimate increasing incidence for each age group and both share characteristics in common. For example, focusing on the 15-24 year olds, both the simple and ageing models show strong similarity between WAIFW2, WAIFW3 and WAIFW4 and WAIFW1 giving initially higher incidence in 2002 but lower incidence in 2006. Similarly for the 25-34 age group both types show similar predictions for the WAIFW1 and WAIFW3 structures and for the WAIFW2 and WAIFW4 structures. For the 35-44 age group, both model groups show only WAIFW2 as differing in form from the other three stuctures. So the form or shape of the incidence growth for both is similar.

However the estimates of incidence are quite different for the two older age groups. For the 15-24 year olds the range of incidence estimates in 2006 for example for the simple model is between 6,375 and 7,051 for the simple models and between 6,786 and 7,359 for the ageing models. But for the 25-34 age-group, the simple model estimates incidence in 2006 as ranging from 6,041 to 6,721 whereas the range of incidence estimates in 2006 in the ageing model are between 4,938 and 5,384. The difference in estimates is even larger for the 35-44 age group ranging from 2,133 to 2,380 for the simple model and from 1,126 to 1,284 for the ageing model. It seems reasonable to assume that users aging and moving into the 35-44 group who have not yet ceased use form an important input to this group.

As the β estimates are more credible for the ageing model, we conclude that this model is more realistic. The model predicts that the prevalence of cocaine use among 15-24 year olds in 2010 will range between 29,490 and 34,883, that the prevalence of cocaine use among 25-34 year olds in 2010 will range between 29,424 and 33,849 and that the prevalence of cocaine use among 35-44 year olds in 2010 will range between 11,547 and 13,540. The estimates for the 15-24 and 25-34 are very similar and those for the 35-44 year olds are approximately 40% of these. The low rate of initiation for this age group would suggest there is a group of users who have used cocaine for a considerable time.

