Changing Crime Mix Patterns of Criminal Careers:

Longitudinal Latent Variable Approaches for Modelling Conviction Data in England & Wales and the Netherlands

Amy Elizabeth Elliott, MSc BA(HONS)

Submitted for the degree of Doctor of Philosophy

Applied Social Statistics

Lancaster University

August 2017

DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis is supported by one original paper published in a handbook and one original paper published in a peer reviewed journal;

- Elliott, A., Francis, B., Soothill, K. & Blokland, A. 2017. Changing crime-mix patterns of offending over the life course: a comparative study in England & Wales and the Netherlands. *In:* Blokland, A. & Van Der Geest, V. (eds.) *The Routledge International Handbook of Life-Course Criminology.* London: Routledge.

- Francis, B., Elliott, A. & Weldon, M. 2016. Smoothing Group-Based Trajectory Models Through B-Splines. *Journal of Developmental and Life-Course Criminology,* 2**,** 113-133.

The ideas, development and writing up of all the papers in support of the thesis were the principal responsibility of myself, the student, working within the mathematics and statistics department, Lancaster University, under the supervision of Professor Brian Franics.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

Amy Elizabeth Elliott

ACKNOWLEDGMENTS

I would like to take this opportunity to thank the many people who have provided me with support and guidance to make this thesis possible.

Firstly, I would like to thank my supervisor, Professor Brian Francis. I am very grateful for all his encouragement throughout the process and for never losing confidence in me. I would also like to thank the late Professor Keith Soothill for all his valuable advice at the beginning of my PhD.

I would also like to thank my PhD colleagues in the Maths and Statistics department at Lancaster University for their assistance and for keeping me motivated right till the very end.

A special thank-you goes to the rest of my friends at Lancaster University, particularly those from the Lancaster Roses Cheerleading Squad.

Finally, the most important people to thank are my family and closest friends. None of this would have been achievable without their care and understanding throughout the many ups and downs. I am eternally grateful for their continuous love and support.

This research was supported by the Economic and Social Research Council [ES/J500094/1]

ABSTRACT

In criminal career research, there has been a great deal of attention paid to the frequency of offending over the life course. This neglects any changes in the patterns and types of offences being committed. However, it is crucial to explore these patterns of offending in detail and various types of crimes being committed, as this will enhance the understanding of criminal activity and the causes of offending behaviour. This is especially true for policy makers, so they can make better informed decisions when deciding how best to target their resources when it comes to tackling crime.

This thesis aims to identify crime mix patterns (different offenders will commit different selection of offences) and how they develop over the life course from two official conviction datasets. The first is the England and Wales Offenders Index (OI). The cohort data of the OI contains the court convictions of offenders from 1963 to the end of 2008 in eight birth cohorts. The other dataset is from the Netherlands Criminal Career and Life-course study (CCLS) which contains data covering the criminal careers of those offenders who were convicted of a crime in the Netherlands in 1977, starting at age 12 and followed up till 2005.

The study will provide a contrasting analysis of the two datasets using a Latent Markov Model approach similar to that published in Francis et al. (2010) where the idea of lifestyle specialisation and short-term crime typologies (crime mixes) over five-year age-periods was introduced for female offenders. This approach will jointly estimate the crime mix patterns and the transition probabilities (offenders move from one pattern to another). The study adds methodological innovation in criminology by the use of B-splines in group based trajectory models and in the modelling of Poisson counts in latent Markov models.

The thesis also contributes to cross-national research. Not only is it important to be able to identify crime mix patterns in both datasets separately but being able to compare and contrast the results from each country will allow for the examination to check if offender's crime mix patterns are the same across jurisdictions. These findings will be of great interest to both criminologists and policy makers. The analysis provides a cross- national understanding of progression in the types of crime mixes offenders are involved in, whether some crime mix patterns are more specialised than others in terms of their long term patterns and whether some crime patterns desist earlier than others.

The results show that each dataset both have versatile and specialist crime mix offending groups but there are also important differences in the makeup of these groups, with regard to the type of offences. These results are discussed in further detail, along with the issues of how best to carry out analyses upon the two datasets. The additional problems encountered when comparing the two datasets and the strategies used to overcome them are explained. Finally, suggestions for future research are given along with encouragement of replicating the methodologies used in this study upon more recent datasets in other jurisdictions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This thesis is concerned with both the patterns and pathways of criminal careers. It develops upon recent ideas in quantitative criminology and associated methodology needed for a new analysis of criminal careers.

1.1 Research Questions

The main research questions and the areas this thesis will focus on, are introduced below.

Patterns of Offending over time

Firstly, this thesis will explore the various methods for modelling longitudinal patterns of offending behaviour with a focus on recidivism. The following questions will be investigated:

- What trajectories of reoffending frequency in the criminal histories can be observed in the data?
- Can offending trajectories be used as predictors for subsequent recidivism?
- Are offenders belonging to a declining trajectory less likely to recidivate than those on an upward trajectory?

Crime Mix Patterns and Pathways

Secondly a more in-depth analysis of the datasets will be undertaken to explore and identify crime mix patterns and pathways to answer the following questions:

- What crime mix patterns can be identified in the data, and how do offenders transit between such groups as they age?

- Can the crime mix pathways be identified in terms of escalation paths and in terms of specialisation groups?
- Which crime mix pathways (most common routes though the crime mix patterns) are most likely to recidivate, and which are most likely to desist?

Cross National Research

Finally, this thesis will contribute to cross national comparative research and provide answers to the following questions:

- What similarities and differences can be identified in the two datasets?
- Are there any crime mix patterns which are more specialised than others in terms of their long term patterns for both datasets?
- Do some crime mix patterns desist earlier than others and are these common for both datasets?

1.2 Proposed Approach

To address the first set of research questions, a group based trajectory model (GBTM) will be used to identify offending frequency trajectories in the data, after the consideration and rejection of other approaches. GBTM is an extension of a finite mixture model to allow for repeated observation. It is a valuable method for modelling the relationship between age and criminal behaviour and it was designed to examine the patterns that develop over time or age. It is also able to help identify clusters of individuals with similar offending trajectories. The outputs of group based trajectory models are usually continuous and tend to show curves that are smooth due to the fitting of a cubic curve to each class over the time period. When fitting the model, it is important to determine the optimal number of classes, one method to do this is by minimizing the Bayesian information criterion which is a penalised log-likelihood measure (penalised by a function of the number of parameters) but other methods will be considered. Once the optimal number of classes has been chosen the posterior group membership probabilities can be used to predict the likelihood that individuals belong to certain trajectory group. This will then aid in identifying the distinctive characteristics of each group to help predict future offending.

To be able to identify crime mix patterns and estimate the transition probabilities in the datasets, this study will develop upon the methodology used in Francis et al. (2010) where the idea of lifestyle specialisation and short-term crime typologies (crime mixes) over five-year age-periods was introduced. Latent Markov Modelling (LMM) will provide the methodology, using Latent Gold software. This jointly estimates the crime mix patterns and the transition probabilities. Latent Markov Modelling (sometimes referred to as Latent Transition Analysis) is an extension of Latent Class Analysis (LCA). LCA categorises individuals into classes or groups based on an unobserved construct. LCA will allow estimation of characteristics of the crime mix offending groups. However, LCA will only provide information on which offences tend to co-occur and not much about how offenders switch and move between the offending classes due to each time period being treated as independent.

LMM is able to combine looking at the offending classes and looking at how individuals transition between the classes over time. It allows estimation of transition matrices so these matrices give an idea to which classes or crime mix offending groups, are likely to transit into another class as an individual ages. It is therefore possible to see when an individual moves from one crime mix pattern to another. Estimation of a LMM model is complex, and the transitions have to be estimated as additional parameters.

1.2.1 Why Are These Questions Important?

These questions are important for several areas; firstly, they contribute to methodology development within criminology. Longitudinal research is important for looking at the development and progression in criminal activity over an extensive period of time, the same results cannot be obtained from cross sectional methods as they may miss important changes. Studying longitudinal data based on criminal activity and convictions has been of great interest to researchers over the past 20 years and has greatly developed the knowledge of criminal careers. This research has helped with gaining an understanding of criminal offending particularly the number of people who commit offences, the amount of offending, the types of crimes committed and the desistance from offending.

The research questions also contribute to development within social statistics as it promotes the use and development of LCA and LMM as a method for criminology.

This research is also important with regard to policy issues, new methods of assessing risk would be highly valuable to policy makers and law enforcement agencies. Also, being able to identify chronic and high rate offenders early on in their criminal careers would be very advantageous.

There has also been a lack of cross national research especially within Europe. The research will be able to contribute to cross national comparisons between England & Wales and the Netherlands.

As many studies upon criminal careers have originated in the USA, there has been a tendency for many of these studies to use arrest data for analysis. This thesis is fortunate enough to be using two complete datasets of official convictions, and the full England & Wales Offenders Index dataset is rarely fully analysed. Normally past analyses upon this dataset have often been on a particular cohort, or on specific individuals based on age or gender.

1.3 Structure of the Thesis

The structure of this thesis will be outlined below. Chapter 2 presents a comprehensive literature review on the criminal career research, concentrating on specific criminal career dimensions most relevant to the aims of the study. The most significant studies are summarised and past research limitations are discussed, revealing the gaps in the literature that require further development.

Chapter 3 begins by explaining the requirement for the use of longitudinal data sources when examining criminal offending patterns over the life-course. It discusses the advantages and disadvantages of using official data and self-report data and explains the reasons for choosing official conviction datasets for analysis in this thesis. The England & Wales Offender's Index (OI) and the Netherland's Criminal Career Life-Course Study (CCLS) conviction datasets that have been acquired for this study are then introduced. The chapter describes the background and features of both the datasets and explains the challenges of comparing datasets. Finally, the chapter finishes with the sampling strategies used for the alignment of the two datasets.

Chapter 4 builds on from Chapter 3 by providing an exploratory analysis of the two aligned dataset samples that will be used in analyses in the preceding chapters. A definition of the most significant criminal history variables, constructed from the datasets, is given before the exploratory analysis of each dataset sample. The chapter ends with a comparative discussion of the similarities and differences between each dataset.

Chapter 5 introduces the methods for modelling longitudinal trajectories of criminal offending. The chapter is focused on identifying distinct offending groups and the changing patterns of offending over time, with the aim of providing answers to the first set of research questions. Three of the main approaches for examining changes over time; Linear Mixed Effects modelling (LME), Group-Based Trajectory Modelling (GBTM), and Growth Mixture Modelling (GMM), are described and the decision on the final model is explained. A further extension to the final model is also applied to the dataset, that provides a flexible approach to modelling the trajectory curves. Finally, the posterior probabilities are estimated from the final models to aid in the prediction of reconviction, which is the focus of the following chapter.

Chapter 6 uses the results from the previous analyses in Chapter 5, to predict reconviction of offenders based on the offending group they were allocated to. The method undertaken is a logistic regression model and the technique is introduced and described. The assigned trajectory offending group membership, is tested to see how useful it is in predicting the likelihood of reconviction. The predictive performance of the logistic regression models is also assessed, along with adding in another covariate to further test the model performance.

Chapter 7 is aimed at exploring the second set of research questions. It provides a more in depth analysis of the datasets to identify the different crime mix patterns and how they develop over time. Latent Markov Modelling is introduced as the methodology to estimate the different crime mix offending groups and also the transition probabilities. Therefore, this allows the identification of the different pathways offenders can take by observing the transitions between the crime mix offending groups from one age period to the next. The initial set up of the LMM is described and then extended to accommodate for Poison count data – using the counts of conviction occasions in each of the offence categories. This is then followed by a discussion around how to deal with missing data and software issues. The LMM results are then presented and evaluated, proceeding with a summary of the chapter in the conclusion.

Finally, Chapter 8 provides a conclusion to the thesis by summarising the main findings from the results in the previous chapters. Potential policy implications from the research are highlighted and discussed. The original contributions the thesis has provided are given and areas for further research are suggested.

2 LITERATURE REVIEW

2.1 Criminal Career Research

As previously mentioned in the introduction, researchers have been concerned with the changes of crime and offending over the life-course since the ground breaking work of the Gleucks in the 1930s (Glueck and Glueck, 1930). The totality of this research is based on studying longitudinal data on criminal activity and convictions, and has provided a rich amount of information to enhance the understanding of criminal offending patterns. In particular the research has discovered information about the relationship between past and future offending, the number of people who commit offences, the amount of offending, the types of crimes committed and the desistance from offending relating to understanding the criminal career (Piquero et al., 2012, Brame and Piquero, 2003). When studying crime over the life course it is important therefore to fully understand the term ‘criminal career’. Put simply, a criminal career is the “characterization of the longitudinal sequence of crimes committed by an individual offender” (Blumstein et al., 1986). This explanation pays particular focus to the aspect of time, and suggests that there is a beginning or starting point of criminal activity, with a period of active offending leading to an ending point when an individual will desist from committing crimes. MacLeod et al. (2012) refer to the dictionary definition of the term career having two meanings “the term ‘career’ specify two different concepts: a course or progress through life (the use of the term in this thesis) or “a way of making a living.” It is important to note that there is a difference between the terms ‘Criminal Career’ and ‘Career Criminal’. A ‘Career Criminal’ actually refers to offenders who make a career out of crime and is a way of making a living for them. However, a ‘Career Criminal’ is also used as a term to describe an individual who committed a large number of crimes and engaged in serious offences (Blumstein et al., 1988).

To study criminal activity over time, criminologists have explored criminal trajectories (pathways showing the long-term patterns of development) to study the criminal activity over time. Gaining knowledge about any changes in frequency and nature of offending is highly important, as any differences indicates the need for various theories to explain the separate trajectories. As the majority of this research has been focused on looking at the *frequency* of offending and how these frequency trajectories change with age, any changes in the *patterns* and *types* of offences being committed tend to be ignored (Francis et al., 2004). It is important to not ignore these changes and to study these patterns of offending behaviour in detail for several reasons. Firstly, it allows us to see what offence typologies appear to be precursors for other types of offences (Francis et al., 2004). For example, knowing what type of offences criminals commit prior to committing more serious crimes like murder or rape. This is very important and advantageous to law enforcement agencies and policy makers, as they can then target these high risk offenders early on in their criminal careers, implementing crime prevention strategies to deter them from committing more serious offences. Secondly, gaining knowledge of crime mix patterns in detail will help with our understanding of offending behaviour and the causes behind it.

Crime mix patterns refer to the types of offences committed by an individual within an age period and the particular offending characteristics or styles these individuals hold. The crime mix patterns can be thought of as different groups or classes, where the individuals belonging to that group all share similar patterns of offending styles. These offending characteristics are not always identifiable or can be measured directly and therefore considered to be ‘hidden’ or ‘latent’. Therefore, some researchers have started to incorporate the use of particular statistical methods (e.g. Latent Class Analysis) to uncover these ‘latent’ groups or classes based on something that is directly measurable, such as the frequency of offences.

The studies by Colman et al. (2009), Eggleston et al. (2004), Piquero et al. (2001), Sampson and Laub (2003), Ward et al. (2010), and Yessine and Bonta (2009) have all discovered latent trajectory groups of offenders based on the frequency of offences committed or number of arrests. However a small number of studies by Francis et al. (2004), Francis et al. (2010), McGloin et al. (2009) _ENREF_100, and Soothill et al. (2008), have found latent classes of offenders based on the type of offences committed. The studies prove that there are different types of offenders who vary in their frequency, types of offences and behaviours and cannot be explained by the general and static theories of criminal offending.

Further research has expanded on the latent class analysis model to examine changes in the latent classes over time by using Latent Transition Analysis (LTA) or Latent Markov Model (LMM). These statistical techniques estimate the transition probabilities across the identified latent classes. Therefore allowing examination of how patterns of criminal behaviour can change and if individuals do change or remain stable in their offending behaviour (McGloin et al., 2009). Studies by Francis et al. (2010), Bartolucci et al. (2007) and McGloin et al. (2009) have utilised these models to discover distinct offending groups that display evidence of both specialist and versatile offending patterns dependent on factors such as age, gender and prior offences.

There are still many areas of criminal careers that need exploring further and many of the research findings on criminal careers are widely debated. This thesis will discuss some of the most argued theories and topics in criminal careers and contribute to some of the areas that need further research and new methodologies. It will also develop upon previously used methodologies such as Latent Transition Analysis

(LTA)[1] in Francis et al. (2010), to look at not only the types of offences and the mix of crime types that offenders commit but also to see how these offenders progress over time, observing if they stick to a specific crime mix pathway and pattern or whether they transition into other crime mix patterns and follow a different pathway. A particular aim of this thesis is to provide a comparative analysis of England & Wales Offenders Index (OI) to the Netherlands Criminal Career and Life Course Study (CCLS). This will provide a new piece of cross national research using official conviction data, which has often been avoided due to the difficulties encountered when comparing crime related data across different jurisdictions.

2.2 Key Studies in Criminal Careers

Amongst all the research upon criminal careers, there are a few key studies that have contributed significantly to the growing interest of the topic. Going back to the 1830's Quetelet in 1831 completed a study upon criminal offences and discovered that offending peaked in late adolescence and males committed more crime than females (Gittens, 2011). This research by Quetelet triggered more studies and research into criminal careers throughout the 19th century.

Two particularly famous researchers of criminal careers were Glueck and Glueck (1930). Their study '500 Criminal Careers' was a longitudinal study of 500 male criminals, contrasted against a matched sample of 500 non-criminal males by age, race and IQ in Massachusetts. They discovered a significant relationship between age and crime, noticing a strong correlation between individual crime rates declining as offenders got older. Another significant finding was discovering the importance of criminal onset age. The Glueck's noticed that the earlier an offender began committing crimes, the longer the length of their criminal career. This meant that the younger an offender began committing crimes, the longer the length of their criminal

[1] Latent Markov Modelling is used in this thesis which is essentially the same as the LTA approach

career. This theory of early criminal onset being highly correlated with a longer criminal career length, has also been confirmed by a number of other studies, (Piquero et al., 2007, Tracy et al., 1990, Farrington et al., 1990, Blumstein et al., 1986, Wolfgang et al., 1972, Farrington and Zara, 2015). In 1993, Sampson and Laub reanalysed the Gluecks data and developed their own life course approach for explaining criminal behaviour. Sampson and Laub further investigated the sample of 500 high risk boys from the Glueck and Glueck study (1930). They had two key aims; firstly, they wanted to try and discover a distinct group of offenders who had a rate of crime that remained the same throughout their life course. Secondly, they wanted to examine the effects of individual differences, childhood characteristics and family backgrounds on the prediction of long term trajectories of offending. They discovered three key findings. Firstly, when they examined the aggregate data, the pattern of offending that was revealed was one of a unimodal peak in crime during adolescence, which was then followed by a decline in mid adulthood. Secondly, they found evidence of heterogeneity in individual age-crime curves across the life span. Thirdly, no strong evidence was found to state if individual differences, childhood characteristics and family background would help in the predictions of long term offending trajectories.

In 2009, Bersani Nieuwbeerta and Laub decided to test the robustness of the Sampson and Laub (2003) findings by replicating and extending the study further. They approached this in three stages. First, they wanted to examine any long-term offending patterns in a cross national context. Second, they wanted to examine if trajectories of offending can be distinguished by risk factors that are identified during adolescence. Thirdly, they proposed a different test of predicting offending trajectories by using Group Based Trajectory Modelling.

Their findings showed that offending trajectories tend to follow a general path and decline over time for all offenders. Even though they used a much larger dataset and sample whilst covering a longer time period, their findings still were consistent with previous results. There was still a difficultly in predicting high rate offenders. They noted that there are limitations with the data, as they could only use 4 risk factors.

A few years later after the Glueck's study, Wolfgang and colleagues in 1972 conducted another longitudinal study upon males in Philadelphia. This study discovered that the majority of all crimes are committed by a small number of 'chronic offenders', this has been confirmed in many studies since (Block et al., 2010, Laub, 2004, Farrington, 2003) and has been described as a major 'turning point' in criminology by Laub (2004). It has led to researchers examining in more depth the characteristics of these chronic offenders, as it has great importance and interest to law enforcement agencies. This is because being able to identify and target this small group of offenders early on in their careers, can help reduce and prevent many more criminal acts.

The Wolfgang et al study led to the development of the National Academy of Sciences Panel on Criminal Careers (Blumstein et al., 1986). From this Blumstein and colleagues, produced probably one of the most crucial readings in criminal career research, 'Criminal careers and "career criminals". It provides an extensive review of all the different aspects of criminal careers. Blumstein et al (1986) states that the majority of studies usually concentrate on aggregated crime rates and ignore individuals. For the criminal career method, a focus needs to be on analysing the criminal activity of specific individuals.

Additional key readings in criminal career research includes the long essay by Piquero et al. (2003). This provides a widespread review of many criminal career studies. Piquero and colleagues explain the criminal career paradigm, suggesting that an individual will have a starting point to the criminal career, often referred to as onset. This will then be followed by a period of offending behaviour, sometimes called persistence. This may then progress to where the individual begins to commit more serious crimes and committing crimes more frequently, called escalation. Finally, there is and ending point where the offender stops committing crime called desistance. The period between onset and desistance is often called ‘duration.’

Brame and colleagues (2004) investigated the relationship between two offending dimensions; frequency and offence switching (e.g. versatility in offending types). Using data from the 1945 Philadelphia Birth Cohort, they tested two model hypotheses. The first model was used to predict the probability of offence switching between violent and non-violent offences varied with the frequency of offending. The second model was used to predict that the two dimensions were independent of each other. The results found that offence frequency and offence switching or versatility, had little correlation. Both high and low rate frequency offenders had similar propensity for offence switching. However, there were limitations with this study. Unfortunately, this study ignores any variability that may occur within the violent or non-violent categories. The offence categories are too broad and will miss more nuanced forms of versatility. This means that many patterns of of offence switching would be missed as an offender who changes offence type within these broad categories would be mistaken as repeating the same offence.

Following on from Brame et al. (2004), Monahan and Piquero (2009) continued to examine the relationship between frequency and variety of offending. From examining young offenders from the Pathways to Desistance Study, Monahan and Piquero used joint-trajectory modelling to examine the relationship between offending frequency and variety. Results showed that a 5 - trajectory group model was discovered for variety of offending, and a 6-trajectory group model was discovered based on frequency of offending. However, joint-trajectory modelling showed a highly correlated relationship between offending frequency and offending variety. For example offenders who displayed high frequency in offending were also involved in a wide variety of offences. Whereas offenders within the low frequency trajectory group had the highest probabilities of belonging to the low variety trajectory group.

More recently the England & Wales Offenders Index (one of the datasets examined in this thesis) has been used in past criminal career research. Francis et al. (2004) used the data to try and discover clusters of criminal activity that co-occurred together in time. A large number of offending categories were used in which 71 binary indicator variables were created to indicate if an individual had committed an offence in a specific time period. The study proposed an alternative approach that concentrated on types of offences as well as frequency, to develop criminal typologies. It also focused on examining criminal offending in smaller time periods (5 years) over the life course. This allowed for assessment of potential changes in offending behaviour over time and to see if there are any patterns of versatility and specialisation. The study identified several distinct latent classes of offenders and gave a new way to define specialisation. They referred to a specialist offenders as one that displayed “stability in offending types” meaning the offender stayed offending within the same latent class they were assigned for subsequent age periods.

The OI dataset was also analysed in the study by Bartolucci et al. (2007). Using a Latent Markov Modelling (LMM) approach, they detected patterns of criminal behaviour examined and how criminal activity changes over time. Again, both the types of offences as well as frequency was incorporated into the analysis. Ten binary offence categories were established and similar to the Francis et al. (2004) study, the offenders’ criminal careers were aggregated into six age periods (of 5 years). This was the first study to use LMM in the analysis of criminal trajectories, therefore supporting the use of more advanced quantitative methodologies in criminology. The results discovered 5 distinct latent classes of offenders and revealed differences in the patterns of criminal activity for both male and female offenders.

Finally, the study by Baker et al. (2013) contributes to the growing evidence of specialisation in offending. Using a number of logistic regression models, they estimated which prior offences could be used to predict future offences across different criminal career typologies. The results concluded that prior offences are the best indicator of future offences, suggesting that the odds of committing any offence over the life course are much more likely if the previous offence committed was the same type.

2.3 Theories of Criminal Careers

Research on criminal careers has prompted several theories to explain crime and criminal behaviours, particularly on why some individuals engage in criminal activity while others do not. Theories also attempt to explain how and why offending behaviour changes over the life course (Blokland, 2005). There is a vast amount of criminological literature to show that those individuals that have previously have offended are very likely to commit further offences in the future. These include static, typological and dynamic frameworks (Paternoster et al., 1997). Despite numerous theories for each of the frameworks, only three of the most influential theories will be discussed below; Gottfredson and Hirschi's self-control theory, Moffitt's dual taxonomy and finally Sampson and Laub's age-graded social control theory

Static theories or general theories of crime propose that any theory of crime should be able to provide an explanation for all types of criminal behaviour. It suggests that there is a general cause for criminal behaviour for all individuals (Paternoster et al., 1997). Static theories include the 'population heterogeneity' (Nagin and Paternoster, 1991) or the 'persistent heterogeneity' (Piquero et al., 2003) explanation which sees that individuals all differ in their latent tendencies to commit crimes. These theories claim that criminal behaviour is the result of an underlying personal propensity that differs from individual to individual, and differences are established early on in life as these individuals begin their journey on a criminal pathway which is unaffected by external events (Blokland and Nieuwbeerta, 2010). This inclination within individuals to commit crime is then time stable. The connection between past and future offending is a result of the “deviant trait” within the individual. External factors and life circumstances have very little influence on the individual’s propensity to offend once the unobserved individual differences are accounted for. Past and future offending are positively related due to the stable criminal tendency, if the individual is likely to offend at one point in time they are just as likely to offend at another point in time.

