Lancaster University
Management School

Identifying Factors Associated with Retention in Substance Abuse Treatment Facilities in Ireland Using Logistic Regression Analysis

Student Name: Norma Madden

Student ID: 31330541

Programme of Study: MSc Operational Research and Management Science

Supervisor Name: Dr Richard Williams

Date of Submission: 11th September 2015

"This dissertation is submitted to the Department of Management Science and the Board of Examiners of Lancaster University in partial fulfilment of the requirements for the degree of Master of Science in *Operational Research and Management Science*. I hereby attest that this dissertation is entirely my own work with and that all sources of information have been fully referenced."

Acknowledgements

Firstly I would like to express my sincere gratitude to my academic supervisor, Dr. Richard Williams, for all of the help and support provided during this research project. His expertise and patience were reflected in his ability to attend to my many queries, his encouragement throughout this process is much appreciated.

Besides my supervisor I would also like to thank Dr. Adam Hindle for providing encouragement and guidance on developing this research project and Dr. Nicos Pavlidis for dissipating all confusions encountered during this project. I would also like to take this opportunity to thank all of the Management Science Faculty for their help and support throughout the year.

I am extremely grateful to the Health Research Board for releasing National Drug Treatment Reporting System data for the purposes of this project, and in particular Dr. Suzi Lyons for taking the time out of her busy schedule to answer my questions and prepare the data. This project was made possible by the substance abuse treatment services in Cork and Kerry, I am deeply grateful for their support and willingness to allow access to their client data for use in this important endeavour.

I also would like to thank my family for assisting me in pursuing this degree, and all of my classmates for making this a memorable and enjoyable year.

Abstract

Substance abuse and dependence have detrimental effects at both the individual and societal levels. Treatment for such problems has been shown to reduce the negative consequences and represents a worthwhile investment. However, rates of retention in substance abuse treatment varies widely. To date no investigation has examined treatment retention across the substance abuse treatment population in Ireland. This study aims to describe the characteristics of service users entering substance abuse treatment programmes in the Cork and Kerry region of Ireland, and to identify the significant factors associated with treatment retention.

The National Drug Treatment Reporting System (NDTRS) was used to identify those service users beginning their first treatment episode between January 1st 2008 and December 31st 2013. Logistic regression analysis was used to ascertain significant factors which lead to retention in substance abuse treatment. Models were developed and assessed for goodness of fit and discriminatory ability between binary outcomes using a range of metrics. An adequate and thorough procedure was followed for examining validity of results.

Results indicated that 47% of service users completed their treatment programme while 53% dropped out prematurely. Furthermore, it revealed factors that are related to treatment retention including treatment modality, frequency of substance use, education level, living status, secondary substance used and the involvement of a concerned family member in the treatment episode. Additionally it was highlighted that the factors leading to treatment retention were the same for those using alcohol or illicit substances, with the exception of higher levels of secondary substance use among illicit substance users. Significant differences were also identified between those entering residential treatment compared to other forms of treatment.

Results are compared and contrasted with the existing substance abuse treatment literature. Study limitations are discussed, along with implications for service providers. Future investigations at the individual programme level are recommended to guide the monitoring, design, implementation, and evaluation of treatment procedures to enhance substance abuse treatment retention.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

This chapter provides a background to the research area and places the study to be undertaken within the context of the relevant issues. It begins by defining substance abuse. This is followed by an overview of the prevalence of substance abuse in Ireland. The financial cost implications of substance abuse are then outlined followed by a briefing on the value of substance abuse treatment and the rates of drop out. The problem to be address by this research is clearly stated and the purpose of the study is outlined. Finally the specific research questions are stated.

Substance abuse represents a significant public health problem that has generated increasing concern in recent years. The World Health Organisation estimates that the use of alcohol results in 3.3 million deaths per year worldwide population (World Health Organisation 2015a) . In addition 15.3 million people are estimated to have drug use disorders, with injecting drug use reported in 148 countries, 120 of which report HIV infection among the injecting. Substance abuse refers to "the harmful or hazardous use of psychoactive substances, including alcohol and illicit drugs", (World Health Organisation 2015b). The use of psychoactive substances have can lead to dependency. Typical manifestations of dependency may include a strong desire to take the drug, difficulties in controlling its use, continued usage despite harmful consequences, prioritising drug use to the detriment of other activities and obligations. An increased tolerance may also develop, sometimes coexisting with a physical withdrawal state (World Health Organisation 2015b). These statistics underlines the importance for individual countries to monitor the severity of substance use and take steps to ameliorate associated problems.

Prevalence of substance abuse in Ireland

The prevalence of substance abuse in Ireland is measured in the general population through the All Ireland Drug Prevalence Survey. Three measurements are employed for this purpose which consists of whether the participant: 1) ever used a drug 2) used a drug in the last twelve months and 3) used a drug in the last thirty days. The most recent survey which took place in 2011 revealed that lifetime use of illegal drugs had increased from 24% to 27% since 2007. Furthermore a 2007 survey found that 6% of respondents reported that they had used an illegal drug in the year prior to the survey. One in five people had used cannabis, while a relatively small proportion of drug users by comparison reported using cocaine (0.5%) or heroin (0.1%), in the previous month. Despite these surveys measuring drug consumption in different ways and to various severity levels, they all indicate that drug misuse is prevalent in the general population in Ireland, and is not just a problem affecting a confined group of society (Connolly & Long 2014).

Cost of substance abuse

The cost implications of substance abuse span many health and social as well as criminal justice systems. One of the key impacts of illicit drug use on society is the negative health consequences experienced by its members placing a substantial financial burden on society. In monetary terms this amounts to US$ 200 billion-250 billion which translates to 0.3 to 0.4 percent of global GDP that would be needed to cover all of the costs related to drug treatment worldwide (United Nations Office of Drugs and Crime 2012). Of course the financial impact extends to a society's lost productivity. A study in the United States suggested that productivity losses were a major source of societal costs and totalled US$98.5 billion in 1998 solely due to the lost productivity of those in prison for drug related crimes. This represented over 69% of total costs to society. The costs associated with drug related crime are also substantial. The cost of crime in 1998 attributed to illicit drug abuse is estimated at US$89 billion (Cartwright 2008). According to the Irish Department of Health, the cost incurred by the Irish society due to alcohol related problems is currently in excess of €3.5 billion a year, and this figure relates to alcohol only. Furthermore, alcohol abuse is also estimated to be a contributory factor in over 50% of all suicides in Ireland every year. The budget for directly drug related public expenditure in 2014 totalled over €240 million in Ireland (Connolly & Long 2014).

Value of substance abuse treatment

Several substance abuse treatment evaluation studies have been carried out over the past 40 years, of particular note are the Drug Abuse Treatment Outcome Study (DATOS) in the United States, National Treatment Outcome Research (NTORS) in the UK and the Research Outcome Study in Ireland Evaluating Drug Treatment Effectiveness (ROSIE). All studies have consistently found that treatment 'works', indicating that when treatment is provided to people requiring services for substance use problems, alcohol and drug use decreases. In addition the crime rate is reduced and other measures of social functioning improve

(Prendergast et al. 2002; Hubbard et al. 2003). Furthermore, many of these studies have reported a positive relationship linking the length of time spent in treatment and favourable outcomes for the individual and society. Studies which examine the costs associated with addiction treatment have shown that the benefits of treatment substantially outweigh the costs of investment in drug treatment Cartwright (2000), through a reduction in public health costs and criminal activity. Therefore it makes sense to optimise the benefit derived for the individual involved as well as wider society. Retaining people in treatment is a first step towards this goal as it is the best means to reduce substance abuse, particularly if the planned duration and necessary components of treatment is undertaken.

Substance abuse treatment drop out

The benefits derived from substance abuse treatment are encouraging, however, there is a worrying trend whereby many service users do not remain in treatment long enough to realise its benefits. The actual percentage of service users who do not complete substance abuse treatment due to choosing to dropout themselves or being forced to leave due to non-compliance varies widely across different treatment modalities and treatment models. This adds to the difficulty in measuring reasons for drop out. Lower estimates of the dropout rates for residential treatment programs are around 20%, while upper estimates can reach 70%. Non-residential treatment tends to fare much worse and often exhibit dropout rates exceeding 60% to 70% (Stark 1992; Wickizer et al. 1994). Overall, approximately 50% of clients involved in substance abuse treatment drop out within the first month. Why is it important to be able to predict the potential dropout from substance abuse treatment? The question has clinical, organisational and economic significance (Craig 1985). Service users who drop out still use up the scarce resources of treatment providers. Furthermore the service user fails to reap of the associated benefits of treatment. For these reasons, treatment retention has emerged as an important outcome measure in the study of substance abuse treatment.

Statement of the problem

Substance abuse and dependence can have damaging effects both at the individual and societal level. Adequate, focused and timely treatment is the best available tool to reduce the aforementioned negative individual and societal impacts. However, the individual requires successful engagement in treatment to realise those benefits. A ‘successful’, treatment episode can be characterised by three distinct phases, each of which presents opportunities for either success or failure. First, a client must contact the treatment service. Second, after contact has been made and a recommendation for treatment received, the client must initiate treatment by beginning to attend treatment sessions. Third, the client must stay in treatment long enough to complete the program, including aftercare (Green et al. 2002). The large body of literature on treatment retention clearly indicates that unplanned discharge, or 'drop out', is a well-recognised phenomenon of drug treatment, with people dropping out before participating fully in elements of treatment that will engender change (Gossop et al. 1999). In an attempt to understand the nature of the relationship between service user characteristics, treatment modality, treatment type and whether the service user completes treatment, a wide range of methodological practices have been utilised.   Many of the studies designed to predict who will be retained in treatment and who will drop out have shown a great variability among different types of treatment programmes and inconsistency among factors which predict treatment drop out (Craig 1985). Therefore, no consistent profile of those likely to complete or dropout has been confirmed. This makes it difficult to generalise findings across countries, treatment modalities, treatment types and service users.

Purpose of the Study

A primary purpose of this study is to develop a profile of service users who enter substance abuse treatment services in the Cork and Kerry region of Southern Ireland. The study will also seek to identify the variables which can significantly predict treatment retention.   From a service provider perspective, it is difficult for treatment centres to examine treatment outcomes without initially identifying who is entering treatment and moreover who is being retained in treatment. The identification of variables that positively and negatively relate to retention provides an important contribution to the development of procedures, enabling treatment providers to detect service users who may be at risk for dropout. Furthermore, exploring the factors which may affect treatment retention in this particular population of service users offers a baseline in determining if and how such information can be useful to the service providers. This has the potential to inform the development of interventions aimed at enhancing treatment retention, which can potentially improve treatment outcomes as the positive relationship between retention and outcomes is well established in the literature. From an empirical perspective, this study will allow the comparison of results to existing research regarding the characteristics of service users and their relationship to treatment retention.

Research Questions

In light of the problem which has been outlined and the purpose of this investigation, this study will address the following research questions:

1) How does the treatment population who completed treatment in residential, non-residential, low threshold community based and primary care/institutions in Cork and Kerry between January 1st 2009 and December 31st 2013 differ from those who dropped out prematurely during the same period based on the following variables?:
Demographic: Gender, Age Group
Socio-Economic: Education Level, Employment Status, Living Status, Family Involvement
Programme Level: Treatment Modality, Source of Referral
Substance: Primary Substance Type, Age of First Use, Frequency of Use, Presence of Second Problem Substance, Second Drug Type, Presence of Tertiary Problem Substance
2) What are the factors significant factors which affect treatment retention?

Study protocols outlining the plan for conducting the study can be viewed in Appendix 1. This document was used to explain the purpose and function of the study as well as ethical considerations to substance abuse services from which data was obtained.

Chapter 2: Literature Review

This chapter explains the importance of treatment completion and positive treatment outcomes. It is organised as follows: first the impact of treatment completion will be outlined, followed by a critical review of the literature relating to treatment retention. The focus will be on demographic factors, treatment modality differences, socio-economic and individual level factors, and intrinsic factors. The final section will place treatment retention in an Irish context.

Impact of Treatment Completion

Treatment completion, often confounded with length of stay in treatment, has been linked to improved substance use and social outcomes (Greenfield et al. 2004). There appears to be general agreement on the positive individual and societal effects of treatment, Messina et al. (2000) sought to ascertain differences of post treatment drug use, employment and criminality between genders. It was found that treatment completion significantly reduced drug use, and criminal activity while increasing employment at 19 month follow up interviews. In fact treatment completion was discovered to be the only significant variable for all three outcomes. The researchers suggest that women required 12 month of structured residential treatment combined with aftercare to derive the aforementioned benefits; however this claim is not substantiated by the research as there was no comparison drawn with shorter durations of stay. Experiments conducted by Zarkin et al. (2002) conflicted to some extent with those of Messina et al. (2000) as both length of stay and treatment completion had a negative effect on the number of crimes, but their effects were not statistically significant. These results suggest that the length of stay captures at most only a small part of the excluded effect of treatment completion. Although the treatment completion findings were mixed, the results indicated that, at least for employment, treatment completion plays an important role (Zarkin et al. 2002).

A meta-analysis conducted by Greenfield et al. (2004) used secondary data sources from three large national research projects to examine if longer stays in residential treatment were justified in light of the high costs incurred. Subjects were limited to females, most of whom were pregnant or had dependent children. This work had some limitations, most notably, it did not differentiate between treatment sites which included onsite care of children and those that did not. In reality this important factor may impact substantially on the time a woman would be able to remain in a residential facility. Its findings did suggest that longer duration of treatment is predictive of post treatment abstinence at 6 and 12 month follow up periods, moreover completion of treatment plan goals appeared to be as important as length of stay once subjects stayed for a minimum threshold period (Greenfield et al. 2004). As the focus was on a very specific cohort and long term effects were not tested, the generalisation of these findings to the overall treatment population are somewhat questionable. Having said that, findings regarding completion of prescribed treatment goals rather than length of stay were echoed by Zarkin et al. (2002). They proposed that completion independent of length of stay is predictive of many positive outcomes (Zarkin et al. 2002). This somewhat agrees with the findings of Greenfield et al. (2004) but differs in that Greenfield placed more emphasis on a minimum length of stay of 3 months as a more valuable component of favourable outcomes.

The cost of treatment is clearly a central concern of treatment providers, therefore deriving the optimal benefit from a single treatment episode should mean the service user is less likely to require further interventions. Luchansky et al. (2000) confirmed that treatment completers were less likely to be readmitted during the following year. This finding means that those completing treatment are less likely to re-present at a later stage and use up valuable resources. In addition to findings stated previously with regard to treatment completers having a higher probability of being employed than non-completers, the longitudinal study by found that, controlling for wages earned before treatment, individuals who completed treatment earned more than others (Luchansky et al. 2000). This was a valuable finding which had not been revealed by previous studies, however it neglected to incorporate educational achievement which may have occurred in the interim period between treatment and follow-up, particularly as the follow-up period was 4 years post treatment. Educational attainment in the interim period would likely have a large influence on earnings regardless of treatment completion status.

Gender Differences in Treatment Outcomes

Several studies examined mixed gender samples of patients enrolled in substance abuse treatment and found predictors of retention and completion specific to men or to women (Mertens & Weisner 2000; Green et al. 2002). However, the ways in which gender affects time spent in treatment should be understood as a series of complex relationships between gender, the treatment process or modality, and personal and social factors (Stark 1992). Women are more likely than men to drop out of substance abuse treatment according to

Simpson & Joe (1993) and Stark (1992). Simpson and Joe (1993) confirmed the previous views of Stark (1992). They attempted to synthesise socio-demographic characteristics with psychosocial and treatment process variables and assess the combined effects on treatment retention specific to gender. In addition, the inclusion of a motivation scale added value to their analysis as it allowed for a holistic assessment of the subjects. Greater severity of drug problems was the only common predictor for both genders but motivation was higher for female than males. Mertens & Weisner (2000) were in general agreement that both men and women who had less severe substance related problems stayed in treatment longer, but that other factors differed. Women who were married, unemployed, or had higher incomes were more likely to stay in treatment. Among men, employer and family suggestions to enter treatment, abstinence goals, and being older predicted treatment retention. In contrast to motivation as a predictor, Mertens & Weisner (2000) proposed that external pressures would be more predicative of retention rather than internal motivation. Their finding that employer pressure caused men to remain in treatment while spousal pressure was more predictive for women makes sense in light of traditional gender roles in society.

Later work by Green et al. (2002) illustrated that external pressures from legal issues led to women's retention in treatment. This work also further confirmed the link between addiction severity and higher rate of drop out as put forth by Mertens & Weisner (2000). A potential weakness detected in the data collection and methodology of the Green et al. (2002) study was the selection of subjects. Participants were recruited by counselling staff and offered a monetary reward for taking part, furthermore subjects were from a specific type of health plan. This may dilute the strength of findings as the study was not based on a random sample and selection bias may have been a factor. A more robust analysis by Claus et al. (2007) sought to ameliorate somewhat the lack of random assignment of subjects to specific treatments by the use of propensity scoring which improves the validity of effect inferences by from non-experimental comparisons of alternate treatments. It was found treatment completion rates for women attending specialised female specific treatment were similar to those of women attending mixed gender traditional treatment programmes (Claus et al. 2007). This finding does not support the general consensus which has prevailed in the literature which argues for specialised treatment for women.

Assessing different treatment modalities by gender highlighted some differences. DATOS (Drug Abuse Treatment Outcome Study) researchers found that men were more likely to drop out of outpatient drug free programs, while women were more likely to be categorised in the low retention group for outpatient methadone treatment (Simpson et al. 1997). A later DATOS article reported greater treatment retention among women than men in non-residential programs, but no direct gender relationship in long term residential modalities (Joe et al. 1999). This is inconsistent with a UK based study by Beynon et al. (2008) who found no difference between men and women in rates of dropout. However there was no distinction between treatment modalities as differences were indicated by Simpson et al. (1997). Despite the large body of literature on gender difference's which focus on different predictors of success specific to gender, the most fundamental question remains answered, specifically the causes for under representation of women in treatment. To be equally represented in treatment, the ratio of males to females in treatment should be similar to the ratio of males to females in problem drug use, while in Europe it is reported as 4:1 higher than the ratio between male and female drug users. To date there are few studies that analyse gender differences in the accessibility of treatment services (United Nations Office of Drugs and Crime 2012).

Age Differences in Treatment Outcomes

Research has generally indicated that age is an important predictor of treatment completion. Older men are more likely than younger men to complete treatment according to Green et al. (2002) and have longer overall stays in treatment if aged 40 and older (Mertens & Weisner 2000). However, it is important to note that the mean age of participants in both aforementioned studies was approximately 36 for women and 39 for men. This represents a potential limitation as the sample may have been biased towards older participants who had longer 'drug use histories', and 'treatment careers'. Similarly, it was reported in a UK study that younger age predicted drop out from treatment and those in the older age groups were more likely than their younger counterparts to re-present at treatment (Beynon et al. 2006). The findings by Beynon et al. (2006) are somewhat contradictory as they indicate older people are more likely to complete treatment but yet also more likely to re-enter treatment at a later stage following successful completion indicating potentially higher rates of relapse among this group. Though this fact was not addressed in the research. Later Beynon et al. (2008) illustrated that the odds of drop out exhibited a significant inverse relationship with age but this effect was mitigated when level of social deprivation of subjects was also considered (Beynon et al. 2008). In addition (Wickizer et al. 1994) found that older clients had higher rates of completion in both residential and non-residential settings. There was agreement by Woodward et al. (2006) who found that probability of treatment completion increased as patient age increased. (Woodward et al. 2006), There was a 36 percent increase in the odds of treatment completion for every 10 year increase in age. It is possible that age has a greater influence on retention in treatment depending on the level of severity of the problem at entry which the aforementioned studies failed to control for. Stark (1992) discovered through a review of several studies from the 1970's and 1980's that results for the effect of age on treatment completion were largely mixed and inconclusive. While it could be argued that those findings may not be as relevant today, hence more recent studies such as those discussed do offer evidence to support the hypothesis and there does appears to be less controversy in the findings.