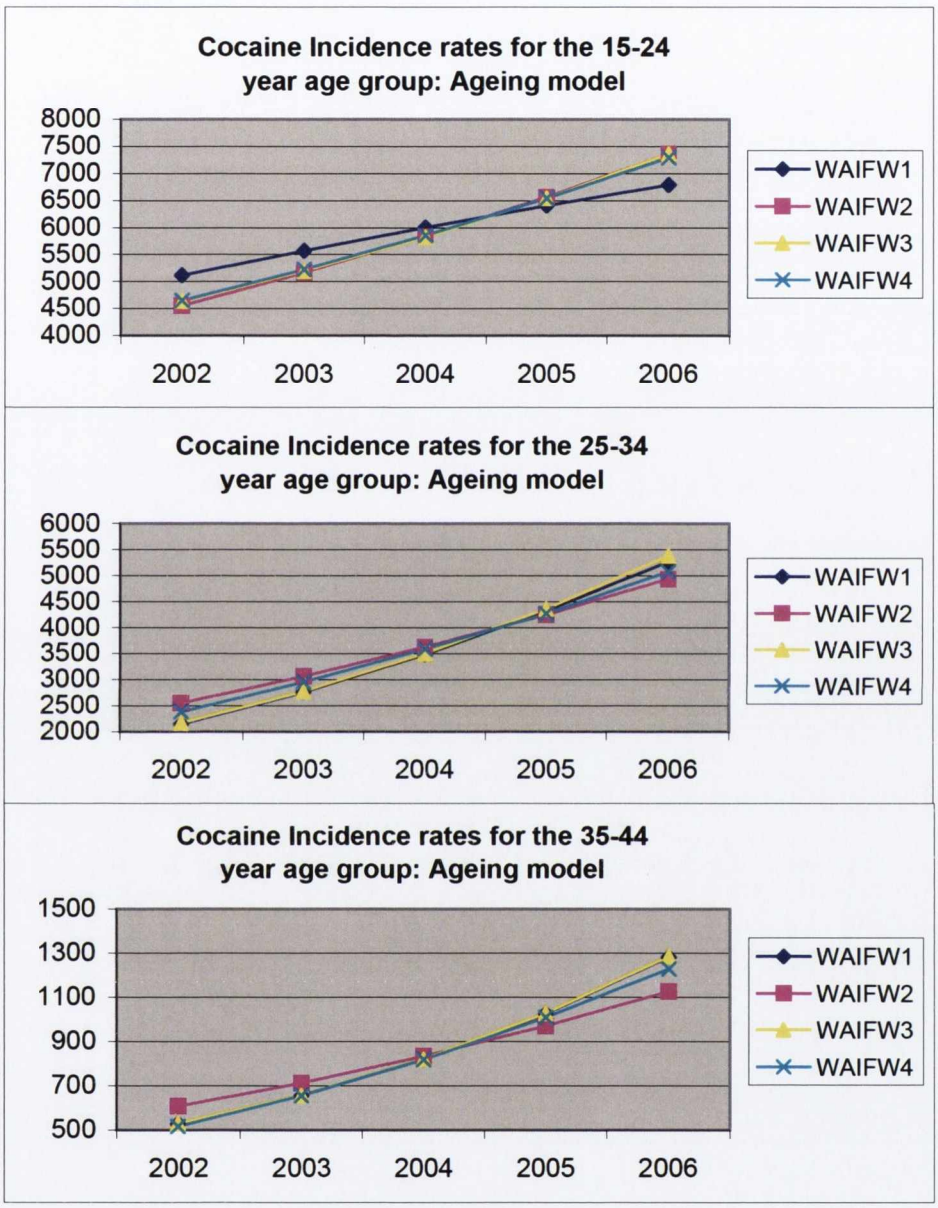

Figure 6.6: Incidence data 2002 and 2006 with an aging component to model

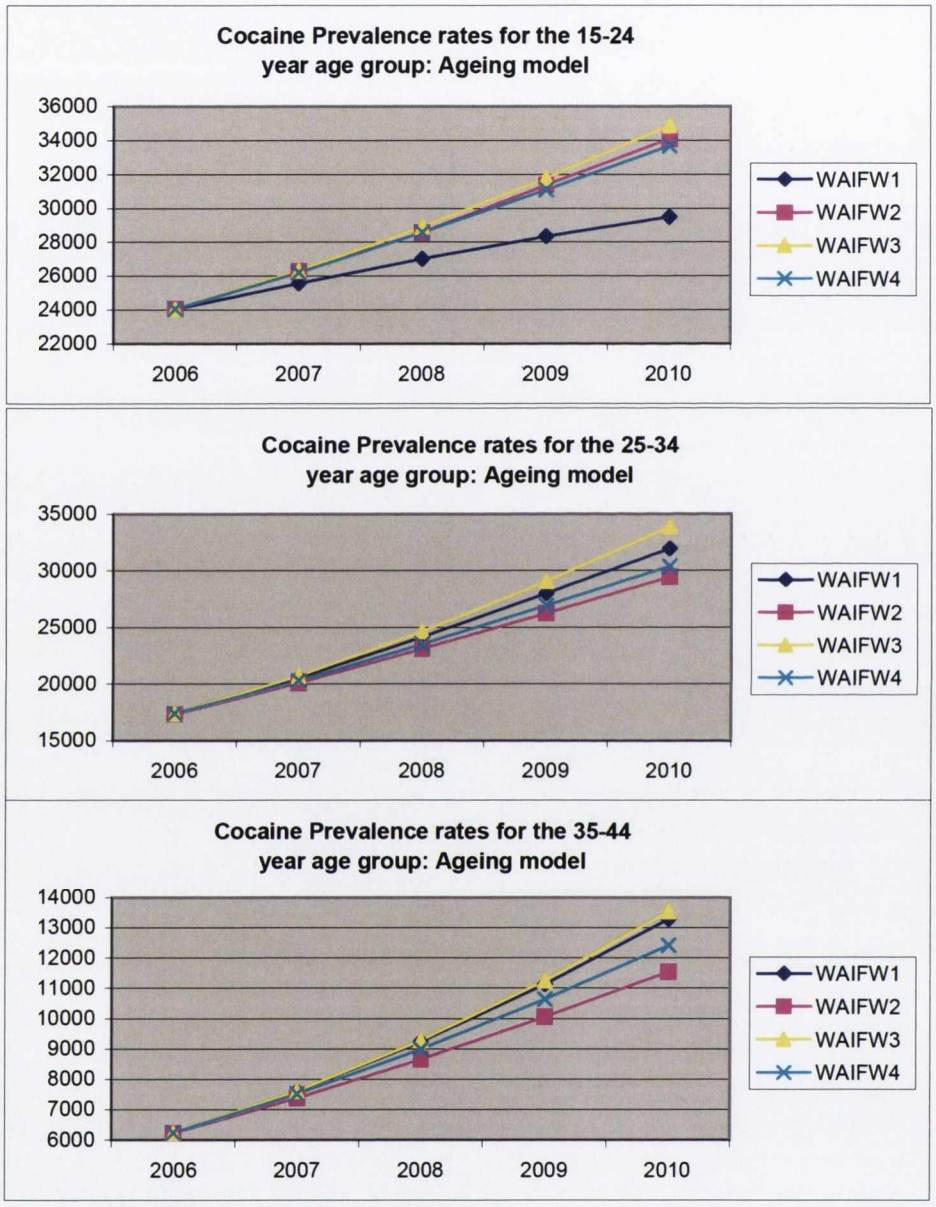

Figure 6.7: Incidence data 2002 and 2006 with an aging component to model

6.5 Chapter Summary

This chapter developed a model which aimed to investigate the levels of social contacts, differentiated by age group, which leads to initiation into cocaine use. It again used last year prevalence rates for the republic of Ireland obtained from two NACD studies (NACD 2006, NACD 2008) and applied these rates to CSO census figures obtained during the same years (CS0 2002, CSO 2006) enabling a model to be built of the dynamics of the change in cocaine use during that period. A dynamic mathematical model was developed to simulate preventative measures and this simple model suggested that effective treatment is more efficient in reducing user numbers than preventative measures.It used the concept of the WAIFW matrix which was proposed by Anderson & May (1991). Four models of social mixing were modelled using structures labelled WAIFW1, WAIFW2, WAIFW3 and WAIFW4. These structures were based on Anderson & May's suggestions and also on what seemed epidemiologically reasonable. Anderson & May's suggested method to calculate the different matrix parameters was followed but the prevalence estimates were considerably higher than expected. An alternative approach was sought. The models were implemented in Berkeley Madonna and validated using the expected User levels in 2006. The β values so derived were however epidemiologically unlikely, particularly the β value representing initiation into cocaine use by the 35-44 age group. This parameter had the highest value compared to the other two age groups in three of the four model structures. An ageing parameter was therefore built into the model, which included the natural flow of the populations from 15-24 to 25-34 and from from 25-34 to 35-44. The model was validated as before but now the β parameters seemed more intuitively plausible. The models produced incidence estimates for each social mixing model. Prevalence predictions were also reported for each model up to 2010. These estimates, were based on the growth in user numbers between 2002 and 2006. Cocaine use may have decreased due to external forces which would imply that these estimates are over-estimates. This can only be judged when the

results of the next NACD Household Survey become available.

Chapter 7

Conclusions and Further Work

7.1 Introduction

This thesis presents recognised statistical methodologies which can be used in the estimation of the prevalence of cocaine use. It then applies these methods, for the first time, in an Irish setting. It also develops and applies dynamical mathematical modelling techniques to the problem of cocaine use and produces extensions to existing models. It applies these newly extended models to cocaine use throughout Ireland. In this chapter the aims and objectives of the thesis are reviewed and results achieved and unaccomplished are described along with suggestions for further research suggested by the study. The objectives of this research were:

1. To use the truncated Poisson, capture-recapture and multiplier methods, recommended by the EMCDDA and the UNODC, to estimate the prevalence of cocaine use among opiate users in treatment in an area of Dublin.

2. Investigate by simulation the effectiveness of interventions aimed at reducing cocaine user numbers in the general population by comparing the relative reduction in cocaine user numbers which could be achieved by pre initiation strategies (preventative measures) as opposed to post initiation measures seeking to reduce the duration of cocaine use through treatment or public

information measures

3. Investigate age differentiated initiation using WAIFW matrices and in so doing produce age differentiated prevalence and incidence estimates where incidence is defined to be the number of new cocaine users in a given year

7.2 Strengths and Limitations

The first objective was addressed within chapter 4. Two methods for point prevalence estimation were critically assessed and applied for the first time to cocaine use in an Irish context. In particular the capture recapture method as proposed by Bishop *et al.* (1975) was examined in detail and estimates obtained were compared with those obtained in SPSS. Whereas Bishop *et al.* used a purely probabilistic approach, SPSS fitted an independent loglinear model to the data. According to Agresti (1996), the SPSS approach to solving the problem of redundancy in such a model is to set the first or last level of each factor equal to zero. The estimates produced using the Bishop *et al.* method and the SPSS method were in fact identical. This was unexpected given the difference in approaches but an in depth analysis of the SPSS program was deemed to be beyond the bounds of the present study belonging more to the realm of computer applications.