One of the most renowned static theories is the prominent General Theory of Crime proposed by Gottfredson and Hirschi’s (1990) self control theory. They believe it is the sole reason for explaining this deviant behaviour and that it is general enough to explain all crime (Buker, 2011). Their theory attempts to explain both criminal behaviour as well as the individual differences in the “propensity” to abstain from offending. This includes all deviant behaviours at various ages (Akers and Sellers, 2004). They also offer an explanation for why offending remains stable and persistent throughout the life course.

Individuals with low self-control are susceptible to criminal behaviour compared to those with high self-control. The self-control theory states that all individuals have an underlying tendency that is responsible for them becoming involved in criminal activities. This hidden trait is classified as self-control. The level of self-control is established very early on in life and remains stable through time (Buker, 2011). This means, that the inclination to commit crime is time stable and is unaffected by external factors. Having determined this level of self-control (or lack thereof), Gottfredson and Hirshi argue that lower levels of self-control increase the likelihood of deviant behaviour which, create short term gains of pleasure and enjoyment (Gibson, 2010). As opportunities for crime are always available, those individuals with low self-control become heavily involved in criminality. It is the individual’s self-control, rather than the opportunities to commit crime, that establish the individual's participation in crime over the life course (Cullen and Agnew, 1999).

The self-control theory claims that all offending behaviour can be explained by a single approach that applies to the entire population. The age-crime relationship is invariant and crime declines similarly with age for all offenders, therefore suggesting it is unnecessary to classify offenders into groups, as the same theory of criminality applies to all (Greenberg, 1991). Paternoster et al. (1997), points out that many of the studies to test the static and general theories, have normally involved samples of high risk offenders using conviction and arrest datasets. This places strong restrictions on the outcome of the results and does not represent the whole population. General theories therefore, are ignoring the fact that subgroups of offenders exist, who follow different patterns and pathways (Blokland et al., 2005, Farrington, 1986). Gottfredson and Hirschi do not expect for any variation or differences in the crime mix to be associated with the frequency or duration of offending (Blokland et al., 2005; Farrington, 1986).

Arguing against general theories of crime is the dual taxonomy proposed by Moffitt (1993). Typological theories assume that there are distinct groups or typologies of offenders with defined pathways over the life course. These differing offending typologies all require a separate explanation, therefore a general theory is unsuitable. Moffitt states varying patterns of offending can define offenders into specific groups. Moffitt suggested there are two types of offenders both following different pathways. The first is the smaller group that contains 'Life Course Persisters' (LCP), who engage in offending at an early age and continue to offend at a steady rate, forming a rather flat trajectory path over the life course. According to Moffitt, the LCP group are more likely to be versatile in their offending and commit more serious offences. The second group are the 'Adolescent Limited' (AL) offenders, consisting of a larger number of offenders, who commit their first offence in their teenage years. However, they peak in early adulthood and begin to decrease their offending and desist. This group are involved in less serious offences compared to the LCP group. The AL group can explain and mimic the age-crime curve. The smaller LCP chronic offenders group do not follow the age-crime curve, and disagree with the general theories of criminality (Sampson and Laub, 2003). The typology theories suggest that persistence or changes in offending behaviour are affected by various causal procedures for different offending typologies. Continuity in offending behaviour could be the result of the criminal propensity which is time-stable, while intermittent offending is due to effects of changing life-circumstances (Paternoster et al., 1997).

Dynamic theories also contrast with the static and general theories of crime. These include the state dependence explanation, which discards the belief that instability and changes in life circumstances is unable to influence or effect criminal behaviour (Paternoster et al., 1997). These theories claim that past offending becomes an incentive for that particular behaviour, so that similar criminal behaviour becomes much more likely in the future (Blokland and Nieuwbeerta, 2010). This then weakens constraints and increases the appeal of criminality (Nagin and Paternoster, 2000). Dynamic theories allow for instability and recognise that criminal behaviours are subject to change. This is due to the influence of external events over the life course. Persistence in offending is driven by the practices established by earlier developmental issues (Thornberry, 2005).

The major proponents of this approach are Sampson and Laub (1993) with the age-graded theory of informal social control. The focus of this theory is that criminal behaviour is due to a lack of social control which can change over the life course. They focus on explaining criminal behaviour changes over a period of time, rather than one fixed time point. Sampson and Laub, suggest that as an individual begins offending, they are weakening their bonds in society and therefore making it more likely to commit future offending. However, the longer the time that elapses from committing the last offence, the weaker the effect and the likelihood to reoffend decreases (Nagin and Paternoster, 1991). Major life events can strengthen or weaken the social bonds to society and in turn influence criminal behaviours. Factors such as gaining employment, marriage, and having a family can inhibit criminal activity by increasing the ties to conventional society. Individuals are likely to desist from crime because they gain routine activity from employment, have less time and opportunity to offend and spend with delinquent peers (Loeber and Farrington, 2012). Persistence in offending is therefore the result of cutting these ties to society and weakening these bonds. For example, some turning point or changes can have negative consequences. For instance, losing a job or a marriage breakdown can contribute towards offending behaviour. This creates low level of social control and the likelihood of offending increases (Sampson and Laub, 1993).

These theories have sparked numerous debates in the criminology field. It has instigated researchers to continue to improve their investigative methodologies to explain and describe the patterns of criminal behaviour over the life course (Blokland and Nieuwbeerta, 2010).

There are many parameters to consider within the criminal career paradigm, however for the purpose of this thesis only three will be focused on in the next sections which are relevant to the aims of the thesis; criminal typologies, crime mix patterns, and specialisation/versatility.

2.4 Criminal Typologies

As previously mentioned, the research on criminal careers has led to the use of typological theories for explaining the different patterns of offending. For many years, researchers have tried to distinguish offenders from non-offenders by classifying them into criminal typologies (Gibbons, 1975, Moffitt, 1993). This was in an attempt to try and understand the causes of crime and criminal behaviour by defining offenders based upon their offending patterns (Vaughn et al., 2008). More recently, offenders are now usually defined by a number of characteristics such as the length of their criminal career, the type and seriousness of their offences, and the mix of crime types they engage in (Sullivan et al., 2009).

The popular typological theory, the dual taxonomy theory, proposed by Moffitt (1993), suggested there are two types of offenders with distinct criminal pathways. Moffitt's original theory unfortunately does not consider many quantitative properties about either of the two groups of offenders. For example, it is unknown what percentage of offenders will be either 'life course persistent' or 'adolescent limited', the frequency of offending for each group at different ages or the length of the criminal careers (MacLeod et al., 2012). Although the proportions in each group has been assessed by group based trajectory modelling on offending data (See D'Unger et al.,1998).

Taking into account the typologies based on frequency of offending is important, but it is also useful to examine the typologies of offence mixes which give more detailed information of the vast amount of offending behaviour (Soothill et., 2008). Classifying *offenders* into categories such as a 'burglar' however, gives them a label and therefore does not allow for changes over time. Considering the most appropriate way forward is to use an approach that embraces dynamic changes; classifying offending patterns within a specific time window is beneficial as it allows for changing crime mix patterns.

2.5 Crime Mix Patterns

Although there appears a lack of research into the types of offences criminals commit over time, there have been some studies that have attempted to investigate this. For example, Gibbons (1972) found 19 role careers of offending types. Unfortunately, most early work on criminal careers did not use quantitative data. One exception to this was by Chaiken and Chaiken (1982), who used the self report data of a large sample of prisoners. They were able to classify the prisoners into 10 types of offending patterns based on 5 crime categories. As noted by Chaiken and Chaiken, several offences that the prisoners committed were missed due to the broad crime categories and may miss if an offender switches to committing a different type of crime. This is also the case for other studies that use broad categories of offending (Armstrong and Britt, 2004, Bartolucci et al., 2007, Brame et al., 2001).

Research has progressed over the years and has attempted to explain the degree of how much criminal careers can be categorised by offending patterns over the life course, and whether different crime mix patterns or if evidence of specialisation or versatility in offending can be identified (Block et al., 2010, Piquero et al., 2007, Piquero et al., 2003).

Crime mix patterns refer to the types of offences committed by an individual within an age period and the particular offending characteristics or styles these individuals hold. The crime mix patterns can be thought of as different groups or classes, where the individuals belonging to that group all share similar patterns of offending styles. There can also be specialised crime mix patterns, where the offenders are inclined to repeat the same type of offences (not necessarily the same offence each time), and there can be versatile crime mix patterns where offenders have no inclination to stick to the same offences.

For example, there could be a group of offenders who are adolescent limited and specialise in shoplifting offences and a group of offenders with an early onset age and are very versatile in their offending, committing a variety of offences.

The Massoglia (2006) study showed changing patterns of offending. Massoglia examined the within individual changes in offending patterns. These changes in criminal activity were referred to as displacement, suggesting that offenders move away from some types of crime and move towards other types of offending as they progress from adolescent to young adults. The authors found evidence that while there are offenders who do desist from criminal activity, there are other offenders who change the patterns of offending instead of terminating their offending activity. This evidence was reaffirmed by the findings in Francis et al. (2010) and shows that using newer methods for examining offending behaviour is proving to be useful in finding new information about the patterns of criminal activity over time.

While there has been more recent studies that have started to consider the changing patterns in the types of offences, many of them have used rather broad categories of offences to examine the changing patterns, (Armstrong and Britt, 2004, Bartolucci et al., 2007, Brame et al., 2001, Britt, 1996). This means that many offenders who do change the type of offence they commit, will not necessarily be noticed and could be mistaken as repeating the same offence, therefore being considered a specialist offender. There are also potential problems for having too many categories, with new evidence found in more recent studies (Francis et al., 2010) showing that there are offenders who may stick to a particular set of offences within a distinct offending domain. These offenders can be considered specialised. Deciding upon an appropriate amount and relevant offence categories is of great importance as it significantly affects the end results.

2.6 Specialisation and Versatility

Investigating crime mix patterns for evidence of specialisation or versatility is important for understanding the processes that cause criminal activity over the life span. It is also of interest to policy makers, as knowing the extent to which offending patterns are specialised can help with tackling certain offences, by being able to focus on a particular type of offender (Nieuwbeerta et al., 2011). Also Farrington et al. (1988) stated, it is important to understand specialisation, as being able to recognise early offence types in an individual’s criminal career would assist in predicting future offences. However, past research has shown that studies have often discovered extensive evidence of versatility in offending patterns (Blumstein et al., 1988, Britt, 1996, DeLisi, 2005, Farrington et al., 1988, Piquero et al., 2003, Wolfgang et al., 1972). Evidence of specialisation in offending was very limited, but the research of specialisation continued due to its importance in policy decisions (Sullivan et al., 2009). Versatility in offending became the dominant view, due to the way past studies considered specialisation. However, Sullivan et al, (2009) pointed out that defining the term specialisation is important because it influences how it is measured and has consequential effects on any findings or results. On the one hand, specialisation can be defined by measuring the amount that an offender keeps committing the same type of offence in a direct successive order (Kempf, 1987) . However, other researchers believe this definition is too restrictive, referring to specialisation as “stability in offending types” (Francis et al., 2004). Therefore an offender is considered a specialist if they display concentration on particular types of offences, which is observed over a specific period of time.

More recently, evidence of specialisation has been discovered for some offenders (McGloin et al., 2009, Francis et al., 2004, Francis et al., 2010, Lussier et al., 2005, Osgood and Schreck, 2007, Sullivan et al., 2006). Piquero et al. (2003) reviewed some studies that found evidence of specialisation when the categories of offending were split into violent and non-violent categories. Unfortunately, as previously mentioned, this methodology ignores any variability that may occur within the violent or non-violent categories: in effect, the categories are too broad and will miss more nuanced forms of versatility. This means that many offenders who do change the type of offence they commit within a category would be mistaken as repeating the same offence and considered a specialist offender.

Methods for assessing specialisation are varied; dominant methods include the Forward Specialisation Coefficient (Farrington et al., 1988) and the Diversity Index

(Piquero et al., 1999); Sullivan et al. (2009) provides a review. An alternative approach to specialisation is through the diversity index, although Francis and Humphreys (2016) has queried the invariance of the diversity index when the same size is small. Francis et al. (2010) contributed to the research on specialisation by proposing the concept of criminal lifestyles. Extending the methodology used previously to identify offence clusters in the Francis et al. (2004) paper, this study proposed the use of latent transition analysis (a form of latent Markov modelling) to examine the transition of female offenders switching offence clusters over time. Offenders can then be seen as specialised if they remain in the same offending cluster from one period of time to another or versatile if they switch clusters. The results showed that some female offenders do switch offending clusters and become more versatile as they age. This study also showed that using latent class analysis and latent transition analysis to examine offending behaviour is worthwhile in gaining insights about the patterns of criminal activity over time.

2.7 Methodologies

Several methodologies have been adopted by researchers to measure the various dimensions in criminal careers. As this thesis is interested in longitudinal latent variable methods, which examine changing patterns of offending behaviour over time, this section will concentrate on two of the most significant statistical methods. The first is group-based trajectory modelling (GBTM) and the second is latent Markov modelling (LMM).

Using developmental offending trajectories to examine criminal behaviour patterns over time has increased in popularity over recent years. Popularised by Nagin (2005), GBTM is a useful statistical method for studying changes in criminal activity over the life course. The technique allows researchers to identify distinct offending groups within the population that following separate offending trajectories or pathways over time (Piquero, 2008). It is a special case of a finite mixture model and assumes the population is made up of a discrete number of groups (Nagin et al., 2016). In response to the criminal career paradigm debates in the 1990s, Nagin and Land (1993) first introduced GBTM to tackle some of the core issues surrounding the age-crime curve (Nagin et al., 2016). In the study, the authors used a technique to analyse the Cambridge Study of Delinquent Development (Farrington and West 1990), which consisted a longitudinal count data in the form of convictions. The features of the GBTM include the grouping together of individuals into distinct clusters or classes that all follow similar development trajectories over time (Nagin et al., 2016). A more detailed description of the GBTM is given in Chapter 5. Nagin and Land, discovered that using the GBTM allowed for the discovery of different offending trajectory groups across the offending population. This clearly contrasted with the single age-crime curve that Gottfredson and Hirschi (1990) supported. Since Nagin and Land proposed the GBTM approach, many other researchers within criminology have applied this technique to discover various offending trajectory groups in other datasets (See Piquero 2008 for a review).

A second method to model longitudinal data is the latent Markov Model. The LMM approach is demonstrated and described in detail in Chapter 7. The LMM, alternatively referred to as Latent Transition Analysis (LTA) (Collins and Lanza, 2010), is a model that can identify crime mix offending patterns and also estimate how offender transition between these crime mix groups over time. The model not only assigns individuals to separate latent classes or states, but also estimates the transition probabilities over a set period of time. LMMs make the assumption that there is a latent or hidden process that ultimately influences the distribution of dependent variable (Bartolucci, 2013). The LTA methodology has been recommended by Graham et al. (1991) for modelling drug abuse and more recently has been incorporated into the field of criminal career research (McGloin et al., 2009, Francis et al., 2010). LMM has also been applied to criminology data in the paper by Bartolucci et al. (2007). However, this study used binary indicator variables to determine if an individual had a conviction, in Chapter 7 the model is extended to incorporate the count of convictions to provide a more comprehensive examination of the crime mix patterns. For a detailed review of the LMM see Bartolucci (2013).

2.8 Limitations of past research

Criminal careers have been extensively researched and many studies have been carried out to test the numerous theories surrounding the subject. However, there are several problems associated with the past research. This next section discusses the main problems identified that directly affect the areas that this study examines.

As previously mentioned above, there are many studies that have used broad categories of offences when looking at patterns of offending over time (Baker et al., 2013, Bartolucci et al., 2007, Moffitt et al., 1996, Piquero et al., 1999). These studies can miss changes in offending types over time and misinterpret specialisation in offending. This is because an offender may commit a different type of offence from one offence to the next, but both the offences come under the same category of offences. Therefore, in the results it appears that the offender is committing the same offences sequentially and will be considered specialised. For example, the studies that have used two broad categories (violent vs non-violent) (Brame et al., 2004, Lynam et al., 2004, Osgood and Schreck, 2007) may consider a person who has committed robbery and then murder as specialised, as both the offences fall under violent category of offending. Being able to separate these offences is important for analysis as noticing these changes in offending is key to understanding the crime mix patterns that occur within the data. As the majority of crimes committed tend to be non-violent offences, these studies that have primarily focused on more serious crime types and violent offences to examine specialisation, have ignored the majority of crimes committed.

In this thesis, a wide range of different offence categories are included in the analyses, therefore identifying crime mix patterns in the conviction datasets being modelled will be more thorough and accurate.

When analysing specialisation in offending, many early studies have used modelling approaches which look at offences committed sequentially such as the Forward Specialisation Coefficient (FSC) (Farrington et al., 1988, Kempf, 1987, Paternoster et al., 1998). The FSC measures the tendency to repeat an offence in succession on a scale from 0 to 1 (See Farrington et al., 1988). An offender is considered specialised, as they always commit the same offence type over and over. However, an offender is also specialised if they commit the same type of offences within a domain of offending. The FSC however, wouldn't consider an offender to be specialised if they for example, they commit burglary then theft then burglary again. This is because they haven't committed the same offence in succession. Another issue with sequential analysis is that it requires data which is date-ordered, which can be a problem with self-report data. Choosing an appropriate methodology to examine crime mix patterns and to provide evidence of specialisation or versatility in offending needs careful consideration.

Many previous studies have only covered criminal careers of young offenders (Glueck and Glueck, 1930, Farrington and West, 1990, Loeber and Snyder, 1990, Tracy, 1990, Wolfgang et al., 1972). This neglects to capture the offending behaviour into adulthood, therefore less is known about the offending behaviour of older offenders. It also fails to catch offenders who are late starters or falsely assumes offenders may have desisted as they may restart offending at a later age due to changes in life circumstances (Andersson et al., 2012). Moreover, the research and knowledge on highly chronic adult offenders is scarce due to them occupying only a small proportion of offending samples (Tarling, 1993). However, it is of great importance to be able to identify and understanding the offending behaviours of chronic adult offenders.

Along with the lack of studies into adult offending patterns, is also the shortage of research into the offending patterns on female offenders. Most criminal career theories and investigations are based upon research of male offenders (Blokland, 2005). This neglects the discovery of different patterns of criminal offending that could be displayed by female offenders. Certain researchers have suggested that females criminal career trajectories differ to those of male offenders (Silverthorn and Frick, 1999). Studies by Block et al. (2010), Blokland et al. (2010), and Francis et al. (2010) have all investigated female offending and contributed the knowledge on female criminal careers. However, there is still a need to study female offending as gaining precise knowledge on the criminal behaviour is important for criminology theories and for criminal justice practitioners (Block et al. 2010). This is because certain interventions aimed at crime reduction may be ineffective on female offenders due to them being based on male offending patterns.

Finally, it is worth mentioning the issue with identifying ‘chronic’ offenders, as defining what constitutes a chronic offender is particularly difficult. Chronic offenders according to Blumstein et al (1986) were offenders who committed a high number of serious crimes over a prolonged time period. Studies such as Wolfgang et al. (1972) consider an offender to be chronic if they commit 5 or more offences or contact with the police. However other researchers believe an offender that commits 10 or more offences is considered chronic (Chaiken and Chaiken, 1982, Ezell and Cohen, 2005, Horney and Marshall, 1991). Deciding upon how many offences to be the cut off between a chronic and non-chronic offender is not straight forward. For example defining that an individual with 5 or more offences or instead 8 or more is chronic, will create two entirely different groups with varying numbers of how many of the population are chronic offenders (Skardhamar, 2009). Although identifying and separating the chronic offenders is important and helpful it still does not tell us that much about the offending behaviours of these particular offenders. Knowing what sort of offences and the patterns and paths they follow is ultimately just as important and will provide a much better understanding of why these offenders commit more offences.

2.9 Summary

This chapter has given an overview of literature on criminal career research. It started with an introduction to some of the early work undertaken and how the criminal career research has evolved and progressed over the years. Reviews of the most influential and relevant studies have been discussed along with key criminology theories. Particular attention has been paid to three areas relevant to this thesis; criminal typologies, crime mix patterns and, specialisation and versatility. Two methods for examining longitudinal patterns of criminal activity have been briefly introduced and a discussion of some of prior research limitations has been given.

The next chapter introduces the dataset acquired for this study, the England & Wales Offender's Index (OI) and the Netherland's Criminal Careers and Life Course Study (CCLS). It discusses the advantages and disadvantages of using different longitudinal datasets along with strategies used to align the two official conviction datasets. Challenges in comparative research are also addressed.

3 DATA SOURCES AND LONGITUDINAL DATASETS

3.1 Introduction

This chapter starts by discussing the need for longitudinal data in examining changes in offending patterns over time, weighing up the advantages and disadvantages of official conviction data and self-report data. Official conviction data is chosen because of its ability to have representative samples of the offending population and consistency over long periods of time. Time can be defined as the length of period under study, such as the number of years from when an offender has their first recorded conviction to the their last recorded conviction.

The features of each dataset will be described, along with a detailed account of the sampling strategies used to align the datasets, which is needed to ensure comparability. The challenges of comparing datasets and the limitations of using official data are discussed.

3.2 Official Data and Self Report Data – Advantages and Disadvantages

To study criminal trends over time, longitudinal datasets of criminal offending are necessary to examine the patterns of offending behaviour that occur. There are various types of data sources that can be utilised which have numerous strengths and weaknesses. Typically, criminal behaviour is examined using the collection of data from three main sources; official data (e.g. police recorded crime), victimisation surveys and self-report offending surveys. As this thesis is primarily concentrated on criminal careers and patterns of offending behaviour, discussion will focus only on official data and self-report surveys.

Self-report surveys, such as the Cambridge Study in Delinquent Development (West and Farrington, 1973), provide a wealth of information on offenders and their criminal activity. Information about the individual characteristics and background factors, such as social environment, social economic status and family situation, can all be gathered from these types of surveys. They are able to capture a larger proportion of the true number of offences committed and provide a more realistic representation of criminal career features (Farrington et al., 2006). Self-report offending can help shed light on the "dark figure" of crime by allowing offenders to reveal the offences that have not been brought to the attention of law enforcement. In contrast to official data which tends to record more serious crime, self-report data can capture information on a much wider range of offences including those which are considered much less serious.

Although there are several benefits to self-report offending surveys, there are unfortunately disadvantages. Firstly, the validity of the information provided is problematic due to offenders either over or under reporting their criminal activity. Some offenders may lie about the offences they have committed, either exaggerating the amount or withholding information about all crimes they have been involved in (Thornberry and Krohn, 2003). Also self-reported offences are based on the individual's own interpretation of their offences. This can lead to misleading information as offenders could be mistaken when they report offences in survey's. Self-report data is subject to respondent biases, where individuals omit less socially acceptable behaviour such as sexual behaviour. It also can be biased due to recall errors. Individuals may have trouble recalling offences they committed in the past, particularly older offenders (Piquero et al., 2014). Self-reported offending surveys are subject to biases, which are affected by lack of respondents, as they rely greatly on voluntary participation.

Often the sample of offenders in a self-report offending survey is small and not representative of the offending population. Self-reported offending surveys are very time consuming and expensive, and often subject to validation issues (Farrington, 2001) . There are also issues with the type of questions being asked. Closed questions are easier for quantitative analysis but can hide or limit information, whereas open questions allow for a more informative answer but are more problematic to analyse quantitatively.

Official data is a popular choice for researchers to examine trends of criminal activity over time and much of the knowledge discovered about criminal careers has used official recorded data sources (Farrington, 2001). Unlike self-report data, official data sources often cover long periods of time making it ideal for studying long term criminal patterns. Such data (e.g. police recorded crime or court conviction records) is systematically recorded and the classification of offences is based upon definitions provided by the criminal justice system.

The sources of official conviction data being used in this thesis provide over 40 years of data on criminal offending and cover a wide range of offences, making it an ideal source of data for studying criminal careers.

Just like self-report data, official data also has its disadvantages. As with all official data on convictions, the true figure of crime is not realistically represented. This is because not all offenders are caught, arrested and found guilty for the crimes they commit. This means many offenders will not receive convictions for all of the offences they have committed.

A further problem is with police recording practices and the discretion of police officers when it comes to recording crimes. In 2002, The Home Office tried to standardise the counting rules for recording crimes to ensure that all crimes from all police forces were being recorded properly and consistently across the country. This means that crimes recorded from 2001/02 to 2003/04 saw an increase overall in the recorded crime rate, but this did not necessarily mean that crime was increasing. The rise in the overall recorded crime rate was probably due to better recording practices by police forces following the implementation of the new counting rules (Simmons et al., 2003). This has potential implications for the analysis of crime over time as it may give misleading results. Therefore comparing recorded crimes prior to the implementation of the new counting rules to crimes recorded after, should be done with caution as they are not directly comparable.

In addition to this, official crime data is also subject to data entry errors when recording, for example some individuals may be recorded as multiple offenders, when they are not. This therefore can impact the data recorded on repeat offenders which is of high value to law enforcement agencies who are trying to target chronic offenders. There is also no control for the decision of victims to come forward to report crimes and this again contributes to the 'dark figure' of crime (Coleman, 1996).

Conviction data, however, is also problematic. There is the issue of pseudo-reconvictions. These are convictions that occur after the recorded conviction which were committed prior to the conviction date, but are related to the offence or offences(Howard and Kershaw, 2000). This can occur when there are several trials against an offender for numerous offences committed over a series of dates.

Therefore, pseudo-reconvictions can give a false representation of the actual rate of reconvictions and make it rather complex when looking at comparisons between reconviction rates for different disposal methods.

3.3 Why Longitudinal Data Over Cross-Sectional Data?

The use of longitudinal data for studying criminal careers, has been criticised by Gottfredson and Hirshi (Gottfredson and Hirschi, 1986, Gottfredson and Hirschi, 1988, Gottfredson and Hirschi, 1990). They claim that longitudinal data is unnecessary as they believe the predisposition to commit crime is stable over the life course. Therefore, it is thought that cross-sectional data is suitable and will return the same results as a longitudinal study (DeLisi and Vaughn, 2007).

Cross-sectional data analysis is often considered a more time efficient and less costly option and reduces the risk of panel attrition, for example when an offender dies before the end of the study period (Menard and Elliott, 1990).