Treatment Modality & Organisational Factors

Treatment dropout rates vary widely across modalities and programs. Non-residential programmes usually report higher dropout rates, as proportions of more than 70% may be observed, but reports on residential dropout rates have ranged from 19% to 67% (Wickizer et al. 1994). However many studies may include clients from different payment structures where state funding may or may not be received. Lack of state funding may be a financial barrier to completion. The relative effectiveness of different types of treatment modality is a critical concern for both service providers and policy makers, this is to some extent due to the significant differences in costs. Hser et al. (2004) found that in residential treatment, large caseload size decreased chances of retention where in non-residential programs, a group therapy focus decreased retention. Given the differential effects of factors in the two modalities, research by Ghose (2008) hypothesized that in addition to a direct effect of modality on post treatment use, modality will interact with other program factors in influencing outcomes. The results of this study indicated that after controlling for time spent in treatment, completing treatment reduced the risk of use after treatment. The strength of the Ghose (2008) study relative to that of Hser et al. (2004) is in its use of stratified random sampling procedures which were used to select a nationally representative sample making the findings more generalizable. Furthermore the use of hierarchical regression modelling techniques to account for variance within and between facilities ensures estimations and inferences are more accurate, therefore findings are more robust. Woodward et al. (2008) views the probability of treatment retention as a link between program structure and operational factors and the individual client demographics. The value of this study compared to other lies in the use of a national dataset with detailed cost and staffing information. The findings indicate, treatment completion rates decline as the number of groups per counsellor increase. That is, if counsellors have to treat too many patients in both group and individual sessions, treatment quality declines and is reflected in lower treatment completion rates. These finding built on the earlier work of Hser et al. (1998) where high clinical staff: client ratios and small caseloads were found to be significantly associated with favourable treatment outcomes.

Socio-economic & Individual Level Factors

Socio-economic factors which may predict treatment retention or drop out have been covered extensively in the literature. As level of education is often used as an indicator of socio-economic status, it follows that higher higher educational attainment would lead to greater retention. Knight et al. (2001) noted education level was a significant predictor of treatment completion for women in residential treatment settings. Sayre et al. (2002) also produced findings which suggest that education level may play an important role in treatment completion for non-residential treatment as those completing programmes had more years of education than those who dropped out of treatment. Although the findings are in agreement with regard to the positive correlation between education and retention, research has neglected to attempt to understand the reasons for this. However, it is a generally accepted notion in public health that higher educational attainment lead to better health outcomes overall. Further indicators of socio economic status include income, Green et al. (2002) found that for both males and females, higher levels of financial resources predicted retention and completion. Another finding which has been replicated in many studies is the positive link between employment and retention in treatment. Simpson & Joe (1993) found that being unemployed increased the odds of dropping out by 3.1. As well as a measure of socio-economic status, employment can also be viewed as a measure of general social functioning, therefore this result is not surprising as employment may signal more stable relationships and greater social support.

Intrinsic Factors

Most research on predicting factors associated with treatment retention or drop out has focused on demographic or organisational factors. Deviating from this, Simpson & Joe (1993) examined motivation as a predictor of dropping out of treatment. Having rigid expectations of immediately quitting drugs forever (versus more modest expectations) increased the odds of early dropout by 1.9. This study used a scale to measure a service users 'desire for help', to ascertain level of motivation. As motivation is a transient state, it is therefore questionable as to its efficacy for use in determining accurate prediction of treatment retention. The results suggested that opiate users with more modest and possibly more realistic expectations have a more favourable chance of remaining in treatment. These results are somewhat limited as only those subjects using opiates were included. It was therefore unsurprising that for this high risk category of drug users, the methadone dose they were receiving was identified as one of the most reliable predictors. However, a follow up study conducted by Joe et al. (1998) which included alcohol and other type of substance users indicated that high levels of pre-treatment motivation was found to be the most important predictor of retention and completion. Ball et al. (2006) built further on this finding by the additional examination of interpersonal problems, and program perception factors. In agreement with Simpson & Joe (1993) the intrinsic obstacles appeared to be more relevant indicators of retention rather than external logistical factors, although more difficult to address (Ball et al. 2006). While the work of Ball et al. (2006) had some limitations in that the sample size included only 24 subjects which limits the power of the findings substantially, its main value lies its evaluation of service user's subjective reasons for early drop.

Treatment Outcomes in an Irish Context

The only major study into treatment outcomes conducted in Ireland was the ROSIE study in 2009 (Research Outcome Study in Ireland Evaluating Drug Treatment Effectiveness). This national, prospective, longitudinal drug treatment study was undertaken in response to acknowledged gaps in the global literature on drug treatment evaluation. It followed similar studies in the USA (DATOS: Drug Abuse Treatment Outcome Study) and (NTORS: National Treatment Outcomes Research Study) in the UK in the 1990's. The overall outcomes from the ROSIE study show drug treatment works and that investment in drug treatment is paying dividends. Significant reductions were shown in the key outcome areas of drug use, involvement in crime and injecting drug use. In addition improvements were seen in employment and training (Department of Community, Rural and Gaeltacht Affairs 2009). Although this was a comprehensive study, it focused exclusively on opiate users, this particular group of high risk users may exhibit certain characteristics which are not representative of the entire treatment population, and as such its findings may differ for other types of substance user.

A follow-up study was conducted using ROSIE data which sought to examine the alcohol consumption and treatment outcomes for those who have undergone treatment for opiate addiction. Analysis revealed that those who abstained from alcohol were less likely to be using heroin, methadone, cocaine or benzodiazepines than non-abstainers (Stapleton & Comiskey 2010). The most recent Irish study also focused on opiate users, but particularly the retention of opiate users in methadone substitution treatment. This research conducted by (Mullen et al. 2012) found that those who attended a specialist addiction treatment service site were two times more likely to leave methadone treatment within 12 months compared with those who attended a general practitioner. The most important predictor of retention in treatment was methadone dose, confirming a similar finding by Simpson & Joe (1993). Other indicators of retention included female gender, males were found to score lower on the readiness for treatment scale (Mullen et al. 2012).

To date no studies on treatment retention rates with regard to substances other than opiates have been conducted in Ireland. The National Drugs Strategy 2009-2016 has acknowledged that there has been an improvement in attracting problem drug users into services and retaining them there (Department of Community, Rural and Gaeltacht Affairs 2009). However, no clear target for retention levels or time in treatment has been specified. Furthermore, most of the existing evidence regarding factors associated with treatment retention is derived primarily within the USA, and to a lesser extent in the UK. This study will attempt to fill this research gap by assessing treatment completion in Irish substance abuse services and identifying factors which lead to treatment retention. In addition findings will be compared to that of similar studies in the UK and USA.

Chapter 3: Methodology

This section details the process undertaken in the analysis of the data set. The overall objective to be investigated is to determine the factors which may lead subjects to complete treatment and if there is a clear difference between the characteristics of those completing and not completing treatment. Initially a background to the database from which the dataset was derived is given, this is followed by an overview of the study population. A description of the variables under consideration is provided. Finally the stages of the preliminary and main analysis are outlined in detail.

Data

The data used for this study was obtained from the National Drug Treatment Reporting System of the Health Research Board. The National Drug Treatment Reporting System (NDTRS) is an epidemiological database on treated drug and alcohol misuse in Ireland which was established in 1990. Treatment is broadly defined as any "activity which aims to ameliorate the psychological, medical or social state of individuals who seek help for their substance misuse problems", (Health Research Board 2015). NDTRS data are reported to the European Monitoring Centre for Drugs and Drug Addiction. All agencies providing treatment as defined by the Health Research Board are requested to complete a structured questionnaire on each person attending their service. The structured questionnaire is completed by the substance misuse health professional in the presence of the service user.

Study Population

For the purposes of this study, agencies providing treatment were limited to those based in the Cork and Kerry region of Ireland. Only data pertaining to those individuals making their first ever contact for treatment from January 2009 to December 2013 was included. The initial data was obtained for 5114 client records. A preliminary examination of the data revealed that 423 cases had been previously treated and were removed to ensure that each individual only entered the study once. Treatment completion status was defined in terms of the discharge categories which are contained within the questionnaire. The categories included: 1) treatment completed, 2) transferred stable, 3) transferred unstable, 4) client did not wish to attend further treatment because he/she considered him/herself to be stable, 5) client refused to have further sessions or did not return for subsequent appointments, 6) premature exit for non-compliance. Those cases in relating to 1 and 2 were categorised as having completed treatment, while 4 to 6 were categorised as not having completed treatment. Following data cleaning, the final study population included 4,339 cases for further analysis. Please see appendix 2 for a more detailed explanation of observations which have been removed from the dataset.

The NDTRS database can provide analysis by place of treatment which provides a direct indicator of the demand for treatment. It can also classify the treatment population by drug(s) used which allows for a description of patterns of problem substance use, and the relationship between substance use and demographic, socio-economic and substance using characteristics. The advantage of using this secondary data source is that it is already in existence, however the disadvantage is the data collection and compilation have not been done with the current research purpose in mind (Sorensen et al. 1996). There are certain metrics which would be valuable in this study which are not available from this dataset; nevertheless it does offer value in researching the stated problem as an initial evaluation. In addition there were several variables available in the dataset which will not be used for analysis, there are two reasons variables were excluded: 1) the number of observations to which the variable was applicable were very small relative to the total number of observations, or 2) previous literature on this research topic had not yielded any meaningful results through inclusion of those variables. The variables which were selected for this analysis included demographic factors such as gender and age range, Socio-economic factors include living with whom, employment status, highest level of education achieved at the time of treatment and family involvement. Treatment modality and source of referral were the only programme level factors available. Finally substance specific variables identifying primary problem substance, frequency of use, age at first use, existence of a secondary problem substance, secondary problem substance and the existence of a tertiary problem substance were also included. The original dataset had coded each individual drug within a drug class, to simplify analysis and interpretation each drug was categorized according to the classification in which it belongs, there were five classifications; alcohol, cannabis, hypnotic/benzodiazepine, opiate and stimulant. Table 1 provides a description of the abbreviated variables in the dataset.

Table 1: Variable Description

Variable	Remarks	Description
Outcome	Binary: 0 No, 1 Yes	Treatment completed or not completed
Gender	Factor: male, female	Gender of subject
Agerange	Factor: 4 levels	Age range of subjects divided into five levels: under 20, 20 to 29, 30 to 39, 40 and over
Living	Factor: 5 levels	With whom the subject is living divided into five levels: alone, alone with children/partner and children (abbreviated to alch/ptch), partner or friends (abbreviated to part/fri), parents and family (abbreviated to par/fam), other which includes homeless, institution and foster care.
Empstatus	Factor: 4 levels	Status of employment of the subject divided into four levels: employed, unemployed, student and training, other which includes housewife/husband, retired and unable to work.
Educlevel	Factor: 5 levels	The level of education obtained by the subject at the time of treatment. The levels include current fulltime students, primary completed or less, junior certificate (completed at approximately age 15), leaving certificate (completed at approximately age 18), and third level which includes any post-secondary education.
Refsource	Factor: 4 levels	The source from which referral to treatment was received. The levels include self, family or friends, health professional which includes other treatment centres, GP, hospitals, accident and emergency staff. Other professionals include courts, probation officers, police and social services.
Treatmod	Factor: 4 levels	Setting in which the subject was treated, residential, non-residential, low threshold community setting, primary care provider and institution
Primdrug	Factor: 5 levels	The main problem substance for which the subject is attending treatment including alcohol, cannabis, hypnotic/benzodiazepine, opiate and stimulant
Primfreq	Factor: 4 levels	Frequency of use of the main problem substance which is either daily, 2 to 6 days per week, once per week or less or no use within the month prior to treatment
Primage	Continuous	Age of first use of the primary problem substance
Secondary use	Factor: no, yes	Presence of a second problem substance
Second Drug type	Factor: 7 levels	Type of drug if a secondary problem substance exists, the levels include alcohol, cannabis, hypnotic/benzodiazepine, opiate, stimulant, other (unspecified, “headshop”, solvents, and antidepressants). If no secondary problem substance is present this is denoted by no drug.
Tertiary use	Factor: no, yes	Presence of a third problem substance
CP	Factor: no, yes	Concerned person such as family member or friend involved with the subjects treatment

Preliminary Analysis

Determining Variable Relevance

Determining the relative strength of variables in predicting the outcome is a critical phase in any model building process. Therefore two methods were used to assess the strength, and the results of each method were compared. An initial analysis was undertaken using SAS (EM) (Statistical Analysis Software (Enterprise Miner)), the variable selection node was used to test each variables relation to the target variable. The R^2 criterion was used to score the dataset using a goodness-of-fit criterion to evaluate variables. A stepwise regression method of selecting variables that stops when the improvement in the value is less than 0.00050 is undertaken. The method rejects variables whose contribution is less than 0.005. The second form of variable relevance evaluation undertaken was using Weight of Evidence (WOE), Information Value (IV) and relevant graphs. WOE measure how much information on having completed or not completed treatment is in each variable, while IV is a measure of divergence between p(x|treatment completed) and p(x|treatment not completed). WOE and IV were calculated manually using the formulas in Equation 1 and 2.

Equation 1

$$WOE = ln\left(\frac{\%\ treatment\ not\ completed}{\%\ treatment\ completed}\right)$$

Equation 2

$$IV = \sum (\%\ treatment\ not\ completed - \%\ treatment\ completed) \times WOE$$

Missing Value Analysis

Most studies carried out on epidemiologic data will have some degree of ‘missingness’, in the dataset. Missing data can cause a loss of predictive power of any model that is built, as valuable information is omitted. Typical statistical procedures exclude records with missing values. Analyses excluding records with missing data can provide biased estimates by using less information and by ignoring possible systematic differences between complete and incomplete records (Van Der Heijden et al. 2006). Larger standard errors result when less information is utilised, and biased estimates will be obtained when the data are not missing completely at random (Little 1988). Therefore a rigorous analysis of missing data was carried out to assess the extent of the problem details of which are outlined in Appendix 3.

Statistical Approach

Selection of the Statistical Approach

Logistic regression was initially proposed in the 1970's as an alternative approach to overcoming the limitations of ordinary least squares regression in handling dichotomous outcomes. It is a maximum-likelihood method that is widely used in studies involving epidemiologic data where often the outcome variable is dichotomous (Peng & So 2002). As the outcome variable in this study is binary, coded as 0 = treatment not completed and, 1 = treatment completed, logistic regression is the most suitable statistical modelling approach. In addition multivariate logistic regression will be used as there is more than one predictor variable. In logistic regression, instead of predicting the value of a variable Y from a predictor variable X, or several predictor variables (Xs), the probability of Y occurring given the known values of X is predicted instead. The logistic regression equation from which the probability of Y is predicted is given by:

$$P(Y) = \frac{1}{1 + \ e^{-(\beta_0 + \ \beta_1 X_{1i} + \ \beta_2 X2i + \cdots + \beta n X ni)}}$$

The resulting value from the equation varies between 0 and 1. A value close to 0 means that Y is very unlikely to have occurred, while a value close to 1 means that Y is very likely to have occurred (Field 2012). The exponentials of the regression coefficient estimates are called odds ratios, which are crucial to the interpretation of logistic regression. The odds ratio is an indicator of the change in odds resulting from a unit change in the predictor. Therefore the odds of completing treatment is the probability of completing treatment divided by the probability of not completing treatment. The odds ratio computed by dividing the odds after a unit change in the predictor by the original odds is the proportionate change in odds (Field 2012). If the value is greater than 1, it indicates that as the predictor increases, the odds of completing treatment increases. Conversely, a value of less than 1 indicates that as the predictor increases, the odds of completing treatment decreases. Odds ratios will be used for interpretation of regression output for the purposes of this study. All regression models and statistical tests were conducted in R Studio.

Model Building Process

Three experimental approaches were taken to model fitting. Firstly a pairwise bivariate analysis between outcome (treatment completed or treatment not completed) on all predictor variables was conducted and assessed independently using odds ratios and chi squared analysis. Variables that had a significance level of P < 0.25 in the bivariate analysis were then entered into the multivariate logistic regression analysis. The cut off value of 0.25 was chosen as more traditional values such as 0.05 can fail in identifying variables known to be important (Bursac et al. 2008). Any value below 0.25 is at least weakly associated with the outcome variable. The Likelihood Ratio Test (LRT) was used to test the null hypothesis that the variable under consideration was not associated with treatment completion.

Secondly automated model building techniques were used including all variables. Multiple automated variable selection techniques have been developed based on decision rules and algorithms. Forward, backward and stepwise selection are three such methods which are commonly used. Each procedure decides which variables to keep and which to drop based on either the F-statistic or the AIC. Each of the three techniques in R uses a criterion based procedure called the Akaike Information Criterion (AIC). The AIC is a goodness of fit measure that favours smaller residual error in the model but penalises the inclusion of further predictors and thus helps to avoid overfitting (Austin & Tu 2004). Appendix 4 offers a more detailed explanation of stepwise variable selection using the AIC.

Finally an iterative process of model building was used. Much of the literature on variable selection has been critical of automated selection techniques, it has been proposed that all model selection techniques are subject to error and no optimal method is known (Greenland 1989). With this in mind and in an attempt to build the most parsimonious model, a further strategy was employed whereby all predictors were initially forced into the model. Each predictor was subsequently systematically removed and the effect was monitored. Individual parameter estimates were tested by the likelihood ratio test and the Wald statistic computed from the ratio of the estimated slope parameter over its standard error. Appendix 5 offers a more detailed explanation of the Likelihood Ratio Test.

Testing Overall Model Fit

Several pseudo R squared measures have been proposed as a descriptive measure of goodness of fit for a logistic regression model, however there appears to be no consensus as to which is best (Field 2012). For the purposes of this study Nagelkerke's R^2 is reported as a measure of explained variation, which is calculated on the log-likelihood scale (Nagelkerke 1991). However, it should be noted for logistic regression models, R^2 values tend to be much lower than for linear models (Hosmer & Lemeshow 2000). Hosmer and Lemeshow proposed a goodness of fit test that they show, through simulation, is distributed as chi-square when there is no replication in any of the subpopulations. This test is only available for binary response models and was interpreted in this study. (Hosmer & Lemeshow 2000). A more detailed explanation of this test is given in Appendix 6.