Nevertheless, the first Irish estimates of the total number of clients in treatment who used cocaine in the years 2006-2010, both detected and undetected, were produced using both the truncated Poisson method and the single source 3-sample capture-recapture method as cross validation. For the years 2006-2010 the number of unique individuals in treatment in whose urinalysis samples cocaine was detected were 1092, 1057, 1097, 1027 and 1044 respectively. Estimates obtained using the truncated Poisson and capture recapture methods ranged from 1,306 to 1,860 for 2006, from 1,293 to 1,930 for 2007, from 1,301 to 1,764 for 2008, from 1,339 to 1,457 for 2009 and from 1,354 to 2,527 in 2010. Assuming that the numbers of cocaine

users in treatment reflect general usage in society at large, these numbers were taken as benchmark figures to which the multiplier could be applied. This multiplier was estimated using the Zelterman, Chao and capture-recapture estimates. The estimates for cocaine use in the general population in the NAHB ranged from 7,126 to 8,335 in 2007, from 7,894 to 8,082 in 2008, from 6,723 to 7,888 in 2009 and from 7,598 to 10,587 in 2010. Rates from the 2006 all-Ireland general population drug prevalence survey were applied to the regional population figure allowing for the first time an estimate of the ratio of cocaine user in treatment to general population user. The ratio of those in treatment using cocaine to those in the general population using cocaine ranges between 1:6 to 2:9.

A limitation of the prevalence estimates produced were that they were confined to those in treatment. It was originally intended to use capture-recapture with at least two data sources to estimate prevalence in the wider population but the permission to use their client's initials and date of birth which was necessary to facilitate this prevalence estimation was denied by front-line staff . Though they acted out of concern for client confidentiality, these concerns could have been addressed. Gannon and Hay (2012) pointed out that in order for a prevalence study to be successful, timely and cost effective issues surrounding research governance need to be addressed. Previous capture-recapture studies carried out in Ireland have required multiple ethics applications to access HIPE data from a number of hospitals around the country. This has led to significant delays and a waste of time and resources for both the researchers and those assessing the study. They recommend that prior to any future prevalence work that the NACD or other government backed body should make 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. As a lone researcher without official NACD backing it was perhaps to be expected that access to the necessary data would not be given.

The second objective was the subject of chapter 5. It used last year prevalence rates for the republic of Ireland obtained from two NACD studies (NACD 2006, NACD 2008) and applied these rates to CSO census figures obtained during the same years (CS0 2002, CSO 2006) enabling a model to be built of the dynamics of the change in cocaine use during that period. A dynamic mathematical model was developed to simulate preventative measures and this simple model suggested that effective treatment is more efficient in reducing user numbers than preventative measures. The steps in model development as outlined in figure 3.1 show that following model validation, the modeller often finds that changes to the model are necessary and the cycle begins afresh. This occurred in this model as a result of submission of a paper about the model to a peer reviewed journal. The original model used only the Everingham & Rydell quit rate after on average 7.5 years but the one of the journal's reviewers alerted us to the Caulkins *et al.* paper which had a revised quit rate of 5.5 years. This resulted in the model being redeveloped and as a consequence the interesting result was found that the average time that people used cocaine was in fact the most important parameter in the model and that interventions which reduced this time were the most significant.

In chapter 6, age dependence was introduced and a variety of models were explored in an attempt to examine the age differentiated mixing patterns leading to drug initiation in Ireland. It made use of the same data as chapter 5. Models were developed and then refined. The model which included a natural ageing movement between the three examined age groups appeared to produce the most epidemiologically feasible results. Each model predicted cocaine prevalence in 2010 to be in the region of 30,000 for both the 15-24 year olds and 25-34 year olds and to be in the region of 12,000 for the 35-44 year olds. The models indicated that incidence rates in 2006 for the three age groups were approximately 7,00, 5,000 and 1,200 respectively. The limitation of the developed models was that the different hypothesised mixing structures failed to show any sizable differences in the either the predicted

incidence or prevalence estimates. If we exclude the WAIFW1 structure, which assumes no inter age group mixing, the difference in the prevalence estimates between lowest and highest for the 15-24 age group was 3.6%, for the 25-34 age group the difference was 15% and for the 35-44 age group the difference was 17%. Differences in incidence estimates varied by only 1.4% for the 15-24 age group, by 9% for the 25-34 age group and by 14% for the 35-44 age group. Although these models have been developed as fully as possible in line with the model development steps shown in figure 3.1, the major weakness is the lack of available data to verify which model most accurately predicted the prevalence in 2010, which would in turn have led to conclusions as to which WAIFW structure most accurately modeled inter age group initiation pattens. In 2010/2011 the NACD commissioned the third all-Ireland general population drug prevalence survey. The first results have been published (Long 2012) but as more results are released, it will be possible in the future to validate the model estimates against the more recent survey results.

7.3 Discussion and Further Work

This thesis has examined cocaine prevalence and incidence at both a local and national level. We have shown that urinalysis results can be analysed to produce results which will be of interest to health care professional and policy maker alike. We have shown that in the local area of study at least, there is no evidence to suggest that cocaine use by those in treatment for opiate use has waned in any appreciable way and on the contrary, estimated prevalence was generally higher in 2010 than in 2009. Estimates of last year prevalence derived by truncated Poisson methods and validated by a single source capture-recapture approach indicate that the point prevalence for each year lies between 1200 and 2000. These estimates assist in quantification of the extent of the cocaine problem and provide an evidence base for both policy and

practice. In the light of the study by Cox & Comiskey (2011), which revealed that those who used cocaine at intake in the ROSIE study were more likely to use cocaine at 1-year follow-up, to commit crime, and to be homeless, these results should be of interest as service providers attempt to adapt to the changing needs of their clients. On the other hand, the methods used and the models developed should be of interest in the broader international context of worldwide drug use. In the course of the study the prevalence of other substances was also noted. It was observed that the prevalence of benzodiazepines in urinalysis appeared to increase among patients over time. This observation could form the starting point for a future study.

Chapters 5 and 6 focused on national cocaine estimates. The models developed in both these chapters could be improved when the NACD general population drug prevalence survey for 2010/2011 become available. A further improvement would be achieved if an Irish cocaine quit rate could be calculated which could replace the United States rates, used in the current study. Both models could also be expanded to include a relapse parameter as has been suggested in the White & Comiskey (2007) model of the drug using career. These suggestions are dependent on the availability of data which ensures their accurate estimation.

In the course of the present study, a need was identified to study the diffusion of drug use in Ireland in general and of cocaine use in particular. A pilot diffusion study was undertaken by the EMCDDA (1999b). For the present study, a body of work was carried out in relation to this but it was decided not to include this work in the final thesis. Nevertheless this gap in the knowledge has been identified and could form the basis for future post-doctoral work.

Appendix A

Glossary

The terms used are defined in terms of cocaine use.

Prevalence

Prevalence can be defined as the number of those in the population who have ever used cocaine or as is more common in Epidemiology as the proportion of the population who have ever used cocaine. A time distinction is made by both the EMCDDA and the NACD where:

- Lifetime prevalence refers to the number or proportion who have ever having used cocaine in their life
- Last year prevalence reports the number or proportion who used cocaine in the year preceding data collection
- Last month prevalence reports the number or proportion who used cocaine in the month preceding data collection

Incidence

Incidence is defined as the number or proportion in the population who used cocaine for the first time in a defined time period preceding data collection. In the present study the term is used to denote use of cocaine in the year prior to data collection for the first time.