However, there are several reasons why longitudinal data sources are more suited to the types of analyses within this thesis compared to cross sectional data sources. Examining patterns in criminal behaviour over time requires longitudinal datasets to provide information on a number of criminal career dimensions, such as age on onset, duration, desistance, frequency and specialisation of offending (Farrington, 1992). Certain offending patterns could easily be missed with cross-sectional data, for example an individual may begin committing less serious offences and escalate to more serious offences before desisting from offending. Using longitudinal datasets of criminal behaviour allows for the examination of stability and continuity of offending patterns over the life span (Farrington, 1992). Longitudinal datasets make it possible to try and predict future patterns of offending using previous offending behaviour (Farrington, 1992).

For this study, two longitudinal datasets containing official conviction data from two countries (England &Wales and the Netherlands) have been obtained. Choosing to examine conviction data from two different countries provides cross national comparisons, which is highly desirable, particularly for law enforcement agencies as it allows investigation into what extent crime prevention strategies work to reduce offending in different environments (Farrington, 2015).

3.4 England & Wales Offenders Index

3.4.1 Description of the Dataset

The first dataset to be discussed is the Offenders Index (OI) database for England and Wales[2]. The full Offenders Index contains the court convictions of all offenders from 1963 to the end of 2008, when the collection of data stopped. The OI was originally created for researchers to statistically analyse the criminal histories of offenders.

A subset of the Offenders Index has been extracted consisting of the criminal histories of offenders in eight birth cohorts (1953, 1958, 1963, 1968, 1973, 1978, 1983 and 1988), each representing a four week sample (around a one in thirteen sample) of all offenders born in a specific year. The same four weeks were chosen for each cohort year (the weeks are 3rd-9th March, 19th-25th June, 28th September - 4thOctober and 17th-23rd December). Court reports submitted by the police are built upon to form the database. The data is collected from age 10[3], with histories followed up until 2008. The OI is a vast database and researchers can gain information on over 200,000 offenders per annum.

The dataset has a hierarchical structure with three levels: the first level - the offender details; the second level – offence details and finally the third level - proceeding details. It is, however, supplied in a flat form, with first and second level variables duplicated for all third level records.

Not all offences form part of the O I- with only offences that are considered ‘standard list’ being included. ‘Standard list’ offences are those which are considered indictable

[2] Only information for offenders convicted in England and Wales is included in the database. Offenders convicted in Scotland and Northern Ireland are not included. For the period 1987 to July 1992 there is incomplete data for the Metropolitan police area.
[3] The age of criminal responsibility changed from 8 years old to 10 years old in February 1964. Therefore, some offenders in the 1953 cohort have convictions under the age of 10 years.

(Crown Court offences) or triable either way (either in the Magistrate' Court or the Crown Court) and the more serious summary (Magistrate's Court) offences. Therefore, many minor offences are excluded.

Between 1995 and 1996, several important new offences were considered 'standard list' offences. This increased the number of convictions that were added to the OI by almost 100,000 annually. These new offences included all common assault categories and several driving offences (driving whilst disqualified, driving with excess alcohol and dangerous driving).

The OI dataset contains only a limited amount of information on everyone. This includes the date of birth, age at conviction, gender, offence code, dates of court appearances, the number of convictions at each appearance and the disposal outcome. As the dataset is anonymised, the name of the offender, the criminal record number and the police force that dealt with the offender are excluded. Instead each offender is given an OI identification number, making it possible to distinguish between the offenders.

3.4.2 Data Security and Protection

Access to the OI is restricted and external researchers requiring data from the OI must gain permission from the Ministry of Justice (MOJ). After a written request was sent to the MOJ, explaining in detail the intended use of the dataset, access was granted subject to a number of security requirements to ensure data protection and confidentiality of offenders. Security measures have been imposed to ensure all rules and regulations surrounding the use of the data are maintained, including storing data upon encrypted devices. The anonymised dataset was saved onto an encrypted and external hard-drive, which was physically secured within a locked office with a security cable. Accessing the hard-drive was made only possible on my laptop, which was only connected when the dataset was being used for analysis. Any results and analysis upon the dataset was always saved onto the encrypted hard-drive. Extreme care was taken to ensure all security rules and regulations were met and adhered to. The presentation of data and results throughout this thesis has maintained all data confidentiality rules.

3.4.3 Advantages of the Offenders Index

There are numerous advantages to the Offenders Index dataset. It is an excellent resource for examining long term patterns of offending, providing criminal histories over a 45-year period. Although access to the dataset is restricted, there is an anonymised subset of the dataset which is publicly available from the Economic and Social Data Service[4] which contains a more limited range of data. The subset of the data available is from the 1953 - 1978 cohorts.

The OI has a user guidebook (Home Office, 1998b), which contains information on how to interpret the data and how it can be analysed using SPSS.

Unlike the subset of data which is publicly available at the UK Data Archive, the dataset being used in this thesis is more up to date and contains all 8 birth cohorts up to 1988.

The definition of offences over time is extremely consistent, with only a few changes to what is considered a 'standard list' offence. However, it is fairly simple to deal with any changes to offences over time by removing those offences which stopped or started being 'standard list.'

As there are eight available cohort years, this dataset provides the opportunity to examine generational changes. An advantage of investigating cohort differences is that it is possible to check if any changes in the patterns of offending are due to

[4] The dataset SN 3935 available at UK Data Archive (http://discover.ukdataservice.ac.uk/catalogue?sn=3935)

generational effects. Possible generational effects include: changes in the criminal justice system, policing, social attitudes towards crime and changes in the social environment.

Each cohort is extremely large containing a huge number of cases and convictions. For example, the 1953 cohort contains 53678 convictions.
Tables 3.1 and 3.2 contain details of the number of convictions and the number of offenders respectively, in the various cohorts. Therefore, it is possible to test the robustness of the dataset by analysing two large samples of the dataset.[5]

Table 3.1 Number of convictions for each OI cohort by gender

Cohort	**1953**	**1958**	**1963**	**1968**	**1973**	**1978**	**1983**	**1988**
Male	47624	59527	79218	79855	69183	52504	40180	18052
Female	6054	7627	10855	10066	9218	7601	4744	2711
Total	*53678*	*67154*	*90073*	*89921*	*78401*	*60105*	*44924*	*20763*

Table 3.2 Number of offenders for each OI cohort by gender

Cohort	**1953**	**1958**	**1963**	**1968**	**1973**	**1978**	**1983**	**1988**
Male	9077	9956	11527	10316	7749	6261	5269	2473
Female	2330	2712	2981	2277	1538	1247	923	548
Total	*11407*	*12668*	*14508*	*12593*	*9287*	*7508*	*6192*	*3021*

The date of when the offence took place is not recorded, only the court appearance is. Since there can be a substantial delay between the offence date and court date in some instances the offender will be a different age when they committed the offence to the age recorded. This can cause inaccuracies in the results when analysing the data giving misleading portrayals of the number of offenders committing crimes at specific ages.

There is no information recorded if an offender appeared at court but was given a fixed penalty fine and nothing is recorded if the prosecution was unsuccessful. Similarly, Police cautions and warnings are not recorded. This can again be viewed as potentially leading to biased results, as for example the fact the offender has appeared at court and been given a fine shows that a crime has taken place and that a punishment has been issued to the offender.

3.4.4 Disadvantages of the Offenders Index

Although there are numerous advantages to using the OI dataset for analysing criminal careers, there are a few limitations to using official conviction data that researchers need to be aware of when interpreting results.

Problems can occur if an offender dies, as this is not recorded and the offender is not removed from the database. Instead it would be viewed as the start of a period of inactivity or non-offending. There is also no information recorded on immigration or emigration and again these offenders would appear to have periods of 'non-offending' misleadingly recorded.

The OI database is created through a matching process which uses the offenders surname, DOB and criminal record number (if available). This is subject to human error and can produce inaccuracies in the dataset especially with female offenders, who may change surname when they get married for example. Francis and Crosland (2002) also found matching problems with common surnames such as Singh.

Even though it is possible to follow the criminal histories of some offenders for a period of over 40 years, this is not the case for offenders who are born in the later cohorts. Those born in 1988 only have conviction histories of up to 20 years, making them unsuitable to use to examine the longitudinal patterns of offending over the life course.

As previously discussed, all longitudinal datasets of criminal convictions are subject to: changes in the law over time, changes in offence categories and the creation of new offences. For the purposes of this thesis, any offences which are subject to changes have either been removed, or had relevant re-coding so that they are consistent over the period of the study.

Due to the lengthy time of the dataset, the results of analysis are also vulnerable to social changes over time. This means that any changes in the patterns of offending could be influenced by changing attitudes towards certain offences. Over time some offences may be viewed as less serious meaning the offender may receive a lesser sentence which is not necessarily recorded as a conviction and therefore does not appear on the database.

3.5 The Netherlands Criminal Careers and Life Course Study

3.5.1 Description of the Dataset

The second dataset acquired is the from the Criminal Careers and Life Course Study (CCLS), which is an extensive study carried out by the Netherlands Institute for the Study of Crime and Law Enforcement (Blokland et al., 2005). The CCLS consists of conviction data from the Netherlands and contains a representative 4% random sample of all offenders convicted in 1977. All the cases in the sample are either ruled upon by a judge or public prosecutor.

In the Dutch criminal justice system, a public prosecutor has the power to decide whether to prosecute each case. They can make the decision to drop the case if they believe there is insufficient evidence to lead to a conviction.

The dataset is a large sample and it follows individuals from age 12 right through till 2002, where some offenders are aged 87 years. There is at a least a 25 year follow up period after the age of the sampled conviction year in 1977 and there is also retrospective data from the age of the offender at 1977 back to age 12.

There is a variety of information provided for each case, including the offenders' gender, ethnicity and employment status. Information is available on the type of offence committed (coded into 28 offence categories) and how the case was dealt with in court.

Before release, some adjustments were made to the data by the Netherlands research team. Drink driving convictions had to be reduced to 2% as they were over-represented in the sample[6]. Some very serious and rare offences[7] were under-represented and therefore were over-sampled in the dataset to represent the original distribution of offences. The dataset has the inclusion of a weight factor to account for the structure of the sample so that the weighted sample is representative of the distribution of the offences when they were tried in 1977 (Blokland, 2005). For each offence in the CCLS dataset, the charge with the highest punishment was coded[8]. By choosing the primary charge for each offender, the severity of offences that individuals are convicted for are slightly overestimated in the dataset (Blokland, 2005).

Extracts from the General Documentation File (GDF) of the Dutch Criminal Records Office were used to create the criminal histories of the offenders from the sample. The criminal cases that are registered by the Public Prosecutor's Office are contained in the GDF. Researchers could use these files to build up the offending history of the 1977 sample up to 2002.

For a small number of the offenders, their information had been recorded more than once due to having more than one conviction in 1977. The most serious conviction in 1977 was therefore chosen and retained in the dataset.

[6] This was done by the Netherlands research team.

[7] Robbery, attempted robbery, public violence, and battery were sampled at 25%. Murder, attempted murder, offences against decency, rape, child molesting, and other sexual assaults were sampled at 100% and drug offences were sampled at 17%.

[8] An offender can receive a primary charge, which will be for the most serious offence they've committed, and several subsidiary charges which are for less severe offences. This prevents offenders from being acquitted in cases where there is sufficient evidence the offender has committed an offence, but not all aspects from the primary charge can be proven beyond reasonable doubt. Blokland, A. 2005. *Crime over the life span; trajectories of criminal behavior in Dutch offenders*, Netherlands Institute for the study of Crime and Law Enforcement, Faculty of Law, Leiden University.

Full offending histories were reconstructed for 5164 offenders, including foreign national individuals. The dataset then excluded offenders who were born outside the Netherlands due to unreliable information on their offending history outside the country. All cases in the dataset resulted in a conviction, prosecutorial fine or policy waiver.

Similar to the OI, no date is recorded in the CCLS of when the offence took place. This means the offence date as recorded in the dataset is actually the date it was registered at the public prosecutor's office. This happens relatively early on during a police investigation, as the registration of the offence is needed to allow investigative powers to be implemented pre-trial.

The year of birth distribution of the offenders varies greatly, from 1912 to 1965, making the offenders in the CCLS dataset much older than the OI dataset. Table 3.3 summarises the distribution of the year of birth of the offenders in the CCLS dataset.

Table 3.3 CCLS Year of Birth Distribution

Year of Birth	**1912-1921**	**1922-1931**	**1932-1941**	**1942-1951**	**1952-1961**	**1962+**
Male	74	251	481	1083	2112	164
Female	14	48	79	153	123	10
Total	*88*	*299*	*560*	*1236*	*2235*	*174*

The dataset sample is extremely small in comparison to the OI. Even just one cohort of the OI is significantly larger than the CCLS sample of offenders. It is also subject to similar problems as the OI such as changes in the law or offence definitions, social changes and pseudo-reconvictions.

3.6 Challenges in Cross National Comparisons

For certain analyses where it has been necessary to compare both the OI dataset and CCLS dataset, a number of challenges have arisen. The next section covers the problems that arise and the solutions that have been used to approach them.

As the aim is to identify similarities or differences in criminal career patterns across the two different datasets, there is a need to control both for known differences in the criminal justice system and also for sample differences.

Firstly, the criminal justice system differences are considered. England and Wales use the adversarial system and the Netherlands use the inquisitorial system. In the adversarial system, the court is there to act as an unbiased referee between the defence and the prosecution. However, in the inquisitorial system, the court is involved in investigating the facts of the case. Additionally, during the 1960s, the Netherlands tended to predominantly use diversionary methods away from court for juveniles when compared to England and Wales. Prison sentences are also usually much shorter in the Netherlands and custodial sentences are less likely than in England and Wales. The age of criminal responsibility is also different: age 12 in the Netherlands and age 10 in England and Wales.

Looking at sampling differences, it should be noted that the different sampling methodologies mean that the year of birth distribution varies greatly between the two datasets. The CCLS dataset has a wider range - from 1912 to 1965 –with all offenders having a conviction in 1977. On the other hand, the OI dataset has one of eight pre-specified years of birth, with convictions in any year from age 10 up to 2008. To reduce any generational differences, offender samples have only been taken from the 1953, 1958 and 1963 cohorts from the OI as they include offenders who can have a conviction in 1977.

There are also differences in offence categories. The OI dataset uses a complex coding system with over 500 offence codes which are documented in the Home Office Offenders Index codebook (Home Office, 1998a). The CCLS dataset classifies offences based on the Dutch legal code and the offence types are registered by the Dutch Ministry of Justice – 28 categories are used.

3.6.1 Alignment of the Two Datasets

The previous section highlighted the problems in ensuring that the two datasets are as similar as possible before analysis. For the analyses performed in Chapters 5, 6, and 7 the datasets are initially aligned using the following strategy. Some further restrictions are then applied which are discussed in the relevant chapters.

Common start age of offending

For the OI sample, all offences prior to age 12 (the age of criminal responsibility in the Netherlands) were excluded.

A conviction in 1977

As the offenders in the Netherlands sample all had at least one conviction in 1977, offenders in the OI sample without a conviction in 1977 were excluded. This meant only the 1953, 1958 and 1963 OI cohorts were used in analyses as they were the only cohorts that included offenders with a conviction in 1977. This is important as offenders from later cohorts are younger and are subject to social and law changes which may affect the number and type of convictions received. It also ensures that the offenders selected from the OI dataset are more directly comparable to those in the CCLS dataset as all offenders are restricted to those with a conviction in 1977.

Matching offence categories

Aligning offence categories across the two datasets was a more complex problem. Coding manuals were examined to determine which offences in the OI dataset best aligned with those in the CCLS dataset. In addition, the categories for both the Netherlands and the England and Wales datasets were collapsed so that similar offences would be incorporated into one offence category. Furthermore, some of the more minor offence categories in the CCLS dataset which did not exist in the OI dataset were omitted.

These adjustments led to 11 common offence categories across the two jurisdictions which were similar in nature. Table 3.4 shows the offence categories used throughout the analysis and more detail on each of these categories can be found in the Appendix.

Table 3.4 The 11 Offence Categories

Offence category
Murder/Violence
Firearms
Authority
Sexual Offences
Blackmail
Robbery
Burglary/Theft
Fraud/Forgery
Criminal Damage
Drugs
Public Order

In Chapters 5 and 6, further data restrictions are placed upon the two datasets for modelling longitudinal trajectories of offending leaving 4420 offenders in the CCLS dataset. For chapter 7, additional data restrictions are imposed leaving the CCLS with 2267 offenders. For the Offenders Index, a random sample of 4420 offenders was taken from the 1953, 1958 and 1963 cohorts for chapters 5 and 6. A random sample of 2267 offenders, from the same three cohorts, was also taken for the analysis in Chapter 7. These matched sample sizes are based upon how many CCLS offenders are left once the necessary restrictions are imposed as this is the smaller dataset. The sample size is smaller in chapter 7 due to increased restrictions placed upon the dataset before proceeding with more complex analyses. More details on the restrictions are discussed in the corresponding chapters.

3.7 Conclusion

This chapter has discussed and focused on longitudinal data sources used for examining changes in offending patterns over time. After weighing up the advantages and disadvantages of various types of longitudinal data, two official convictions datasets have been presented and described – The Offenders Index for England and Wales and the Criminal Careers Life-Course study for the Netherlands.

Although there are many benefits to using two different datasets there are limitation as well. The issues that are arise when comparing different datasets have been discussed along with a brief strategy outlined on how to tackle these problems of aligning the datasets.

The next chapter contains exploratory analysis of both datasets, giving a more detailed description of them and highlighting their differences and similarities.

4 EXPLORATORY ANALYSIS OF THE DATASETS

4.1 Introduction

This chapter introduces and explores the two datasets. Section 4.2 provides a definition of significant criminal history variables which are constructed from the datasets. This is followed by section 4.3, which present an exploratory analysis of the Offenders Index and the CCLS datasets. The chapter concludes with a discussion of the similarities and differences between the datasets.

4.2 Definition of Criminal History Variables

There are several criminal history variables which will be constructed from the datasets that need explanation before proceeding with the exploratory analysis of the datasets.

Conviction and Conviction Occasion – A conviction is when an offender is found guilty of a criminal offence. Offenders may have several convictions at one court appearance, and this is called a conviction occasion.

Frequency of offending – This is measured by the total number of conviction occasions an individual has over the time period under observation. In more detail, each conviction occasion can have one or more convictions; and conviction occasions are counted to measure frequency. Frequency of offending therefore is counting the number of sentencing dates, where multiple convictions for offences may be made.

This definition of frequency is used as convictions will tend to be grouped together and the use of convictions introduces considerable overdispersion into the data.

Offenders will carry out criminal offences until one is detected. This detection will in turn tend to lead to the detection of other earlier criminal offences, and the cluster of offences will be brought to court at the same time.

Age of onset – this is the age of the offender at the first recorded conviction. This may not be the actual onset age of offending as not all offences lead to a conviction and not all offences committed by individuals are recorded. In addition, even if the offender is convicted, there will still be a time lag between offence and conviction. Some offenders may receive their first conviction for an offence they committed at a younger age due to the time processes in bringing an offence to court or they may have begun offending at an earlier age and not been caught by law enforcement agencies. This means the true age of onset is likely to be a younger age than what is recorded in the data.

Participation in offending – this is defined to be the distinction between those who commit crime and those who do not. Every individual in both datasets has participated in offending, having at least one conviction. Again, as conviction histories are used to measure participation, some of those recorded as non-offending will be offenders who have not been captured by the Criminal Justice System.

Reconviction – This is defined to be at least one conviction occasion after the target conviction occasion within an analysis-defined time period. In this study, short term recidivism (2-year time horizon) and long term recidivism (10-year time horizon) are both examined. Recidivism is a general word meaning variously rearrested, reoffending or reconviction. In this study the term recidivism is taken to mean reconviction and the two terms are used interchangeably.

Desistance – This refers to termination from offending. However, it is important to realise that a period of inactivity can sometimes be mistaken for desistance from offending.

Duration – This refers to the time between onset and desistance from offending. For this study, however, this is the time between the first recorded conviction occasion and the last known conviction occasion.

4.3 Exploratory Analysis of the Datasets

The original Offenders Index dataset received from the Ministry of Justice contained conviction data on offenders (those convicted in an England and Wales court) from eight separate birth cohorts. In total, there is data available on the criminal histories of 77,184 offenders. A small portion of the offenders in the sample are female (19%).

There are over 500 offence classifications for the standard list offences as categorised by the Home Office, (see OI Codebook Home Office, 1998). Therefore, a significant amount of data cleaning was required before progressing with any analysis. The eight OI datasets were recoded so that the offences were re-categorised into 38 categories (See Appendix).

The original CCLS dataset received from the NSCR contained 4597 offenders, with all having at least one conviction in 1977. Only a very small proportion of the offenders in the sample are female (9%). Offences are organised into 28 offence categories defined by the Dutch Ministry of Justice (See Appendix). Some pre-processing was applied to the dataset to ensure any data input errors were eradicated, as follows: firstly, any individuals that did not have a conviction in one of the 24 offence categories were removed[9,10]. Then, only cases with an age at conviction between 12 (the age of criminal responsibility in the Netherlands) and 90 and that have a year of birth between 1912 and 1965 are kept. This left 4420 offenders in the CCLS dataset, with female offenders accounting for 9% (379) of the

[9] The dataset received had 28 offence categories but 4 of these were removed for exploratory analysis – misdemeanours, other, traffic and unknown.
[10] Due to data input errors, some offenders in the database did not have any recorded convictions in any of the 28 categories.

total number of offenders.

Unlike the England and Wales Offenders Index, where data was given for each conviction occasion, the CCLS data was supplied on a year by year format, with the convictions in each calendar year summarised for each offender. Overall the 4420 participated in 32,224 conviction years.

As this a comparative study, it was necessary to create a sample dataset of the OI that is more closely aligned with the CCLS dataset before continuing with any analyses. Date of birth was aligned first. As discussed in Chapter 3, the CCLS dataset is a subset of all those convicted in 1977. Any offender in the Offenders Index birth cohorts from 1968 on would not have a conviction in the year 1977. So only the first three OI cohorts (1953, 1958 and 1963) were used.

Secondly only those offenders who had a conviction in 1977 were selected. Next, a random sample of 4420 offenders was taken from across the three cohorts – this sample size matched the size of the CCLS dataset once restrictions had been imposed, as mentioned in Chapter 3. Finally, all criminal convictions before age 12 were ignored; again, aligning with the age of criminal responsibility in the Netherlands.

4.3.1 Participation and Frequency of Offending

In the OI sample, all offenders had participated in criminal offending and had at least one court conviction, and these 4420 offenders were convicted on 25,135 occasions over the period of the study (From age 12 up until 2008). This sample consisted of 80% males (3525 offenders) and 20% females (895 offenders). The mean number of conviction occasions per person was 5.7, and the modal number of conviction occasions per person was one. When individuals with only one conviction occasion were excluded, the mean number of conviction occasions increased to 8.7 and the total number of conviction occasions became 23,394.

Including all 4420 offenders, 56% of individuals have 1-2 conviction occasions which is 13% of the total number of conviction occasions. However, when examining the most chronic offenders 5% (236) of individuals are responsible for 42% (10,509) of all conviction occasions.

In the CCLS sample, once the above-mentioned restrictions were applied, 4420 individuals in the sample were discovered to have participated in criminal offending in 1977. The mean number of convictions per person therefore comprised of 9.8 offences and most offenders had only 1 conviction. In addition, 39% (1737) of the 4420 offenders had between 1 and 2 conviction occasions, which is 4% of the total number of conviction occasions. In contrast, for the most chronic offenders, 5% (219) of individuals in the sample were responsible for 32% (13979) of all conviction occasions.

4.3.2 Reconviction

Over one third of the offenders in the OI (39%) had only one conviction occasion, and they are referred to as the non-recidivists. The remaining 61% of offenders are all reconvicted within the period under analysis with some extremely chronic offenders having over 100 conviction occasions.

Male offenders accounted for 85% of the reconvicting offenders, and only 15% were female offenders. The prevalence of reconviction is shown in Table 4.1; this shows the proportion of offenders who have at least one other conviction occasion after their first recorded conviction occasion.

Many of the CCLS offenders (16%) had only one conviction occasion, and they are referred to as the non-recidivists. The remaining 84% of offenders all reconvicted within the period under analysis with some extremely chronic offenders having over

197 conviction occasions. Male offenders accounted for 3493 of the recidivating offenders, and only 222 were female offenders. The prevalence of reconviction is shown in Table 4.3 which shows the proportion of offenders who have at least one other conviction occasion after their first recorded conviction.

Table 4.1 Prevalence of reconviction of OI offenders

	Non-recidivists		**Recidivists**	
	N	***%***	***N***	***%***
Male	1257	36%	2268	64%
Female	484	54%	411	46%
Total	*1741*	*39%*	*2679*	*61%*

Table 4.2 Conviction frequency distribution of OI offenders

# conviction occasions per individual	**# individuals**	**% individuals**	**# of conviction occasions**	**% convictions occasions**
1 to 2	2470	55.8%	3199	12.7%
3 to 4	710	16%	2395	9.5%
5 to 10	672	15.2%	4638	18.5%
11 to 20	332	7.5%	4754	18.9%
21 to 40	152	3.4%	4444	17.7%
41 +	84	1.9%	5705	22.7%
Total	*4420*	*100.00%*	*25135*	*100.00%*

Table 4.3 Prevalence of recidivism of CCLS offenders

	Non-recidivists		**Recidivists**	
	N	***%***	***N***	***%***
Male	548	14%	3493	86%
Female	157	41%	222	59%
Total	*705*	*16%*	*3715*	*84%*

Table 4.4 Conviction frequency distribution of CCLS offenders

# conviction occasions per individual	**# individuals**	**% individuals**	**# of conviction occasions**	**% convictions occasions**
1 to 2	1737	39.30%	1761	4.07%
3 to 4	616	13.94%	2108	4.87%
5 to 10	904	20.45%	6410	14.82%
11 to 20	560	12.67%	8192	18.94%
21 to 40	384	8.69%	10811	24.99%
41 +	219	4.95%	13979	32.31%
Total	*4420*	*100.00%*	*32224*	*100.00%*

As Table 4.2 and Table 4.3 and Table 4.4 shows, the offending frequency is not distributed uniformly for either dataset. Most offenders only have 1-2 conviction occasions, accounting for nearly 56% of all OI offenders and for just under 40% of all CCLS offenders. Many of the reconvicting offenders have only 1-3 subsequent conviction occasions. There are however a small proportion of offenders who are responsible for a large proportion of all conviction occasions.