Evaluating Predictive Accuracy of Candidate Models

Final candidate models were evaluated against each other using a range of methods. Firstly a classification based approach was used, this assesses how good the model is at accurately separating cases into two bins, one where the event occurs (y=1) and one where it does not occur (y=0) (Esarey & Pierce 2011) . The percent correctly predicted sorts the model predictions into two categories, $\hat{Y}$=1 and $\hat{Y}$=0, on the basis of $\hat{p}$. A threshold of 0.5 was set, when $\hat{p} > 0.5$, $\hat{Y}$=1, otherwise $\hat{Y}$=0. A misclassification table indicated the number of true and false positives and true and false negatives for all models. A receiver operator curve (ROC) was constructed for each final model to illustrate the degree to which the predictions agree with the data. The ROC curve is a plot of sensitivity versus 1 minus specificity. Sensitivity denotes the proportion of cases which were correctly classified as treatment completed or the fraction of true positives. Specificity on the other hand denotes the proportion of cases correctly classified as treatment not completed or the fraction of false positives. Therefore this curve shows the best model as the one associated with the greatest sensitivity and the lowest 1 minus specificity.

The c-statistic was also calculated, which is equivalent to the area under the receiver operating characteristic curve and is a central measure of performance in many studies on predicting binary outcomes (Austin & Steyerberg 2014). A detailed explanation of the c-statistic can be viewed in Appendix 7. A similar measure to the c-statistic which was also examined is Somers Dxy rank correlation between predicted probabilities and observed outcomes. When the value of Dxy = 0, the model is making random predictions. When Dxy = 1, the model discriminates perfectly. For binary outcomes the AUC and Dxy are related by, AUC = Dxy/2 + 0.5. In addition, measures that quantify the overall accuracy of predictions were examined. The Brier score, or average prediction error score, provides a measure of the agreement between the observed binary outcome and the predicted probability of that outcome. It is calculated as (*yi pi*) 2 / *n*, where *y* denotes the observed outcome and *p* the prediction for subject *i* in the data set of *n* subjects. For adequate models the Brier score ranges from 0 (perfect) to 0.25 (worthless) (Steyerberg et al. 2001). Being mainly a relative measure, a lower score points to a superior model; the actual value of the score seems of limited value (Rufibach 2010).

Model Diagnostics

Residuals were used to: 1) assess observations for which the model is a poor fit, and 2) isolate observations which are exerting a high level of influence on the model. The deviance, studentized and standardised residuals were examined to assess 1) and Cook's Distance, Leverage and DFbetas were examined to assess 2). In addition marginal model plots as, a variation of the basic residual plot, were also assessed (Cook & Weisberg 1997).

Collinearity Diagnostics

Due to 13 of the 14 variables being categorical, visually checking for correlation patterns among predictors was not possible. Therefore the VIF function in the Car package in R was used to test the variance inflation factor. The results provided are for the Generalised Variance Inflation Factor (GVIF). The GVIF is calculated if any terms in an unweighted linear model have more than 1 degrees of freedom. These are interpretable as the inflation in size of the confidence ellipse or ellipsoid for the coefficients of the term in comparison with what would be obtained for orthogonal data (Fox & Monette 1992).

Model Validation

Validation of any model is an important step in ensuring that the model is likely to perform as expected. While the split sampling approach which involves fitting a model to a 'training' dataset and using the model to test a 'validation' dataset is often used to validate predictive models, this approach does not make efficient use of all of the information contained in the dataset (Austin & Steyerberg 2014). Split-sample validation results in the validation of a model fit to a 'training' dataset, but it does not validate the model fit to the complete dataset. Harrell et al. (1996) instead recommend the use of bootstrap resampling as a method of internal validation. This involves taking a large number of samples with replacement from the original sample. Bootstrapping provides nearly unbiased estimates of predictive accuracy that are of relatively low variance (Steyerberg et al. 2001). In addition the entire dataset can be used for model development. This approach gives optimism corrected estimates of the c-statistic along with other model performance statistics. Optimism is defined as true performance minus apparent performance, where true performance refers to the underlying population of observations, and apparent performance refers to the estimated performance in the sample (Steyerberg & Harrell 2015). This approach was implemented using the validate function in Harrell's RMS package in R.

Chapter 4: Results

This section provides results of the preliminary and main analysis as set out in chapter 3. Descriptive characteristics provide an overview of the sample population according to treatment completion status. A graphical profile of each variable and its importance in prediction of the outcome variable is provided along with the weights of evidence and information value statistics. The results of missing value analysis and approach used to address missing values is detailed. Finally, results of the model building process, regression diagnostics and model validation are outlined in detail.

Part 1: Preliminary Analysis

Data Description

Table 2 presents the descriptive characteristics of the sample by programme level factors. Non-residential treatment accounts for 53.5% of those attending treatment for the first time, while this group also accounts for 67.5% of those not completing treatment. The largest proportion of cases is self-referred to treatment. Referrals from health professionals only account for 16.2% of the overall referrals. A larger number are referred from other professionals made up of courts, probation, police and social services at 22.7%.

Table 2: Characteristics of those completing and not completing treatment by programme level factors

Explanatory Variable	Total		Completed		Not Completed	
N (%)	4339	(100)	2048	(47)	2291	(53)
Referral Source						
Self	1620	(37.3)	722	(35.3)	898	(39.2)
Family/friends	1003	(23.1)	577	(28.2)	426	(18.6)
Health Professional	702	(16.2)	323	(15.8)	379	(16.5)
Other Professional	985	(22.7)	419	(20.5)	566	(24.7)
Missing	29	(0.7)	7	(0.3)	22	(1.0)
Treatment Modality						
Residential	1012	(23.3)	906	(44.2)	106	(4.6)
Non-residential	2323	(53.5)	776	(37.9)	1547	(67.5)
Low threshold	824	(19.0)	325	(15.9)	499	(21.8)
Primary care/Institution	180	(4.1)	41	(2.0)	139	(6.1)

Sample characteristics by demographic and socio-economic factors are given in table 3. Overall 68.5% were male. The unemployment rate was 45.1% and 40.1% had ceased school attendance at the junior certificate level or lower. The 20 to 29 age range represents the largest proportion of subjects at 30.1% followed by those under 20 years at 26.7%.

The number of subjects who had family or friends involved in their treatment was evenly split between those who did and did not. Furthermore the largest proportion of subjects (46.2%) reported living with parents or family while only 12.9% reported living alone.

Table 3: Characteristics of those completing and not completing treatment by demographic and socio-economic factors

Explanatory Variable	Total		Completed		Not Completed	
N (%)	4339	(100)	2048	(47)	2291	(53)
Gender						
Male	2973	(68.5)	1339	(65.4)	1634	(71.3)
Female	1354	(31.2)	703	(34.3)	651	(28.4)
Missing	12	(0.3)	6	(0.3)	6	(0.3)
Age Range						
under 20	1157	(26.7)	524	(25.6)	633	(27.6)
20 to 29	1304	(30.1)	565	(27.6)	739	(32.3)
30 to 39	808	(18.6)	367	(17.9)	441	(19.2)
40 and over	1064	(24.5)	590	(28.8)	474	(20.7)
Missing	6	(0.1)	2	(0.1)	4	(0.2)
Living						
Alone	558	(12.9)	249	(12.2)	309	(13.5)
Children/partner	916	(21.1)	469	(22.9)	447	(19.5)
Partner/Friends	473	(10.9)	206	(10.1)	267	(11.7)
Parents/Family	2004	(46.2)	955	(46.6)	1049	(45.8)
Other	351	(8.1)	155	(7.6)	196	(8.6)
Missing	37	(0.9)	14	(0.7)	23	(1.0)
Employment Status						
Employed	990	(22.8)	560	(27.3)	430	(18.8)
Unemployed	1956	(45.1)	812	(39.6)	1144	(49.9)
Student/training	934	(21.5)	462	(22.6)	472	(20.6)
Other	414	(9.5)	199	(9.7)	215	(9.4)
Missing	45	(1.0)	15	(0.7)	30	(1.3)
Education Level						
Primary or less	544	(12.5)	200	(9.8)	344	(15.0)
Second Level (Junior)	1198	(27.6)	521	(25.4)	677	(29.6)
Second Level (Leaving)	1198	(27.6)	633	(30.9)	565	(24.7)
Third Level	271	(6.2)	171	(8.3)	100	(4.4)
Current	701	(16.2)	373	(18.2)	328	(14.3)
Missing	427	(9.8)	150	(7.3)	277	(12.1)
Family/Friends involved						
Yes	2141	(49.3)	1364	(66.6)	777	(33.9)
No	2154	(49.6)	662	(32.3)	1492	(65.1)
Missing	44	(1.0)	22	(1.1)	22	(1.0)

Table 4 represents the sample by substance specific factors. Alcohol accounts for 62.4% of primary problem substances, with cannabis making up another 23%. The remaining 14.6%

is divided between hypnotics and benzodiazepines, opiates and stimulants. The percentage of those using either daily or two to six days per week amounts to almost 70% of the sample. The mean age for commencing use of the main problem substance was 16 years (SD 4.8) across the population. Almost 43% report the presence of a secondary problem substance while 21.7% reported a tertiary problem substance. Of those with a secondary problem substance cannabis and alcohol are the most prevalent with usage at 15.3% and 13.2% respectively. Hypnotics and benzodiazepines account for 4.8% or 208 cases where there is presence of a secondary problem substance, while the same drug class accounts for only 4.1% of primary problem substances.

Table 4: Characteristics of those completing and not completing treatment by substance use factors

Explanatory Variable	**Total**		**Completed**		**Not Completed**	
N (%)	**4339**	**(100)**	**2048**	**(47)**	**2291**	**(53)**
Primary Problem Substance						
Alcohol	2708	(62.4)	1351	(66.0)	1357	(59.2)
Cannabis	999	(23.0)	449	(21.9)	550	(24.0)
Hypnotic/Benzodiazepine	180	(4.1)	66	(3.2)	114	(5.0)
Opiates	296	(6.8)	114	(5.6)	182	(7.9)
Stimulants	156	(3.6)	68	(3.3)	88	(3.8)
Frequency of Use						
Daily	1225	(28.2)	478	(23.3)	747	(32.6)
2 to 6 days per week	1783	(41.1)	848	(41.4)	935	(40.8)
Weekly or less	564	(13.0)	285	(13.9)	269	(11.7)
No use past month	716	(16.5)	406	(19.8)	310	(13.5)
Missing	51	(1.2)	21	(1.0)	30	(1.3)
Mean age of first use: primary substance (SD)	16	(4.8)	16	(4.6)	16	(4.7)
Secondary Drug Use						
Yes	1854	(42.7)	861	(42.0)	993	(43.3)
No	2485	(57.3)	1187	(58.0)	1298	(56.7)
Second Drug Type						
Alcohol	574	(13.2)	268	(13.1)	306	(13.4)
Cannabis	662	(15.3)	313	(15.3)	349	(15.2)
Hypnotic/Benzo	208	(4.8)	88	(4.3)	120	(5.2)
Opiate	41	(0.9)	16	(0.8)	25	(1.1)
Stimulant	279	(6.4)	139	(6.8)	140	(6.1)
Other	90	(2.1)	37	(1.8)	53	(2.3)
No Drug	2485	(57.3)	1187	(58.0)	1298	(56.7)
Tertiary Drug Use						
Yes	943	(21.7)	452	(22.1)	491	(21.4)
No	3396	(78.3)	1596	(77.9)	1800	(78.6)

Variable Relevance

An initial visual inspection of each attribute's percentage of treatment completed and treatment not completed was undertaken. If the IV (Information Value) is <0.02, the predictor is considered to contain no predictive power. Those predictors with an IV of 0.02 to 0.05 are denoted as being weak. Weak to moderate predictors have and IV between 0.05 and 0.1. Moderate predictors are from 0.1 to 0.3 and strong predictors have an IV >0.3. Figure 1 shows the proportion of treatment completed and not completed according to the presence of a secondary problem substance and the type of substance if present. Both the graphs and the WOE (Weights of Evidence) table show that the numbers completing or not completing treatment are very similar for both attribute's. The contribution of each class of the attribute towards the information value is extremely low implying this variable may be a poor predictor of the outcome. The information value of the secondary use is 0.0007, therefore presence of a secondary problem substance appears to be of little relevance to classifying treatment outcomes. The IV of second drug type is 0.0052 also indicating that it is a weak predictor. However WOE for hypnotic/benzodiazepine, opiate and stimulant is 0.1980, 0.3341 and 0.2472 respectively. This indicates the variable second drug type may be moderately relevant.

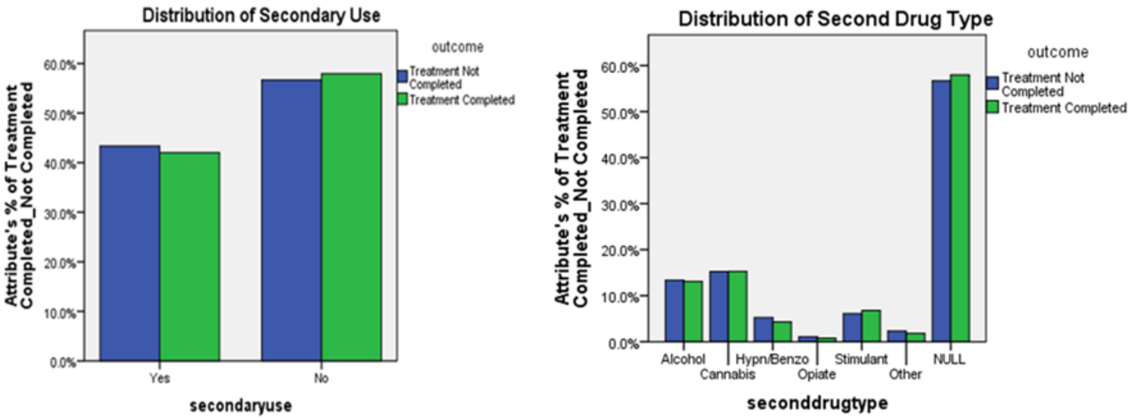

Figure 1: Distribution of Secondary Use and Second Drug Type

Figure 2 represents the distributions of the primary drug of choice and frequency of use. The proportion of those completing treatment is higher when alcohol is the main problem substance. All other drugs have lower rates of completion. The WOE of hypnotics/benzodiazepines and opiates are 0.4344 and 0.3556 respectively, which indicated that they contain relevant information. The IV for *primdrug* overall is 0.0259. Therefore primary drug of choice appears to be moderately relevant to treatment outcome. The IV for frequency of use is 0.0612. The WOE for daily use is 0.3343 increasing the relevance of the variable, all other rates of use have WOE which are much lower.

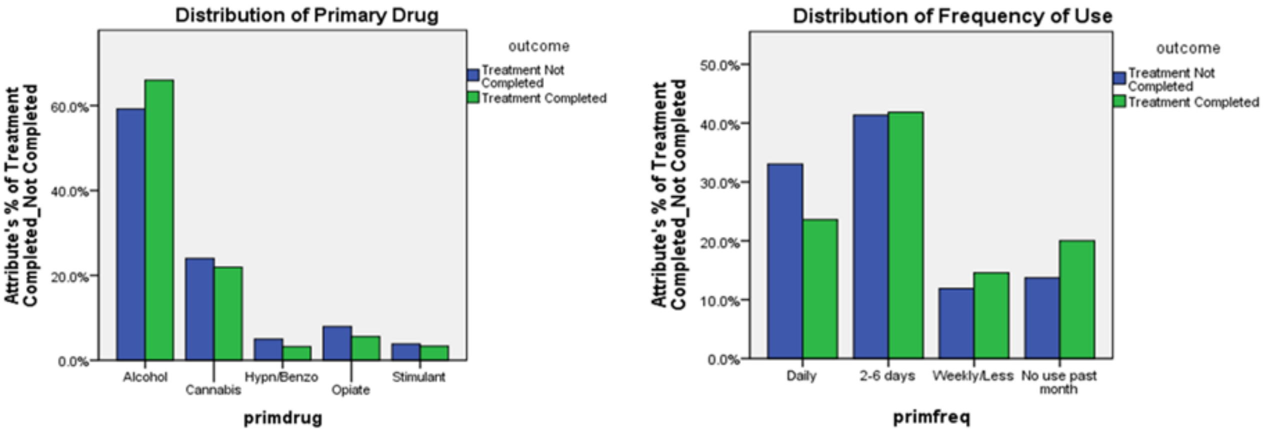

Figure 2: Distribution of Primary Drug and Frequency of Use

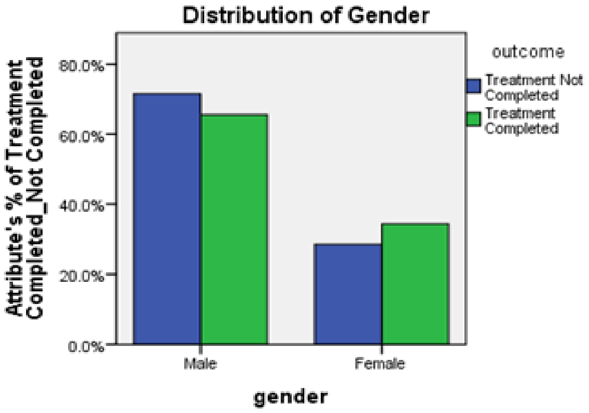

Figure 3: Distribution of Gender

The IV for gender is weak at 0.0179, Figure 3 does indicate there is a small difference in completion rates for male and female. Figure 4 illustrates the contribution of education level and employment status to treatment completion rates. Both variables appear to show significant differences in each class for percentage of the outcome variable. Those who are educated to primary/less and junior certificate level have WOE of 0.4301 and 0.1497 respectively. The overall IV for education level is 0.1019 indicating is a moderate to strong predictor. Employment status is a weaker predictor with an IV of 0.0612. The WOE for unemployed is highest at 0.2306 which is also shown by the graph.

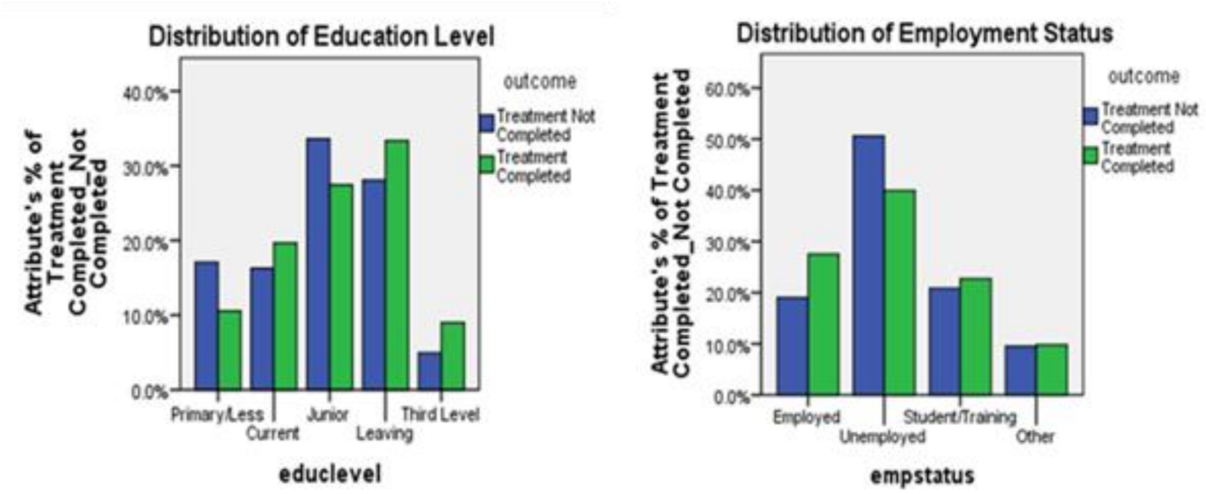

Figure 4: Distribution of Education Level and Employment Status

Two of the strongest predictors as demonstrated by IV and WOE graphs in figure 5 are family involvement in treatment (CP: Concerned Person) and treatment modality. Their IV's are 0.4505 and 1.1293 respectively. This is evidenced in the graphs in Figure 5 as the percentage of treatment completed and not completed is very pronounced for all classes within the variables. WOE for those answering no to concerned person involvement is 0.7004 indicating this has valuable information. Similarly within the treatment modality variable the WOE for the class which identifies those being treated in primary care or institutions is 1.1087 and the figure for the class non-residential is 0.5777 showing that they are potentially very strong predictors.