Point Estimate

A point estimate of a population parameter is a single value of that parameter inferred from a sample statistic. Following the NACD household surveys in 2002/2003 and 2006/2007, point estimates of the lifetime, last year and last month proportions using cocaine were inferred from the sample data collected along with 95% confidence intervals for these proportions.

Appendix B

Berkeley Madonna

Berkeley Madonna is a program that numerically solves systems of ordinary differential equations (ODEs) and difference equations. It was originally developed in 1996 by Robert Macey, George Oster and Tim Zahnley of the University of California, Department of Molecular and Cellular Biology. The program is used mainly for Dynamic Simulation such as SIR virus problems, birth and death population problems, chemical reaction kinetics or any problem involving a rate based flow over time. It uses numerical methods such as Euler, Runga Kutta 2 and 4, and the stiff and auto methods to solve the differential equations. As the equations are solved using a specified time step size, values of both variables and parameters can be tracked over time as well as viewed in the graphical display option. The program also allows the estimation of parameters based on text input through its curve fit menu. The program allows a sensitivity analysis to be performed through the available option which allows the plotting of a family of curves. This facilitates the investigation of dynamic changes caused by small variation in specific parameter values in the model.

Appendix C

Ethical Approval DAIS

THE UNIVERSITY OF DUBLIN
TRINITY COLLEGE
SCHOOL OF MEDICINE
FACULTY OF HEALTH SCIENCES

Trinity College, Dublin 2, Ireland
Tel: +353 1 896 1476
Fax: +353 1 671 3956
email: medicine@tcd.ie

Professor Dermot Kelleher, MD, FRCPI, FRCP, F Med Sci
Head of School of Medicine
Vice Provost for Medical Affairs

Ms Fedelma McNamara
School Administrator
email: medschadmin@tcd.ie

Prof Catherine Comiskey,
Director of Research,
School of Nursing and Midwifery,
Trinity College Dublin,
D'Olier St.
D 2

Tuesday, 20 July 2010

Study: Statistical analysis of secondary historical anonymised data from the Drugs AIDS Information System (DAIS)

Dear Applicant (s),

Further to a meeting of the Faculty of Health Sciences Ethics Committee held in June 2010, we are pleased to inform you that the above project has been approved without further audit.

Yours sincerely

Prof. Orla Sheils
Chairperson
Faculty of Health Sciences Ethics Committee

References

Agresti A. (1996) *An Introduction to Categorical Data Analysis.* John Wiley & Sons, New York.

Aitkin M.A. (1989) *Statistical Modelling in GLIM.* Oxford University Press, Oxford.

Akaike H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6), 716-723.

Almeder C., Caulkins J.P., Feichtinger G. & Tragler G. (2001) Age-Specific Multi-State Drug Initiation Models: Insights from Considering Heterogeneity. *Bulletin on Narcotics* **53**, 105-118.

Anderson R.M. & May R.M. (1991) *Infectious Diseases of Humans.* Oxford University Press, Oxford.

Archibald C.P., Jayaraman G.C., Major C., Patrick D.M., Houston S.M. & Sutherland D. (2001) Estimating the size of hard-to-reach populations: a novel method using HIV testing data compared to other methods. *AIDS* **15** S41-8.

Babbad H.R., Nokes D.J., Gay N.J, Miller E., Morgan-Capner P & Anderson R.M. (1995) Predicting the impact of measles vaccination in England and Wales: model validation and analysis of policy options. *Epidemiology and Infection* **114**, 319-344.

Bailey N.T.J. (1957) *The Mathematical Theory of Epidemics.* Hafner Publishing Company, New York.

Bailey N.T.J. (1975) *The Mathematical Theory of Infectious Disease of Humans and its Applications.* Griffin, London.

Barnes B. & Fulford G.R. (2002) *Mathematical Modelling with Case Studies. A Differential Equation Approach using Maple.* Taylor and Francis, London.

Barry S. & Lawlor E. (2005) The Treatment of Cocaine Addiction - with particular reference to the Irish Experience. Final report prepared for the Joint Oireachtas Committee on Arts, Sport, Tourism, Community, Rural & Gaeltacht Affairs. Retrieved from `http://www.oireachtas.ie/documents/committees29thdail/committeereports2005/Cocaine-Report.pdf` on 22 September 2011.

Bartlett M.S. (1957) Measles periodicity and community size. *Journal of the Royal Statistical Society* **120**, 48-70.

Bayer R. (1991) The Great Drug Policy Debate: What means this thing called Decriminalization? *The Millbank Quarterly* **69**(3), 341-363.

Behrens D.A. & Tragler G. (2001) The dynamic process of dynamical modelling: the cocaine epidemic in the United States of America. *Bulletin on Narcotics* **53**(1)&(2), 65-76.

Bellerose D., Carey A.M., Lyons S. & Long J. (2009) Trends in treated problem cocaine use in Ireland, 2002 to 2007. *Health Research Board Trends Series HRB Trends Series 6. Health Research Board, Dublin.* Retrieved from `http://www.hrb.ie/uploads/tx_hrbpublications/HRB_Trend_Series_6.pdf` on 20th July 2011.

Berkeley Madonna. (1996) *Berkeley Madonna Official Website.* Retrieved from `http://www.berkeleymadonna.com/` on 22 September 2011.

Berridge V. (1978) War Conditions and Narcotics Control: The Passing of Defence of the Realm Act Regulation 40B. *Journal of Social Policy* **7**(3), 285-304.

Best D., Day E. & Morgan B. (2006) Addiction careers and the natural history of change. Retrieved from `http://www.misbrugsnet.dk/fagligt/self_change/nta_addiction_careers_2006.pdf` on 20th July 2011.

Beynon C., Bellis M.A., Millar T., Meier P., Thomson R. & Jones K.M. (2001) Hidden need for drug treatment services: Measuring levels of problematic drug use in the North West of England. *Journal of Public Health Medicine* **23**, 286-291.

Bishop Y., Fienberg S. & Holland P. (1975) *Discrete Multivariate Analysis.* The MIT Press, Massachusetts.

Böhning D., Suppawattanabodee B. & Kusolvisitkul W. (2004) Estimating the Number of Drug Users in Bangkok 2001: A Capture-Recapture Approach Using Repeated Entries in One List. *European Journal of Epidemiology* **19**(12), 1075-1083.

Brisson M., Edmunds W.J., Gay N.J., Law B. & De Serres G. (2000) Modelling the impact of immunization on the epidemiology of varicella zoster virus. *Epidemiology and Infection* **125**, 651-669.

Brogan R. (2008) Battling a pandemic: A history of drugs in the United States. In West K. *Junior Drug Awareness: Crack and Cocaine.* Infobase Publishing, New York, pp. 6-11.