4.3.3 Age of Onset

As previously mentioned, the age of onset in this study refers to the age at which an individual receives their first conviction. This does not necessarily mean it is the same age when an individual begins their criminal career.

Just over half of offenders in the OI dataset received their first conviction between the ages of 12 and 20 years, with 13% of individuals receiving their first conviction before the age of 15 years. It can be seen that there are fewer individuals in the Netherlands who received their first conviction between 12-14 years, which may suggest that younger offenders are more likely to be convicted in the England and Wales criminal justice system compared to the Dutch criminal justice system.

There is a substantial proportion of OI offenders who receive their first convictions at later ages with just over 35% of offenders having their first recorded conviction occasion after 24 years of age.

Table 4.6 shows the distribution of convictions by age for the CCLS. Like the OI dataset, the majority of offenders in the CCLS dataset received their first conviction between the ages of 12 and 20 years. However, only 10% of the CCLS offenders received their first conviction before the age of 15, compared to OI dataset where 13% received their first conviction before age 15. This could be due to the diversion of younger juvenile offenders away from court and towards other forms of disposal such as to caution in the Dutch criminal justice system. There are still a proportion of offenders who receive their first convictions at later ages. Table 4.6 also shows that 22% of individuals had their first conviction at age 27 years or over. A proportion of individuals are therefore late-starters (Thornberry, 2005) and begin their criminal careers at an older age.

Table 4.5 Distribution of onset age for OI offenders

Age of onset	**%**	**# of offenders**
12 to 14	12.9%	574
15 to 17	23.5%	1039
18 to 20	20.1%	887
21 to 23	8.3%	368
24 to 26	8.7%	385
27 to 29	6.0%	266
30 to 32	4.3%	190
33 to 35	3.8%	170
36 to 38	3.6%	157
39 to 41	4.4%	193
42 +	4.3%	191
Total	*100.0%*	*4420*

Table 4.6 Distribution of onset age of CCLS offenders

Age of onset	**%**	**# of offenders**
12 to 14	9.9%	437
15 to 17	26.9%	1188
18 to 20	21.7%	959
21 to 23	12.6%	558
24 to 26	7.2%	318
27 to 29	5.2%	231
30 to 32	4.4%	194
33 to 35	2.9%	130
36 to 38	2.0%	88
39 to 41	1.7%	74
42 +	5.5%	243
Total	*100.0%*	*4420*

4.3.4 Offence Categories

In order to compare the CCLS dataset with the OI dataset, the offence categories were compared and common categories were identified. This led to 11 offence categories being identified as common across the two datasets, as shown in Table 3.4 in the previous chapter. When an individual receives a conviction at court, they are prosecuted for the most serious offence which is called the ‘principal conviction’. Offenders can have between 1 and 25 convictions (of different offences) per court appearance, averaging around 1.5. All other offences were placed into the ‘Other’ category (these consist mainly of administrative offences such as absconding from bail).

Table 4.7 shows the frequency of the classification of offences for principal convictions of offenders in the OI dataset. The most common first time offence that individuals in the Offenders Index dataset are convicted of are ‘Burglary & Theft’ offences, comprising of just under 40% of all first-time convictions. Excluding the ‘Other’ offence category, the second most common offence category was ‘Fraud/Forgery’ (11%) closely followed the ‘Murder/Violence’ category (10%). The least common first time offences were for ‘Blackmail’ which made up just 0.11% of all first-time convictions.

As the 11 offence categories have been selected that most closely align with the offence categories in the OI several of the offence categories that are contained in the CCLS dataset are excluded from this and placed together in the “other” offences category.

The most common first principal offence that individuals in the Netherlands dataset are convicted of are ‘Burglary & Theft’ offences, comprising of just under 45% of all first-time convictions. The second most common offence category was the ‘Murder/Violence’ category, which accounted for just over 11% of all first-time convictions. The least common first principal offences were for ‘Blackmail and Robbery’ which together made up less than 2% of all first-time convictions.

Table 4.7 Frequency of first principal convictions by offence category for OI offenders

Offence category	Frequency	%
Murder/Violence	445	10.07%
Firearms	42	0.95%
Authority	61	1.38%
Sexual Offences	72	1.63%
Blackmail	5	0.11%
Robbery	29	0.66%
Burglary/Theft	1729	39.12%
Fraud/Forgery	465	10.52%
Criminal Damage	418	9.46%
Drugs	179	4.05%
Public Order	22	0.50%
Other	953	21.56%
Total	*4420*	*100%*

Table 4.8 Frequency of first principal convictions by offence category in CCLS

Offence category	Frequency	%
Murder/Violence	503	11.38%
Firearms	59	1.33%
Authority	105	2.38%
Sexual Offences	377	8.53%
Blackmail	13	0.29%
Robbery	48	1.09%
Burglary/Theft	1962	44.39%
Fraud/Forgery	308	6.97%
Criminal Damage	265	6.00%
Drugs	167	3.78%
Public Order	115	2.60%
Other	498	11.27%
Total	*4420*	*100%*

Examining the total frequency of conviction occasions over the entire period of the study Table 4.9 and Table 4.10 shows similar results. For the OI, again the ‘Burglary & Theft’ category is the most common offence category accounting for 37% of all conviction occasions. Blackmail, Robbery and Public Order are again the least common of all conviction occasions.

For the CCLS, again the ‘Burglary & Theft’ category is the most common offence category accounting for almost half of all convictions. Blackmail and Robbery, again, account for the lowest frequency of convictions.

Table 4.9 Frequency of conviction occasions by offence category for OI offenders

Offence category	Frequency	%
Murder/Violence	2238	8.90%
Firearms	292	1.16%
Authority	353	1.40%
Sexual Offences	391	1.56%
Blackmail	20	0.08%
Robbery	162	0.64%
Burglary/Theft	9359	37.23%
Fraud/Forgery	2856	11.36%
Criminal Damage	1920	7.64%
Drugs	1177	4.68%
Public Order	122	0.49%
Other	6245	24.85%
Total	*25135*	*100%*

Table 4.10 Frequency of conviction occasions by offence category of CCLS offenders

Offence category	Frequency	%
Murder/Violence	5943	13.7%
Firearms	938	2.2%
Authority	1207	2.8%
Sexual Offences	2041	4.7%
Blackmail	243	0.6%
Robbery	788	1.8%
Burglary/Theft	21299	49.2%
Fraud/Forgery	3093	7.2%
Criminal Damage	2769	6.4%
Drugs	2370	5.4%
Public Order	1322	3%
Other	1248	2.9%
Total	*43261*	*100%*

4.3.5 Duration and Last Known Conviction

The age at last conviction does not indicate termination from offending, only the last known offence in the period of the study. Each individual in the dataset has a last known offence, even if they only have the one convicted offence.

The duration of the criminal career is measured as the time in years from age of first conviction to age at last conviction. The actual duration of the criminal career may possibly be longer. Individuals may have been criminally active before their first conviction, and not caught by law enforcement, and/or they may continue offending after the last known conviction. Therefore, the examination of the duration of criminal careers needs to be interpreted with caution.

In the OI, the exposure length of the offenders is different in nature compared to those in the CCLS dataset. Those born in 1953 have criminal histories collected from 1963 to 2008 (up to age 55), those born in 1958 have criminal histories collected from 1968 to 2008 (up to age 50) and finally those born in 1963 have criminal histories collected from 1973 to 2008 (up to age 45).

In the CCLS the exposure length of the offenders is different again, and varies by age. Those aged 12 in 1977 have their criminal history collected from 1977 to 2002 (up to age 37) whereas those born earlier have a longer exposure length. Therefore, if an offender was born in 1912 they would have their entire criminal activity over the life course recorded (from age 12 till 90 years or death).

On average the age of the last known conviction occasion was 27.4 years for the OI offenders. The duration of offending ranged from zero years (offenders with only one conviction occasion) to 55 years (1963-2008). The average duration length of a criminal career, excluding those with only one convictions occasion, is 12.8 years. A summary of duration of offending for offenders in the OI dataset can be found in Table 4.11.

In the CCLS, on average the age of last known conviction occasion was 40.9 years. This was similar for both male (40.9 years) and female (40.3 years) offenders. The duration of offending ranged from zero years (offenders with only one conviction occasion) to 64 years. For 705 (16%) of offenders, the first conviction occasion was also their last known conviction. The average duration length of a criminal career in this Netherlands sample, excluding careers with only one conviction year, is 21.6 years.

Table 4.11 Criminal career duration in years of OI offenders

Duration	frequency	%
0-9 years	3323	75.18%
10-19 years	558	12.62%
20-29 years	397	8.98%
30-39 years	138	3.12%
40+ years	4	0.09%
Total	*4420*	*100%*

Table 4.12 Criminal career duration in years for CCLS offenders

Duration	frequency	%
0-9 years	1318	29.82%
10-19 years	803	18.17%
20-29 years	1368	30.95%
30-39 years	721	16.31%
40+ years	210	4.75%
Total	*4420*	*100%*

4.4 Conclusion

This chapter has provided an exploratory analysis of the two official conviction datasets. It has introduced some significant criminal history variables which have been constructed from the datasets.

The construction and compilation of the datasets have been described along with strategies for aligning them. Neither of the full datasets are fully explored for this chapter due to the restrictions placed upon them. Although measures have been taken to try and align the datasets as closely as possible these will never be perfect and this must be kept in mind when interpreting any results.

After exploring the restricted datasets, there are some similarities and differences worth summarising. The Offenders Index had 25,135 conviction occasions and the CCLS had 32,224 conviction occasions, significantly more than the OI sample. Participation in offending was highest for males in both datasets and male offenders are much more likely reconvict.

For the OI, the average offender was convicted on 5.7 occasions and for the average CCLS offender this was 9.8. Offence frequency was not evenly distributed for either dataset. In the OI dataset, 5% of all offenders were responsible for 42% of all conviction occasions. In the CCLS dataset, 5% of all offenders were responsible for 32% of all conviction occasions. This may suggest that the OI have more chronic offenders even though the overall number of conviction occasions is lower than the CCLS. In both datasets, the majority of offenders were recidivists. In the OI, 61% of offenders were reconvicted and in the CCLS 84% of offenders were reconvicted. However, the majority of these offenders were only recorded to have one subsequent conviction occasion.

The age of onset was relatively similar for both datasets. However, it appears that more OI offenders start offending at earlier ages, with 13% of offenders receiving their first conviction before the age of 15, suggesting that younger offenders are more likely to be convicted in the England and Wales criminal justice system. It is apparent that both datasets have offenders that are considered ‘late starters’ with an onset age above 25 years. Although these could be genuine offenders with a late onset to offending, it also could be due to these offenders avoiding contact with law enforcement agencies till later on in their criminal careers.

When examining the principal type of offences that offenders have received convictions for, there are many similarities between the two countries. The majority of both datasets have been convicted of ‘Burglary & Theft’ offences – 37% in OI and 49% in CCLS. This is not surprising as this is the largest offence category and contains the most sub-classes of offences (See Appendix). It is also one of the least serious offence types and contains offences such as shoplifting, theft from a person, theft from a machine and burglary of a dwelling.

Making fair comparisons between the two datasets is always going to be challenging and not without obstacles. The differences in law and jurisdictions between the two countries are bound to have an impact on the number and type of convictions recorded. Although it may appear that the CCLS offenders have participated in a higher number of criminal offences, this does not necessarily mean that these offenders are more chronic than the OI offenders. The CCLS dataset has a slightly higher number of male offenders than the OI dataset, and plenty of research provides evidence that males commit more crimes than females (Cauffman, 2008). There are also many other variables not collected such as marital status, ethnicity, social class and urban/rural residency. Many of these variables may be associated with the likelihood to reoffend, and the types of reoffending. However, alignment between the two jurisdictions has been achieved as far as possible.

In the next chapter, models for the examination of changing frequency of convictions over time are introduced. The chapter introduces and focus on three methods of modelling longitudinal trajectories of criminal offending.

5 MODELLING LONGITUDINAL TRAJECTORIES OF CRIMINAL OFFENDING

5.1 Introduction

Modelling longitudinal patterns of offending or developmental trajectories of individual offenders, has been a growing research area in criminology. It typically involves analysis of the data using longitudinal latent variables techniques such as Latent Class Analysis, Latent Markov Models, Structural Equation Modelling, Growth Curve Models and Latent Class Regression. This chapter focuses on longitudinal trajectories of offending, which are concerned with finding latent or hidden groups which represent changing frequencies of offending over time. Many studies have concluded that distinct 'offending groups' can be identified usually ranging from three to five groups (Moffitt, 1993, D'Unger et al., 1998, Nagin and Land, 1993, Francis et al., 2004, Blokland et al., 2005, Bushway et al., 2009).Common groups identified normally consist of trajectories which can be named as 'adolescent limited,' 'low-rate chronic', 'high-rate chronic' and 'late onset'. Such studies have shown that identifying offending trajectory groups, shows the heterogeneity in the overall offending population does exist and requires these latent variable approaches to detect unobserved latent groups within the data.

The aim of using these techniques is to identify the differences between and within-individuals and how they change as offenders age. This chapter will focus on three of the main approaches for examining change over time whilst allowing for within individual changes. The three methods are: Linear Mixed Effects modelling (LME), Group Based Trajectory Modelling (GBTM) and Growth Mixture Modelling (GMM). All three types of model have been used within the field of criminology and take different approaches within the same broad class of underlying models (Francis and Liu, 2015). Each approach is discussed and the decision on the final chosen model is explained. In section 5.2, 5.3 and 5.4 the three model approaches are discussed. These are then applied to the two datasets described in section 5.6 An extension to the model is described in section 5.16 Which, provides a more flexible approach to modelling the trajectory curves.

5.2 Linear Mixed Effects Modelling

Pioneered by Laird and Ware (1982), LME models are a common modelling approach in longitudinal data sets where there is variation between-individuals through a random effects term and within individual dependence among the repeated observations. They are flexible models which are very good at handling data imbalances in longitudinal datasets, where the number of observations per individual is not the same. LME models include random effects, which are multivariate Normally distributed, along with the inclusion of the fixed effects. This therefore allows for analysis to be performed upon the between-individual (random effects) and within-individual (fixed effects) variation in the repeated observations over time. The estimated within-individual changes over time can be referred to as growth curves or latent developmental trajectories, which can vary in their characteristics from person to person. The estimated growth parameters are responsible for explaining the changes in the average responses from the population and can predict the individual trajectories changes over time. The LME model takes a division of the regression parameters which randomly vary from individual to individual into fixed and random effects, which results in a single trajectory for the entire population and individuals vary around this trajectory. This is therefore taking into account the natural heterogeneity from the entire population or sample being examined. The idea is that individuals have their own developmental trajectories with a subject-specific mean response over time, making the subset of regression parameters viewed as random. This mean response from individuals is a mix of attributes which are assumed to be shared by all individuals in the population (fixed effects) and individual-specific features which are exclusive to each individual (random effects). Incorporating random effects allows the covariates to be measured as functions of time among the repeated responses. These random effects can be interpreted as exhibiting the natural heterogeneity that occurs in the population from the factors which are not measured.

Let Y_{ij}be the response for individual i $(i = 1, \dots, N)$ at the jth time occasion $(j = 1, \dots, n_i)$, where n_i denotes the number of responses observed from individual i. Assume also that the Y_{ij} are continuous and Normally distributed. Let p p and q denote the number of fixed effects and random effects parameters respectively. Define $\boldsymbol{X}_{ij} = (X_{ij1}, X_{ij2}, \dots, X_{ijp})$ to be the fixed effects covariates for individual i at time occasion j and $\boldsymbol{Z}_{ij} = (Z_{ij1}, Z_{ij2}, \dots, Z_{ijq})$ to be the random effects covariates. Also define $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$to be a p-vector of unknown regression coefficients for the fixed effects. Assume $e_{ij} \sim Normal(0, \sigma^2)$ and the $\boldsymbol{a}_i = (a_{i0}, a_{i1}, \dots, a_{iq})$ parameters are multivariate Normally distributed with mean zero and variance-covariance matrix Φ:

$$\boldsymbol{a_i} \sim MVN(\boldsymbol{0}, \Phi).$$

Then the general form of the LME can be defined as follows:

$$Y_{ij} = \beta_0.X_{ij0} + \beta_1.X_{ij1} +, \dots, +\beta_p.X_{ijp} + a_{i0}.Z_{ij0} + a_{i1}.Z_{ij1} +, \dots, +a_{iq}.Z_{ijq} + e_{ij}. \tag{5.1}$$

In the LME model these are usually identical – thus fixed effect covariates of age and age-squared also mean that are random effects covariates of age and age-squared.

Measuring longitudinal patterns of criminal careers or offending requires non-linear trajectories so it is necessary to use polynomial growth curves. Therefore, the first column of X_i is a vector of ones and the other columns are the polynomial time transformations of a chosen order p . For example, in a cubic model, the entries of $\boldsymbol{X}_{ij}$ would be 1, t_{ij}, t_{ij}^2 and t_{ij}^3, where t_{ij} is a time measurement such as age for individual i time occasion j.

The typical structure of the model includes the estimation of the intercept and slopes, at both the individual and group level, which are represented via time effect covariates such as age. The random effects addition to the model represents the variance of the intercept and the growth parameters. The LME model makes the assumption that the variation in the responses is accountable to the variation within-individual and to variation between-individuals. The within-individual variation is the deviation between the individual observations Y_{ij} and the linear trajectory. The betas in the fixed effects part of the model are used to define the trajectory pathway $\beta_0 + \beta_1.X_{i1}+,\dots,+\beta_p.X_{ip}$for individual i where X_{ij} is (typically) the jth power of t_i. Each individual has their own intercept and slope and the within-individual variation is reflected in the deviance between the observations and individual trajectories.

For the datasets in this chapter, an extended form of the LME model is used, which allows the response to be a count variable. The aim is to estimate the mean number of the counts of convictions over time for the entire population, as well as obtaining predictions of individual counts of convictions over time.

5.2.1 Extending the model for count data

Standard LME models are limited to using a continuous dependent variable. The model therefore needs to be extended to accommodate for a count dependent variable and to allow a sample of the regression coefficients to vary randomly between individuals. A *generalised linear mixed effect model* (GLMM) can be used for this. The GLMM is an extension of a Generalised Linear Model (GLM) to longitudinal data, by building upon the LME approach (Fitzmaurice et al., 2004). In a GLMM, the assumption is made that any of the responses by individuals' are independent observations from an exponential family of distributions. For example, if the response variable Y_{ij} is a count, then the Poisson distribution is usually a sensible choice. As the response variable from the two conviction datasets is the total number of convictions, the Poisson distribution will be the chosen distribution for the models in this chapter.

5.2.2 Linear Mixed effects model for count data

Let Y_{it} be the observed number of convictions for offender i in time period t. It is assumed that the polynomial used to represent the mean trajectory is cubic in this development; this assumption can easily be changed to other orders of polynomial.

The GLMM model for count data can be written as;

$$Y_{it} \sim Poisson\ (\lambda_{it})$$

With
$$\log(\lambda_{it}) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \alpha_{0i} + \alpha_{1i} t + \alpha_{2i} t^2 + \alpha_{3i} t^3$$

Where
$$\boldsymbol{\alpha_i} = \begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \\ \alpha_{2i} \\ \alpha_{3i} \end{bmatrix} \sim MVN\left(\mathbf{0}, \quad \begin{bmatrix} v_{00} & v_{01} & v_{02} & v_{03} \\ v_{01} & v_{11} & v_{12} & v_{13} \\ v_{02} & v_{12} & v_{22} & v_{23} \\ v_{03} & v_{13} & v_{23} & v_{33} \end{bmatrix} \right)$$

and where MVN represents the multivariate normal distribution.

v_{00}, v_{11}, v_{22}, and v_{33} are unknown variances, and $v_{01}, \dots v_{23}$ are the off diagonal unknown covariance terms.

The above model is a log-linear regression model that includes intercepts and slopes that are allowed to vary randomly. The model assumes a Poisson distribution for counts, which are conditional on the random effects.

5.3 Group-Based Trajectory Models

Group based trajectory modelling (GBTM), also known as latent class growth analysis (LCGA) and, in the context of count data, semi-parametric mixed Poisson modelling (SMPM), is a type or extension of a finite mixture model to allow for repeated observations over a period of time. Developed by Nagin and Land (1993) GBTM is designed to examine patterns that develop over time so it is a highly useful method for modelling the relationship between age and criminal behaviour (Nagin and Piquero, 2010). Unlike LME and GLMM models, where there is no group structure, the main goal of a GBTM is to find different groups of individuals with similar ‘pathways’ that show the long term patterns of development by taking into account the variation of within and between individuals. These ‘pathways’ are referred to as trajectories and observing these trajectories within the two conviction datasets will display the long term patterning of criminal activity over the life course. Typical GLMM or LME growth models have made the assumption that individuals come from a single population and that one mean trajectory is acceptable to be representative and approximate the entire population. Another assumption made by growth models is that individuals are all influenced in exactly the same way by the covariates that affect the growth factors (Jung and Wickrama, 2008). GBTM, on the other hand makes the assumption that the underlying population consists of a finite number of groups, which are unknown, but each group has their own distinct trajectory and estimated mean values which vary for each group over time. Therefore, GBTM differ in their parameter assumptions from LME models by not assuming a multivariate normal distribution of random effects and instead make the assumption that individual specific parameters vary between groups of individuals but are within each group (Bushway et al., 2009). The use of GBTM within the field of criminal career research has become very popular as they can model the relationship between age and criminal behaviour and reveal distinct offending groups that are 'hidden' in aggregated datasets. It also has the flexibility of modelling various types of longitudinal datasets (e.g. cohort, panel, event history data etc) and a range of different data types such as count, binary and continuous.

In GBTM it is assumed that there are several 'hidden' or 'latent' groups within the population that follow different developmental trajectories. Using GBTM will allow any hidden groups of offenders that follow different criminal pathways over time to be discovered. Different shaped trajectories for each of the identified distinct offending groups can be revealed, which would otherwise be hidden if analysis was performed on aggregated data and not at the individual level. Offending typology theories such as those suggested by (Moffitt, 1993) can be tested and checked to see if any of the identified offending trajectories are 'adolescent limited' or 'life-course persistent'

As described in section 5.4 the Growth Mixture model (GMM) extends the GBTM model by allowing for random effects. GBTM is therefore a restricted model within the GMM framework and fixes the variance and covariance estimates for the growth factors or trajectories within each group or class to zero. In other words, if all the variances and covariance's in the random part of a GMM are set to zero, the model would produce the same results as a GBTM. The assumption in GMM's is that all the individual trajectories within each latent class are homogenous. This is the same framework developed upon by Nagin and Land (1993) in GBTM. The groups in these models are latent classes and individuals in the sample all have a probability of latent group membership or trajectory group. The variability between individuals is represented via the differing individual probabilities of latent group membership.

There is no inclusion of random effects. The homogeneity is then assumed to be "within" each distinct latent class.

5.3.1 The structure of the basic Group-Based Trajectory Model

Unlike LME, GBTM challenges the assumption of normality in the random effects. Instead, it is assumed that are discrete groups of developmental trajectories, which can capture the overall variation. The group differences shown by the discovered groups of individuals (all sharing common developmental trajectories), can possibly help explain the individual-level heterogeneity.

In the basic model *N* offenders are observed repeatedly over a number of *T* time periods. For each offender *i,* it is observed that $\mathbf{y}_i = (y_{i1,}\, y_{i2,\dots,} y_{it,\dots,} y_{iT})$ measured at $\mathbf{t}_i = (t_1, t_2, \dots, t_t, \dots, t_T)$

It is assumed in this development that the observed counts of convictions have no missing data and all offenders are observed at identical time periods, resulting in a 'balanced' dataset. However, both of these restrictions are not required and can be lifted.

Assuming there are *K* latent classes or trajectory groups in the data, then;

$$
\begin{aligned}
P(\boldsymbol{y}_i) &= \sum_{k=1}^{K} \pi\,(k) P(\mathbf{y}_i|k) \\
&= \sum_{k} \pi(k) \prod_{t=1}^{T} P(y_{it}|k)
\end{aligned}
$$

The latent class sizes are represented by $\pi\,(k)$, where $\sum_k \pi(k) = 1$.

The likelihood L is then;

$$L = \prod_i P(\mathbf{y}_i) = \prod_i \left(\sum_k \pi\,(k) \prod_{t=1}^{T} P(y_{it}|k) \right)$$

As the data being modelled is count data then a Poisson distribution is usually specified and can be written as;

$$y_{it} \sim Poisson(\lambda_{tk}) \text{ with } P(y_{it}|k) = \frac{\lambda_{tk}^{y_{it}} e^{-\lambda_{tk}}}{Y_{it}!}$$

With a different set of means λ_{tk} for each class k. Finally, the trajectories are assumed to be modelled by a polynomial of order q in t

$$\log(\lambda_{tk}) = \beta_{0k} + \beta_{1k}t + \beta_{2k}t^2 + \cdots + \beta_{qk}t^q$$

Usually, $q = 3$ and cubic polynomials are used to represent each of the k trajectories. For K classes and $q = 3$, the model estimates $4K + (K - 1)$ parameters, made up of 4 β parameters for each trajectory and $(K - 1)$ parameters for the $\pi(k) \quad (as\ \sum_k \pi(k) = 1)$.