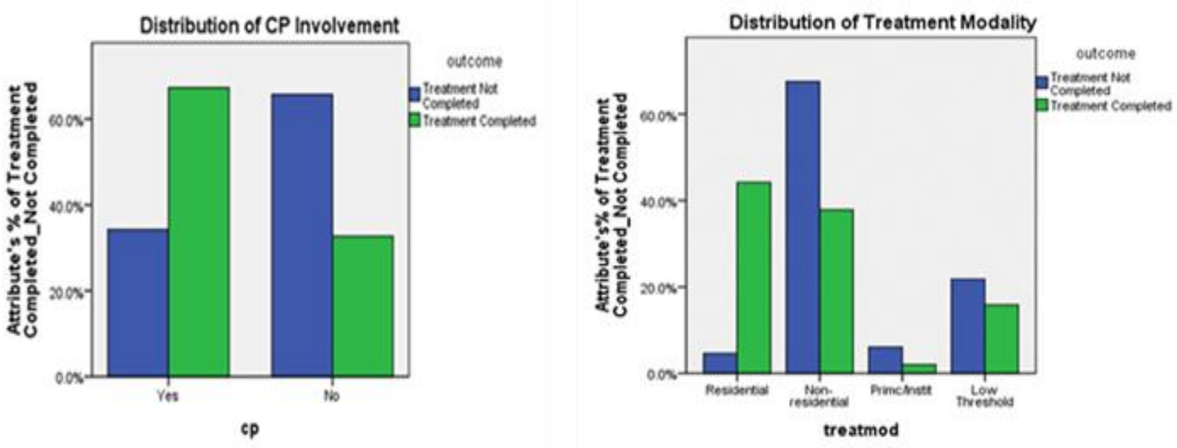

Figure 5: Distribution of CP Involvement and Treatment Modality

Age range appears to be a weak to moderate predictor, IV is 0.0375, although the WOE graph in figure 6 does show some disparity between percentages completing and not completing treatment, this is particularly evident in the 20 to 29 and 40 and over class of the variable. Similarly source of referral has a low IV of 0.0587 indicating it holds limited important information for predicting the outcome variable. WOE for self-referral is 0.1060 and 0.1885 for other professional.

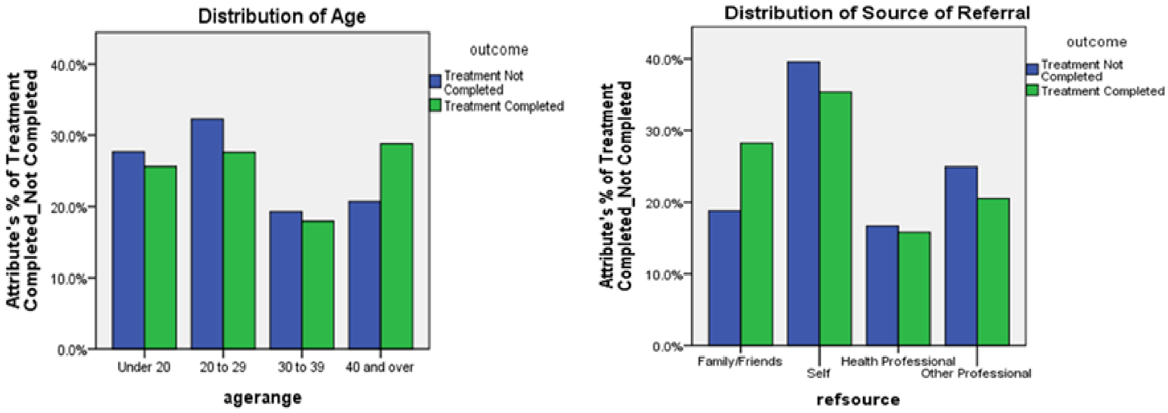

Figure 6: Distribution of Age Range and Source of Referral

The living variable which denotes with whom the subject is living appears to be a weak predictor with an IV of 0.0118. The WOE for the class for living alone is 0.1037 while it is 0.1472 for living with partner/friend and 0.1225 for living other.

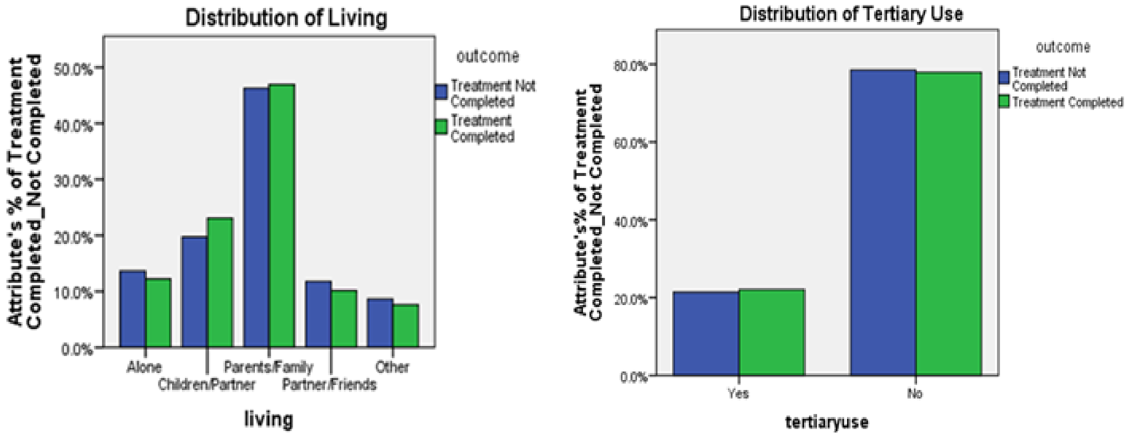

Figure 7: Distribution of Living Status and Tertiary Use

These WOE are low indicating that the variable contains very limited important information towards the outcome variable. The tertiary use variable which identifies those subjects with a third problem substance is one of the weakest predictors as illustrated by both the IV and the WOE graph. The overall IV for the variable is 0.0002. The graph illustrates that there is little difference between those completing and not completing treatment in each of the classes.

The distribution of age of first use from Figure 8 shows that the shape of the distributions is very similar whether the outcome is treatment completed or not completed.

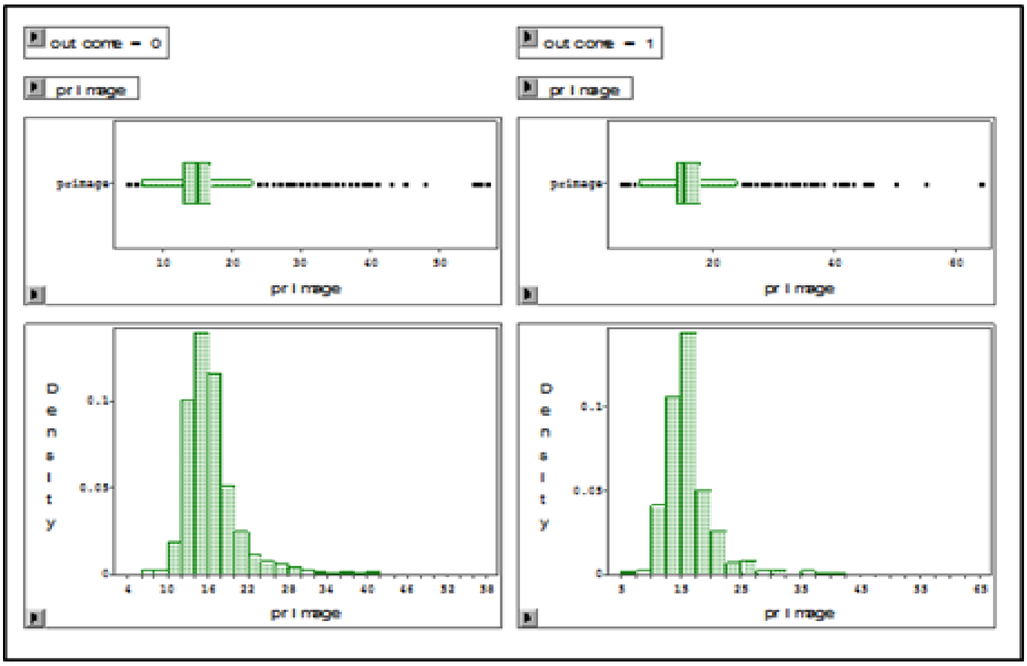

Figure 8: Distribution of Age of First Use

Also, the only class within the variable with a high WOE figure is missing which denotes when the value was missing from the data set, the WOE is 0.6725. Further, the variable is discretized into 10 bins and the information value is 0.0357. Thus, age of first use of the problem substance appears to be less important to classify the target variable.

The variable relevance exploration was undertaken to highlight those variables which have the most important information in classifying the target variable, treatment outcome. Table 5 shows the predictor variables ranked by level of importance according to their information value. Predictive power is presented in the third column the final column shows the results for each variable based on the SAS EM (Statistical Analysis Software (Enterprise Miner)) variable selection node analysis. This analysis is in agreement with the information value for the first five predictors which it indicated to include. It also agrees on the last five predictors which have very weak information values and which SAS (EM) indicated should be excluded. *Refsource* which has a weak to moderate IV was excluded according to SAS (EM) variable selection, while *agerange* which has a lower IV was included. Also variable selection in SAS (EM) indicated that *primdrug* should be included while its IV was quite low, however as evident by figure 2, certain classes within the *primdrug* variable have high WOE indicating there is somewhat important information within the variable.

Table 5: Variable Relevance Results

Variable	Description	IV	Predictive Power	SAS EM
Treatmod	Type of treatment	1.1293	strong	Include
CP	Involvement of family member	0.4505	strong	Include
Educlevel	Highest level of education	0.1020	moderate	Include
Primfreq	Frequency of use	0.0612	weak to moderate	Include
Empstatus	Status of employment	0.0612	weak to moderate	Include
Refsource	Source of referral	0.0588	weak to moderate	Exclude
Agerange	Age group	0.0376	weak	Include
Primage	Age of first using substance	0.0358	weak	Exclude
Primdrug	Primary problem substance	0.0260	weak	Include
Gender	Gender	0.0179	no predictive power	Exclude
Living	Living with whom	0.0118	no predictive power	Exclude
Seconddrugtype	Second problem substance	0.0052	no predictive power	Exclude
Secondaryuse	Presence of a second drug	0.0007	no predictive power	Exclude
Tertiaryuse	Presence of a third drug	0.0002	no predictive power	Exclude

Missing Value Analysis

There were 689 cases with at least one missing value accounting for 15.88% of the total number of cases. 884 values were missing across the entire sample, representing 1.358% of all values. Of the 15 variables which include the target variable, 9 variables had missing values. This is visually represented by figure 9.

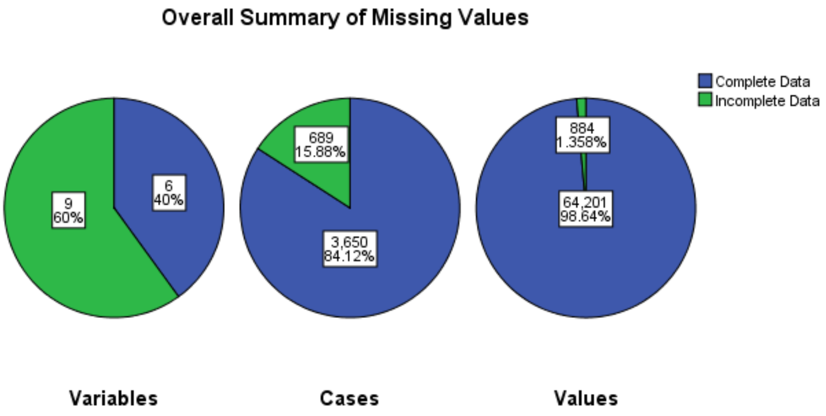

Figure 9: Summary of Missing Values

Each variable was subsequently investigated individually to ascertain the percentage missing within the variable. *Gender, agerange, living* and *refsource* had less than 1% missing. *Empstatus, primfreq* and *cp* have less than 2% missing. *Primage* and *educlevel* had the highest proportion of missing values with 5.37% and 9.84% missing respectively. While 35.13% of values missing in *educlevel* were treatment completers, the balance 64.87% were not treatment completers. These proportions are similar for the *primage* missing values, with 31.33% completing treatment and 68.67% not completing treatment. A significant Little's Missing Completely At Random test, χ^2 = 876.3241, *p* = <.001, revealed that the data were not missing completely at random. Therefore it is possible that they are cases of missing at random or non-ignorable missing as opposed to missing completely at random. For this reason the Expectation-Maximization (EM) Algorithm was used to impute values for the continuous variable *primage*. Missing data were imputed using Missing Values Analysis within SPSS 20.0. The EM algorithm involves an interactive procedure in which it uses other variables to impute a value (Expectation), then checks whether that is the value most likely (Maximization). If not, it re-imputes a more likely value until it reaches the most likely. This method is superior to mean imputation because it preserves the relationship with other variables, which is important for building regression models. The missing indicator method was also considered which involves creating new binary variables which are coded no for not missing and yes for missing values if the value was missing in the original dataset. The original variable would then be imputed with the mode. This method was dismissed as it has been shown to provide an overestimated ROC area. This is because of the inclusion of significant but clinically meaningless missing-indicator variables in the final model (Van Der Heijden et al. 2006). Categorical variables which had missing values were treated by introducing a new value called 'missing'. The effect of adding this new value is that if the fact that the value is missing contains important information about the value itself (non-ignorable missing), the introduction of the new 'missing', value will capture this. If the values are in fact missing completely at random, the cases with this value will not exhibit any deviation from those which do not have the 'missing', value (Gries & Schneider 2010)

Part 2: Regression Analysis

Bivariate Analysis

Bivariate analyses showed that treatment completion was significantly associated with ten variables denoted in table 6. Firstly, being male reduces the odds of completing treatment by 24% (odds ratio 0.76, 95% confidence interval 0.67 to 0.86) compared to being female (reference group). Secondly, those in the age group 40 and over are significantly more likely to complete treatment compared to those in the age groups under 20, 30 to 39 and 20 to 29(the reference group) (odds ratio 1.63, 95% confidence interval 1.38, 1.92). The odds of treatment completion were significantly higher among those educated to third level (third level refers post-secondary education e.g. university). With all other predictors the same a service user educated to third level has odds of completing treatment 1.5 times higher, with 95% confidence interval 1.43 to 1.90. Living alone, compared to living with a partner and children significantly reduces the odds of treatment completion (odds ratio 0.77, confidence interval 0.62, 0.95). Being unemployed or a student significantly reduces the odds of completing treatment by 45% (odds ratio 0.55, confidence interval 0.47, 0.64) and 25% (odds ratio 0.75, confidence interval 0.63, 0.90) respectively compared to being employed. For those service users who attended treatment at residential treatment centres the odds of completing treatment were significantly greater than the odds of completion for those attending non-residential treatment. They have an odds ratio of completion 13 times higher, with 95% confidence interval 10.31 to 16.83.

With respect to the primary substance profile, the odds of completing treatment was significantly lower for individuals who were hypnotic/benzodiazepine users (odds ratio 0.82, 95% confidence interval 0.71 to 0.95), and for opiate users (odds ratio 0.63, 95% confidence interval 0.49 to 0.80). Those whose frequency of use is weekly or less and those who hadn’t used for one month prior to treatment have a higher odds of completing treatment and are significantly more likely to complete treatment compared to those using 2 to 6 days per week. Those who use daily are less likely to complete treatment (odds ratio 0.71, 95% confidence interval 0.61 to 0.82). Finally, the odds of treatment completion varied according to the involvement of a concerned family member or friend. Compared with a service user who did not have a concerned person involved with their treatment, those who did had odds of completing treatment 1.76 times higher, with 95% confidence interval 0.96 to 3.2.

Table 6: Factors Affecting Substance Abuse Treatment Retention, Bivariate Analyses

			Standard	Wald chi-		Odds	
Parameter		**Estimate**	**Error**	**square**	**Pr>ChiSq**	**ratio**	**(Conf. Int)**
Gender	Male	-0.276	0.0657	-4.2	< 0.001	0.76	(0.67, 0.86)
	Missing	-0.0768	0.5799	-0.13	0.8946	0.93	(0.29, 2.97)
	Female	**reference**					
Age Range	under 20	0.0795	0.0813	0.98	0.3282	1.08	(0.92, 1.27)
	30 to 39	0.0848	0.0901	0.94	0.3466	1.09	(0.91, 1.3)
	40 and over	0.4874	0.0832	5.86	< 0.001	1.63	(1.38, 1.92)
	Missing	-0.4247	0.8675	-0.49	0.6244	0.65	(0.09, 3.36)
	20 to 29	**reference**					
Living	Alone	-0.2639	0.1078	-2.45	0.0144	0.77	(0.62, 0.95)
	Partner/Friends	-0.3074	0.1139	-2.7	0.0069	0.74	(0.59, 0.92)
	Parents/Family	-0.1419	0.0798	-1.78	0.0754	0.87	(0.74, 1.01)
	Other	-0.2827	0.1262	-2.24	0.0251	0.75	(0.59, 0.96)
	Missing	-0.5445	0.3453	-1.58	0.1148	0.58	(0.29, 1.13)
	Children/partner	**reference**					
Employment Status	Unemployed	-0.6069	0.0788	-7.7	< 0.001	0.55	(0.47, 0.64)
	Student/training	-0.2856	0.0916	-3.12	0.0018	0.75	(0.63, 0.9)
	Other	-0.3415	0.1174	-2.91	0.0036	0.71	(0.56, 0.89)
	Missing	-0.9573	0.3227	-2.97	0.003	0.38	(0.2, 0.71)
	Employed	**reference**					
Education Level	Primary or less	-0.6709	0.1168	-5.75	< 0.001	0.51	(0.41, 0.64)
	Second Level (J)	-0.3905	0.0955	-4.09	< 0.001	0.68	(0.56, 0.82)
	Second Level (L)	-0.0149	0.0953	-0.16	0.8756	0.99	(0.82, 1.19)
	Third Level	0.4079	0.1469	2.78	0.0055	1.5	(1.13, 2.01)
	Missing	-0.7419	0.1265	-5.86	< 0.001	0.48	(0.37, 0.61)
	Current	**reference**					
Referral Source	Self	-0.5215	0.0811	-6.43	< 0.001	0.59	(0.51, 0.7)
	Health Prof	-0.4633	0.0991	-4.68	< 0.001	0.63	(0.52, 0.76)
	Other Prof	-0.6041	0.0907	-6.66	< 0.001	0.55	(0.46, 0.65)
	Missing	-1.4485	0.4386	-3.3	0.001	0.23	(0.09, 0.53)
	Family/friends	**reference**					
Treatment Modality	Residential	2.5744	0.125	20.6	< 0.001	13.12	(10.31, 16.83)
	Non-residential	-0.2611	0.0838	-3.12	0.0018	0.77	(0.65, 0.91)
	PrimC/Institution	-0.7921	0.1915	-4.14	< 0.001	0.45	(0.31, 0.65)
	Low threshold	**reference**					
Primary Substance	Cannabis	-0.1985	0.0743	-2.67	0.0076	0.82	(0.71, 0.95)
	Hypno/Benzo	-0.5421	0.1594	-3.4	0.0007	0.58	(0.42, 0.79)
	Opiate	-0.4634	0.1255	-3.69	0.0002	0.63	(0.49, 0.8)
	Stimulant	-0.2534	0.166	-1.53	0.1268	0.78	(0.56, 1.07)
	Alcohol	**reference**					
Frequency of Use	Daily	-0.3488	0.0754	-4.63	< 0.001	0.71	(0.61, 0.82)
	Weekly or less	0.1899	0.0967	1.96	0.0496	1.21	(1, 1.46)
	No use past month	0.3674	0.0891	4.12	< 0.001	1.44	(1.21, 1.72)
	Missing	-0.259	0.2884	-0.9	0.3692	0.77	(0.43, 1.35)
	2 to 6 days p/w	**reference**					
CP Involved	Yes	0.5627	0.3048	1.85	0.0649	1.76	(0.96, 3.2)
	No	-0.8126	0.3051	-2.66	0.0077	0.44	(0.24, 0.81)
	Missing	**reference**					

Secondary drug use, second drug type, tertiary drug use and age of first use were not significantly associated with treatment outcome in bivariate analysis.