Brugal M & Domingo-Salvany A. (2004) Prevalence of problematic cocaine consumption in a city of Southern Europe, using capture-recapture with a single list. *Journal of Urban Health* **81**(3), 416-427.

Buster M.C.A., van Brussel G.H.A. & van den Brink W. (2001). Estimating the number of opiate users in Amsterdam by capture-recapture: The importance of case definition. *European Journal of Epidemiology* **17**, 935-942.

Butler S., & Mayock P. (2005) "An Irish solution to an Irish problem": Harm reduction and ambiguity in the drug policy of the Republic of Ireland. *International Journal of Drug Policy* **16**, 415-422.

Byrne R. (2001) Opiate-related deaths investigated by the Dublin city and county coroners in 1998, 1999 and 2000. Unpublished. *The Addiction Research Centre, Trinity College Dublin.*

Calkins R.F. & Aktan G.B. (2000) Estimation of heroin prevalence in Michigan using capture-recapture and heroin problem index methods. *Journal of Drug Issues* **30**, 187-204.

Cameron D. & Jones I.G. (1983) John Snow, the Broad Street Pump and Modern Epidemiology. *International Journal of Epidemiology* **12**, 393-396.

Cartmell L.W., Aufderheide A.C., Springfield A., Weems C. & Arriaza B. (1991) The frequency and antiquity of prehistoric coca-leaf-chewing practices in northern chile: radioimmunoassay of a cocaine metabolite in human-mummy hair. *Latin American Antiquity* **7**(3), 260-268.

Caulkins J., Behrens D.A., Knoll C., Tragler G. & Zuba D. (2004) Markov chain modeling of initiation and demand: The case of the US cocaine epidemic *Health Care Management Science* **7**, 319-329.

Central Intelligence Agency (2010) *The World Factbook.* Retrieved from `https://www.cia.gov/library/publications/the-world-factbook/fields/print_2086.html` on 11 August 2011.

Central Statistics Office (2002) *Persons, males and females classified by single year of age, 2002.* Retrieved from `http://beyond2020.cso.ie/Census/TableViewer/tableView.aspx?ReportId=1359` on 26 November 2010.

Central Statistics Office (2006) *Persons, males and females classified by single year of age, 2006.* Retrieved from `http://census.cso.ie/Census/TableViewer/tableView.aspx?ReportId=76969` on 26 November 2010.

Chao A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783-791.

Chernin E. (1988) Sir Ronald Ross, malaria, and the rewards of research. *Medical History* **32**(2), 119-141.

Choi Y. & Comiskey C.M. (2003) Methods for providing the first prevalence estimates of opiate use in Western Australia. *International Journal of Drug Policy* **14**, 297-305.

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

Comiskey C.M., Kelly P., Leckey Y., McCulloch L., ODuill B., Stapleton R.D. & E. White E. (2009) *The ROSIE Study Drug Treatment Outcomes in Ireland.* The Stationery Office, Dublin.

Comiskey C.M., Dempsey O. & Snel A. (2011) *The National Prevalence of Most at Risk Populations (MARP) for HIV/AIDS.* Unpublished Report, Serbia.

Connolly J. (2005) Arrest referral in the north inner city. *Drugnet Ireland* **14**, 13-14. Retrieved from `http://www.drugsandalcohol.ie/11176/1/Drugnet_14.pdf` on 22 September 2011.

Connolly J. (2008) CSO report: drug-crime statistics. *Drugnet Ireland Issue 26 13-16.* Retrieved from `http://www.drugsandalcohol.ie/12083/1/Drugnet_26.pdf` on 20 July 2011.

Connolly J., Foran S., Donovan A., Carew A. & Long J. (2008) Crack cocaine in the Dublin region: an evidence base for a Dublin crack cocaine strategy. HRB Research Series 6. Dublin: Health Research Board. Retrieved from `http://www.hrb.ie/uploads/tx_hrbpublications/HRB_Research_Series_6.pdf` on 20 July 2011.

Connolly J. (2010) Drug-related crime statistics. *Drugnet Ireland* **36** 23-24. Retrieved from `http://www.drugsandalcohol.ie/14707/1/Drugnet_Issue_36_final.pdf` on 20 July 2011.

Cormack R.M. (1992) Interval estimation for mark-recapture studies of closed populations. *Biometrics* **48**(2), 567-576.

Cox G. & Comiskey C.M. (2007) Baseline characteristics of patients attending treatment for opiate use in Ireland. *Drugs: Education, Prevention and Policy* **14**(3), 217-230.

Cox G. & Comiskey C.M. (2011) Does concurrent cocaine use compromise 1-Year treatment outcomes for opiate users? *Substance Use and Misuse* **46**(9), 1206-1216.

Crosscare. (2011) *Crosscare Drug and Alcohol Programme.* Retrieved from `http://www.drugs.ie` on 22 September 2011.

Darroch J. & Ratcliff D. (1980) A note on capture-recapture estimation. *Biometrics* **36**, 149-153.

Davies A.G., Cormack R.M. Richardson A.M. (1999). Estimation of injecting drug users in the City of Edinburgh, Scotland, and number infected with human immunodeficiency virus. *International Journal of Epidemiology* **28**, 117-121.

Davison A. (2003) *Statistical Models.* Cambridge University Press, Cambridge.

De Alarcon R. (1969) The spread of heroin abuse in a community. *Bulletin of Narcotics* **21** 17-21.

Del Rio Vilas V. & Böhning D. (2008) Application of one-list capture-recapture models to scrapie surveillance data in Great Britain. *Preventive Veterinary Medicine* **85**, 253-266.

Department of Community, Rural and Gaeltacht Affairs (2009) National Drugs Strategy (interim) 2009-2016 Retrieved from `http://www.drugsandalcohol.ie/12388/1/DCRGA_Strategy_2009-2016.pdf` on 11 August 2011.

Department of Health (2005) *Review of the Methadone Treatment Protocol* Retrieved from `http://www.dohc.ie/publications/methadone_review.html` on 10 September 2011.

Domingo-Salvany A. (1996) Estimating the prevalence of drug use using the capture-recapture method: an overview in *Estimating the Prevalence of Problem Drug Abuse in Europe.* EMCDDA Scientific Monograph Series. Retrieved from `http://www.emcdda.europa.eu/attachements.cfm/att_44760_EN_Monograph1.pdf` on 13 August 2011.

Domingo-Salvany A., Hartnoll R.L., Maguire A., Brugal M.T., Albertin P., & Cayla J. (1998) Analytical considerations in the use of capturerecapture to estimate prevalence: Case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *American Journal of Epidemiology* **148**, 732-740.

Doweiko H.E. (2006) *Concepts of Chemical Dependency*, 6^{th} edn. Thompson Brooks/Cole, Belmont, California.