5.3.2 Dealing with Overdispersion and Intermittency

There is often the occurrence of overdispersion when modelling count data, particularly conviction data. Overdispersion sometimes is the result of when there is an unusual amount of excess zeros in the data, causing the variance of the dependent count variable (number of convictions) to be greater than what the Poisson model estimates. It can also be caused by clustering, when offences are clustered within conviction occasions. When an offender is brought to court, often they are charged for more than one offence, as investigation leads to the discovery of more offences, and offenders may ask for other offences to be taken into consideration. In a Poisson model distribution, the mean and variance are equal and are not independently adjustable. This would be perfectly acceptable if the rate of events (or convictions) period of time was constant and that these events occurred independently of each other. Instead a Poisson model makes the assumption that events or criminal offences occur independently of each other over time (D'Unger et al., 1998). This assumption is unrealistic when modelling the events of criminal offences, as time plays a very important factor in the occurrences of criminal acts. Past research shows that the more convictions an individual has the more likely they will commit further offences (Wolfgang et al., 1972). Also further research shows that as more time passes the probability of re-offending becomes less likely (Nagin and Paternoster, 1991). Dynamic theories, such as the state dependence explanation,(Sampson and Laub, 1991, Sampson and Laub, 2003, Nagin and Paternoster, 1991) claim that past offending becomes an incentive for future offending, increasing the probability of offending behaviour in the future (Blokland and Nieuwbeerta, 2010). Individuals may commit subsequent crimes or periods of criminal activity occur in spells not independent of each other. Criminal behaviours are subject to change due to external events and certain life circumstances can influence the behaviours of individuals. Sampson and Laub (2003) claim that as an individual begins offending, they are weakening their bonds in society and increasing the probability to commit future offences. However, the longer the time that elapses from committing the last offence, the likelihood to reoffend decreases (Nagin and Paternoster, 1991).

A more improved model to account for this overdispersion can be used to obtain a better model fit for the data. The negative binomial distribution can be used as this contains an extra error term which comes from a gamma distribution. The conditional distribution of conviction counts is then dependent on the distribution of this extra error term. The negative binomial distribution occurs when this extra error term is the result of the logarithm of a random variable drawn from a gamma distribution. Using the negative binomial model instead of the standard Poisson model increases the flexibility of the model as the variance is no longer constrained to be equal to the mean, but can be larger, allowing greater spread in the distribution This then still allows the model to still work under the assumption of the Poisson model for rare events as well as using the gamma distribution to allow for individual rate of offending to be gamma distributed throughout the population instead of being constrained to the same mean rate of offending (D'Unger et al., 1998).

The GBTM model can be extended to use a negative binomial distribution:

$$y_{it} \sim NegBin(\lambda_{tk,}\theta_k) \text{ with } P(y_{it}|k) = \frac{\Gamma(y_{it}+\theta_k^2)}{y_{it}!\,\Gamma(\theta_k^2)}\left(\frac{\theta_k}{\theta_k+\lambda_{kt}}\right)\theta_k\left(\frac{\lambda_{kt}}{\theta_k+\lambda_{kt}}\right)y_{it} \tag{5.2}$$

In equation(5.2 $\lambda_{tk,}$represents the mean of the kth trajectory at time period t. The scale parameter for the kth trajectory is represented by θ_k.

To model overdispersion, it is often convenient to reparametrise so that $\tau_k = \frac{1}{\theta_k}$ represents the overdispersion of the kth trajectory, and the variance of $y_{it}|k$ equal to $\lambda_{tk,}(1+\tau_k)$. Therefore if $\tau_k$is then zero, there is a Poisson variability for that trajectory.

Polynomial smoothing is applied to the means of each K trajectories via a log-linear model. As before the value of the time axis in time period t be x_t. For cubic smoothing in the kth trajectory:

$$\ln(\lambda_{tk}) = \beta_{0k} + \beta_{1k}t + \beta_{2k}t^2 + \beta_{3k}t^3 \tag{5.3}$$

For K latent trajectories with cubic smoothing, $(K-1)$ of the $\pi(k)$terms need to be estimated (as the $\pi(k)$ sum to one) along with $4K$ β parameters for a Poisson model or $4K$ β parameters and three θ parameters for a negative binomial model.

5.4 Growth Mixture Models

The third type of model considered is the Growth Mixture Model (GMM) (for a detailed overview see Muthén and Sheden,1999). These models can be thought of as a combined method of both LME and GBTM, by estimating groups but allowing the parameters for each group to vary. The GMM approach is used to identify several latent class groups, whilst describing the within changes of each latent class group. The model framework allows for *post-hoc* classification and explanation of the differences within the latent class groups (Ram and Grimm, 2009). GMM differs from the two previously discussed modelling approaches, however these models can be viewed both as extensions of LME models, as they are able to handle heterogeneous populations, and also as an extended model of the GBTM, allowing for within latent class variation for each individual trajectory (Francis and Liu, 2015). The assumption that all individuals come from a single population with the same parameters is relaxed in GMM's and the growth parameters are allowed to vary across the latent class groups. One of the main differences between GMM and GBTM is the within latent class variability. In GBTM, the variance and covariate estimates for the latent trajectories (growth parameters) within each latent class are fixed to zero. All individual trajectories in each latent class are assumed to be homogenous (Nagin 1999). GMM provides a much more flexible modelling framework, allowing for variability within-individual trajectories in each of the latent class groups.

5.4.1 Structure of the Growth Mixture Model

Using the structure of the GBTM model as described in the last section, random effects are added to the model. Thus, assuming there are K latent classes or trajectory groups in the data, then; as before

$$P(\boldsymbol{y}_i) = \sum_{k=1}^{K} \pi\,(k) P(\mathbf{y}_i|k)$$

$$= \sum_k \pi(k) \prod_{t=1}^{T} P(y_{it}|k)$$

Assuming that the y_{it} are counts of offences collected over time, as before, the model distribution can be defined as;

$$y_{it|k} \sim Poisson(\lambda_{tk}) \quad \text{or} \quad y_{it|k} \sim Negative\ Binomial(\lambda_{tk,}\theta_k)$$

The Poisson form should be used when there is no overdispersion, and the negative binomial form should be used when overdispersion is present.

The model, however has random effects terms added to it. For the example of K groups and cubic trajectories, the model becomes:

$$\log(\lambda_{it|k}) = \beta_{0k} + \beta_{1k}t + \beta_{2k}t^2 + \beta_{3k}t^3 + \alpha_{0ki} + \alpha_{1ki}t + \alpha_{2ki}t^2 + \alpha_{3ki}t^3$$

Where $$\boldsymbol{\alpha_{ki}} = \begin{bmatrix} \alpha_{0ki} \\ \alpha_{1ki} \\ \alpha_{2ki} \\ \alpha_{3ki} \end{bmatrix} \sim MVN\left(\mathbf{0}, \quad \begin{bmatrix} v_{00k} & v_{01k} & v_{02k} & v_{03k} \\ v_{01k} & v_{11k} & v_{12k} & v_{13k} \\ v_{02k} & v_{12k} & v_{22k} & v_{23k} \\ v_{03k} & v_{13k} & v_{23k} & v_{33k} \end{bmatrix} \right)$$

With a separate variance-covariance matrix for each class. The model can be simplified by constraining the K variance-covariance matrices to be equal.

GMM can be fitted using MPlus software using the MIXTURE and TWOLEVEL MIXTURE commands. It can also be fitted in R using the `lcmm` package, although the form of the outcome variable is limited to continuous and ordinal, and does not include count data.

In conclusion, the three model types can be summarised inTable 5.1 below. The next section describes the approach taken for the modelling in this thesis, and explains the choice made between the three models.

Table 5.1 Summary of the three models

Model	Form of the trajectory	Random Effects present?
Linear Mixed Effects	Single Fixed trajectory	Random Effects
Group Based Trajectory	Multiple Trajectories	No Random Effects
Growth Mixture Modelling	Multiple Trajectories	Random Effects

5.5 Choice of model strategy

The LME and GBTM models have been chosen to be fitted to the two datasets, and the results of this fitting will be discussed in the following sections. This choice has been made for the following reasons. Firstly, LME and GBTM are models that aim to estimate the continuous distribution of individual trajectories. Secondly, they are a good choice for modelling the longitudinal offending patterns of the two conviction datasets and showing the changes over time in distinct group trajectories. Finally, there are various software packages available such as R, SAS, Latent Gold and MPLUS and are relatively straightforward to fit.

Earlier in this chapter, the GMM modelling framework and how it can be fitted using M-Plus software has been previously discussed. However, using a GMM model is not necessary or appropriate for the following statistical analysis due to a number of reasons. The estimations tend to be quicker and model convergence is more likely in GBTM over GMM. Nagin (2005) argues that adding the random effects to the group-based approach to relax the assumption of homogeneity within the groups is just adding unnecessary complexity. GMM models are more susceptible to model specification errors due to the complexity of having to estimate more parameters to account for the heterogeneity within and between latent trajectory groups (Bauer and Curran, 2003). GMM are not suitable because they are limited practically due to software limitations. Although GMM can be fitted using M-Plus software, these models are often unstable and struggle to reach convergence. Moreover, other implementations cannot deal with count data (Proust-Lima et al., 2017). In summary, the GMM models tend to be so unstable which makes them less practical over the GBTM models. Models such as GBTM are relatively more stable and using a simpler model is perfectly adequate to measure longitudinal patterns of offending over time for the datasets used in this thesis.

5.6 Preparation of Datasets

Before performing any analyses, the sample data sets needed to undergo some restructuring and new variables created to make it possible to examine any longitudinal patterns of criminal behaviour. The correct format is what is known as long form data (Wickham, 2014), where data on each observational unit (offender) extends over many rows, with each row within an observation representing a different age group (two years in the analysis here). Both data sets in their standard form were in the wrong format and needed to be reshaped.

For the CCLS, dataset there is a new row of data for every age (starting at age 12 up to their age in 2002 or death). There can be multiple convictions on one row if an individual has received a conviction for more than one offence at the same age[11].

The OI dataset on the other hand contains a separate row for each conviction. Individuals can have several rows of data all at the same conviction date; each row contains new information about every separate conviction received at that court appearance. It does not contain a row of data for each age like the CCLS dataset. For the OI dataset, a binary indicator was constructed for each age period stating 1 if the conviction occurred in each of the 2 year age periods or 0 if not

The two datasets were then aggregated using the unique ID variable as the break variable and aggregating the new indicator age variables, taking the sum of each one within each case. The sex of the offender variable was also added when aggregating

---

[11] To ensure that the correct number of convictions at each age were computed correctly, the indicator age variable was multiplied by the total number of convictions at each age only if the indicator variable equalled 1 (stating that they had an offence within the 2 year age period).

the datasets. The aggregated datasets were then changed from the short fat format to a long thin format by using the restructuring option in SPSS. In the long thin format, the datasets contained 17 rows[12] (one for each of the 2 year age periods) for each ID/Offender, and a variable "convictions" giving the number of convictions. Five new variables age, age^2, age^3, age^4 and age^5 were added which represent the age of the offender, quadratic age, cubic age, quartic age and quintic age. These are used as predictor variables when modelling the number of convictions as a function of age. Using the quadratic, cubic, quartic and quantic of age allows a more accurate measure of the trajectory shape.

5.7 Application of LME and GBTM approaches to the OI and CCLS datasets

In the above section the three different modeling strategies for analysis of longitudinal data were discussed. The next section will show how both LME and GBTM models can be applied to the two conviction datasets and how they are fitted by the chosen statistical software packages. In both cases, models are fitted using cubic polynomials in age. The cubic assumption is common in these models, although other orders of polynomials can be fitted. This assumption is relaxed in section 5.15.

5.8 Fitting the Generalised Linear Mixed Effect Model to count data

The variables included in the GLMM are the count of 'convictions' or more strictly – conviction occasions as the dependent, with explanatory variables; 'age', 'age^2' , and 'age^3' which represent the linear, quadratic and cubic polynomials of the offenders age. The age of offenders starts at age 12 and goes up to age 35 in 2 year intervals (totalling 13 different age periods). To fit the GLMM model to the two datasets, the lme4 package (Bates et al., 2015) in R with the glmer function was used. In R the

[12] In the models that follow in this chapter only the first 13 age periods, from age 12 up to age 35, are included.

GLMM model is specified with the following code:

```
fitOI=glmer(convictions~poly(age,3) + (poly(age,3)|OINumber),
family=poisson, data=OI)
```

Here, the age variables have been fitted as orthogonal polynomials in the model specification, which improves the convergence and numerical stability. The columns in the design matrix for the orthogonal polynomial are independent and scaled. The term in the model formula;

```
(poly(age,3)|OINumber)
```

need further explanation. The specifies that the random effets structures in the model are determined by the polynomial curve (poly(age,3) - each respondent (OINumber) has its own random cubic curve, and the parameters of the cubic curve come from a multivariate normal distribution.

The fixed and random parameter estimates of the model are given below in Table 5.2 and Table 5.3 for each dataset.

Table 5.2 GLMM model output for OI Data

OI dataset *N = 4420*	AIC	BIC	Log Likelihood	Deviance	Residual Degrees of Freedom
	75270.6	75394.4	-37621.3	75242.6	51190
Random Effects	**Variance ν**		**Standard deviation (square root of variance)**	**Correlation**	
Intercept	1.6		1.3		
age	13328.9		115.5	0.68	
age^2	18832.9		137.2	-0.10 -0.80	
age^3	27721.2		166.5	-0.20 -0.42 0.41	
Fixed Effects	**Estimate β**	**Std. Error**	**z value**	**P-value**	
Intercept	-2.37	0.02	-106	<0.0001	
age	-134.7	1.5	-91.4	<0.0001	
age^2	-81.85	1.72	-47.5	<0.0001	
age^3	116.03	0.97	119.8	<0.0001	

Table 5.3 GLMM Model Output for CCLS data

CCLS dataset *N = 4420*	AIC	BIC	Log Likelihood	Deviance	Residual Degrees of Freedom
	147329.2	147457.9	-73650.6	147301.2	72508
Random Effects	**Variance ν**		**Standard deviation (square root of variance)**	**Correlation**	
Intercept	1.292		1.137		
age	30515.8		174.7	0.21	
age^2	11046.7		105.1	-0.17 -0.20	
age^3	10916.4		104.5	-0.19 -0.55 0.05	
Fixed Effects	**Estimate β**	**Std. Error**	**z value**	**P-value**	
Intercept	-1.35	0.02	-68.7	<0.0001	
age	-0.94	2.07	-0.45	0.65	
age^2	-124.7	2.1	-59.4	<0.0001	
age^3	92.3	1.51	60.92	<0.0001	

5.9 Results from GLMM

From examining the GLMM model outputs in Table 5.2 and Table 5.3 , the p-values for fixed effects in the OI dataset, it can be seen that all four model terms are highly significant showing that a cubic term for age is needed in the model. The p-values for the CCLS fixed effects show that the intercept and quadratic and cubic age terms are highly significant. There is some evidence of instability in the CCLS model fit, as the random effects variances are very high. This is a concern, and changing the parameters of the fit (the algorithmic method and convergence criterion) did not improve this. However, the fitted mean trajectory and examples of specific trajectories are stable, which suggests that perhaps the variances are simply representing substantial variability between offenders. The standard output from the software reports the correlation matrix of the random effects terms.

However, the function VarCorr can be used to construct the variance/covariance matrix. The correlation matrices for both datasets do not indicate multi-collinearity, with no correlation greater than 0.80 in magnitude.

Table 5.4 Variance/Covariance Matrix for GLMM OI Data

OI	**Intercept**	**age**	**age^2**	**age^3**
Intercept	1.603	98.66	-18.202	-39.98
Age	98.66	13328.9	-12745.65	-8091.7
age^2	-18.202	-12745.65	18832	9476.05
age^3	-39.98	-8091.7	9476.05	27721

Table 5.5 Variance/Covariance Matrix for GLMM CCLS Data

CCLS	**Intercept**	**age**	**age^2**	**age^3**
Intercept	1.292	42.41	-20.02	-22.09
Age	42.41	30515.75	-3611.75	-10093.02
age^2	-20.02	-3611.75	11046.7	586.077
age^3	-22.09	-10093.02	586.077	10916.4

The population mean predicted rate of convictions can be plotted against age for both datasets. This is obtained in R through the `predict()` function and by setting all random effects terms to zero, and is thus the trajectory for an average offender.

Figure 5.1 Average rate of convictions over time for OI and CCLS

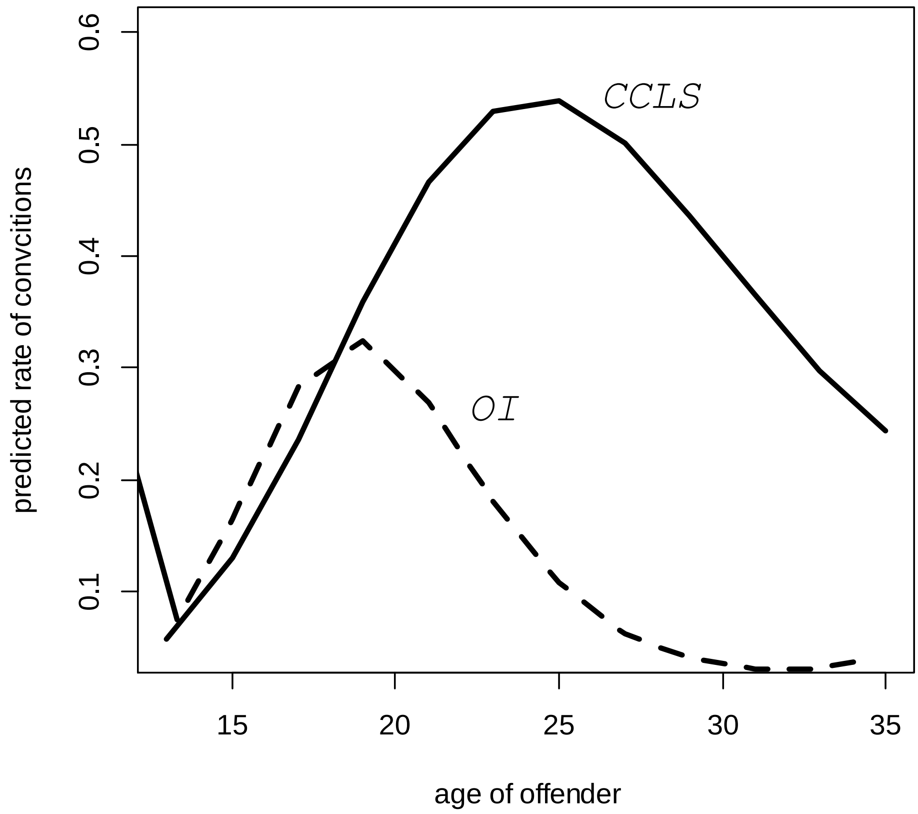
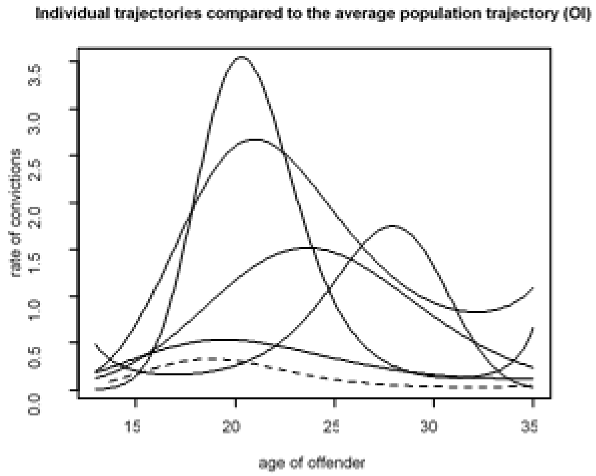
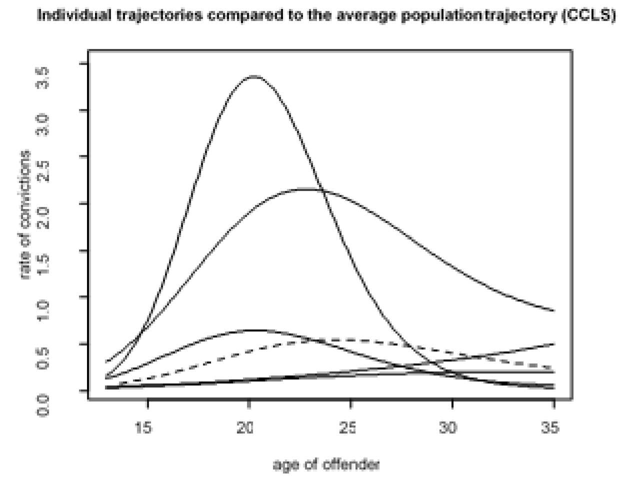


Figure 5.1 , shows that the OI offenders peak at an earlier age but have a lower rate of convictions in comparison to the CCLS offenders. The peak of convictions is at a later age for the CCLS offenders (mid 20s) which is unexpected as the majority of research suggests that the peak age of offending is late adolescence (Farrington, 1986). This may be because a large number of offenders are diverted away from the courts at young ages in the Netherlands. Towards the end of the curve for the OI trajectory, there is a slight uptick which may suggest late onset offending. However, there is also a known problem that occurs when fitting polynomial curves over age, that could explain this (see section 5.14 below for detailed discussion on the issues with polynomial curves).

From the fitted GLMM models it is also possible to plot individual trajectories against the population average. Again, using the `predict()` function in R, the predicted trajectory for specific individuals was plotted against age. The individual responses were firstly placed into a new data frame from which the predict function uses to look for variables with which to predict. The default predictions are the log-odds and specifying `type = ″response″` gives the predicted probabilities. Once the predicted probabilities are extracted, they can then be plotted in R using the `plot()` function.

Figure 5.2 Selected individual trajectories (solid lines) plotted against population average trajectory (dashed line) for both OI and CCLS datasets.





The OI individual trajectories show quite a bit of variability around the average population trajectory. There is an example of one individual that follows a very similar path to the average for the population but at a slight higher rate. Two other individuals exhibit high rates of conviction but are adolescent peaked. Another individual is a mid-rate offender showing a peak at around 25 years. Finally, there is an individual who has a late onset age and peaks in their late 20's. A few of the trajectories show quite a lot of changes in direction displaying upticks towards the end. This could be a true representation of the individual's offending behaviour and they are actually increasing their rate of convictions at a later age. However, it could also be one of the issues that can occur when fitting polynomials (again see section

5.14).

Again the individual trajectories for the CCLS show some substantial deviations from the population average. One individual is a very high rate adolescent limited offender and another individual follows the same shape as the population average but at a higher rate. There is also another individual with adolescent peak but at a rate similar to the population average. A couple of individuals display late onset offending. Unlike the OI trajectories, there are no upticks towards the end, therefore it would seem the fitting of polynomial curves in this model has been more suitable for this dataset.

5.10 Conclusion of LME & GLMM approach

An advantage of using mixed effects models, which have assumed normal distributions for the random effects, is that the individual and group level information is incorporated all under one model. The model parameters assume a continuous distribution for the population based on a multivariate normal distribution (Nagin 2005). However, as Verbeke and Lesaffre (1996) point out, this assumption of a multivariate normal distribution in the random effects can extremely influence the parameter estimates. There by extending the LME to a GLMM deals with the issue of modelling count data as they can also be applied to non normally distributed outcomes such as counts with a Poisson distribution (Gibbons et al., 2010).This then deals with the correlation amongst the repeated responses at the group level.

Although LME and GLMM are suitable for modelling longitudinal patterns of both continuous and categorical outcomes, there are a number of drawbacks of using this modelling technique. The fitting of GLMM in statistical software has computational challenges, as stronger parametric assumptions are required (Bushway et al., 2009). The assumptions made by LME and GLMM that all individuals follow the same pattern within the population can be violated as the differences that can occur in the individual trajectories may not be able to be described by a single explanation

(Nagin, 2005). Unfortunately if all individuals do not follow the same growth pattern, then LME or GLMM may not have the flexibility to model the variation between individual trajectories. This means that any individuals who have very different or distinct patterns of offending may not be captured in the overall estimated population trajectory (Bushway et al., 2009). Both the LME and GBTM aim to model individual heterogeneity in the developmental trajectories but differ in their assumptions about the population trajectories distributions. GBTM provides a distribution free alternative model and allows for non parametric assumptions about the distribution of the unobserved heterogeneity within the population. In contrast to LME, the GBTM make the assumption that the continuous distribution can be approximated by a discrete number of fixed points (Bushway et al., 2009). The model parameters are unrestricted and estimated non-parametrically using maximum likelihood. The idea of GBTM is to then identify distinct trajectories within the population, estimate the shape of the trajectories, as well as to examine the assignment of individuals to trajectories, and the effect of covariates on trajectory membership (Francis et al., 2016).

5.11 Fitting the GBTM to Count Data

Group-based trajectory models can be fitted in various software, such as SAS using the PROC TRAJ procedure (Jones et al., 2001), in MPLUS where it is referred to as Latent Class Growth Analysis (LCGA) using the TWOLEVEL MIXTURE command, in R using the lcmm package (Proust-Lima et al., 2017), and finally via Latent Gold software (Vermunt and Madigson, 2005) using the latent regression option. For this study, Latent Gold software has been chosen to fit the GBTM models. Latent Gold is a good choice for fitting GBTM because it is very flexible as it is able to fit dependent variables of varying data types such as ordinal, continuous, counts, binary etc. It also allows the inclusion of not just predictor variables (modelling changing class profiles over time) but also covariates (modelling changing class sizes) to be specified in the model. Finally, it is fast and well-written, and does not suffer from convergence problems. The population does not need to be homogenous, which is a typical assumption of other regression programs.

The GBTM are fitted using the Expectation Maximisation (EM) algorithm and once approximate convergence has been achieved, a switch is made to the Newton-Raphson (NR) algorithm. This ensures that the model gains advantages from both algorithms by using the stability of the EM and the rapid speed of the NR once the EM is close to a final solution. The higher the EM tolerance the faster the switch to NR. By default the maximum number of EM iterations is set at 250 and the maximum number of NR iterations is set at 50[13]. The number of random sets (starting values) are defaulted to 10 in the software but I have set them at 100 for Poisson models and 50 for the negative binomial models. Increasing the sets of random starting values for the model parameters reduces the chances of model convergence to a local solution. The Bayes Constants are all set at 1 by default in the software but they are all changed to be zero. These constraints are provided in the software to ensure that parameter estimates of the profile probabilities do not hit zero, essentially adding an extra proportion of an observation $(1/K)$ to each class with "conservative" or non-zero values. Setting the Bayes constants to zero ensures that the solution is full maximum likelihood.