Multivariate Stepwise Analysis

Automated variable selection using forward, backward and stepwise methods was in agreement on the optimal model for the data. Logistic regression analysis results in table 7 show that treatment completion continued to be significantly more likely among those attending residential treatment centres compared to those attending other types of treatment in multivariate analyses (adjusted odds ratio 12.06, 95% confidence interval 9.12 to 16.05). Treatment completion was significantly more likely among less regular users compared to those using weekly or less (Adjusted odds ratio 1.96, 95% confidence interval 1.59 to 2.24). Those having an education level of primary or less continued to be significantly less likely to complete treatment (adjusted odds ratio 0.44, 95% confidence interval 0.3 to 0.64). In addition being unemployed also continued to be significantly associated with treatment drop out (adjusted odds ratio 0.65, 95% confidence interval 0.53 to 0.79). Those who did not have a family member or friend involved in their treatment continued to be less likely to complete treatment (adjusted odds ratio 0.5, 95% confidence interval 0.26 to 0.96). Other explanatory variables (primary substance, referral source, age range and gender) were not significantly associated with treatment outcome following simultaneous adjustments for other variables.

Iterative Model Building

The process of building the model step by step initially including all predictors and then systematically removing one at a time yielded some interesting results. The variable second drug type was not significant in a bivariate analysis indicating that it is not related to the treatment outcome. However it was found that this variable became significant when the variable treatment modality was added to the model. This suggests that there may be an interaction effect between the two variables. Some additional investigation showed that of those cases where alcohol was the primary problem substance, only 23% indicated a second problem substance. Conversely of those cases whose primary problem substance was an illicit drug, 63% indicated a second problem substance. The data was therefore subdivided into cases whereby alcohol was the primary problem substance and illicit drugs were the primary problem substance.

Table 7: Factors Affecting Substance Abuse Treatment Retention, Multivariate Analysis

Parameter		Odds ratio (Conf. Int)
Intercept		1.45 (0.68, 3.11)
Living	Alone	0.86 (0.66, 1.13)
	Partner/Friends	0.88 (0.66, 1.15)
	Parents/Family	1.01 (0.82, 1.26)
	Other	2.04 (1.51, 2.76)
	Missing	1.03 (0.48, 2.14)
	Children/partner	**reference**
Employment Status	Unemployed	0.65 (0.53, 0.79)
	Student/training	0.91 (0.65, 1.29)
	Other	0.94 (0.7, 1.25)
	Missing	0.86 (0.42, 1.72)
	Employed	**reference**
Education Level	Primary or less	0.44 (0.3, 0.64)
	Second Level (Junior)	0.66 (0.47, 0.92)
	Second Level (Leaving)	0.69 (0.48, 0.98)
	Third Level	0.95 (0.6, 1.5)
	Missing	0.61 (0.41, 0.91)
	Current	**reference**
Treatment Modality	Residential	12.06 (9.12, 16.05)
	Non-residential	0.8 (0.66, 0.96)
	Primary care/Institution	0.45 (0.29, 0.69)
	Low threshold	**reference**
Frequency of Use	Daily	0.91 (0.76, 1.08)
	Weekly or less	1.37 (1.1, 1.71)
	No use past month	1.96 (1.59, 2.41)
	Missing	0.74 (0.37, 1.44)
	2 to 6 days per week	**reference**
CP Involved	Yes	0.93 (0.49, 1.8)
	No	0.5 (0.26, 0.96)
	Missing	**reference**

Logistic regression was conducted on each of the two datasets. All variables which were significant in the full dataset were also significant in the two sub datasets. The only difference was that second drug type was significant for illicit drug cases but not for alcohol cases. All other predictors which were shown to contribute to the model using stepwise selection were shown to be relevant through the iterative process.

Collinearity

The generalised variance inflation factor (GVIF) showed that employment status and education level had the highest GVIF at 5.656 and 4.494 respectively. It is possible that these two variables are correlated as the possibility of being employed is likely higher for those with higher levels of education. Both variables are indicators of socio-economic status, therefore including both in the model may be redundant as they don't contribute independent pieces of information.

Final Candidate Models

Three models were selected as the final candidate models. Model 1 was derived from using bivariate analysis and subsequently retaining predictors which were still significant in multivariate analysis. The variables including; *Living, Educlevel, Empstatus, Treatmod, Primfreq* and *CP*. Model 2 was derived from the iterative model building process, whereby the variable second drug type was included even though it was not shown to be related to the outcome variable in bivariate analysis. The variables include; *Living, Educlevel, Empstatus, Treatmod, Primfreq, CP* and *Seconddrugtype*. Model 3 was derived from model 2 but with the exclusion of the employment status variable due to its high GVIF value. The variables include; *Living, Educlevel, Treatmod, Primfreq, CP* and *Seconddrugtype*.

Testing Overall Model Fit

Table 8 shows the results of goodness of fit tests for each of the three final models. The Hosmer Lemeshow tests the hypothesis that there is no significant difference between the observed and predicted values. The P-value for model 1 and model 2 and model 3 of 0.3835, 0.4355 and 0.5442 respectively. Therefore the null hypothesis cannot be rejected. The large p-values signify that there is no significant difference between the observed and predicted values of the outcome of any of the models. This indicates that all three models fit quite reasonably. The R^2 values, although quite low do offer a comparison between the three models, model 2 has the highest R^2 value at 0.360.

Table 8: Goodness of Fit Statistics

Overall Model Fit: Goodness of Fit Statistics			
Tests	Model 1	Model 2	Model 3
Inferential Test			
Hosmer Lemeshow	p = 0.3835	p = 0.4355	p = 0.5442
Descriptive Measures			
Nagelkerke R^2 Index	0.356	0.360	0.356

Evaluation of Final Models

Table 9 depicts assessments of the predictive power of the models. The misclassification rate is 27% for model 2 and model 3, the value is 28% for model 1. This indicates that both model 2 and model 3 are slightly better as classifying the outcome correctly. Somers Dxy rank correlation indicates that the predictive ability of all three models is above the 0.5 level. The statistic for model 2 is slightly higher at 0.587 compared to 0.579 for model 1 and 0.583 for model 3. The Brier Score is similar for all three models with models 2 and 3 performing slightly better at 0.180, which is closer to zero than 0.181 indicating model 2 and model 3 are superior in terms of calibration, the statistical consistency between the predicted probability and the observations.

Table 9: Predictive Power Tests

Model Comparison Results: Predictive Power Tests			
Tests	Model 1	Model 2	Model 3
Misclassification Rate	28%	27%	27%
Somers Dxy	0.579	0.587	0.583
Brier Score	0.181	0.180	0.180
c-statistic	79.0	79.4	79.2

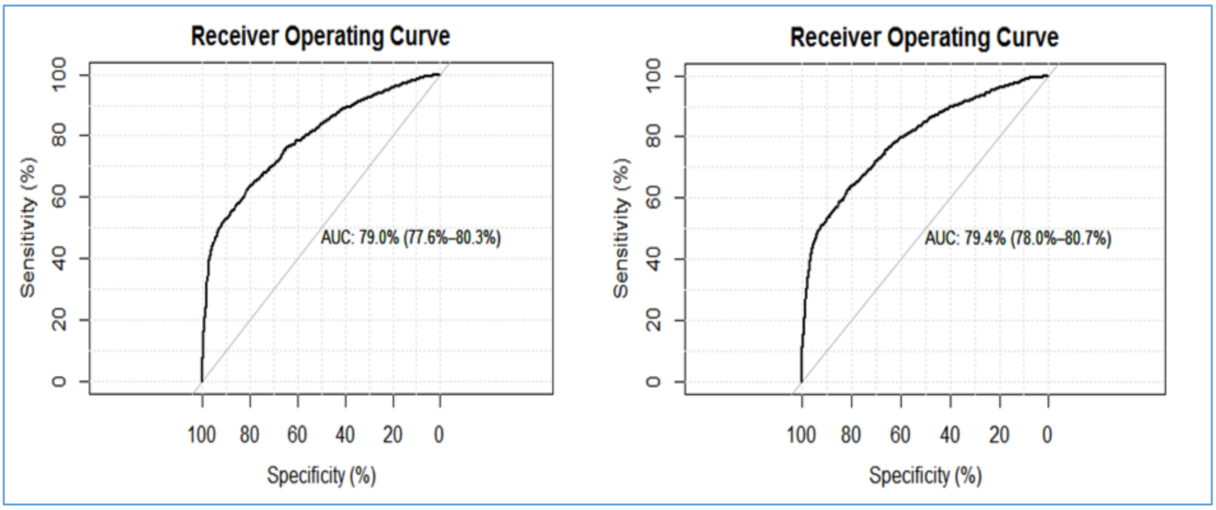

Figure 10: Receiver Operator Curve Model 1 and Model 2

The area under the receiver operator curve given by the c-statistic indicates that model 2 and model 3 are slightly better at classifying the outcome variable as the c-statistic is 79.4 and 79.2 respectively compared to 79 for model 1. The Receiver Operator Curve (ROC) for each model illustrated in figure 10 and figure 11 further demonstrate this along with 95% confidence intervals for the c-statistic.

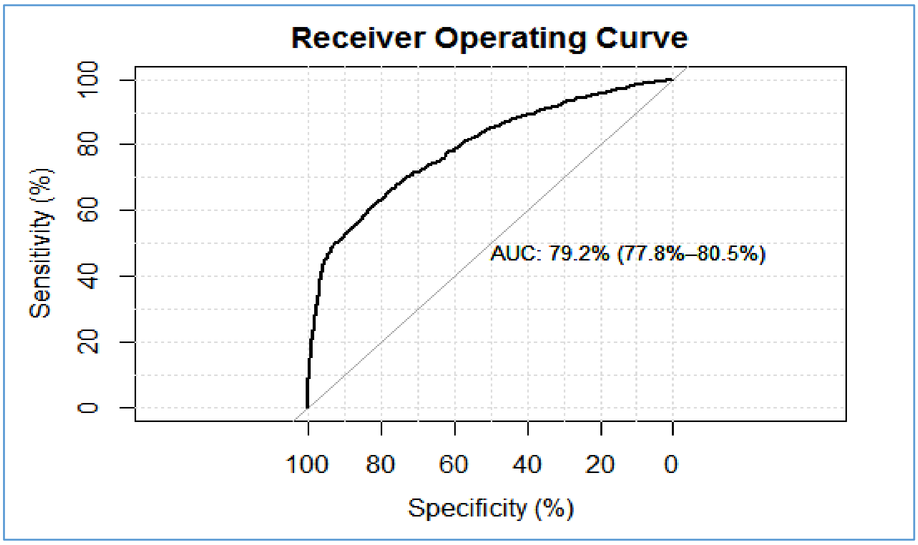

Figure 11: Receiver Operator Curve Model 3

Selection of Optimal Model

The analysis thus far has shown that all three candidate models perform reasonably well on the metrics which were tested. It does appear that model 2 and model 3 consistently outperformed model 1. A final measure to evaluate the models is to plot the conditional distribution of the response given the fit of the model. This plot shows that if the lines are similar, the model is reproducing the data in that direction, if they differ it is an indication that the model may be miss-specified.

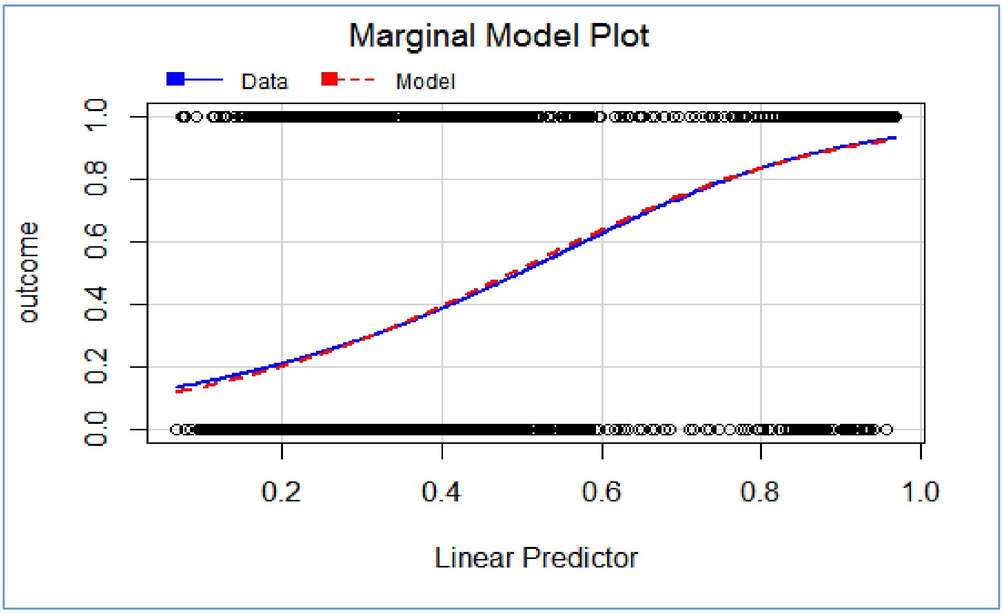

Figure 12: Marginal Model Plot - Model 1

Figure 12 shows the outcome plotted against the linear predictor. The real data is represented by the solid blue line and the model is represented by the dashed red line. . The data dips below the solid line from about 0.1 to 0.2, it then goes above the solid line from 0.5 to 0.6. There is only a slight deviation between the blue line representing the data and the red line representing the model meaning that the model relating the parameter to the predictors has a reasonably close resemblance to the true relationship.

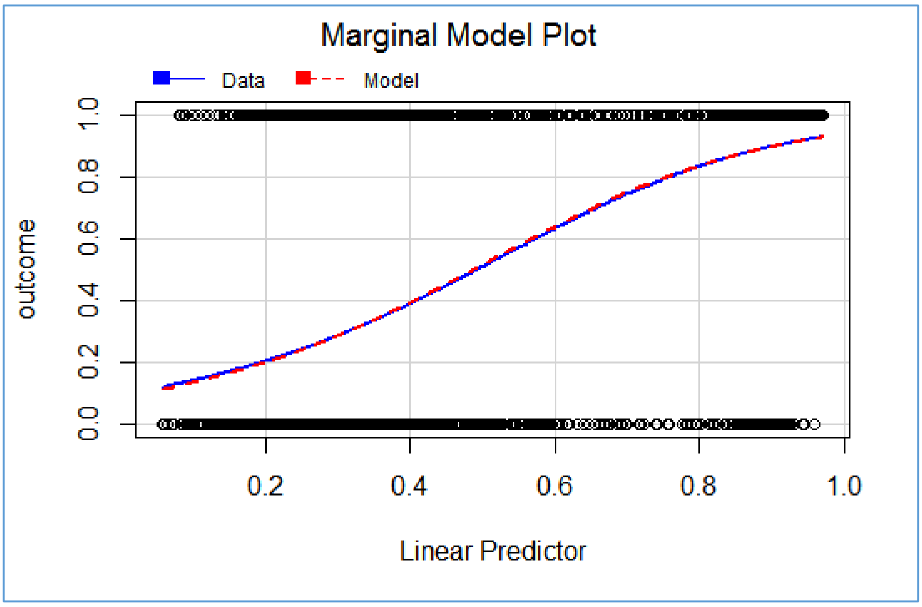

Figure 13: Marginal Model Plot - Model 2

Similarly Figure 13 shows this relationship for model 2. We can see that there is an improvement on the fit of model 1. The lines representing the model and the data are more closely aligned with less visible deviation between the data and the model. Finally the marginal model plot for model 3 was examined and is shown in figure 14. It is evident that the model fits the data perfectly as both lines are exactly aligned to one another. The red line representing the model is also more visible coming through the blue line. There does not appear to be any deviation indicating that the model closely resembles the true relationship. Model 3 performed well in terms of goodness of fit statistics and predictive power. Although it was slightly lower than model 2 on several of the metrics, figure 14 illustrates that it is an excellent fit to the data. Furthermore collinearity diagnostics show that GVIF values are all close to 1, indicating that there is no collinearity between variables. The inclusion of the second drug type predictor in both Model 2 and Model 3 has shown to improve performance in both models over Model 1. For this particular dataset which includes both alcohol and illicit drug cases, the inclusion of second drug type produces significantly stronger models. While the GVIF for employment status is not exceptionally high at 5.656, and moreover its inclusion in Model 1 and 2 did not have a substantial effect on standard errors, the marginal model plot does show a slightly better fit to the data with its exclusion from Model 3. Therefore in the interests of building the most parsimonious model to explain the outcome variable, Model 3 appears to be the best choice.