European Monitoring Centre for Drugs and Drug Addiction (1999a) *Methodological Guidelines to Estimate the Prevalence of Problem Drug Use on the Local Level.* Retrieved from `http://www.emcdda.europa.eu/attachements.cfm/att_1422_EN_guidelines_local_pdu.pdf` on 11 December 2009.

European Monitoring Centre for Drugs and Drug Addiction (1999b) *Geographic spread and GIS (Geographic Information Systems).* Retrieved from `http://www.emcdda.europa.eu/html.cfm/index1374EN.html` on 22 September 2011.

European Monitoring Centre for Drugs and Drug Addiction (2001) *Modelling drug use: methods to quantify and understand hidden processes.* Re-

trieved from http://www.emcdda.europa.eu/attachements.cfm/att_80903_EN_Monograph6.pdf on 14 August 2011.

European Monitoring Centre for Drugs and Drug Addiction (2010) *Annual report on the state of the drugs problem in Europe.* Retrieved from http://www.emcdda.europa.eu/publications/annual-report/2010 on 15 September 2011.

European Monitoring Centre for Drugs and Drug Addiction (2011a) *Treatment demand indicator Tables: TDI-19.* Retrieved from http://www.emcdda.europa.eu/html.cfm/index103055EN.html?type=stats&stat_category=w100&stat_type=w87&order=stat_reference on 7 July 2011.

European Monitoring Centre for Drugs and Drug Addiction (2011b) *Table DRD-108. Drug-induced deaths.* Retrieved from http://www.emcdda.europa.eu/attachements.cfm/att_130972_EN_Table-DRD-108-part0.xls on 15 September 2011.

European School Project Survey on Alcohol and other Drugs (2001) *The 1999 ESPAD report.* Retrieved from http://www.epid.ifc.cnr.it/AreaDownload/Report/ESPAD/EspadEurope/The_1999_ESPAD_report.pdf on 20 July 2011.

European School Project Survey on Alcohol and other Drugs (2005) *The 2003 ESPAD report.* Retrieved from http://www.espad.org/documents/Espad/ESPAD_reports/The_2003_ESPAD_report.pdf on 20 July 2011.

European School Project Survey on Alcohol and other Drugs (2009) *The 2007 ESPAD report.* Retrieved from http://www.espad.org/documents/Espad/ESPAD_reports/2007/The_2007_ESPAD_Report-FULL_091006.pdf on 20 July 2011.

Everingham S. & Rydell C. (1994) *Modelling the Demand for Cocaine.* Retrieved from http://www.rand.org/pubs/monograph_reports/2005/RAND_MR332.pdf on 22 September 2011.

Frischer M., Heatlie H. & Hickman, M. (2006). Prevalence of problematic and injecting drug use for Drug Action Team areas in England. *Journal of Public Health* **28**, 3-9.

Galvin S. & Campbell, M. (2010) Cocaine-related admissions to an intensive care unit: a five-year study of incidence and outcomes. *Anaesthesia* **65**(2), 163-6.

Gannon M., Hay G & McKell J. (2011) *Research into methods and data sources for the estimation of prevalence of problematic opiate and cocaine use in Ireland* Retrieved from `http://www.nacd.ie/publications/prevalence_of_problematic_opiate_use.html` on 23 February 2012.

Garavan N., Pankiewicz J., Bloom A., Jung-Ki C., Sperry L., Ross. T.J., Salmeron B., Risinger R., Kelley D., & Stein E.A. (2000) Cue-induced cocaine craving: Neuroanatomical specificity for drug users and drug stimuli. *American Journal of Psychiatry* **157**, 1789-1798.

Giaquinto F. (2007) *Court appointed drug education - the perspective of the voluntary sector.* Retrieved from `http://www.crosscare.ie/images/uploads/CourtAppointedDrugEducation.pdf`on 27 January 2011.

Hall W.C., Talbert R.L. & Ereshefsky L. (1990) Cocaine abuse and its treatment. *Pharmacotherapy* **10**(1), 47-65.

Hamer W. (1906) Epidemic disease in England. *Lancet* **1**, 733-739.

Haney M. (2008) Neurobiology of stimulants. In *Textbook of Substance Abuse Treatment*, 4^{th} edn. (Galanter M. & Kleber H.D., eds), American Psychiatric Press, Inc., Washington, DC. pp 143-155.

Harris D.S., Everheart E.T., Mendelson J. & Jones R.T. (2003) The pharmacology of cocaethylene in humans following cocaine and ethanol administration. *Drug and Alcohol Dependence* **72**, 169-182.

Hartnoll R., Mitcheson M., Lewis R. & Bryer S. (1985) Estimating the prevalence of opioid dependence. *Lancet* **325**(8422), 203-205.

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

Hay G. (2000) Capture-recapture estimates of drug misuse in urban and non urban settings in the north east of Scotland. *Addiction* **95**, 1795-1803.

Hay G. & Smit F. (2003) Estimating the number of drug injectors from needle exchange data. *Addiction Research & Theory* **11**(4), 235-243.

Hay G., Gannon M., MacDougall J., Eastwood C., Williams K. & Millar T. (2010) Opiate and crack cocaine use: A new understanding of prevalence. *Drugs: education, prevention and policy* **17**(2), 135-147.

Health Research Board (2006) *Latest European drug facts, figures and analysis released today Health Research Board puts the Irish drug situation in context.* Retrieved from `http://www.hrb.ie/about/media/media-archive/press-release/release/75/` on 11 January 2007.

Health Research Board (2010) *2010 National Report (2009 data) to the EMCDDA by the Reitox National Focal Point. Ireland: new developments, trends and in-depth information on selected issues.* Retrieved from `http://www.drugsandalcohol.ie/14714/1/Ireland_National_Report_2010_(2009_data).pdf` on 11 August 2011.

Health Service Executive (2007) *HSE LHO population by 5 year Age Groups.* Retrieved from `http://www.hse.ie/eng/staff/FactFile/Census_2006/Census_2006_LHO_by_5_year_Age_Groups` on 11 December 2009.

Health Service Executive (2011) *Community Awareness of Drugs.* Retrieved from `http://www.hse.ie/eng/services/Find_a_Service/Mental_Health_`

Services/healthlines/VolGrp/VolOrg/AWARE_DRUGS.html on 22 September 2011.

Hethcote H. (2000) The Mathematics of Infectious Disease. *SIAM Review* **42**(4), 599-653.

Hickman M., Taylor C., Chatterjee A., Degenhardt L., Frischer M., Hay G. & Tilling K. (2002) Estimating drug prevalence: Review of methods with special reference to developing countries. *UN Bulletin on Narcotics, LIV(1 and 2)*, 15-32.

Hook E.B. & Regal R.R. (1995) Capture-recapture methods in epidemiology: methods and limitations. *Epidemiologic Reviews* **17**, 243-264.

Hope V.D., Hickman M.K. & Tilling K. (2005) Capturing crack cocaine use: estimating the prevalence of crack cocaine use in London using capturerecapture with covariates. *Addiction* **100**, 1701-8.