Models are firstly fitted by estimating a one class Poisson regression model with cubic polynomial over age. Then more models are estimated with increasing numbers of classes, up to 8 classes in total. Estimating a one-class regression model to start is useful as it provides a good starting base in which to compare $k > 1$-class models. Also, if there are any problems that arise these can be dealt with before proceeding to extend the model any further.

[13] These parameters values were used when fitting the models in Latent Gold. If convergence was not achieved then the EM and NR iterations was be increased until convergence was achieved. See Vermunt and Magidson (2005) user guide for more details.

Technically, in the Latent Gold software, the ID variable is the case ID, the number of convictions is the dependent variable which is declared to be a count variable. To specify a cubic polynomial for each trajectory, the first three age variables are declared to be “predictors” and set to be numeric. The number of classes is then set to be from 1 – 8. No restrictions are imposed for these models and instead are added post-hoc to estimate new models after viewing the results.

5.12 Determining the Optimal Number of Trajectories

Choosing the best fitting model requires consideration of a number of factors Jung and Wickrama (2008) suggest that not only should the model fit indices be taken into account when deciding the number of classes, but also the “research question, parsimony, theoretical justification and interpretability” should also be part of decision in choosing the best fitting model to the data.

The aim is to carefully select the model with an appropriate number of classes that most closely describes the data (Yessine and Bonta, 2008). In GBTM, using the standard log likelihood ratio test to compare models is not suitable. This is because a k group model is nested within a $k + 1$ group model. To obtain the k group model from the k + 1 group model involves setting some parameters to zero, and these are then on the boundary of the parameter space. This violates the asymptotic assumptions of the Likelihood ratio test. Instead, there are a number of model statistics available to help in deciding the optimal number of classes, such as the BIC (Raftery, 1995, Schwarz, 1978). Various authors have suggested methods for determining the number of groups or classes in the data. Nagin (2005), in his book, recommends the use of BIC based on his extensive simulation work. Other authors have suggested alternative methods. For example, Nylund et al. (2007) examined the performance of AIC, CAIC, BIC and adjusted BIC, as well as the bootstrap test for a wide range of latent class and growth mixture models, and concluded that the bootstrap test often works well. Aitkin et al. (2015) has recently suggested a Bayesian method based on draws from the posterior deviance which is found to perform very well on small and medium samples.

From experience, the BIC has worked well for smaller sample sizes, however it does not always work well on large samples such as the two conviction datasets being used in this study. Nagin (2005) found that the BIC can sometimes be unsuitable as it will not always neatly identify an optimal number of classes. In fact the BIC can continue to increase as more groups are added. McLachlan and Peel (2000) suggested that using the ICL-BIC (Integrated Completed Likelihood Bayesian Information Criterion) as a good alternative information criterion when using general mixture models. The ICL-BIC is the BIC plus twice the entropy of the model:

$$ICL - BIC = -2\log L + \log(n)\, q + 2\sum_{ik} \hat{p}_{ik} log \hat{p}_{ik}$$

where q represents of the number of parameters in the model, and n represents the number of individuals or offenders. Nagin found that using this criterion was too cautious and found fewer groups that were actually present in the data. Although the ICL-BIC has not been examined extensively there have been a few researchers that have found that it performs well (McLachlan and Peel, 2000, Morgan and Beaujean, 2014). Morgan and Beaujean (2014) support the use of the ICL-BIC and found that it performed well in their study. Both the BIC and ICL-BIC will be examined to help with deciding upon the optimal number of groups.

5.13 Results from GBTM

Following the above methods described above, a number of Poisson and then negative binomial latent trajectory models were fitted to the Offenders Index and CCLS data shows the BIC and ICL-BIC values from these models.

There are several features to notice from examining Table 5.6 and Table 5.7. Firstly, in the OI Poissonmodels, the BIC and AIC continues to decrease from one to twelve classes and no minimum value of either is obtained. This is a common characteristic of trajectory models when there are a large number of cases. A third column showing the ICL-BIC values, show a similar pattern of continuing to decrease from one totwelve classes.

In the negative binomial models, the BIC does minimise at 10-classes for the cubic, 5-classes at the quartic and at 9-classes at the quintic models. The AIC also minimises at 10-classes for the quartic model. The ICL-BIC does mimimise in all the OI negative binomial models at 2 classes, which suggests that accounting for the overdispersion allows a better solution to be obtained. The best fitting model based on the ICL-BIC is the quartic negative binomial model with 2 classes.

For the CCLS data, again the Poisson models, the BIC and AIC values continue to decrease with the exception of the quintic model – where the BIC minimises at 9-classes. The ICL-BIC also minimises for Poisson quintic model at 8-classes. In the negative binomial models the BIC and AIC both minimise at 11-classes for the cubic model and at 8-classes for the quintic model. The ICL-BIC minimises at 4-classes for the cubic, quartic and quantic polynomial negative binomial models. The lowest value occurs for the quintic 4-class model at 106046.4. However, this is only slightly lower than the quartic model ICL- BIC value of 106050.4. This suggests that the most optimal model is one with 4-classes and either quartic or quintic polynomial model should be used.

From examining the CCLS results there are similar features to the OI results, however the main difference is that the ICL-BIC actually minimises at 4 classes for the negative binomial models not 2 classes. The best fitting model based on the ICL-

BIC is the quintic negative binomial model with 4 classes.

There is not one right way to decide upon the optimal number of classes and choosing the best model is complex, as previously discussed above. Using previous experience and taking into account previous research, two classes may not be suitable and may hide certain offending groups. Typically, four class solution has often been found in previous studies (Nagin and Land, 1993, Nagin et al., 1995, D'Unger et al., 1998, Blokland et al., 2005, Blokland and Nieuwbeerta, 2005). Utilising the ICL-BIC values and recalling upon past studies, a 4-class negative binomial model using a quartic polynomial is chosen, as using increasing number of polynomials can have several issues. This is a compromise between the two datasets to make them more comparable.

Table 5.6 Information criterion from GBTM for OI data.

Offenders Index N = 4422		**Poisson**			**Negative Binomial**		
		BIC	**AIC**	**ICL-BIC**	**BIC**	**AIC**	**ICL-BIC**
Cubic	**1 Class**	109307.7	109282.3	109307.7	72634.49	72602.7	72634.49
	2 Class	85628.8	85571.57	85746.86	68083.52	68013.57	***68545.15***
	3 Class	81644.51	81555.48	81966.18	67544.34	67436.24	71152.25
	4 Class	79804.21	79683.39	81815.79	66970.17	66823.92	70639.13
	5 Class	78323.49	78170.88	80498.44	66810.09	66625.69	70829.17
	6 Class	77153.28	76968.88	79663.57	66728.97	66506.42	71092.55
	7 Class	76294.25	76078.05	78956.38	66732.18	66471.48	73098.76
	8 Class	75660.2	75412.21	78426.07	66610.66	66311.8	73215.56
	9 Class	75420.15	75140.37	78323.57	66553.08	66222.43	73785.86
	10 Class	74818.69	74507.12	77845.05	***66550.59***	66181.79	75214.59
	11 Class	74650.27	74306.90	77755.98	66554.73	66154.13	74776.62
	12 Class	73980.03	73604.87	77788.68	66571.40	66126.30	75359.52
Quartic	**1 Class**	109068.3	109036.5	109068.3	72583.45	72545.29	72583.45
	2 Class	85398.98	85329.04	85514.44	68015.14	67932.48	***68477.05***
	3 Class	81427.28	81319.19	81742.83	67536.83	67409.65	71197.96
	4 Class	79787.51	79641.26	81815.23	66981.14	66809.46	70683.83
	5 Class	78302.59	78118.19	80437.18	66830.13	66613.93	70860.41
	6 Class	77139.41	76916.86	79575.53	66749.99	66489.28	71129.55
	7 Class	76394.81	76134.11	78874.71	66628.64	66323.42	72861.19
	8 Class	75608.67	75309.81	78320.49	66627.22	66277.49	73259.55
	9 Class	75057.21	74720.20	77915.42	75420.15	75140.37	78323.57
	10 Class	74700.10	74324.94	77660.66	74818.69	74507.12	77845.05
	11 Class	74198.30	73784.98	77789.51	74650.27	74306.90	77755.98
	12 Class	73825.04	73373.57	77291.26	73980.03	73604.87	77788.68
Quintic	**1 Class**	109072.2	109034.1	109072.2	72589.89	72545.38	72589.89
	2 Class	85412.54	85329.87	85528.17	68031.08	67935.7	***68492.88***
	3 Class	81447.36	81320.19	81763.73	67553.82	67407.57	71262.13
	4 Class	79799.87	79628.18	81811.1	67003.05	66805.94	70719.97
	5 Class	78329.35	78113.15	80463.82	66859.55	66611.56	70902.66
	6 Class	77148.11	76887.41	79578.69	66786.04	66487.18	71157.81
	7 Class	76219.6	75914.39	78771.53	66689.27	66339.54	73154.71
	8 Class	75549.67	75199.95	78198.6	66728.78	66328.18	73262.89
	9 Class	74997.72	74603.48	77762.67	***66684.42***	66232.96	74924.41
	10 Class	74633.27	74194.53	77556.53	66701.65	***66199.31***	74924.21
	11 Class	74245.46	73762.20	77784.03	66763.87	66210.67	75007.27
	12 Class	74091.87	73564.10	77722.89	66825.12	66221.05	75939.63

Figures in bold italic are the lowest ICL-BIC value in the column.

Table 5.7 Information criterion from GBTM for CCLS data.

CCLS N = 4422		Poisson			Negative Binomial		
		BIC	AIC	ICL-BIC	BIC	AIC	ICL-BIC
Cubic	**1 Class**	153556	153530.5	153556	114706	114674.2	114706
	2 Class	120784.8	120727.6	121039.8	106205.5	106135.6	106929.2
	3 Class	114694.2	114605.2	115422.3	104775.3	104667.2	106409.8
	4 Class	111892.6	111771.8	112895.4	104229.1	104082.8	***106341.3***
	5 Class	111069.9	110917.3	112716.7	103707.2	103522.8	107350.8
	6 Class	109875.7	109691.3	111910.7	103491	103268.4	108044
	7 Class	109569.1	109352.9	112560.9	103257.6	102996.9	108487.2
	8 Class	108531.7	108283.7	112194.4	103248.9	102950	108121.6
	9 Class	108182.2	107902.4	111825.2	103144.3	102800.9	108368.7
	10 Class	107921.5	107609.9	111588.2	103257.6	102876.1	108978.7
	11 Class	107632.8	107289.5	111428.3	***103041***	***102621.4***	109051.2
	12 Class	107239.4	106864.2	111115	103097.4	102639.6	109021.8
Quartic	**1 Class**	153264.3	153232.5	153264.3	114624.2	114586	114624.2
	2 Class	120492.7	120422.8	120747.5	106091.4	106008.7	106810.4
	3 Class	114411.2	114303.1	115135.2	104656.2	104529	106274
	4 Class	111778.5	111632.2	112785.4	103875.6	103703.9	***106050.4***
	5 Class	110447.2	110262.8	111958.4	103687.1	103470.9	107415.4
	6 Class	109330	109107.4	111262.2	103492.9	103232.2	106877.4
	7 Class	108933.8	108673.1	112196.2	103386.2	103081	108337.8
	8 Class	108235	107936.2	111025.7	103197.9	102848.2	108641.7
	9 Class	107986.2	107649.2	111854.5	103045.5	102651.3	108420
	10 Class	107657.1	107281.9	111671	102966.8	102528	108928.8
	11 Class	107365.9	106952.6	111711.2	102952	102468.7	108930.6
	12 Class	106926.2	106474.8	111318.7	102945.6	102417.8	108894.8
Quintic	**1 Class**	153272.5	153234.4	153248.2	114631.7	114587.2	114619.1
	2 Class	120508.8	120426.1	120731.8	106106.3	106010.9	106804.7
	3 Class	114435.4	114308.2	115121.5	104679.3	104533	106266.9
	4 Class	111801.6	111629.9	112778	103904.8	103707.7	***106046.4***
	5 Class	110455.6	110239.4	111957	103717.1	103469.1	106481.8
	6 Class	109322	109061.3	111255.3	103393.4	103094.5	106816.3
	7 Class	108671.3	108366	110869.2	103252.2	102902.5	106997.9
	8 Class	108287.1	107937.3	***110698.6***	***103164.4***	***102763.8***	108575.5
	9 Class	***103111.4***	102660	108538.5	107863.2	107468.9	110585.8
	10 Class	103162.6	102660.3	109087.6	107578.5	107139.8	110428.9
	11 Class	103231.4	102678.2	108953	107199.1	106715.8	111023
	12 Class	103170.4	102566.4	109979.8	107239	106711.2	111577.2

Figures in bold italic are the lowest ICL-BIC value in the column.

Figure 5.3 OI plot of estimated trajectories for the quartic polynomial negative binomial 4-class model

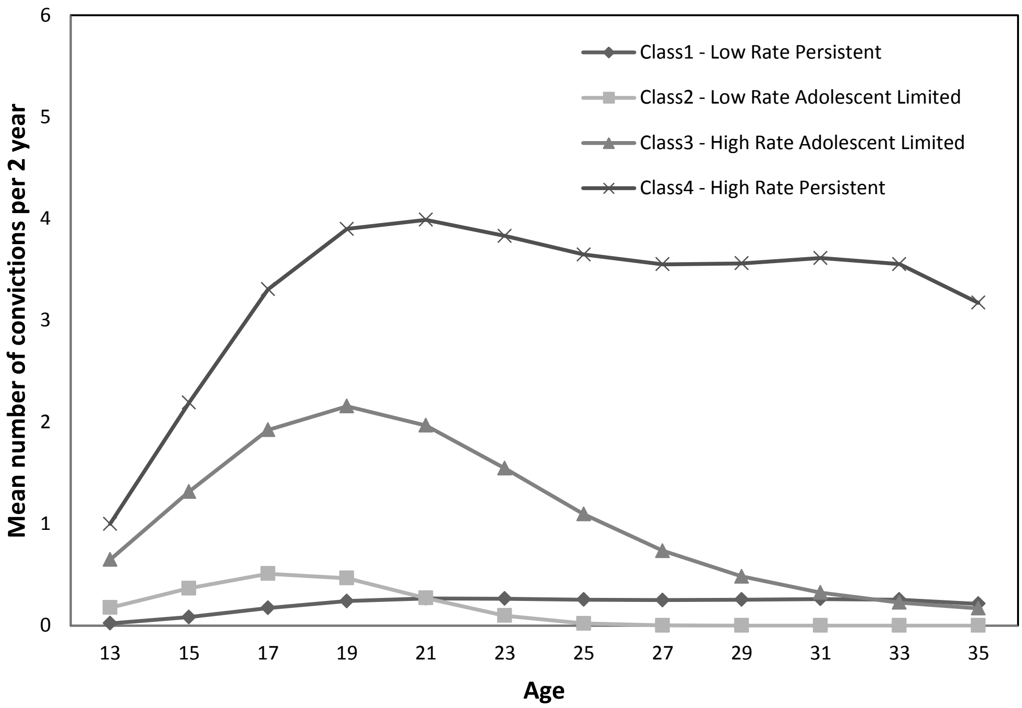
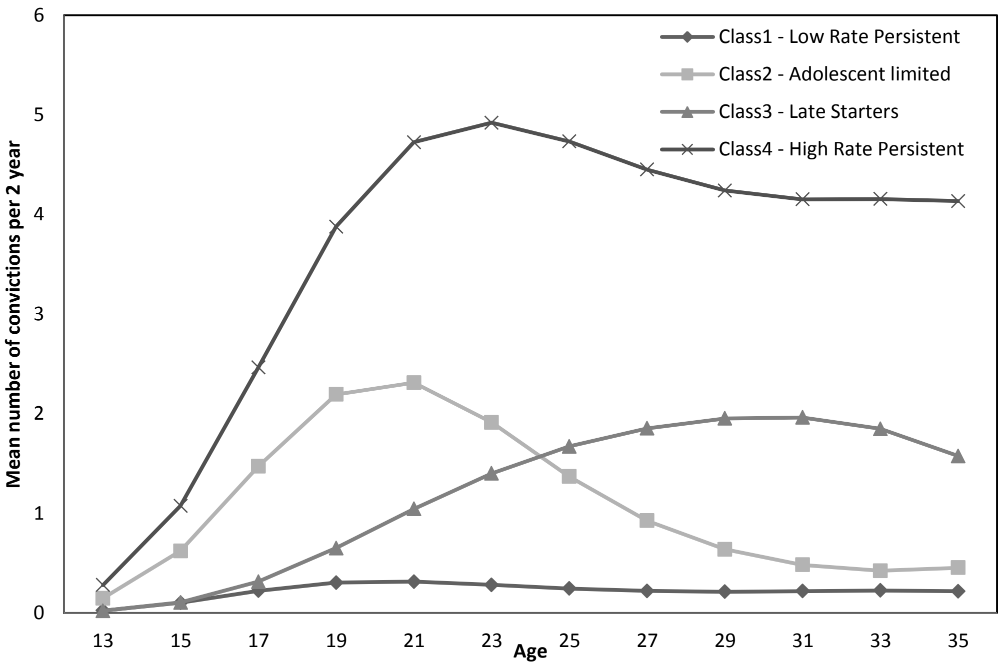


Figure 5.4 CCLS plot of estimated trajectories for the quartic polynomial negative binomial 4-class model



From observing Figure 5.2 and Figure 5.3 the plotted estimated trajectory groups can be labelled as in Table 5.8.

Table 5.8 Offending trajectory groups and class sizes

Trajectory Group	**OI**	**Sum of expected # of convictions**
Class 1	Low Rate Persistent (43%)	2.5
Class 2	Low Rate Adolescent Limited (42%)	1.9
Class 3	High Rate Adolescent Limited (11%)	12.6
Class 4	High Rate Persistent (4%)	39.3
	CCLS	
Class 1	Low Rate Persistent (60%)	2.6
Class 2	Adolescent limited (17%)	12.9
Class 3	Late starters (15%)	14.4
Class 4	High Rate Persistent (8%)	43.2

The model fits were evaluated by comparing the mean observed counts m_{it} for each trajectory with the fitted mean trajectory counts $\hat{\lambda}_{tk}$, using the posterior probabilities of class membership as weights:

$$m_{kt} = \sum_{i} \frac{p_{ik} y_{it}}{n}$$

This then helps provide information on how well the fitted trajectories fit the data when assuming the classes are specified correctly.

Figure 5.5 OI observed plotted trajectories of quartic negative binomial 4 class model

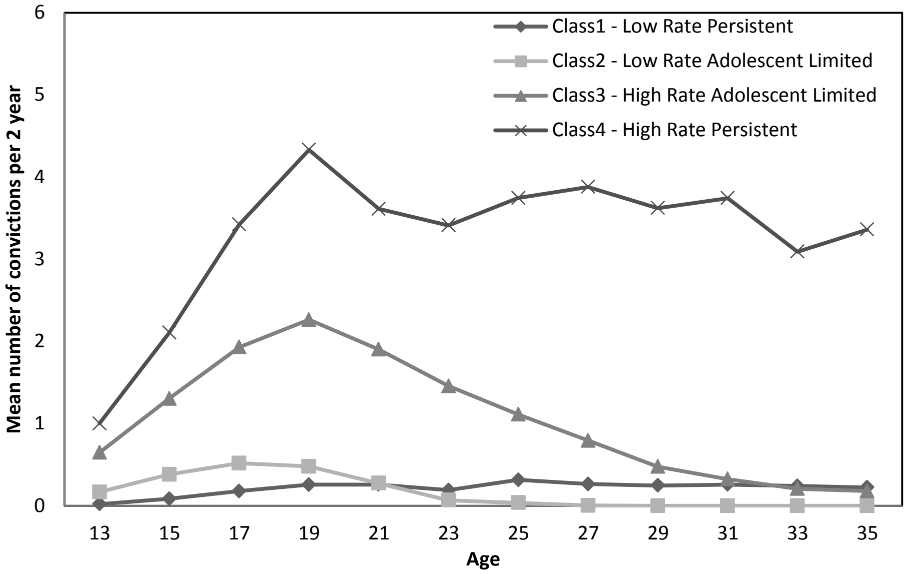
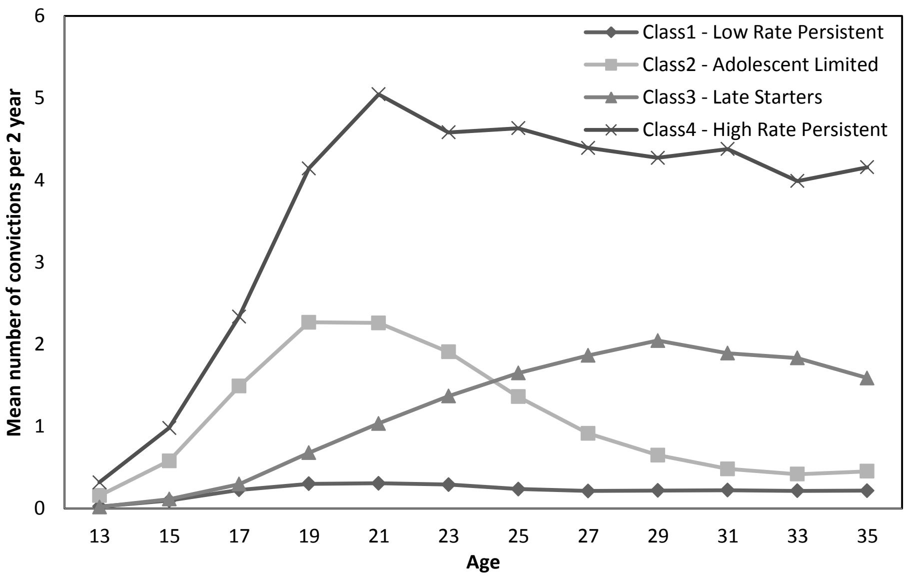


Figure 5.6 CCLS observed plotted trajectories of quartic negative binomial 4 class model



5.14 Limitations of Polynomial Regression and the use of Cubic Splines

Using cubic or quartic polynomials to produce trajectories is popular due to being able to produce smooth curves and interesting shapes. However, the use of polynomial curves can lead to modelling difficulties. Unusual and unexpected changes in the direction of the plotted estimated trajectories can sometimes occur when using polynomial curves, which are not supported by the data. When using polynomials to estimate a curve through a given set of data points, it would seem reasonable to assume increasing the degree of the polynomial (or number of interpolating points) would reduce the error in the polynomial interpolation. However this is not always the case and actually increasing the degree of the polynomials does not always improve the accuracy of the interpolating polynomial (Epperson, 1987). Several researchers have illustrated that when using polynomial curves, the estimated trajectories show a pattern of increase, followed by a decrease, followed by an increase or uptick towards the end of the observed period (Blokland et al., 2005, Blokland and Nieuwbeerta, 2005, Nieuwbeerta et al., 2011). This problem can be presented in Figure 5.7 and Figure 5.8. These are examples from the literature, by Marshall (2006) estimated juvenile trajectories for indigenous (native Australian population) and non-indigenous juveniles, and found there was an upward turn in their high rate trajectory at age 19 after an earlier increase and decrease before age 19. Bushway et al. (2003) also model trajectories of offending behaviour from age 13 to age 22, and estimate an uplift for three of their trajectories. The two papers take different approaches to these trajectory shapes; Marshall comments in the text on the change of shape without suggesting a reason, whereas Bushway et al. (2003) also comment, but suggest such behaviour to be evidence of intermittency. It is clear that when trajectories are estimated which show a number of changes of direction, then authors are sometimes uncertain how to interpret these shapes and whether such changes in direction are real.

Figure 5.7 Example of cubic polynomial trajectories from the literature showing uplifts on some trajectories at the end of the age scale. This example is from Marshall (2006) with an uplift for the 'high' group.