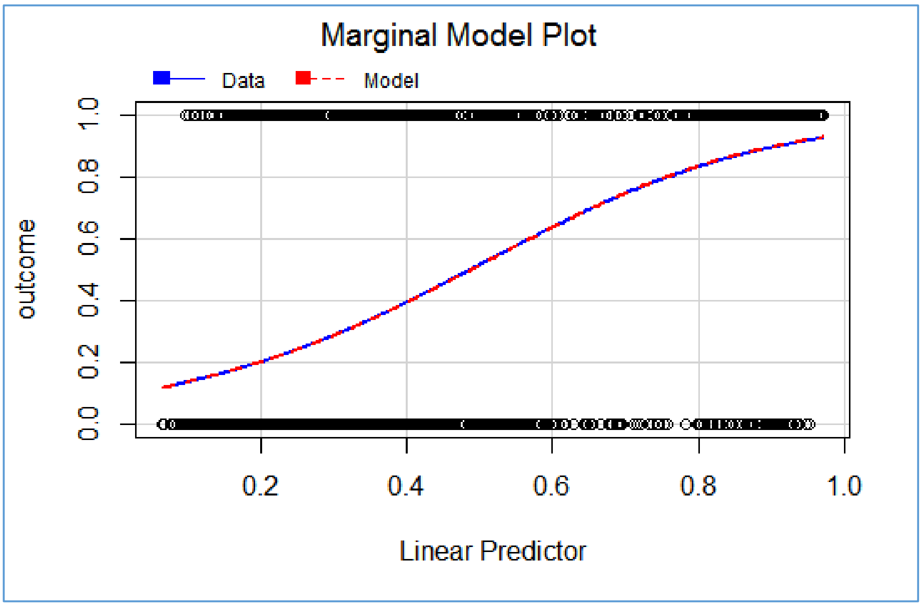

Figure 14: Marginal Model Plot - Model 3

Regression Diagnostics

Residuals

As non-constant variance is always present in the logistic regression model, response outliers are difficult to diagnose. Therefore the residual analysis focused on the detection of model inadequacy in the covariate space only. This suggests that if the model is correct and there is no significant incorporation of outliers, a LOWESS smooth of the plot of residuals against the linear predictor should result in approximately a horizontal line with zero intercept. Any significant departure from this suggests that the model may be inadequate and potential outliers may be having a dramatic effect on the fit of the model (Sarkar et al. 2011). The LOWESS smooth of the deviance residuals plotted against the linear predictor for model is shown in figure 15. The LOWESS smooth approximates a line slightly away from zero, a LOWESS span of 0.8 was used. This graph indicates the potential presence of outliers and influential observations, although it is not a useful measure to detect the specific observations. Furthermore it is evident that there are observations lying outside of 2 in absolute value.

Leverage Values, Outliers and Influential Observations

Table 10 displays cases which had the highest values of Studentized Residuals, Leverage values and Cook's Distance. There is no value of Cook's Distance which is greater than 1. In addition all values of DFbeta were examined, there were no values greater than 0.4, only 3 values were greater than 0.3. Values of Dfbetas less than one are generally considered acceptable and do not signify highly influential observations in the data.

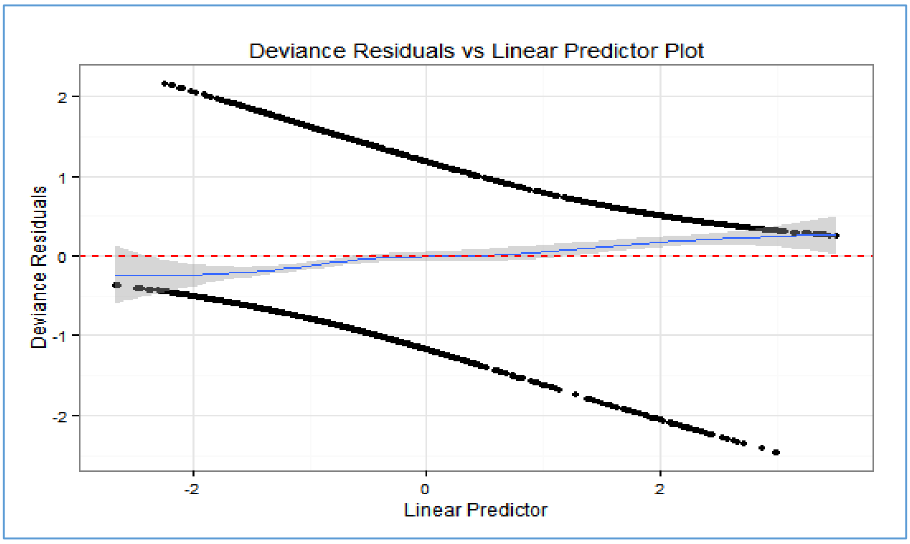

Figure 15: Deviance Residuals versus Linear Predictor

There are no observations with residual values greater than 2.5 in absolute value; however those which are close to 2.5 were examined. Three of the observations have leverage values greater than 0.05 which were also investigated. Figure 15 shows Cook's distance plotted against the leverage values. The highest concentration of cases lies in the lower left quadrant of the graph, showing that the majority of cases are low on both parameters. The observations close to or outside of 0.06 on leverage and outside of 0.006 on Cook's distance are evident in the graph.

All nine cases were removed individually and the effect on parameter estimates was examined. The removal of each case individually did not alter the results substantially. Therefore all 9 cases were removed and systematically added back into the dataset and the effect examined as each new case was added back in. It was found that the removal of case 735 had a substantial effect by changing the sign of the coefficient on the variable CP (concerned person involved in treatment). This categorical variable has 3 levels; yes, no and missing. The yes level changed from a negative to a positive coefficient. The original data for each case was further examined. Case 735 had 4 missing values case 807 had 3 missing values which may account for their significant influence on the coefficients. Case 490, 605, 780, 1313, 2036 and 4224 and 4322 did not have more than one missing value.

Table 10: Cases With Highest Values of Residuals, Leverage and Cook's Distance

Case Number	Studentized Residuals	Leverage	Cooks Distance
490	-2.4727	0.0021	0.0405
605	-2.4736	0.0012	0.0305
735	1.1543	0.0602	0.0483
780	-1.1455	0.0599	0.0477
807	1.0851	0.0579	0.0437
1313	1.9638	0.0276	0.0777
2036	-2.4302	0.0021	0.0383
4224	1.8807	0.0328	0.0773
4322	1.8047	0.0352	0.0739

All cases which were removed exhibited a unique pattern of characteristics of the outcome and predictor variables, indicating that the model was not fitting them adequately.

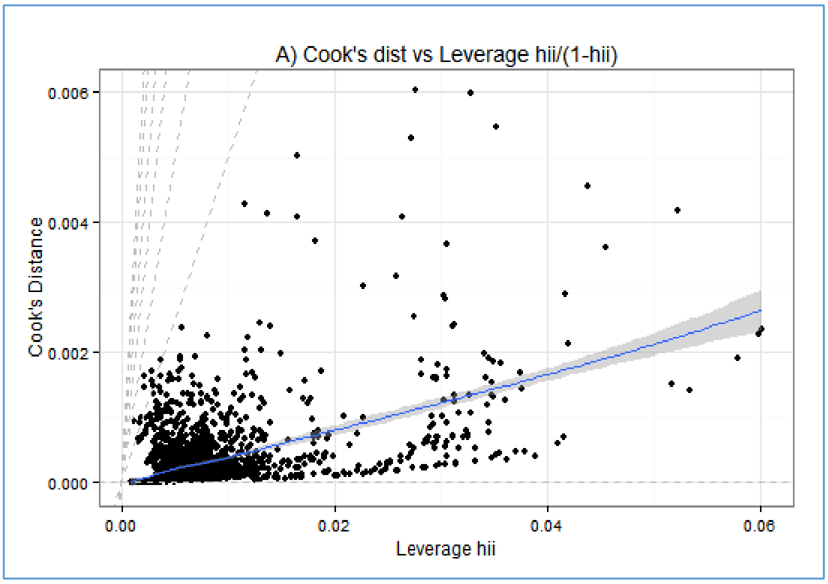

Figure 16: Cook's Distance versus Leverage Values

The most notable change to the coefficients with the removal of the nine cases was on the CP variable. The coefficient for the yes level within the variable changed from -0.09475 to +0.01059. The "no", level of the variable also changed from significant to insignificant. The coefficients for the living variable (level: missing) also changed from a positive to a negative sign, +0.04131 to -0.0040. The coefficient for level "parents/family", within the living variable increased by approximately 20%. The most marked effect on the education level coefficients was on third level education, with a change from -0.0583 to -0.01059. Within the second drug type variable, "opiate", decreased by about 30%. This level of the variable also became more significant than it had been previously. The treatment modality variable saw a decrease in the primary care/institution coefficient, this level of the variable became slightly less significant. The significance level of all other variables remained the same as before the influential cases were deleted. Figure 17 shows the Cook's distance plotted against leverage values following removal of the nine influential cases. It can be seen that no observations are lying outside of or close to 0.06 on leverage.

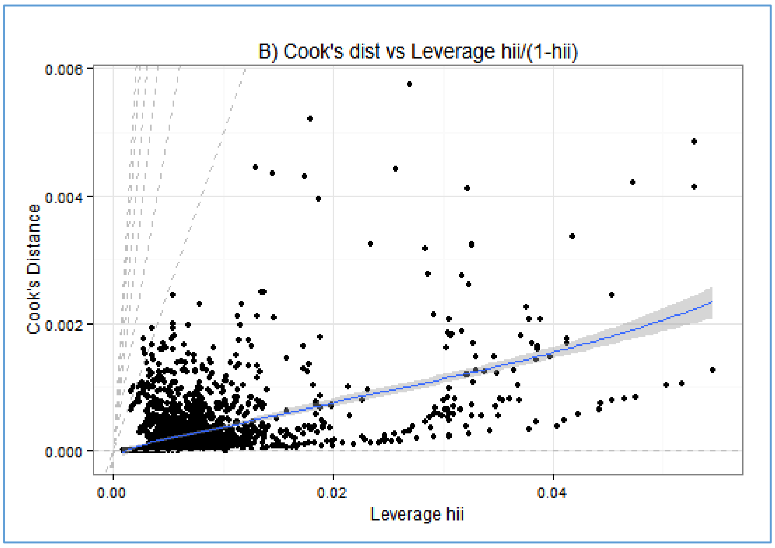

Figure 17: Cook's Distance versus Leverage Values (Influential Observations Removed)

Logistic regression was conducted on the new dataset excluding the nine influential cases. The model performed slightly better on several of the performance metrics. The R^2 value increased to 0.361. The c-statistic increased to 0.794 from 0.792. The value for Somers Dxy is 0.588 increased from 0.583. The Brier score reduced from 0.180 to 0.179. The misclassification rate of the model is remained at 27%. Therefore removal of the influential cases had a positive impact on the overall fit of the model. Results for this model 3 can be viewed in Appendix 8, while Appendix 9 shows results with influential cases removed.

Final Model Validation

Bootstrap resampling validation using the 'boot', method in the RMS package within R Studio was used instead the split-sample approach in order to maximise the sample size for model development, 200 bootstrap repetitions were taken (Steyerberg et al. 2001). Table 11 summarises the estimates of four common model performance statistics: the area under the receiver operating characteristic curve (AUC), the Somer's Statistic (Dxy), Nagelkerke's R^2, and the Brier Score. The optimism corrected estimates on all parameters indicate that the model still performs well under internal validation procedures. A logistic model with an optimism corrected AUC of greater than 0.7873 has a reasonable capability of discriminating between the two states of the binary outcome variable.

Table 11: Optimism Corrected Model Validation Statistics Using Bootstrap Resampling

Estimate	Optimistic Estimate	Optimism Correction	Optimism Corrected Estimate
AUC	0.794	0.0067	0.7873
Dxy	0.588	0.0130	0.5746
Brier Score	0.1795	-0.0024	0.1819
R-squared	0.361	0.0109	0.3499

Further Analysis

The results show that treatment modality is the strongest predictor of treatment retention, to investigate this finding further, a logistic regression analysis was performed to ascertain if there were differences between those entering residential treatment compared to other types of treatment. The outcome variable treatment modality was set to 1 = residential, or 0 = other forms of treatment. The results indicate that those aged 40 and over have an odds ratio of entering residential treatment 2.61 times higher with a 95% confidence interval between 2.1 and 3.24. Those educated to third level have an odds ratio of 3.08 of entering residential treatment with a confidence interval between 1.94 and 4.99. Those educated to leaving certificate level having an odds of 2.5 times, with a confidence interval between 1.72 and 3.92. A significant negative association was revealed between those using daily and the outcome variable, (odds ratio 0.69, confidence interval 0.56 to 0.85). Poly drug use was also shown to be significantly associated with attending residential treatment, those with a second problem substance have an odds ratio of 1.65, with a confidence interval between 1.32 and 2.07, while those with three problem substance have an odds ratio of entering residential treatment of 2.36, with a confidence interval between 1.84 and 3.03. All primary drugs were significantly negatively associated with the outcome variable. The Likelihood Ratio Test resulted in chi-squared 856.53 on 19 degrees of freedom, p <0.001. The c-statistic is 0.785 demonstrating that the model has achieves reasonable discrimination between the outcome, those entering residential treatment and those entering other forms of treatment. A brier score of 0.146 shows that the model provides a good measure of agreement between observed and predicted probabilities as the value is close to zero. The Hosmer Lemeshow goodness of fit test indicated that this model is a good fit, p-value 0.3148. Therefore the null hypothesis that there is no significant differences between the observed and predicted values cannot be rejected, indicating that the model is a good fit to the data. Please see Appendix 10 for detailed results.

Chapter 5: Discussion, Limitations & Conclusion

This section discusses the implications of results presented in chapter 4. It also highlights the strengths and shortcomings of the research undertaken. Recommendations are put forward for service providers and suggested orientation for further research which build on the findings of this study is outlined. Finally a conclusion is given.

Previous research has demonstrated that retention in substance abuse treatment is positively associated with favourable post treatment outcomes such as, reduced substance use, reduced crime levels, and improvement in social functioning. Rates of retention can be improved by better understanding factors associated with premature drop out. The aims of the present study were to identify and assess factors associated with substance abuse treatment retention. While a number of factors associated with treatment retention are identified, results also highlight that drop out and retention is an index that captures the complexity of many factors. Results show that overall in the Cork and Kerry region of southern Ireland, retention rates of 47% were achieved. This is largely in line with previous studies which found retention rates of approximately 50% (Stark 1992).

Findings were contrasted with previous research, although previous studies may have used differing interpretations and methods to measure variables, it is nevertheless a worthwhile comparison. With respect to demographic characteristics, studies in the USA have reported longer durations of treatment for older drug users and higher retention rates (Wickizer et al. 1994). Evidence from a UK-based retrospective cohort study showed that younger drug users (aged 10 to 19 years) were more likely to drop out than their older counterparts (Beynon et al. 2006). Here, bivariate analysis showed that the odds of retention was significantly higher for those aged 40 and over. However, the multivariate analysis showed that age was not significantly associated with treatment retention once other covariates were simultaneously adjusted for. Further analysis to ascertain if there were significant differences in groups entering residential treatment as opposed to those entering other treatment modalities showed that those aged 40 or over were more likely to enter residential treatment and therefore more likely to be among those completing treatment as completion rates for this group were close to 90%. With regard to gender differences in retention rates, studies have shown largely mixed results, some highlight significant differences while others fail to find any difference. Bivariate analysis in this study suggested males had a higher odds ratio of drop out than females, this result was not significant when other variables entered the model. The ratio of males to females in this study was 2:1, this is somewhat encouraging as European wide ratios for males and females entering treatment are reported as 4:1 (United Nations Office of Drugs and Crime 2012).

There is a large body of evidence which shows the relationship between socio-economic status and health. Over 45% of the study population were unemployed and over 40% were educated to junior certificate level which equates to leaving school at approximately aged 15. Low employment levels and low educational attainment generally coexist with poorer health outcomes and poorer substance abuse treatment retention and outcomes specifically (Sayre et al. 2002; Knight et al. 2001; Simpson & Joe 1993). Of those employed 57% completed treatment while only 41% of those unemployed completed treatment. It was demonstrated that being unemployed was negatively associated with treatment completion, providing further support for previous findings which have shown a positive correlation between employment and treatment completion. It is also worth noting that this data relates to a period from 2008 to 2013 when unemployment rates in Ireland increased from 5% to almost 15%, therefore this finding may not hold if tested on data from a different period.

There is some evidence in the literature which indicates that social support for the person undergoing treatment is an important factor in determining treatment retention and outcomes (Dobkin et al. 2002). Researchers have found functional support to play an important role in the prevention of premature termination (Westreich et al. 1997). Stable living arrangements and involvement of a family member in the treatment episode offered indicators of social support in this study. Multivariate analysis revealed that the only significant level of this variable was (other), this level of the living status variable refers to prisoners and homeless people. This group had an odds ratio of retention 1.97 times higher (with a confidence interval between 1.46 and 2.67).   This is perhaps not a surprising result as people accessing services within prison may be more motivated for change, also homeless service users may have reached ‘rock bottom’, with their substance use whereby they are more likely to be retained in treatment. Involvement of a family member in the service user’s treatment episode was shown to be a significant indicator of completion in bivariate analysis but was not significant in the final model. Regarding referral source, previous findings have suggested a negative association between those referred through the criminal justice system and treatment completion (Beynon et al. 2006; Wickizer et al. 1994). Here the route of referral was not significantly associated with treatment drop out in the multivariate analysis.

Bivariate analysis indicated that cannabis, hypnotic/benzo and opiates were negatively associated with treatment completion. Following adjustment for other variables, primary use of these drugs was not found to be significant, a result similar to that reported by Wickizer et al. (1994). The separation of the dataset into observations reporting: 1) alcohol, and 2) illicit substances indicated that the optimal models to explain the relationship between the outcome and independent variables were the same for both groups. The only exception was the second drug type variable which was significant for the model constructed for illicit drugs but not for that constructed for alcohol. This finding is not surprising considering that a secondary problem substance was reported by over 60% of illicit drug users and only 30% of alcohol users. This also highlights the fact that although treatment retention is the result of many complex factors, it does not depend specifically on drug of choice.

The data used for this study lacked any clear indicators as to the severity of the substance use problem, such as the Addiction Severity Index frequently used in studies in the USA. The best available proxy measure was frequency of use of the primary problem substance. The results indicate that those using less often had a significantly higher odds of completing treatment. This may be the case due to less severe addiction problems, as generally speaking, more frequent drug use and a higher degree of drug dependence have been linked to treatment dropout (Green et al. 2002; Mertens & Weisner 2000). An alternative explanation of this finding is that this particular group who were using less frequently were generally more ready for treatment indicated by a reduction in use prior to entry. In addition polysubstance use may also act as a proxy measure of problem severity. Poly substance use was prevalent in the sample with more than 40% of the population reporting at least two problem substances. The presence of a secondary or tertiary problem substance were not shown to be significant predictors of treatment completion in bivariate or multivariate analysis. Further analysis did reveal that those reporting polysubstance use were more likely to enter residential rather than other forms of treatment. This may indicate that more complex drug usage tends to lead in the direction of residential treatment. Cases whereby presence of a second problem substance was reported were significant and showed interesting results. Secondary use of hypnotics, benzodiazepines and opiates were significantly negatively related to retention. It was noted in the data description that the percentage of those reporting hypnotics/benzodiazepines as secondary problem substances was higher than those reporting it as a main problem substance. Duration of drug usage may also indicate severity of the problem, however it was not available in the dataset, and it is possible that initiation of use at a younger age would indicate a more severe problem. However, age of first use was not shown to be related to the outcome in bivariate or multivariate analysis. Duration of use may be a useful piece of data to collect from service users for future research.

The most notable finding was with regard to treatment modality. Residential treatment exhibited substantially higher completion rates relative to other forms of treatment. Attendees at residential treatment had completion rates of 89.5% compared to 33% for non-residential, 39% for low threshold and 23% for primary care and institutions. Those attending residential treatment had an odds 13.5 of retention times those in other in other treatment settings. This finding has been reported in many studies but attempts to understand this have drawn limited conclusions (Wickizer et al. 1994). It has been postulated that strong clinical norms and peer expectations characteristic of residential programmes make it more difficult to drop out of treatment prematurely. In addition dropping out of other forms of treatment merely involves not presenting for a scheduled appointment, leaving residential treatment involves a more deliberate action (Wickizer et al. 1994). Further analysis revealed that those entering residential treatment were significantly more likely to be older, have attained higher level of education and report polysubstance use. Furthermore, those reporting daily use were significantly less likely to enter residential treatment. This may indicate that although those entering residential treatment have more complex substance use histories, they are more ready for change indicated by lower usage at time of entry therefore making them more likely to complete treatment. It has also been put forth that the effectiveness of programs may be influenced by client characteristics, for example, some clients respond more favourably to certain modalities according to Hser et al. (1998), indicating that an approach whereby service user is matched to treatment type may yield better results.