Horgan J. (2011) *Presentation on the Findings from the 1st Bulletin of the 2010/2011 General Population Survey.* Retrieved from http://www.nacd.ie/publications/1st_results_genpop_survey.html on 14 March 2012.

Horvitz D. & Thompson D. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663-685.

House M. (1990) Cocaine. *The American Journal of Nursing* **90**, 40-45.

Hser Y. (1993) Population estimation of illicit drug users in Los Angeles County. *J Drug Issues* **23**, 323-334.

Irish Youth Justice Service (2011) *Garda Diversion Programme.* Retrieved from http://www.iyjs.ie/en/IYJS/Pages/WP08000061 on 22 September 2011.

Kandel D. (1985) On processes of peer influences in adolescent drug use: a developmental perspective. *Adv Alcohol Substance Abuse* **4**, 139-163.

Karan L.D., Hallar D.L. & Schnoll S.H. (1998) Cocaine and stimulants. In *Clinical Textbook of Addictive Disorders*, 2^{nd} edn. (Frances R.J. & Millar S.I., eds), The Guilford press, New York pp. 168-201.

Karch S. (1996) *The Pathology of Drug Abuse*, 2^{nd} edn. CRC Press, New York.

Karch S. (1999) Cocaine: history, use, abuse. *Journal of the Royal Society of Medicine* **92**, 393-397.

Karch S. (2002) *The Pathology of Drug Abuse* 3^{rd} edn. CRC Press, New York.

Kelly A., Carvalho M. & Teljeur C. (2003) Prevalence of Opiate Use in Ireland 2001 - 2001. A 3-source capture-recapture study. Report to the NACD. Stationary Office Dublin.

Kermack W. & and McKendrick A. (1927) Contributions to the mathematical theory of epidemics, part 1. *Proc. Roy. Soc. London Ser. A* **115**(1927), 700-721.

Klein H. (1985) Coca production in the Bolivian Yungas in the colonial and early national periods. In *Proceedings of the Conference The Coca Leaf and Its Derivatives - Biology, Society and Policy* (Pacini D. & Franquemont C., eds.), Cultural Survival, Cambridge, MA, pp 53-64.

Koob G., Ahmed S.H., Boutrel B., Chen S.A., Kenny P.J., Markou A., O'Dell L.E., Parsons L.H. & Sanna P.P. (2004) Neurobiological mechanisms in the transition from drug use to drug dependence. *Neuroscience and Biobehavioral Reviews* **27**(8), 739-749.

Langmuir A.D. (1976) William Farr: Founder of Modern Concepts of Surveillance. *International Journal of Epidemiology* **5**(1), 13-18.

Larson A., Stevens A. & Wardlaw, G. (1994) Indirect estimates of hidden populationsCapturerecapture methods to estimate the numbers of heroin users in the Australian Capital Territory. *Social Science Medicine* **39**, 823-831.

Long J. (2012) Results from the third general population survey in Ireland. Drugnet Ireland, **40**, 6-8. Retrieved from `http://www.drugsandalcohol.ie/16866/` on 1 May 2012.

Mackintosh D.R. & Stewart G.T. (1979) *Journal of Epidemiology and Community Health* **33**, 299-304

Mastro T.D., Kitayaporn D., Weinger E.G., Vanichensi S., Uneklabh T., Uneklabh C., Choopanya K., & Limpakarnjarnat K. (1994). Estimating the number of HTV-infected injection, drug users in Bangkok: a capture-recapture method. *American Journal of Public Health* **84**(7), 1094-1099.

Mayock P. (2001) Cocaine use in Ireland: an exploratory study. In *A Collection of Papers on Drug Issues in Ireland.* (Moran R., ed.), The Health Research Board, Dublin. pp. 83-152.

McCrystal P., Mayock P. & Hannaford S. (2011) A study of cocaine use in Northern Ireland 2009. *Department of Health, Social Services and Public Safety Northern Ireland.* Retrieved from `http://www.tcd.ie/childrensresearchcentre/assets/pdf/Publications/Cocaine_Use.pdf` on 20 July 2011.

McKendrick A. (1926) Applications of mathematics to medical problems. *Proc. Edinburg Math. Soc.* **44**, 98-103.

Morgan M. (2003) *The Walk Tall Programme An Evaluation.* Retrieved from `http://www.drugsandalcohol.ie/5048/1/293-WalkTall.pdf` on 22 September 2011.

Muench H. (1959) *Catalytic Models in Epidemiology.* Harvard University Press, Cambridge, Massachusetts.

Musto D. (1989) America's First Cocaine Epidemic. *The Wilson Quarterly* **13**(3), 59-64.

National Advisory Committee on Drugs (2003) *An Overview of Cocaine Use in Ireland.* Retrieved from http://www.nacd.ie/publications/CocaineReport.pdf on 20 July 2011.

National Advisory Committee on Drugs (2006) *Drug Use in Ireland & Northern Ireland: 2002/2003 Drug Prevalence Survey: Health Board (Ireland) and Health Social Services Board (Northern Ireland) Results (Revised): Bulletin 2: Confidence Intervals.* Retrieved from http://www.nacd.ie/publications/prevalence_b2confidenceintervals.html on 20 July 2011.

National Advisory Committee on Drugs (2007) *An Overview of Cocaine Use in Ireland: II.* Retrieved from http://www.nacd.ie/publications/Overview_Cocaine_Use_2.pdf on 20 July 2011.

National Advisory Committee on Drugs (2008) *Drug Use in Ireland & Northern Ireland.* Retrieved from http://www.nacd.ie/publications/44598NACDConfidenceTables3.pdf on 20 July 2011.

National Advisory Committee on Drugs (2009) *Drug Use in Ireland & Northern Ireland 2006/7: Regional Drugs Task Force Areas (RDTFs) Results.* Retrieved from http://www.nacd.ie/publications/Final2006-7_CIs_B2_Res.pdf on 20 July 2011.

National Drug-Related Deaths Index (2011) *Drug-related deaths and deaths among drug users in Ireland* Retrieved from http://www.hrb.ie/uploads/tx_hrbpublications/NDRDI_web_update_2008_FINAL.pdf on 15 September 2011.

Nie N., Bent D. & Hull H. (1970) *SPSS: Statistical Package for the Social Sciences* McGraw-Hill, New York.

Ó Súilliobháin C. (2009) *Taking the piss: is HSE drug testing wasting millions?.* Retrieved from http://www.imt.ie/opinion/guests/2009/

09/taking-the-piss-is-hse-drug-testing-wasting-millions.html on 22 September 2011.

Pennings E., Leccese A. & de Wolff A. (2002) Effects of concurrent use of alcohol and cocaine. *Addiction* **97**, 773-783.

Petersen R.C. (1977) Cocaine: an overview. In *National Institute on Drug Abuse Research Monograph # 13.* Department of Health, Education and Welfare Public Health Service, United States.