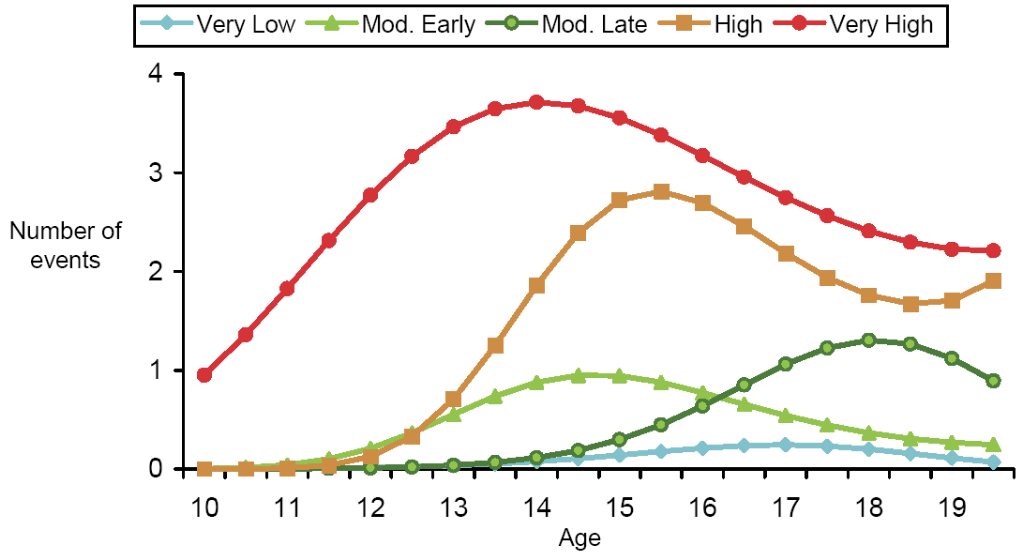
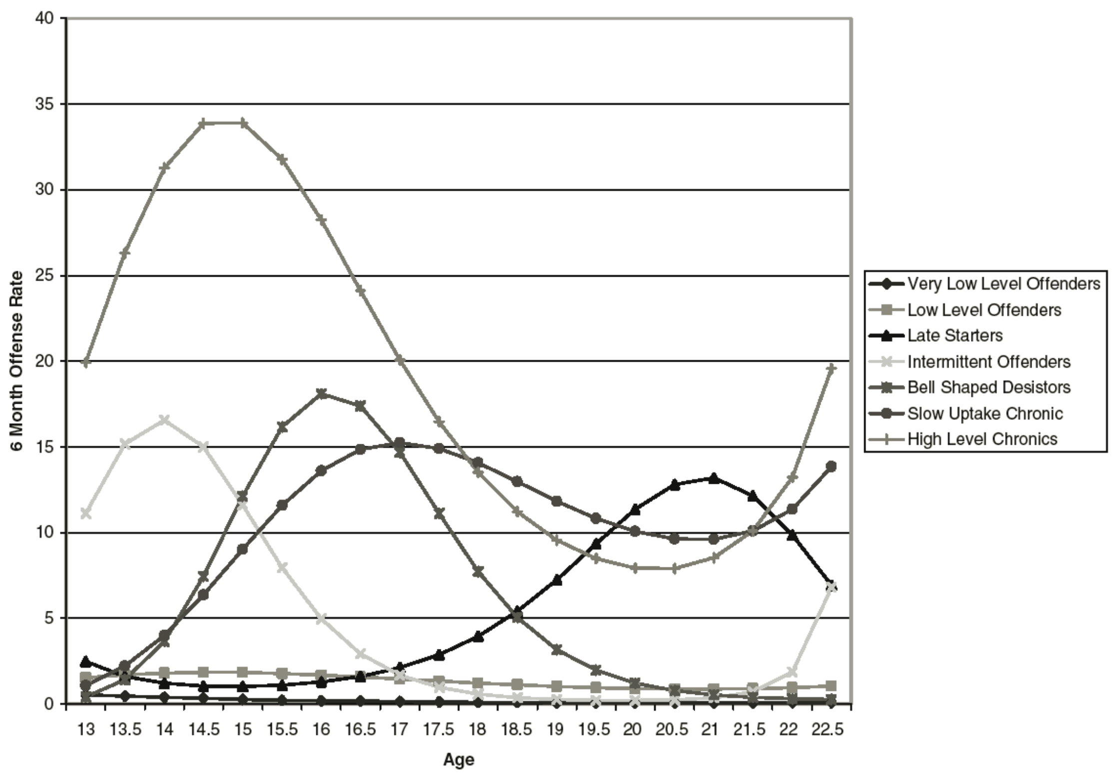


Figure 5.8 Example of cubic polynomial trajectories from the literature showing uplifts on some trajectories at the end of the age scale. This example is from Bushway et al (2003) using the Rochester Youth Development Study. Such uplifts may be spurious.



Fitting polynomial curves is “non-local”, meaning that a data point in one part of the time axis can influence the shape of the curve in a distant part of time axis. This is generally not a desirable characteristic and something researchers would want to avoid. Cubic polynomials are not able to generate curves that are randomly shaped. It is not possible for a trajectory curve to be produced that rises steeply in the first part of the time axis, and then exhibits a constant rate thereafter. To gain more variation in the shape of the curves would mean increasing the degree of the polynomial, which does not always improve the accuracy, and the function behaves abnormally at the extremes of the curves (Liu, 2015).

There are a few approaches that can be used to deal with the issues caused by using cubic polynomials. Firstly age can be fitted as a categorical factor. This would mean age is fitted as a stepwise function where the levels of the steps are constant within each time period, and jump between the time periods. Even though this approach offers local fitting it unfortunately does not use information provided by the previous time period, so in effect the fitting is too local. This approach unfortunately can be problematic as it normally requires a large number of parameters to be estimated for each trajectory. Secondly, higher order polynomials could be used. Sweeten (2014) reports that PROC TRAJ, the SAS software add on for group based trajectory modelling (Jones and Nagin, 2007), allows polynomials up to order five to be fitted. While using high order polynomials may allow more flexibility in the shape of the trajectory, the method still fails to solve the problem of the non-locality of polynomial curves, where a data point at a low age can have a large effect on the fitted curve at a high age. A third option is to use cubic B-splines, which have not previously been used in GBTMs. Employing cubic B-splines provides a flexible approach to estimating curves (Silverman, 1985). Fitting of cubic B-splines is relatively straightforward and together with the flexibility of shape, makes it suitable for group-based trajectory models. (See Francis, Elliott and Weldon, 2016 for a discussion of smoothing GBTM through B-splines and which is a publication arising from this thesis).

5.15 Smoothing Trajectories and B-Splines

As previously discussed, polynomial models are subject to a number of issues. Using a more flexible form of the smooth function, equation 5.3 can be replaced with the following:

$$\ln(\lambda_{tk}) = f_k(t)$$

This now changes the model to now estimate k different smoothers or smooth functions which change over time for each separate trajectory. There are a number of approaches that can be used for estimating the smoothers, such as, kernel smoothers or running line smoothers (see Hastie and Tibshirani 1991 for a detailed account of these different approaches). For this study, regression smoothers have been utilised in the method for smoothing. Fortunately, implementing regression smoothers in statistical software is fairly simple and can be used without having to write a program specifically for it. B-spline regression is a method that can estimate the data by a set of cubic polynomial regressions that are fitted to adjoining sections of the data. Essentially the regression smoothers are piecewise cubic polynomials that meet at a number of points called knots (these are the points where the sequential cubic polynomials touch). This set of piecewise cubic polynomials are smooth and continuous at each of the knot points (the first and second derivatives agree). Deciding on the placement of knots is often cited as a disadvantage to the use of regression splines, however, the fact that the number and places of knots can vary allows a great deal of flexibility and a vast range of functions to be estimated.

The regression splines can be fitted by adding a basis to the original design matrix of the cubic polynomial. The design matrix of the cubic polynomial is defined by;

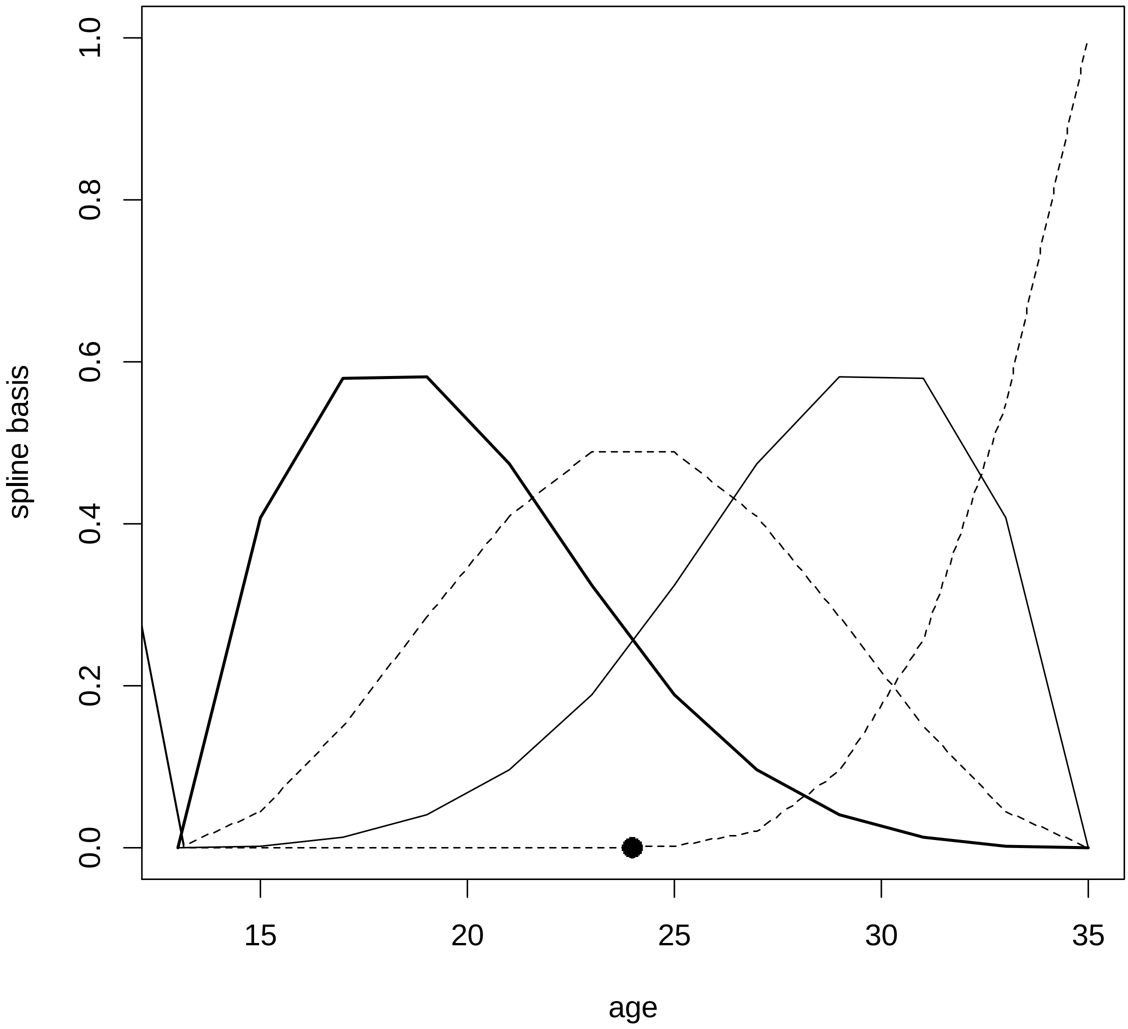
$[\mathbf{1}, \mathbf{T}, \mathbf{T^2}, \mathbf{T^3}]$, these are the column vectors containing the values $1, t_{it,} t_{it}^2$ and t_{it}^3, then the regression spline basis can be defined by:

$$c_{it}^h = \begin{cases} (t_{it} - z_h)^3 & \text{If } t_{it} > z_{h,} \\ 0 & t_{it} \leq z_h. \end{cases}$$

Here extra columns c_h have been added for each knot h, where z_h represents the location of the knot. For example, if a regression spline had three knots, this would results in three extra columns in the design matrix and contain seven degrees of freedom for each trajectory. The extended design matrix is referred to as the basis of the spline.

For this study B-splines have been chosen as the regression spline to be fitted as they are quite flexible, relatively easy to use and fairly simple to incorporate into the approximation process. Often traditional splines calculated from polynomials suffer from being numerical unstable because there may be large values within the design matrix and the columns can be highly correlated (Keele, 2007). To avoid this collinearity and to obtain numerical stability, the B-spline is an orthogonal transformation of the simple regression spline defined above. Choosing the power of the spline is something that needs to be considered. In the field of criminology, many studies modelling the age-crime curve use cubic polynomials as they produce the smooth bell-shaped curve distribution (Liu 2015). As seen in Figure 5.9 the columns of the basis can be presented graphically. It shows the B-Spline basis for 1 knot with 4 degrees of freedom over the range of x from age 13 to age 35. No basis extends over the entire range of x, ensuring that the fit is local. Specifying the number and placement of knots is required when fitting the B-splines. Typically, they are placed at selected quantiles of the x-axis dependent on how many knots are chosen. Rodriguez (2001) suggests that the most appropriate placement of knots is at the areas where $f(t)$ is changing more rapidly. However, it is sometimes appropriate to examine information criterion such as the BIC and ICL-BIC, coupled with the knowledge of the subject and data, to decide the most suitable location and the amount of knots. A different approach to deal with the number and location of knots is to treat it as a model selection problem (Berk, 2008). The aim becomes to choose the number of knots that minimises the model fit statistics, by basically treating knot selection as regressor selection. There are some methods that can assist with choosing the appropriate amount of knots. He and Ng (1999) use a method called the three-step selection, which compares the change in the Akaike information criterion when adding or removing knots. Osborne et al. (1998) use a method called LASSO (Least Absolute Shrinkage and Selection Operator) which is a regression method which involves penalising the absolute size of the regression coefficients.

Figure 5.9 B-Spline basis with one knot and 4 degrees of freedom, with age axis used in the example. The knot point is indicated as a dot on the age-axis



5.15.1 Fitting the BGBTM in Statistical Software

It is possible to use standard software to fit the B-spline group-based trajectory model (BGBTM) providing that the software has a flexible enough user interface to allow user-specified design matrices for the trajectories. In general, it is recommended that a package such as R can easily be used to calculate the cubic B-spline basis using the `bs()` function in library `splines`, and the B-spline basis can then be provided as time varying covariates in the package of choice. Thus, the R-code

```
library(splines)
xx=bs(age, df=7)
```

can be used to generate the B-spline basis for age with four knots. SAS through the PROC TRAJ procedure is a popular method of fitting group-based trajectory models, and allows time-varying covariates. Thus, specifying the B-splines as time-varying covariates and setting the polynomial order to zero will fit a similar model to the one in this chapter (it is similar but not identical as there is still an intercept for each trajectory in the model). MPLUS in contrast does not appear to have the flexibility required. Other alternative software packages considered for trajectory fitting include Latent Gold and R. The package `lcmm` in R can fit group-based trajectory models for continuous and ordinal data, but not for count data. In this chapter, having the focus on count data, Latent Gold is used, which is a general package for a wide variety of latent class models. As done previously group based trajectory models can be fitted using the Latent Class regression option, which required the data to be in long form. R is used to calculate the B-spline basis, and the variables which make up this basis are added to the dataset and become the “predictors” in the regression. The predictor effects are specified as class dependent (that is there is a separate trajectory for each class or group) through the model tab. Latent Gold has the facility to fit either the Poisson or the negative binomial by setting the type of the dependent variable either to “count” or to “overdispersed count”. The fitted trajectories can be examined by requesting “estimated values” on the output tab. Table 5.9 shows the various information criterion from the BGBTM for both datasets.

5.15.2 Results from GBTM using B-Splines

Table 5.9 Information criterion from BGBTM for both OI and CCLS data samples.

BGBTM Negative Binomial		OI BIC	AIC	ICL-BIC	CCLS BIC	AIC	ICL-BIC
4df	**1-Class**	72579.2	72541.04	72579.2	114619.1	114581	114631.7
	2-Class	68009.87	67927.21	***68472.35***	106085.7	106003	106824.8
	3-Class	67420.69	67293.52	71076.74	104650.3	104523.2	106292.4
	4-Class	66889.56	66717.88	70659.75	103871.3	103699.6	***106083.9***
	5-Class	66765.76	66549.56	70905.78	103543.9	103327.7	107480.9
	6-Class	66645.88	66385.17	72778.86	103316.3	103055.6	108154.5
	7-Class	66576.21	66271	72949.91	103185.1	102879.9	108429.9
	8-Class	66534.13	66184.4	73054.28	103049.5	102699.7	108620.9
5df	**1-Class**	72590.36	72545.85	72590.36	114633	114588.5	114633
	2-Class	68032.87	67937.49	***68494.32***	106107.6	106012.3	106826.3
	3-Class	67446.05	67299.8	71110.66	104680.1	104533.9	106292.1
	4-Class	66920.17	66723.05	70694.68	103905.7	103708.6	***106083.2***
	5-Class	66798.91	66550.92	70964.61	103574	103326	106543.2
	6-Class	66696.82	66397.96	72964.66	103352.4	103053.6	106911.7
	7-Class	66641.11	66291.39	73133.85	103226	102876.3	107084.2
	8-Class	66578.82	66184.58	73736.43	103103.2	102702.6	108630.4
6df	**1-Class**	72590.67	72539.8	72590.67	114631.5	114580.6	114631.5
	2-Class	68038.76	67930.66	***68501.44***	106112.4	106004.3	106830.7
	3-Class	67460.85	67295.52	71122.04	104691.4	104526.1	106303.4
	4-Class	66940.68	66718.13	70707.55	103928.4	103705.8	***106102.4***
	5-Class	66827.5	66547.72	70998.03	103605.5	103325.7	106569.1
	6-Class	66719.45	66382.44	72887.89	103386.5	103049.5	106911.1
	7-Class	66665.48	66271.25	73060.36	103262.2	102868	107083.5
	8-Class	66629.33	66177.87	73156.03	103146.6	102695.2	108659.8
7df	**1-Class**	72594.43	72537.20	72594.43	114639.5	114582.2	114639.5
	2-Class	68040.15	67919.33	***68503.03***	106127.6	106006.8	106846.1
	3-Class	67478.38	67293.98	71122.58	104713.9	104529.5	106323.6
	4-Class	66969.64	66721.65	70728.86	103959.6	103711.6	***106127.4***
	5-Class	66856.89	66545.31	72733.99	103644.8	103333.3	106602.1
	6-Class	66759.75	66384.59	72785.58	103434.5	103059.3	106952.5
	7-Class	66707.06	66268.31	72791.34	103314.7	102875.9	107127.6
	8-Class	66678.62	66176.29	73162.48	103206.7	102704.3	108711.9
8df	**1-Class**	72605.24	72541.65	72605.24	114647.9	114584.4	114647.9
	2-Class	68058.79	67925.26	***68524.04***	106144.6	106011.1	106863.2
	3-Class	67495.71	67292.23	71141.87	104736.7	104533.3	106348.2
	4-Class	66987.78	66714.36	70751.56	103988.4	103714.9	***106153.3***
	5-Class	66886.94	66543.57	71057.71	103680.2	103336.8	106636.3
	6-Class	66778.7	66365.39	72933.56	103476.2	103062.9	106996.6
	7-Class	66736.06	66252.8	73150.19	103364.5	102881.2	107185.1
	8-Class	66743.81	66196.97	73903.87	103268	102714.8	108770.5

Figures in bold italic are the lowest ICL-BIC value in the column.

Figure 5.10 OI plot of estimated trajectories for the B-Spline 4df negative binomial model

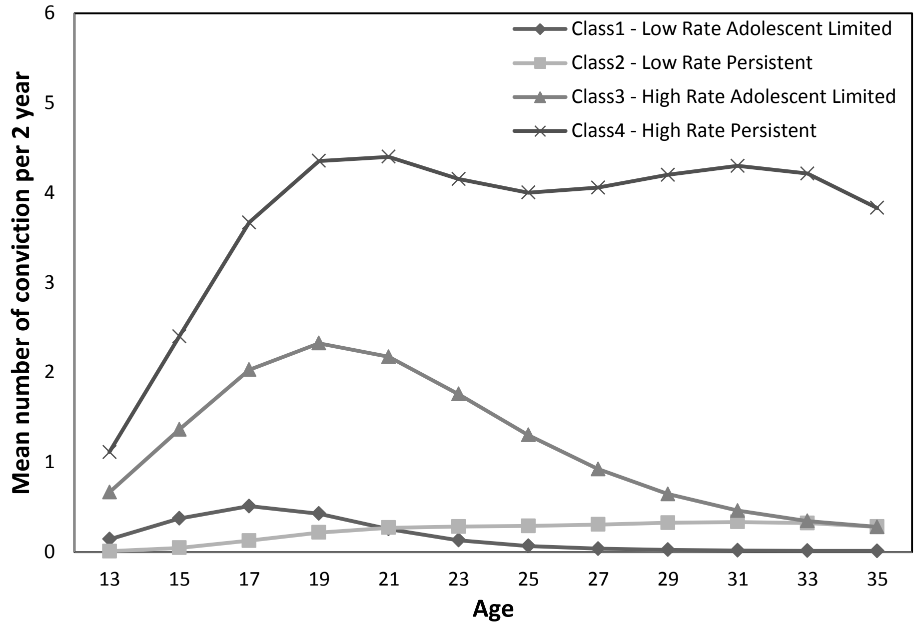
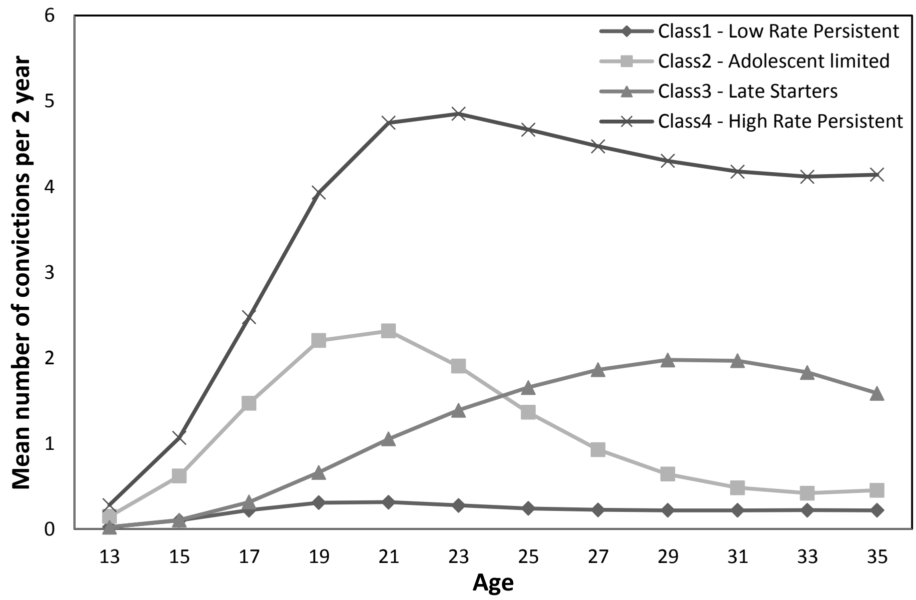


Figure 5.11 CCLS plot of estimated trajectories for the B-Spline 4df negative binomial model



Again from observing the trajectories plotted in Figure 5.10, Figure 5.11, Figure 5.12, and 5.13 the trajectory groups from the BGBTM models are labelled the following in Table 5.10:

Table 5.10 Offending trajectory groups and class sizes for BGBTM models

Trajectory Group	OI	Sum of expected # of convictions
Class 1	Low Rate Adolescent Limited (55%)	2
Class 2	Low Rate Persistent (31%)	2.8
Class 3	High Rate Adolescent Limited (11%)	14.3
Class 4	High Rate Persistent (3%)	44.7
	CCLS	
Class 1	Low Rate Persistent (60%)	2.5
Class 2	Adolescent Limited (17%)	1.9
Class 3	Late Starters (15%)	12.6
Class 4	High Rate Persistent (8%)	39.3

The trajectory groups are still labelled the same as the GBTM models, however the class sizes differ, more so for the OI dataset. The largest group for the OI dataset is now the ‘Low rate adolescent limited’ instead of the ‘Low rate persistent’ group.

Figure 5.12 OI observed plotted trajectories of B-spline 4df negative binomial model

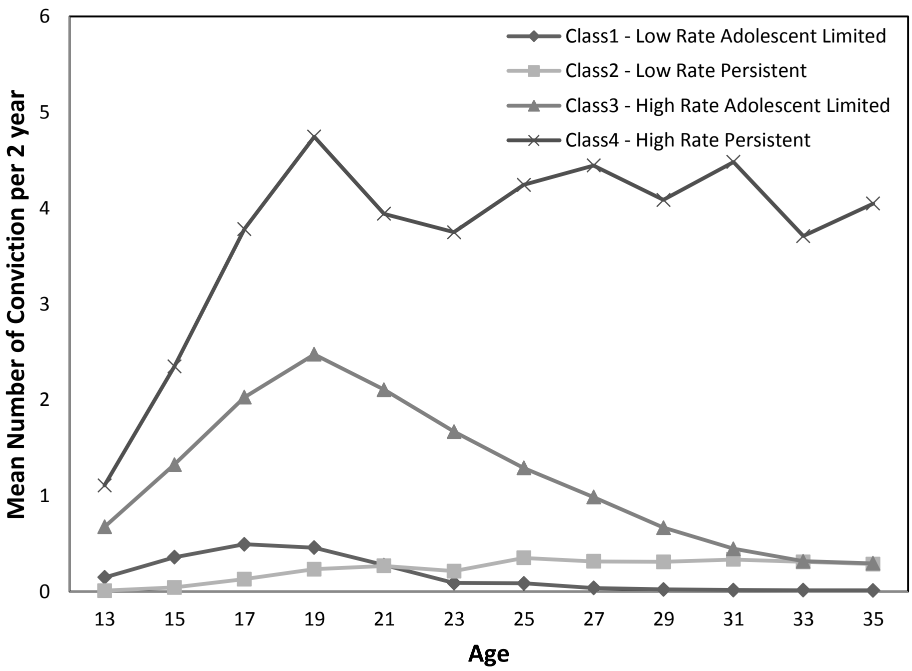
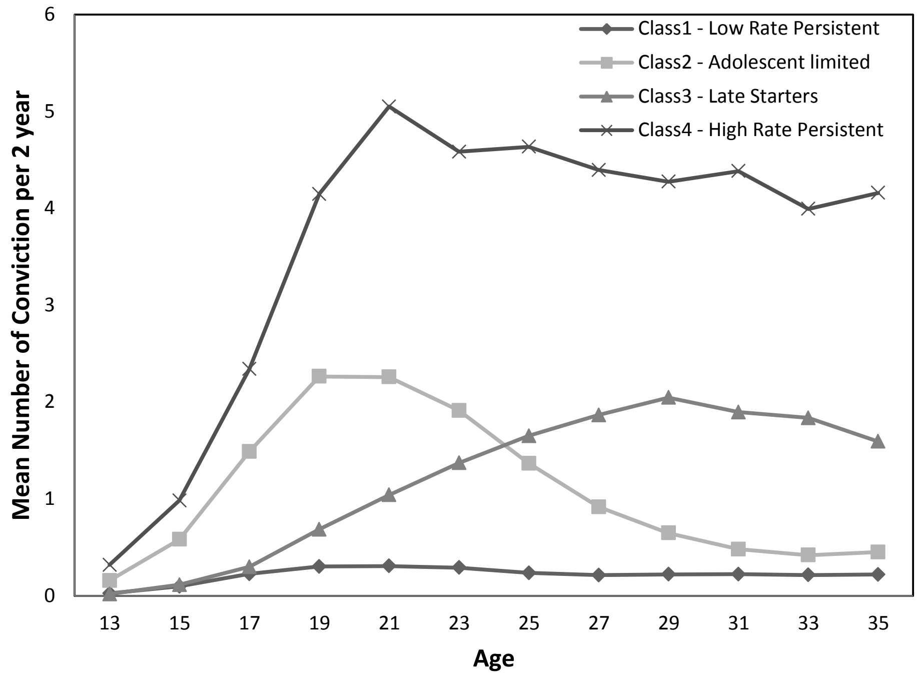


Figure 5.13 CCLS observed plotted trajectories of B-spline 4df negative binomial model



5.16 Posterior probabilities

After estimating the parameters for the model, the posterior probabilities of trajectory group membership can then be estimated for each individual.