Pre-screening variables through examination of Weights of Evidence (WOE) and Information Value (IV) proved a useful preliminary analysis tool in this study. Five of the variables that were included in the final candidate models were shown to be of relative importance in predicting the outcome variables during pre-screening. Although there were only fourteen candidate variables in this study, this practice showed that investigating variable relevance prior to model building is a worthwhile endeavour, particularly if there are large numbers of potential candidate variables, the technique offers a mechanism to discard the weakest predictors. However, it should also be noted that two of the predictors which were classed as having no predictive power at the preliminary stage ended up in the final candidate models. Therefore this method should be used with caution and not serve as an alternative but instead an additional measure to prior knowledge of potentially important variables, established through a thorough review of pertinent literature. In addition reporting of results to service providers to ensure they are sensible would confirm that results are not just spurious.

Bagley et al. (2001) criticized many medical studies using logistic regression for their failure to report validation analysis, regression diagnostics or goodness of fit measures. One of the strengths of this study was outlining of the model fitting procedures in detail along with in-depth diagnostic procedures. In addition the use of a validation method further enhanced the proposed reliability of the model, an optimism corrected estimate of the c-statistic of 0.7873 indicates that the final model has a reasonable capability of discriminating between the two states of a binary outcome variable. Performing multivariate analysis based on variables selected through the bivariate analysis method is a commonly used procedure by many investigators (Sun et al. 1996). However it has been criticized because potentially important variables may be wrongly rejected when the relationship between the outcome and independent variable is confounded by another variable. This fact was illustrated in this study, the second drug type variable was not significant in bivariate analysis but was found to be significant during the iterative model building process and in the full model fit.

Problems associated with the use of routine data, such as, inconsistencies in the manner in which information is recorded, are somewhat mitigated in this study due to the use of nationally agreed reporting protocols, a standardised coding framework and data validation checks which are performed by the Health Research Board. Furthermore, the large sample size increases the power of the results and reduces the likelihood of a Type II error. However, it should also be noted that the study relies on self-reported data and the accuracy of such reports cannot be verified. The standardisation of the drug treatment and data reporting across Ireland mean that the results from the Cork and Kerry region can be generalised to other areas of the country.

Limitations

The advantage of using secondary data is that it provides a ready data source to explore answers to interesting research questions, however one of the major limitations within this dataset was the lack of a unique identifier for subjects. This effectively means that the system tracks treatment episodes as opposed to individuals; therefore this study was confined to new attendees to substance misuse treatment services to ensure that each individual was only entered into the study once. Had this measure not been undertaken the same individual could have been entered into the study several times. Due to this constraint previously treated service users were not included; therefore, the data does not reflect the profile of all service users. If all attendees had been included, the age profile for example may have been older, as it is more likely that older attendees would already have had earlier treatment episodes.   Anecdotal information suggests that multiple prior admissions to substance abuse treatment indicate “bottoming out”, which should result in greater retention. This effect could not be captured in this study. Yet, research results are mixed regarding the association between prior substance abuse treatment and retention, with some finding a positive relationship and some finding no relationship (Stark 1992; Mertens & Weisner 2000). On the other hand this did allow for treatment modality as a predictor of treatment completion to be evaluated. If re-attendees were included, it is likely that an individual’s entire treatment episode could be comprised of treatment within different modalities, rendering the inclusion of this variable impractical. A further limitation of the study is the National Drug Treatment Reporting System itself. It is a difficult undertaking to obtain detailed information on all treatment contacts made to all treatment agencies. Although close contact is maintained between the Health Research Board and all treatment agencies to encourage full compliance there is no way of ensuring that this compliance is 100%.

Finally, acquisition of the data was a slow process, all treatment providers concerned had to be briefed on the nature of the study and provided with study protocols. Those who were in agreement to giving consent for release of data pertaining to their services were requested to contact the Health Research Board directly. This caused considerable delays in beginning analysis and limited the breadth of analysis that could be undertaken within the constrained time scale. As mentioned previously there was a lack of adequate measures of service user addiction severity in the data, there may well have been a large variation on the severity of substance use problems within the study population which may account for retention rates. Therefore frequency of use had to be used as a proxy measure in the absence of any explicit measure.

As well as data specific limitations, several methodological limitations of this study have to be addressed. Firstly interaction terms were not considered for this study. Although there was detection of a possible interaction between treatment modality and second drug type, measures were taken to explore this and the variable was ultimately included in the final model. Second drug type for example was not independently related to the outcome as shown in bivariate analysis. However it only became significant once the treatment modality variable was entered into the model. The decision to exclude interactions was due to the large number of possible interactions for fourteen predictor variables. If all were examined, this would have amounted to an additional 2^{14} potential predictors to be examined.

Missing data in the study were shown to be not missing completely at random, therefore complete case analysis was inappropriate as it is expected to produce biased results when data are not MCAR (Van Der Heijden et al. 2006). Instead single imputation using the Expectation Maximisation (EM) algorithm was used for the continuous predictor and a constant ‘missing’, was used for categorical predictors. The addition of another level to nine of the predictor created additional complexity in the variables and made interpretation of coefficients more difficult. A superior method for the treatment of missing values would have been multiple imputation using the Markov Chain Monte Carlo (MCMC) algorithm as it is more powerful than the Expectation Maximisation (EM) algorithm. Also one of the shortcomings of the EM algorithm is the assumption of a parametric a priori distribution.

Internal validation was conducted through the use of bootstrap resampling, which is the preferred approach for internal validation of prediction models (Steyerberg & Harrell 2015). However, the ultimate test for models developed on a sample dataset is to test them on unseen data, this would indicate if model predictions hold true for example in different treatment centres or for service users entering treatment more recently. This was not possible within the confines of this study but it is a procedure which should be undertaken for future research.

Impact of Findings and Orientation for Further Research

The findings of this study illustrate the complexity of substance abuse treatment and factors which lead to retention or drop out. There is a clear advantage to monitoring retention rates as it allows benchmarking against rates reported both nationally and internationally. It can also operate as marker of operational effectiveness and well as a gauge of the overall capacity of local services to meet the needs of the treatment population. Service providers should focus on four key areas in order to increase retention rates:

1) Establish and monitor performance against retention targets.

2) Investigate programme level factors which may influence retention rates, such as counsellor to service user ratio, caseload per counsellor and provision of auxiliary support services.

3) Ensure auxiliary services for family members are in place to ensure adequate social support is provided to the service user.

4) Identify those with higher levels of substance usage at the earliest stage possible and take steps to retain them in treatment, this may require the provision of additional auxiliary support services or more frequent treatment contact.

Clearly the environment in which treatment is received along with the various process measures which comprise of the entire treatment episode, have a considerable bearing on the outcomes. Further research is required to investigate the link between organisational level variables, such as scope of services provided, programme policies and processes, caseload sizes and their interrelationship with individual level factors such as those examined in this study. Since the introduction of the Health Identifiers Act 2014, it is one of the aims of the National Drugs Strategy to put in place such a unique identifier to facilitate the development of reporting systems. The absence of such an identifier to date has been a key constraint for undertaking in-depth research of this national database as individual histories cannot be tracked (Connolly & Long 2014). This development will facilitate the use of this rich data source as a means of answering research questions relevant to policy makers and service providers.

Conclusion

Despite the limitations outlined above, this study was a successful initial step in describing the treatment population accessing services in the Cork and Kerry region for the first time. Furthermore, it identified factors that are related to treatment retention, including treatment modality, frequency of substance use, education level, living status, secondary substance used and the involvement of a concerned family member in the treatment episode. Additionally it was highlighted that the factors leading to treatment retention were the same for those using alcohol or illicit substances, with the exception of higher levels of secondary substance use among illicit drug users. The study also identified the significant differences between those entering residential treatment compared to other modalities. Those of older age, having higher levels of educational attainment and polysubstance users were more likely to enter residential treatment. It is envisaged that these results will serve as a catalyst for future investigations within this treatment population as policy makers and service providers strive to design, implement, and evaluate treatment procedures aimed at improving treatment retention and outcomes.

Bibliography

Austin, P.C. & Steyerberg, E.W., 2014. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Statistical Methods in Medical Research*, pp.1–13. Available at: http://smm.sagepub.com/cgi/doi/10.1177/0962280214558972.

Austin, P.C. & Tu, J. V, 2004. Bootstrap Methods for Developing Predictive Models. *The American Statistician*, 58(2), pp.131–137.

Bagley, S.C., White, H. & Golomb, B. a., 2001. Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology*, 54(10), pp.979–985.

Ball, S. a., Carroll, K.M., Canning-Ball, M. & Rounsaville, B.J., 2006. Reasons for dropout from drug abuse treatment: Symptoms, personality, and motivation. *Addictive Behaviors*, 31(2), pp.320–330.

Beynon, C.M., Bellis, M. a & McVeigh, J., 2006. Trends in drop out, drug free discharge and rates of re-presentation: a retrospective cohort study of drug treatment clients in the North West of England. *BMC public health*, 6, p.205.

Beynon, C.M., McMinn, A.M. & Marr, A.J.E., 2008. Factors predicting drop out from, and retention in, specialist drug treatment services: a case control study in the North West of England. *BMC public health*, 8, p.149.

Bursac, Z., Gauss, C.H., Williams, D.K. & Hosmer, D.W., 2008. Purposeful selection of variables in logistic regression. *Source code for biology and medicine*, 3, p.17.

Cartwright, W.S., 2000. Cocaine medications, cocaine consumption and societal costs. *PharmacoEconomics*, 18(4), pp.405–413.

Cartwright, W.S., 2008. Economic costs of drug abuse: Financial, cost of illness, and services. *Journal of Substance Abuse Treatment*, 34(2), pp.224–233.

Claus, R.E., Orwin, R.G., Kissin, W., Krupski, A., Campbell, K. & Stark, K., 2007. Does gender-specific substance abuse treatment for women promote continuity of care? *Journal of Substance Abuse Treatment*, 32(1), pp.27–39.

Connolly, J. & Long, J., 2014. TO THE EMCDDA by the Reitox National Focal Point.

Cook, R.D. & Weisberg, S., 1997. Graphics for Assessing the Adequacy of Regression Models. *Journal of the American Statistical Association*, 92(438), pp.490–499. Available at: http://www.tandfonline.com/doi/abs/10.1080/01621459.1997.10474002.

Craig, R.J., 1985. Reducing the treatment drop out rate in drug abuse programs. *Journal of substance abuse treatment*, 2(4), pp.209–219.

Department of Community Rural and Gaeltacht Affairs, 2009. *Drug Addiction Treatment and Rehabilitation*,

Dobkin, P.L., Civita, M. De, Paraherakis, A. & Gill, K., 2002. Social Support Outpatient. , pp.347–356.

Esarey, J. & Pierce, A., 2011. Assessing Fit Quality and Testing for Misspecification in Binary Dependent Variable Models ∗. , 18, pp.1–31.

Field, A., 2012. *Discovering Statistics Using IBM SPSS Statistics* 4th ed., London: Sage.

Fox, J. & Monette, G., 1992. Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, 87(417), pp.178–183. Available at: http://www.tandfonline.com/doi/abs/10.1080/01621459.1992.10475190\nhttp://www.jstor.org/stable/pdfplus/2290467.pdf?acceptTC=true.

Ghose, T., 2008. Organizational- and individual-level correlates of posttreatment substance use: A multilevel analysis. *Journal of Substance Abuse Treatment*, 34(2), pp.249–262.

Gossop, M., Marsden, J., Stewart, D. & Rolfe, A., 1999. Treatment retention and 1 year outcomes for residential programmes in England. *Drug and Alcohol Dependence*, 57(2), pp.89–98.

Green, C.A., Ph, D., H, M.P., Polen, M.R., A, M., Dickinson, D.M., Ii, C., Ii, N., Lynch, F.L., Ph, D., H, M.S.P., Bennett, M.D. & A, M., 2002. Gender differences in predictors of initiation , retention , and completion in an HMO-based substance abuse treatment program. , 23, pp.285–295.

Greenfield, L., Burgdorf, K., Chen, X., Porowski, A., Roberts, T. & Herrell, J., 2004. Effectiveness of long-term residential substance abuse treatment for women: findings from three national studies. *The American journal of drug and alcohol abuse*, 30(3), pp.537–550.

Greenland, S., 1989. Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health*, 79(3), pp.340–349.

Gries, D. & Schneider, F.B., 2010. *Guide to Intelligent Data Analysis*,

Harrell, F.E., Lee, K.L. & Mark, D.B., 1996. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4), pp.361–387.

Health Research Board, 2015. Health Research Board. Available at: http://www.hrb.ie/health-information-in-house-research/alcohol-drugs/ndtrs/ [Accessed August 19, 2015].

Van Der Heijden, G.J.M.G., T. Donders, a. R., Stijnen, T. & Moons, K.G.M., 2006. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. *Journal of Clinical Epidemiology*, 59(10), pp.1102–1109.

Hosmer, D.W. & Lemeshow, S., 2000. *Applied logistic regression*,

Hser, Y.I., Anglin, M.D. & Fletcher, B., 1998. Comparative treatment effectiveness: Effects of program modality and client drug dependence history on drug use reduction. *Journal of Substance Abuse Treatment*, 15(6), pp.513–523.

Hser, Y.-I., Evans, E., Huang, D. & Anglin, D.M., 2004. Relationship between drug treatment services, retention, and outcomes. *Psychiatric services (Washington, D.C.)*, 55(7), pp.767–774.

Hubbard, R.L., Craddock, S.G. & Anderson, J., 2003. Overview of 5-year followup outcomes in the drug abuse treatment outcome studies (DATOS). *Journal of Substance Abuse Treatment*, 25(3), pp.125–134.

Joe, G.W., Simpson, D.D. & Broome, K.M., 1998. Effects of readiness for drug abuse treatment on client retention and assessment of process. *Addiction (Abingdon, England)*, 93(8), pp.1177–1190.

Joe, G.W., Simpson, D.D. & Broome, K.M., 1999. Retention and patient engagement models for different treatment modalities in DATOS. *Drug and Alcohol Dependence*, 57(2), pp.113–125.

Knight, D.K., Logan, S.M. & Simpson, D.D., 2001. Predictors of program completion for women in residential substance abuse treatment. *The American journal of drug and alcohol abuse*, 27(1), pp.1–18.

Little, R., 1988. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83(404), pp.1198–1202.

Luchansky, B., Brown, M., Longhi, D., Stark, K. & Krupski, A., 2000. Chemical dependency treatment and employment outcomes: Results from the “ADATSA” program in Washington State. *Drug and Alcohol Dependence*, 60(2), pp.151–159.

Mertens, J.R. & Weisner, C.M., 2000. Predictors of substance abuse treatment retention among women and men in an HMO. *Alcoholism, clinical and experimental research*, 24(10), pp.1525–1533.

Messina, N., Wish, E. & Nemes, S., 2000. Predictors of treatment outcomes in men and women admitted to a therapeutic community. *The American journal of drug and alcohol abuse*, 26(2), pp.207–227.

Mullen, L., Barry, J., Long, J., Keenan, E., Mulholland, D., Grogan, L. & Delargy, I., 2012. A National Study of the Retention of Irish Opiate Users in Methadone Substitution Treatment. *The American Journal of Drug and Alcohol Abuse*, 38(6), pp.551–558.

Nagelkerke, N.J.D., 1991. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), pp.691–692.

Peng, C.-Y.J. & So, T.-S.H., 2002. Logistic Rrgression Analysis and Reporting: A Primer. *Understanding Statistics*, 1(1), pp.31–70.

Prendergast, M., Podus, D., Chang, E. & Urada, D., 2002. The effectiveness of drug abuse treatment: a meta-analysis of comparison group studies. *Drug and Alcohol Dependence*, 67(1), pp.53–72.

Rufibach, K., 2010. Use of Brier score to assess binary predictions. *Journal of Clinical Epidemiology*, 63(8), pp.938–939. Available at: http://dx.doi.org/10.1016/j.jclinepi.2009.11.009.

Sarkar, S.K., Midi, Habshah, Rana, S., 2011. Detection of Outliers and Influential Observations in Binary Logistic Regression: An Empirical Study. *Journal of Applied Sciences*, 11, pp.26–35.

Sayre, S.L., Schmitz, J.M., Stotts, A.L., Averill, P.M., Rhoades, H.M. & Grabowski, J.J., 2002. Determining predictors of attrition in an outpatient substance abuse program. *The American journal of drug and alcohol abuse*, 28(1), pp.55–72.

Simpson, D.D. & Joe, G.W., 1993. Motivation as a predictor of early dropout from drug abuse treatment. *Psychotherapy: Theory, Research, Practice, Training*, 30(2), pp.357–368. Available at: http://apps.webofknowledge.com.offcampus.lib.washington.edu/full_record.do?product=WOS&search_mode=GeneralSearch&qid=8&SID=2ApC9gnN@JE9Oa24pLg&page=1&doc=1.

Simpson, D.D., Joe, G.W. & Brown, B.S., 1997. Treatment retention and follow-up outcomes in the Drug Abuse Treatment Outcome Study (DATOS). *Psychology of Addictive Behaviors*, 11(4), pp.294–307.

Sorensen, H.T., Sabroe, S. & Olsen, J., 1996. A framework for evaluation of secondary data sources for epidemiological research. *International journal of epidemiology*, 25(2), pp.435–442.

Stapleton, R.D. & Comiskey, C.M., 2010. Alcohol usage and associated treatment outcomes for opiate users entering treatment in Ireland. *Drug and Alcohol Dependence*, 107(1), pp.56–61.

Stark, M.J., 1992. Dropping out of substance abuse treatment: A clinically oriented review. *Clinical Psychology Review*, 12(1), pp.93–116.

Steyerberg, E.W. & Harrell, F.E., 2015. Prediction models need appropriate internal, internal–external, and external validation. *Journal of Clinical Epidemiology*, (August), pp.48–51. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0895435615001754.

Steyerberg, E.W., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y. & Habbema, J.D.F., 2001. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54(8), pp.774–781.

Sun, G.W., Shook, T.L. & Kay, G.L., 1996. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *Journal of Clinical Epidemiology*, 49(8), pp.907–916.

United Nations Office of Drugs and Crime, 2012. *2012 Wolrd Drug Report*,

Westreich, L., Heitner, C., Cooper, M., Galanter, M. & Guedj, P., 1997. Perceived social support and treatment retention on an inpatient addiction treatment unit. *The American journal on addictions / American Academy of Psychiatrists in Alcoholism and Addictions*, 6(2), pp.144–149.

Wickizer, T., Maynard, C., Atherly, a., Frederick, M., Koepsell, T., Krupski, a. & Stark, K., 1994. Completion rates of clients discharged from drug and alcohol treatment programs in Washington State. *American Journal of Public Health*, 84(2), pp.215–221.

Woodward, A., Das, A., Raskin, I.E. & Morgan-Lopez, A. a., 2006. An exploratory analysis of treatment completion and client and organizational factors using hierarchical linear modeling. *Evaluation and Program Planning*, 29(4), pp.335–351.