Pickles W.N. (1948) Research in General Practice. *British Medical Journal* **2**(4574), 469-471.

Pike B. (2008) *Development of Irelands drug strategy 20002007. HRB Overview Series 8.* Retrieved from http://www.hrb.ie/uploads/tx_hrbpublications/Overview_8.pdf on 22 September 2011.

Platt L., Hickman M., Rhodes T., Mikhailova L., Karavashkin V., Vlasov A., Tilling K., Hope V., Khutorksoy M., & Renton A. (2004) The prevalence of injecting drug use in a Russian city: Implications for harm reduction and coverage. *Addiction* **99**, 1430-1438.

Plowman T. (1985) Coca chewing and the botanical origins of Coca in South America. In *Proceedings of the Conference The Coca Leaf and Its Derivatives - Biology, Society and Policy* (Pacini D, & Franquemont C., eds.), Cultural Survival, Cambridge, MA.

Reinerman J.& Zimmer L., (eds.) (1997) *The Social Pharmacology of Smokeable Cocaine: Not All It's Cracked Up to Be.* University Of California Press, Berkeley.

Renshaw E. (1993) *Modelling Biological Populations in Space and Time.* Cambridge University Press, Cambridge.

Repetto M. & Gold M. (1997) Cocaine and crack: neurobiology. In *Substance Abuse : A Comprehensive Text Book* (3rd edn.) (Lowinson J., Ruiz P., Millman R.B. & Langrod J.G., eds), Lippincott Williams & Williams, New York. pp 195-217.

Reynolds S. & Fanagan S. 2008. *Trends in treated problem drug use in Ireland, 2001 to 2006.* Retrieved from `http://www.hrb.ie/fileadmin/Staging/Documents/In_House_Research_Info_Systems/ADRU/Publications/2008/Trends2.pdf` on 27 January 2011.

Roberts M. & Heesterbeek H. (1993) Bluff your way in epidemic models. *Trends in Microbiology* **1**(9) 343-348.

Rossi C. (2004) Operational models for epidemics of problematic drug use: the MoverStayer approach to heterogeneity. *Socio-Economic Planning Sciences* **38**(1), 73-90.

Roy M. (2009) Characteristics of Cocaine Dependent Patients Who Attempt Suicide. *Archives of Suicide Research* **13**(1), 46-51.

RTE (2010) *Dramatic increase in "head shops".* Retrieved from `http://www.rte.ie/news/2010/0126/head.html` on 3 September 2011.

Salloum I.M., Daley D.C, Cornelius J.R., Kirisci L & Thase M.E. (1996) Disproportionate lethality in psychiatric patients with concurrent alcohol and cocaine abuse. *American Journal of Psychiatry* **153**, 953-955.

Sánchez E., Villanueva R.J., Santonja F.J. & Rubio M. (2011) Predicting cocaine consumption in Spain: A mathematical modelling approach. *Drugs: education, prevention and policy* **18**(2) 108-115

Saris A. & O'Reilly F. (2010) *A Dizzying array of substances: An ethnographic study of drug use in the canal communities.* Retrieved from `http://www.drugsandalcohol.ie/13503/1/Canal_communities_A_Dizzying_Array_of_Substances.pdf` 10 September 2011.

Sattenspiel L. (1990) Modeling the spread of infectious disease in human populations. *American Journal of Physical Anthropology* **33**(S11), 245-276.

Schwarz G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics* **6**(2), 461-464.

Seber G. (1982) *The Estimation of Animal Abundance.* The Blackburn Press, New Jersey.

Shearer J., Johnston J., Fry C.L., Kaye S., Dillon P., Dietze P. & Collins L. (2007) Contemporary cocaine use patterns and associated harms in Melbourne and Sydney, Australia. *Drug and Alcohol Review* **26**, 537543.

Simon H. (1959) Review of the book "Elements of Mathematical Biology" by A. Lotka. *Econometrica* **27**, 493-495.

Soper H. (1929) Interpretation of periodicity in disease-prevalence. *J. R. Statistical Society* **92** 34-73.

Strang J. (1993) Cocaine in the UK-1991. *British Journal of Psychiatry* **162**, 1-13.

United Nations Office on Drugs and Crime (2003) *Estimating Prevalence: Indirect Methods for Estimating the Size of the Drug Problem.* Retrieved from `http://www.unodc.org/pdf/iran/links/UNODC/gap_toolkit_module2.pdf` on 11 December 2009.

United Nations Office on Drugs and Crime (2011) *World Drug Report 2011.* Retrieved from `http://www.unodc.org/documents/data-and-analysis/WDR2011/World_Drug_Report_2011_ebook.pdf` on 15 September 2011.

Volkow H., Wang G.-J., Fischman M.W., Foltin R.W., Fowler J.S., Abumrad N.N., Vitkun S. Logan J., Gatley S.J., Pappas N., Hitzemann R. & Shea C.E. (1997) Relationship between subjective effects of cocaine and dopamine transporter occupancy. *Nature* **386**, 827-830.

Vroegop M., Franssen E.J., van der Voort P.H.J., van den Berg T.N.A., Langeweg R.J. & Kramers C. (2009) *The emergency care of cocaine intoxications.* Retrieved from `http://www.njmonline.nl/getpdf.php?t=a&id=10000440` on 15 August 2011.

Vynnycky E. & Edmunds W.J. (2008) Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures. *Epidemiology and Infection* **136**, 166-179.

Vynnycky E. & White R. (2010) *An Introduction to Infectious Disease Modelling.* Oxford University Press, Oxford.

White E. & Comiskey C.M. (2007) Heroin epidemics, treatment and ODE modelling. *Mathematical Biosciences* **208**(1), 312-24.

Whittle P. (1955) The outcome of a stochastic epidemic - a note on Bailey's paper. *Biometrica* **12**, 116-122.

Wiessing L., Hartnoll R. & Rossi C. (2001) The epidemiology of drug use at the macro level: indicators, models and policy-making. *Bulletin on Narcotics* **53**, 119-133.

Wilson R. & Collins M. (1992) Capture-recapture estimation with samples of size one using frequency data. *Biometrika* **79**, 185-216.

Wolff K. & Farrell J. (1999) A review of biological indicators of illicit drug use, practical considerations and clinical usefulness. *Addiction* **94**(9), 1279-1298.

World Health Organisation (2008) *Towards a Global View Of Alcohol, Tobacco, Cannabis and Cocaine Use: Findings from the WHO World mental Health Survey.* Retrieved from `http://www.akzept.org/pdf/presse_pdf/nr22/journal_pmed08.pdf` on 11 August 2011.

Youtube (2008) Cocaine Factory - Lonely Planet Travel Video. Retrieved from http://www.youtube.com/watch?v=UclSEhNdUK8 on 25 July 2011.

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