Woodward, A.M., Raskin, I.E. & Blacklow, B., 2008. A profile of the substance abuse treatment industry: organization, costs, and treatment completion. *Substance use & misuse*, 43(5), pp.647–679.

World Health Organisation, 2015a. World Health Organisation. *World Health Organisation*. Available at: http://www.who.int/substance_abuse/facts/en/ [Accessed August 19, 2015].

World Health Organisation, 2015b. World Health Organisation. *World Health Organisation*. Available at: http://www.who.int/topics/substance_abuse/en/ [Accessed August 19, 2015].

Zarkin, G. a., Dunlap, L.J., Bray, J.W. & Wechsberg, W.M., 2002. The effect of treatment completion and length of stay on employment and crime in outpatient drug-free treatment. *Journal of Substance Abuse Treatment*, 23(4), pp.261–271.

Appendix 1: Study Protocols

Study Protocols

MSc Operational Research & Management Science

Drug & Alcohol Services Research Project

Introduction

The research project to be undertaken will focus on the application of operational research methodologies to drug and alcohol treatment data in order to obtain insights into the treatment population and the factors which may lead to a service user completing treatment. This study has the potential to offer important insights for substance misuse treatment providers in the Cork and Kerry region as a collective study of this nature has not been undertaken, whereby statistical methods are applied to this particular dataset. The group of participating service providers may include the following;

Southern Regional Drug & Alcohol Task Force funded drug and alcohol projects (Tier 2)
Cork Local Drug & Alcohol Task Force funded drug and alcohol projects (Tier 2)
Health Service Executive non-residential treatment services (Tier 3)
Matt Talbot Adolescent Services (Residential treatment Tier 4)
Tabor Lodge Treatment Centre (Residential treatment Tier 4)
Talbot Grove Treatment Centre (Residential treatment Tier 4)
Cuan Mhuire, Farnannes (Residential treatment Tier 4)

Aims

- Develop an understanding of the factors which affect treatment outcomes.
- Develop a detailed demographic profile of service users, patterns of substance use and how this relates to treatment outcomes.
- Outline recommendations for strategic decision making with regard to service delivery aimed at increasing treatment retention.

Primary Objective

Measure treatment outcomes relative to demographic profile and substance use history of the substance use treatment population of Cork and Kerry between **January 1st 2008 and December 31st 2013**, in order to provide insights for treatment providers to be utilised as an evidence base for strategic decision making on service delivery.

Research Questions

1) How does the treatment population who completed treatment in residential, non-residential, low threshold community based and primary care/institutions in Cork and Kerry between January 1st 2009 and December 31st 2013 differ from those who dropped out prematurely during the same period based on the following variables;

Demographic: Gender, Age Group
Socio-Economic: Education Level, Employment Status, Living Status, Family Involvement
Programme Level: Treatment Modality, Source of Referral
Substance: Primary Substance Type, Age of First Use, Frequency of Use, Presence of Second Problem Substance, Second Drug Type, Presence of Tertiary Problem Substance

2) What are the factors significant factors which affect treatment retention?

Methods

Data: The community drugs projects of the Cork Local & Southern Regional Drug & Alcohol Task Forces' and Health Service Executive treatment services and voluntary treatment services feed into the National Drug Treatment Reporting System. The National Drug Treatment Reporting System (NDTRS) is an epidemiological database on treated drug and alcohol misuse in Ireland. Data is collected on every treatment which is carried out by the service. Treatment is broadly defined by the NDTRS as any activity which aims to ameliorate the psychological, medical or social state of individuals who seek help for their substance misuse problems.

The study population will include all treated cases in Cork and Kerry from January 1st 2009 to December 31st 2013 who are attending the service for the first time. Only those attending for the first time will be included to ensure that individuals are entered into the study only once**. (I.e. only data pertaining to treatments where 1 First treatment was ticked for D23 on the NDTRS form is required)** Data will be required on the following parameters, coded by a Health Research Board generated unique identifier for the selected cases.

A) Administrative Details
2a Centre Number
2b Type

B) Demographic Details
4 Gender
5 Age
7a Living with whom
7b Living Where
11 Employment Status
12a Age left primary or secondary school

12b Education: highest level completed

C) Referral/Assessment Details
14 Main reason for referral, specified main drug/problem
15a Source of referral
15b if client was referred from another treatment centre, please give reason for referral

D) Treatment Details
23 Type of contact with this centre (Only data relating to cases where 1 First treatment is ticked is required)

E) Substance Use
All data required from 24a to 28h

F) Injecting Risk Behaviour
All data required from 29a to 30

G) Activity Details
33a Treatment interventions provided

H) Exit Details
34 Outcome for main treatment intervention
35 If outcome for main treatment intervention is premature exit from treatment site (Q34 code 6), main reason for non-compliance
36 Clients condition at discharge or when last seen
39 Please specify the number of family members or significant others involved in this treatment

Analysis: The data will be configured for uploading into R statistical analysis software. Exploratory data analysis will be conducted using both graphical and non-graphical methods. The demographic, substance use, treatment intervention details and outcome for main treatment intervention will be summarised for the subject population. Multivariate logistic regression analysis will be used to identify significant contributors to treatment outcomes for the main treatment intervention. Once significant variables are identified final models will be constructed. The model building process will conclude with a series of goodness of fit tests and diagnostic statistics designed to identify outlying observations and to assess the model’s fit and performance. Final results will be documented.

Ethical Considerations

- Participants attend addiction treatment providers voluntarily where consent is obtained to collect NDTRS data. This consent also includes transfer of the data to the Health Research Board where it is held in a national central database and is used for research purposes. Data is anonymised prior to transfer to the Health Research Board.

- Norma Madden (student) and Dr Richard Williams (supervisor) will have access to the data for the duration of the project from May 18th 2015 to October 31st 2015.
- This project does not require direct access to human participants as secondary data will be used. This secondary data will be anonymised prior to the student acquiring it, therefore human participants will not be identifiable.
- The data will be transferred to the student through an email which will be encrypted.
- All data and associated models will be stored on the student's laptop which will be encrypted using software recommended by Information Systems Services, Lancaster University.
- All data will be destroyed no later than October 31st 2015. This date has been chosen because it allows time for the dissertation to be graded and the student to receive final grades.
- A research ethics form has been completed by the student. This form which outlines the sensitivity of the data, how it will be used and who will have access to it has been approved by the host institution (Lancaster University).
- The student has undertaken and passed information security training provided by Lancaster University.
- A letter of agreement will be co-signed by the HSE Drug and Alcohol Services and Lancaster University stipulating confidentiality practices to be adhered to for the duration of the project.

Project Timeline

Date	Event
12th June 2015	Data Received
15th June 2015	Data Analysis & Model Building
7th August 2015	Dissertation Writing Begins
11th September 2015	Submit Dissertation
31st October 2015	All Data Destroyed

Final Project Deliverable

The final project deliverable is a dissertation submitted for MSc Operational Research and Management Science, Lancaster University.

Appendix 2: Observations Removed from Dataset

Clients who had the following status at the time of termination of treatment were excluded from the study population as failure to complete treatment was deemed to be outside of the clients control: 7) Released from prison but not linked to other treatment site, 8) Died, 9) Sentenced to prison, 10) General medical transfer or medical issue, 11) No longer lives in the area, 12) Mental health transfer, 13) Prison to prison transfer. This amounted to a further removal of 284 records. Further preliminary analysis of the data revealed that 68 cases indicated the primary problem substance as either “headshop drugs”, solvents, inhalants, antidepressants or unspecified substances. As the number for each of these drugs was very low relative to the others these cases were removed.

Appendix 3: Missing Data Analysis

Missing data analysis begun with a procedure whereby the degree of ‘missingness’, in each variable, in each case and across the sample was measured. Cases with missing values were examined in terms of their relationship to the outcome variable to ascertain if for example those cases had a higher percentage of either level of the outcome, or if patterns could be observed. In addition data was tested to ascertain the type of ‘missingness’, by use of Little’s missing completely at random test. This tests the null hypothesis that data are not missing completely at random. A significant p-value therefore indicates that data are not missing completely at random and certain procedures for missing value treatment are not adequate. A significant result implies that data may be missing at random or non-ignorable missing. The three types of missing values are defined as:

Missing completely at random: This means that special circumstances or special values in the data lead to higher or lower chances for missing values. In this instance the value is not missing as such but is just not available in the dataset. This may be due to accidental deletion or another random event whereby we cannot see the value in question. It can therefore be concluded that the missing values follow the same distribution as the known values of the variable.

Missing at random: In this case the probability for a missing value depends of the value of the other attributes Y but is conditionally independent of the true value of X given Y. Therefore the missing values of X do not follow the same distribution as the measured values of X.

Non-Ignorable missing: This refers to situations where the occurrence of missing values directly depends on the true value, and the dependence cannot be resolved by other attributes.

Appendix 4: Use of Akaike Information Criterion for Stepwise Variable Selection

In choosing between competing models, the use of AIC would lead to the selection of the model with the lowest AIC. The use of AIC was considered a superior stopping rule to P-value as large P-values may result from collinearity indicating variables are not significant predictors when this may not be the case. In the forward method, an initial model is defined that only contains the constant (b_0), each variable other than those already included is added to the current model, one at a time, and the one that can best improve the objective function, that is the AIC in this case, is retained. In the backward step, each variable already included is deleted from the current model, one at a time, and the one that can best improve the objective function is discarded. The algorithm continues until no improvement can be made by either the forward or the backward step. All three methods were tried to allow for comparison.

Appendix 5: Description of the Likelihood Ratio Test (LRT)

The likelihood ratio test is based on the difference in deviances that is the deviance of the null model minus the deviance with the predictor(s) in the model. The null model is a representation of when there is no relation between the predictors and the response variable. It is an intercept only model where the data is represented entirely as random variation.

Appendix 6: Description of Hosmer Lemeshow Goodness of Fit Test

First, the observations are sorted in increasing order of their estimated event probability. The observations are then divided into G groups. The Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson chi-square statistic from the 2×G table of observed and expected frequencies, for the G groups. The distribution of the statistic is approximated by a chi-square with (G-2) degrees of freedom (Hosmer & Lemeshow 2000).

Appendix 7: Description of the C-Statistic

When outcomes are binary, the c-statistic is the probability that a randomly selected subject who completed treatment has a higher predicted probability of completing treatment than a randomly selected subject who did not complete treatment. C-statistic or AUC values range from 50 percent which is the case where the model does not have the power to accurately discriminate cases those who complete treatment and those who do not, above and beyond a 50/50 random chance to 100 percent, which is the idealised situation where the model achieves perfect discrimination between completed and not completed.

Appendix 8: Logistic Regression Results Model 3

Table 12: Final Results of Logistic Regression Analysis – Model 3

			Standard	Wald chi-		Odds	
Parameter		**Estimate**	**Error**	**square**	**Pr>ChiSq**	**ratio**	**(Conf. Int)**
Intercept		0.4448	0.3675	1.21	0.2261	1.56	(0.76, 3.21)
Living	Alone	-0.2053	0.1348	-1.52	0.1278	0.81	(0.62, 1.06)
	Partner/Friends	-0.1365	0.1404	-0.97	0.331	0.87	(0.66, 1,15)
	Parents/Family	-0.0275	0.1069	-0.26	0.7974	0.97	(0.79, 1.2)
	Other	0.6637	0.1538	4.32	0	1.94	(1.44, 2.63)
	Missing	0.0413	0.3703	0.11	0.9112	1.04	(0.49, 2.13)
	Children/partner	**reference**					
Education Level	Primary or less	-0.957	0.1435	-6.67	0	0.38	(0.29, 0.51)
	Second Level (J)	-0.5916	0.1147	-5.16	0	0.55	(0.44, 0.69)
	Second Level (L)	-0.5181	0.1186	-4.37	0	0.6	(0.47, 0.75)
	Third Level	-0.1627	0.1869	-0.87	0.3841	0.85	(0.59, 1.23)
	Missing	-0.6492	0.1452	-4.47	0	0.52	(0.39, 0.69)
	Current	**reference**					
Treatment Modality	Residential	2.5667	0.1449	17.72	0	13.02	(9.84, 17.36)
	Non-residential	-0.2118	0.0941	-2.25	0.0244	0.81	(0.67, 0.97)
	PrimC/Institution	-0.6745	0.2076	-3.25	0.0012	0.51	(0.34, 0.76)
	Low threshold	**reference**					
Frequency of Use	Daily	-0.1096	0.092	-1.19	0.2336	0.9	(0.75, 1.07)
	Weekly or less	0.3129	0.1124	2.78	0.0054	1.37	(1.1, 1.7)
	No use past month	0.6802	0.1066	6.38	0	1.97	(1.6, 2.43)
	Missing	-0.3139	0.3462	-0.91	0.3645	0.73	(0.36, 1.42)
	2 to 6 days p/w	**reference**					
Second Drug Type	Cannabis	-0.1912	0.1349	-1.42	0.1566	0.83	(0.63, 1.08)
	Hypnotic/Benzo	-0.6756	0.2026	-3.34	0.0009	0.51	(0.34, 0.75)
	No Drug	-0.0527	0.1121	-0.47	0.6384	0.95	(0.76, 1.18)
	Opiate	-1.0772	0.4449	-2.42	0.0155	0.34	(0.14, 0.8)
	Other	-0.2911	0.2689	-1.08	0.2789	0.75	(0.44, 1.26)
	Stimulant	-0.4153	0.1847	-2.25	0.0245	0.66	(0.46, 0.95)
	Alcohol	**reference**					
CP Involved	Yes	-0.0948	0.331	-0.29	0.7747	0.91	(0.48, 1.75)
	No	-0.7274	0.3307	-2.2	0.0278	0.48	(0.25, 0.93)
	Missing	**reference**					

Appendix 9: Logistic Regression Results Model 3 (Influential Observations Removed)

Table 13: Logistic Regression Results – Model 3 (Influential Observations Removed)

			Standard	Wald chi-		Odds	
Parameter		**Estimate**	**Error**	**square**	**Pr>ChiSq**	**ratio**	**(Conf. Int)**
Intercept		0.3267	0.3768	0.87	0.3859	1.39	(0.66, 2.9)
Living	Alone	-0.2005	0.1354	-1.48	0.1386	0.82	(0.63, 1.07)
	Partner/Friends	-0.1255	0.1409	-0.89	0.3731	0.88	(0.67, 1.16)
	Parents/Family	-0.0084	0.1075	-0.08	0.9375	0.99	(0.8, 1.22)
	Other	0.6778	0.1545	4.39	0	1.97	(1.46, 2.67)
	Missing	-0.004	0.3894	-0.01	0.9918	1	(0.45, 2.11)
	Children/partner	**reference**					
Education Level	Primary or less	-0.9445	0.144	-6.56	0	0.39	(0.29, 0.51)
	Second Level (J)	-0.5774	0.115	-5.02	0	0.56	(0.45, 0.7)
	Second Level (L)	-0.5092	0.119	-4.28	0	0.6	(0.48, 0.76)
	Third Level	-0.13	0.1878	-0.69	0.4888	0.88	(0.61, 1.27)
	Missing	-0.634	0.1458	-4.35	0	0.53	(0.4, 0.71)
	Current	**reference**					
Treatment Modality	Residential	2.6059	0.1464	17.8	0	13.54	(10.2, 18.11)
	Non-residential	-0.214	0.0944	-2.27	0.0233	0.81	(0.67, 0.97)
	PrimC/Institution	-0.7022	0.2095	-3.35	0.0008	0.5	(0.33, 0.74)
	Low threshold	**reference**					
Frequency of Use	Daily	-0.1081	0.0924	-1.17	0.2422	0.9	(0.75, 1.08)
	Weekly or less	0.3275	0.1127	2.91	0.0037	1.39	(1.11, 1.73)
	No use past month	0.6929	0.107	6.47	0	2	(1.62, 2.47)
	Missing	-0.3541	0.3569	-0.99	0.3211	0.7	(0.34, 1.39)
	2 to 6 days p/w	**reference**					
Second Drug Type	Cannabis	-0.2055	0.1354	-1.52	0.1292	0.81	(0.62, 1.06)
	Hypnotic/Benzo	-0.6923	0.2036	-3.4	0.0007	0.5	(0.33, 0.74)
	No Drug	-0.057	0.1124	-0.51	0.6118	0.94	(0.76, 1.18)
	Opiate	-1.4371	0.4678	-3.07	0.0021	0.24	(0.09, 0.59)
	Other	-0.3536	0.2722	-1.3	0.1938	0.7	(0.41, 1.19)
	Stimulant	-0.4336	0.1856	-2.34	0.0195	0.65	(0.45, 0.93)
	Alcohol	**reference**					
CP Involved	Yes	0.0106	0.3413	0.03	0.9753	1.01	(0.52, 1.99)
	No	-0.6324	0.341	-1.85	0.0637	0.53	(0.27, 1.05)
	Missing	**reference**					

Appendix 10: Logistic Regression Results (Dependent Variable – Treatment Modality)

Table 14: Logistic Regression Results (Dependent Variable – Treatment Modality)

			Standard	Wald chi-		Odds	
Parameter		**Estimate**	**Error**	**square**	**Pr>ChiSq**	**ratio**	**(Conf. Int)**
Intercept		-1.7847	0.2182	-8.18	0	0.17	(0.11, 0.25)
Age Range	under 20	-1.9175	0.1863	-10.29	0	0.15	(0.1, 0.21)
	30 to 39	0.2984	0.1124	2.65	0.008	1.35	(1.08, 1.68)
	40 and over	0.9589	0.1106	8.67	0	2.61	(2.1, 3.24)
	20 to 29	**Reference**					
Education Level	Primary or less	0.1102	0.2314	0.48	0.6339	1.12	(0.71, 1.77)
	Second Level (J)	0.3409	0.2124	1.61	0.1085	1.41	(0.94, 2.15)
	Second Level (L)	0.9432	0.2104	4.48	0	2.57	(1.72, 3.92)
	Third Level	1.1264	0.241	4.67	0	3.08	(1.94, 4.99)
	Missing	-0.9193	0.2722	-3.38	0.0007	0.4	(0.23, 0.68)
	Current	**Reference**					
Primary Substance	Cannabis	-0.4899	0.1351	-3.63	0.0003	0.61	(0.47, 0.8)
	Hypno/Benzo	-0.7722	0.234	-3.3	0.001	0.46	(0.29, 0.72)
	Opiate	-0.7298	0.1822	-4.01	0.0001	0.48	(0.33, 0.68)
	Stimulant	-0.6872	0.218	-3.15	0.0016	0.5	(0.32, 0.76)
	Alcohol	**Reference**					
Frequency of Use	Daily	-0.3709	0.1082	-3.43	0.0006	0.69	(0.56, 0.85)
	Weekly or less	-0.1921	0.1315	-1.46	0.144	0.83	(0.64, 1.07)
	No use past month	-0.2148	0.1116	-1.93	0.0542	0.81	(0.65, 1)
	Missing	0.4005	0.3841	1.04	0.297	1.49	(0.68, 3.11)
	2 to 6 days p/w	**Reference**					
Secondary Drug Use	Yes	0.5026	0.1147	4.38	0	1.65	(1.32, 2.07)
	No	**Reference**					
Tertiary Drug Use	Yes	0.857	0.1277	6.71	0	2.36	(1.84, 3.03)
	No	**Reference**					