The posterior probability of trajectory group membership is given by:

$$\hat{p}_{ik} = \hat{P}(k|\mathbf{y}_i) = \frac{\hat{\pi}(k)\hat{P}(\mathbf{y}_i|k)}{\sum_{k=1}^{K} \hat{\pi}\,(k)\hat{P}(\mathbf{y}_i|k)} \tag{5.4}$$

where

$$\hat{P}(\boldsymbol{y}_i|k) = \prod_t \hat{P}(y_{it}|k)$$

for the chosen discrete distribution. These posterior probabilities of class membership will be used in the next section where the probability of reconviction after age 35 will be examined.

5.17 Conclusion

Using GBTM and BGBTM it was possible to assess whether different groups of offenders following separate trajectories could be identified. The results presented showed that the four class solution discovered four distinct offending trajectory groups for both datasets which are labelled in Table 5.8 and Table 5.10 and presented graphically in Figure 5.3, Figure 5.4, Figure 5.5, Figure 5.6, Figure 5.7, Figure 5.10, Figure 5.11, Figure 5.12 and Figure 5.13. Like other writers for example Piquero et al. (2007) the trajectories can relate to Moffit's taxonomy (Moffitt, 1993) and all the figures have trajectories that are representative of the age-crime curve (Farrington, 1986), showing a peak around late adolescence followed by a decline in offending from early 20's. Both datasets produce similar results, with both having trajectories that which can be labelled adolescent limited. The shape of this trajectory appears to be quadratic. However, the remaining trajectories are not polynomial shaped. All the figures display a 'Low rate persistent' offending group which appears to coincide with Moffitt's life course persistent group. This has a nearly constant trajectory shape from around age 20 to age 35. Another group found to be common in both datasets was 'High rate chronic' was also non-polynomial in shape. From peaking at 21 for the OI data at around 2 convictions per year, the trajectory flattens off at age 27 and remain at around 1.7 convictions per year. Similarly, for the CCLS this trajectory has a peak at 23 at around 2.4 convictions per year, the trajectory flattens off at 27 and remains at around 2 convictions per year. The OI adolescent limited group is split into two, a low rate and high rate group, with more offenders belonging to the low rate group. The CCLS data has a fourth trajectory labelled 'Late Starters' which show offenders with a later onset age and continuing to offend at a lower rate than the high rate group. The trajectory group sizes and labels remain identical for the CCLS dataset for both the GBTM and BGBTM models. However, there is slight variation in the group sizes between the models for the OI data. For the quartic polynomial trajectory model the largest group at 43% is labelled the ‘Low rate persistent’ group, with the ‘Low rate adolescent limited’ group at 42%. In the B-Spline model the largest group becomes the ‘Low rate adolescent limited’ group at 55%, with the ‘Low rate persistent’ group being 31%. However, the trajectory shapes remain very similar for both polynomial and spline models.

As pointed out by (Brame et al., 2001), these substantive results would also support general life course theories such as the cumulative disadvantage theory of Patterson (1993) and indeed Gottfredson and Hirshi’s general theory of crime (1990).

The disadvantage of using polynomial trajectories in generating up ticks not supported by the data has been highlighted in section 5.14 showing examples from Marshall (2006) and Bushway et al. (2003) where upticks towards the end of the trajectories were displayed. The B-Spline offers strong advantages for these problems. However, no matter which method is used, researchers must also take care that their trajectories to not extend beyond the range of the data. Additionally, if the B-spline method detects upticks, examination of the number of convictions in the region of the uptick will help determine the criminological importance of such a result.

Another disadvantage of polynomials that was identified is the data points in the early part of time axis can influence the shape of the curve in the later part of the time axis. While this can be seen mathematically, no empirical studies have been undertaken on GBTMs to investigate this problem.

An alternative to B-splines would be to use the observed mean counts and may be an easier way of understanding the underlying shaper of the trajectories. The procedure would be to fit polynomial trajectories and then to plot the observed mean counts at each age point rather than the fitted curves. This idea is suitable for those without access to B-spline software., but it would be an approximate method as the assignment weights for the means are determined by the polynomial trajectory model, rather than the better B-spline model. It would give similar result to that obtained by fitting ‘age’ as a factor.

The model is valuable to those wishing to interpret trajectories either as real subgroups or as approximations to reality with individual variability around each trajectory. It also of use to those who simply want to use trajectories to account for heterogeneity in the population through non-parametric discrete mixing distribution. For these applications, mixtures of splines will provide more flexibility than mixtures of polynomials.

The model can be extended in various ways. The earlier assumption of count data can be relaxed, so that trajectory models can be fitted to binary or continuous observations—this will simply change the distributional assumption and the link function. For continuous data, the lcmm package in R can be used to fit the models. Missing observations can be incorporated easily as the likelihood can be constructed for only those time points that are observed. This would essentially use a full information maximum likelihood analysis and would assume a missing at random process. Other extensions to the model presented here are possible but would require programming work. For example, the Zero-inflated Poisson (ZIP) model, which allows for intermittency, is popular in criminological applications as it is included in PROC TRAJ. However, the full form of the ZIP model is not available in Latent Gold. Where the Nagin and Land ZIP model specifies a different immune proportion for each time point and for each trajectory, the Latent Gold model ZIP model assumes a constant immune probability over all trajectories and time points. Other forms of smoothing technology could also be used in place of B-splines, and this would also require programming. The results with other smoothers would be expected to be similar.

In conclusion, the use of B-spline trajectories is recommended. They provide a more flexible way of fitting trajectories, and the results of such analyses are not constrained to the often unrealistic shapes of polynomial curves. Some software can already be used to fit these models. The assumption of negative binomial counts may be more realistic in many examples than the more common assumption of Poisson counts.

The next chapter uses the posterior probabilities estimated from the models to aid in predicting reconviction over two different time periods – two years and ten years – using the trajectory group membership as a predictor in future reoffending.

6 PREDICTING RECONVICTION

6.1 Introduction

In this chapter, the results from the GBTM and BGBTM developed in the last chapter are used to form a model to assist in predicting reconviction of offenders based on their assigned offending group membership. Logistic regression modelling is introduced as a method to predict the probability of reconviction within a specified time period. This statistical model was popularised by David Cox (1966), such that the dependent variable is binary and the probability of the dependent variable responses are based on one or more independent variables. This is then able to measure the relationship between a binary dependent variable and a set of independent variables through the estimated probabilities via a logistic function. It is a useful technique for predicting the likelihood of reconviction.

The main aim of this chapter is to consider the likelihood of reconviction of a 35 year old with a detailed offending history over time back to age 12. The assigned offending group trajectories are based on the offending history of each individual. These offending group trajectories will be tested to see if they are informative in predicting the likelihood or probability of reconviction in the short and long term. It is not a requirement that each offender has been convicted at age 35 – they may have offended at any time between age 12-35 years.

After deciding upon the optimal number of classes in the previous GBTM models and BGBTM models in the previous chapter, the posterior group membership probabilities from these models can be used to predict the likelihood that individuals belong to a certain trajectory group. This can then be used to predict future offending. The likelihood of reconviction for the quartic polynomial overdispersed 4 class model

(Quartic4) is examined, followed by examining the B-spline with 4 degrees of freedom overdispersed 4 class model (B-Spline4).

To identify the likelihood of reconviction based upon the trajectory group assignment, a follow up period which is referred to as “length" needs to be decided. It is important to define this follow up period, as the longer the period the greater the proportion of offenders who may be reconvicted, but the less relevant trajectory membership may be. Varying the length of the follow up time will reveal how chronic certain offending group trajectories are. If the follow up time is only short (12-24 months) and predicted likelihood of reconviction from the logistic regression is high for a certain trajectory group, then this trajectory group would be considered high risk as they are likely to reoffend in a short space of time. If the follow up time is longer (10 or more years) and the predicted likelihood of reoffending is low for a certain trajectory group, then this trajectory group would be considered low-risk as a long period of time has passed and the risk of reoffending has remained low suggesting that many of the offenders belonging to this trajectory group may have desisted from crime. If a certain trajectory group has very low likelihood of reconviction over 1-2 years, coming to the same conclusion that the members of this trajectory group have desisted from crime is questionable. It may mean that they have a substantial likelihood of reconviction after 5 or more years. Therefore, choosing and examining the length of follow up time needs to be made with consideration and the result need to be interpreted with caution.

Two different lengths of follow up time are used: 2 years and 10 years. This then allows for a thorough examination of the reconviction probabilities and allows an investigation as to whether the likelihood of reconviction changes over time for each of the trajectory groups. It is naturally expected that the longer term follow up period will result in higher probabilities of reconviction.

For both datasets, as the trajectory analysis terminated at age 35, the first follow up period covers the ages of 36-37 years and the second is from 36-45 years. In SPSS, two follow up reconviction variables were computed (one for each follow-up time), coded 1 if the offender received a conviction within the follow up period or 0 if they did not. When fitting the GBTM and BGBTM models in Latent Gold, the posterior probabilities can be produced by choosing the classification posterior option on the “classPred” tab before estimating the regression model with the optimal number of classes. Equation 5.4 in the previous chapter gives the equation for extracting the posterior probabilities. The posterior probabilities of trajectory group assignments are exported into an SPSS file. This file is in long form and has to then be aggregated to produce a single record per case. The highest posterior probability was taken across the classes to determine class membership for each individual – this is termed the modal class in Latent gold

The reconviction variable then needs to be merged with the aggregated posterior probability file, so that each offender has a mean trajectory group membership assignment and a reconviction indicator variable. To predict the probabilities of reconviction, a logistic regression model needs to be performed. This can be done by various software such as SPSS, SAS and R. For this study R is chosen using the glm() function.

6.2 The logistic regression model

Logistic regression models have been used for predicting risk of reconviction in previous studies such as the Copas and Marshall (2002) paper. The author’s used a logistic regression method for calculating an offender’s likelihood of reoffending and the risk they posed, using the offender’s group reconviction scale (OGRS) as a predictor in the model. Furthermore, the paper by Tollenaar and van der Heijden (2013) compares the classical methods of prediciting recidivism to more recent prediction techniques from modern statistcis. They came to the conclusion that tradictional methods such as the logistic regression model, peformed equally as well to more modern methods.

Logistic regression models are a special case of generalized linear models. Regression models for a binary response variable describe the proportions of the population. The response of a binary variable is coded 1 or 0 and normally referred to as a “success” or “failure” outcome. A success of the population proportion is represented by the probability $P(y = 1)$ for a random individual. The probability will vary dependent on the predictor variables.

A simple linear model for a single predictor variable can be defined by:

$$P(y_i = 1) = \alpha + \beta x_i$$

Where the probability of a success is a linear function of x_i. However, any probabilities will either be below 0 or above 1 for very small or large values of x but the probabilities must fall between 0 and 1. To ensure that the probabilities fall between 0 and 1 for all possible values of x the following formula can be used:

$$\log e \left[\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right] = \alpha + \beta x_i \qquad (6.1)$$

Where $P(y_i = 1)/[\,1 - P(y_i = 1)]$ represents the odds. If $p_i = P(Y_i = 1)$ then it is also assumed that $y_i \sim Bernoulli(p_i)$. This distribution has changing variance, unlike the Normal distribution.

The above formula used the log of the odds or logit transformation as the link function. This model can be abbreviated to:

$$logit[P(y_i = 1)] = \alpha + \beta x_i$$

This is now the logistic regression model, the logit follows the straight line model with x_i, the probability $P(y_i = 1)$ follows a curve and the parameter β indicates whether the curve goes up and down as the value of x_i increases.

To calculate the probability of success from the parameter estimates the logistic regression model can be written as:

$$P(y = 1) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \quad (6.2)$$

In the above equation e is raised to a power and represents the antilogarithm of that number. The values of $P(y = 1)$ are estimated at specific values of x.

The model can be expressed in terms of the odds by applying antilogarithms to both sides of Equation (6.2):

$$\frac{P(y = 1)}{1 - P(y = 1)} = e^{\alpha+\beta x} = e^{\alpha}(e^{\beta})^{x}$$

Thus an increase of 1 unit in x will cause a multiplicative increase of e^{β} in the odds of the outcome. Finally, the model can be extended to deal with more than one explanatory variable. With p explanatory variables, equation (6.1) can be rewritten as:

$$\log\left[\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right] = \alpha + \sum_{p=1}^{P} \beta_p x_{ip}$$

where the x_{ip} can be continuous covariates or dummy variables constructed from factors.

6.2.1 Fitting the logistic regression model

Reconviction in this study is treated as a "success", and model the probability of reconviction with a single explanatory factor – the predicted modal group membership from the trajectory analysis.

Recall from the previous chapter that after fitting the GBTM models for both datasets in the previous chapter, the quartic polynomial negative binomial regression model (quartic4) and the B-spline negative binomial model with 4 degrees of freedom (B-Spline4) with 4 classes were chosen as the best fitting models. The posterior probabilities of the 4 class solution models were extracted and used in finding the most likely trajectory group for each offender. The justification for this is that the trajectories contain a summary of the criminal history variables, and therefore may produce simpler models. The assigned trajectory group was then used as a factor in the predictor variable in a logistic regression model, with a "reconviction within 2 years" indicator variable as the dependent variable. Logistic regression models were then repeated for "reconviction within 10 years" as the dependent variable. The predicted probabilities for each of the 4 offending trajectory groups were calculated using Equation (6.2). The results of the logistic regression models can be seen below in Table 6.1 and Table 6.2. In all of these tables, as common with the rest of this thesis, the first category is treated as the reference. In this case, class 1 is the largest category.

Table 6.1 Logistic Regression estimates for OI dataset for various time horizons and trajectory models

OI *N = 4420*	2 year Quartic4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-2.77	0.12	-23.91	<0.0001
Class 1 – Low Rate Persistent	0.00			
Class 2 – Low Rate Adolescent Limited	-0.22	0.15	-1.48	0.14
Class 3 – High Rate Adolescent Peaked	1.03	0.18	5.75	<0.0001
Class 4 – High Rate Persistent	2.63	0.19	13.54	<0.0001
	2 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-3.04	0.09	-33.0.	<0.0001
Class 1 – Low Rate Adolescent Limited	0.00			
Class 2 – Low Rate Persistent	0.38	0.16	2.4	0.0165*
Class 3 – High Rate Adolescent Peaked	1.45	0.16	9.12	<0.0001
Class 4 – High Rate Chronic	3.18	0.2	15.9	<0.0001
	10 year Quartic4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-1.66	0.07	-22.26	<0.0001
Class 1 – Low Rate Persistent	0.00			
Class 2 – Low Rate Adolescent Limited	0.47	0.09	5.30	<0.0001
Class 3 – High Rate Adolescent Peaked	0.92	0.13	7.18	<0.0001
Class 4 – High Rate Persistent	2.44	0.18	13.32	<0.0001
	10 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-1.26	0.05	-27.19	<0.0001
Class 1 – Low Rate Adolescent Limited	0.00			
Class 2 – Low Rate Persistent	-0.34	0.1	-3.57	<0.0001
Class 3 – High Rate Adolescent Peaked	0.70	0.11	6.36	<0.0001
Class 4 – High Rate Chronic	2.27	0.21	11.02	<0.0001

Table 6.2 Logistic regression estimates for the CCLS dataset for various time horizons and trajectory models

CCLS *N = 4420*	2 year Quartic4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-1.56	0.05	-30.27	<0.0001
Class 1 – Low Rate Persistent	0.00			
Class 2 – Adolescent limited	0.77	0.1	7.89	<0.0001
Class 3 – Late Starters	1.66	0.1	17.44	<0.0001
Class 4 – High Rate Persistent	2.66	0.13	19.91	<0.0001
	2 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-1.56	0.05	-30.30	0.0001
Class 1 – Low Rate Persistent	0.00			
Class 2 – Adolescent limited	0.78	0.01	7.94	<0.0001
Class 3 – Late Starters	1.68	0.1	17.65	<0.0001
Class 4 – High Rate Persistent	2.67	0.13	19.91	<0.0001
	10 year Quartic4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-0.04	0.04	-1.07	0.28
Class 1 – Low Rate Persistent	0.00			
Class 2 – Adolescent limited	0.63	0.09	7.03	<0.0001
Class 3 – Late Starters	1.80	0.12	15.03	<0.0001
Class 4 – High Rate Persistent	2.34	0.19	12.38	<0.0001
	10 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-0.05	0.04	-1.15	0.25
Class 1 – Low Rate Persistent	0.00			
Class 2 – Adolescent limited	0.63	0.09	7.08	<0.0001
Class 3 – Late Starters	1.83	0.12	15.17	<0.0001
Class 4 – High Rate Persistent	2.34	0.19	12.36	<0.0001

From the above estimates in Table 6.1 and Table 6.2, an offender belonging to trajectory group k has the estimated probability of reconviction equal to:

$$\hat{P}(y = 1) = \frac{e^{(\alpha+\beta_k)}}{1 + e^{(\alpha+\beta_k)}}$$

So for example, an OI offender who belongs to trajectory $k = 2$, the estimated probability of ten-year reconviction from the b-spline model is equal to:

$$\hat{P}(y = 1) = \frac{e^{(-1.26-0.34)}}{1+e^{(-1.26-0.34)}} = 0.167$$

Therefore the estimated probabilities of reconviction can be calculated for each trajectory class and each follow-up time.

Table 6.3 Predicted probabilities of recidivism for Offenders Index and CCLS for various time horizons and trajectory models

OI (group sizes are in brackets Quartic4/B-Spline4)	2 year		10 year	
	Quartic4	B-Spline4	Quartic4	B-Spline4
Class 1 – Low Rate Persistent (43%) (Quartic) / Low Rate Adolescent Limited (55%) (B-Spline)	0.059	0.046	0.161	0.221
Class 2 – Low Rate Adolescent Limited (42%) (Quartic) / Low Rate Persistent (31%) (B-Spline)	0.048	0.065	0.235	0.167
Class 3 – High Rate Adolescent Limited (11%/11%)	0.149	0.169	0.324	0.365
Class 4 – High Rate Persistent (4%/4%)	0.464	0.535	0.687	0.732
CCLS (group sizes are in brackets Quartic4/B-Spline4)	**2 year**		**10 year**	
	Quartic4	**B-Spline4**	**Quartic4**	**B-Spline4**
Class 1 – Low Rate Persistent (60%/60%)	0.174	0.173	0.490	0.489
Class 2 – Adolescent Limited (17%/17%)	0.313	0.313	0.643	0.643
Class 3 – Late Starters (15%/15%)	0.527	0.531	0.853	0.857
Class 4 – High Rate Persistent (8%/8%)	0.750	0.751	0.909	0.909

From Table 6.3 the OI offending group (England & Wales) most likely to be reconvicted within two years in the quartic4 model is Class 4 (High Rate Persistent) which is nearly 8 times more likely than Class 1 (Low Rate Persistent). For the B-Spline4 model the probabilities are slightly higher than the Quartic4 model (for Class 4 only). The probability of reconviction for Class 1 and Class 2 are both low and similar percentages ranging between 5-7% probability of reconviction within two years. This would be expected for the 'adolescent limited' group as once offenders reach adulthood, it is anticipated that the majority of offenders will reduce and desist from offending (Agnew, 2006, Moffitt, 1993).

For the 10 year follow up, the overall probabilities of reconviction are much higher due to the lengthier time span increasing the chances of being arrested and reconvicted. However, what is not known is whether the chances of reconviction are higher within the first five years or after so the probabilities should be interpreted with this in mind. The probabilities of reconviction for Class 4 are still the highest with

(69/73%) chances of being reconvicted within 10 years. The other offending group probabilities have all increased substantially with all of them being more than double the probability of being reconvicted within 2 years. However, the lowest probabilities still belong to the Class 1 and Class 2. In general, all the predicted probabilities for each group differ between the two models for both the 2 year and 10 year follow up periods. The offending group sizes also differ between the two models which may explain some of the differences in the predicted probabilities.

The CCLS analysis show much higher probabilities of reconviction for both follow up periods in comparison to the OI. The offending group with the highest likelihood of reconviction in 2 years is Class 4 (the High Rate Persistent group, the same result as the OI) which is not surprising. The chance of reconviction for this group is extremely likely being 75% within 2 years and just over 90% in 10 years. The lowest likelihood for reconviction is Class 1 (the Low Rate Persistent group) this is still much higher when compared to the OI equivalent group. Unlike the OI, the Quartic4 and B-Spline4 model probabilities are virtually identical.

The big differences in reconviction probabilities between the two datasets could be due to a number of reasons. As already mentioned in Chapter 3, it is very difficult to directly compare conviction statistics from different countries and it should be done with caution. It is highly likely that there are some differences between the reconviction probabilities between the two datasets, even though their four offending trajectory groups are of similar classifications, they are still slightly different and vary in size. The OI has more offenders belonging to its lower rate offending groups and the CCLS has more offenders belonging to its higher rate offending groups. As noted by Farrington and Zara (2015) “Recidivism in European countries is likely to reflect different legal systems and criminal laws adopted in each country, making interpretation of reoffending and conviction data and comparison between countries complex”. It could be that the England and Wales have better deterrence strategies than the Netherlands therefore reducing the amount of offenders that are likely to reoffend. The Netherlands may have better strategies when it comes to targeting repeat offenders and therefore gain more reconvictions. In the Dutch criminal justice system, juvenile offences tend to be dealt with away from court and the police usually opt for diversionary measures when dealing with young offenders, leading to a big reduction in the number of court orders placing juveniles in institutions (Junger-Tas and Block, 1988). Therefore, offenders that are dealt with in court in the Netherlands are more likely to be for more serious offences. Another possible reason why the Netherlands reconviction probabilities are higher could be down to the way the dataset has been structured. As previously mentioned in chapter 3, each offences in the CCLS dataset, the charge for the highest pusnishment was coded. This could mean that there are more serious offenders in the CCLS dataset sample, compared to the OI sample, who are more likely to reoffend. This was previously observed from exploring the datasets in Chapter 3. It is already known that the CCLS dataset has higher mean number of offences than the OI dataset. Therefore it is expected that the recidivism possibilities will be higher to some groups in the CCLS.

6.2.2 Adding covariates into the model

It might be thought that the prediction model could be improved by including other covariates not related to the shape of the trajectory. These include (but are not limited to) gender, type of last conviction and whether a custodial sentence was given. As an example, gender has been included in the logistic regression models for the two datasets in Tables 6.4, 6.5, and 6.6.

Table 6.4 Logistic Regression Model (including Gender as a covariate) estimates for OI for various time horizons and trajectory models

OI *N = 4420*	2 year Quartic4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-2.89	0.18	-16.11	<0.0001
Class 1 – Low Rate Persistent	0.00			
Class 2 – Low Rate Adolescent Limited	1.0	0.18	5.49	<0.0001
Class 3 – High Rate Adolescent Limited	2.60	0.20	13.28	<0.0001
Class 4 – High Rate Persistent	0.15	0.17	0.86	0.39
Gender – Male	0.27	0.16	1.7	0.09
	2 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-3.17	0.17	-18.86	<0.0001
Class 1 – Low Rate Adolescent Limited	0.00			
Class 2 – Low Rate Persistent	0.39	0.16	2.44	0.01
Class 3 – High Rate Adolescent Peaked	1.42	0.16	8.79	<0.0001
Class 4 – High Rate Chronic	3.15	0.2	15.65	<0.0001
Gender – Male	0.17	0.17	0.96	0.34
	10 year Quartic4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	- 1.8	0.11	-16.8	<0.0001
Class 1 – Low Rate Persistent	0.00			
Class 2 – Low Rate Adolescent Limited	0.47	0.09	5.3	<0.0001
Class 3 – High Rate Adolescent Limited	0.89	0.13	6.9	<0.0001
Class 4 – High Rate Persistent	2.41	0.18	13.1	<0.0001
Gender – Male	0.19	0.1	1.9	0.06
	10 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-1.3912	0.0907	-15.34	<0.0001
Class 1 – Low Rate Adolescent Limited	0.00			
Class 2 – Low Rate Persistent	-0.34	0.1	-3.49	0.00
Class 3 – High Rate Adolescent Peaked	0.68	0.11	6.02	<0.0001
Class 4 – High Rate Chronic	2.24	0.21	10.88	<0.0001
Gender – Male	0.17	0.1	1.7	0.09

Table 6.5 Logistic Regression Model (including Gender as a covariate) estimates for the CCLS for various time horizons and trajectory models

CCLS *N = 4420*	**2 year Quartic4**			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-1.56	0.05	-30.27	<0.0001
Class 1 – Low Rate Persistent	0.00			
Class 2 – Adolescent limited	0.75	0.1	7.6	<0.0001
Class 3 – Late Starters	1.64	0.1	17.09	<0.0001
Class 4 – High Rate Persistent	2.63	0.13	19.62	<0.0001
Gender – Male	0.27	0.16	1.7	0.09
	2 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-1.8	0.15	-11.86	
Class 1 – Low Rate Persistent	0.00	0.00	0.00	0.00
Class 2 – Adolescent limited	0.75	0.1	7.63	<0.0001
Class 3 – Late Starters	1.66	0.1	17.29	
Class 4 – High Rate Persistent	2.64	0.13	19.63	p<0.0001
Gender – Male	0.26	0.16	1.67	0.09
	10 year Quartic4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-0.13	0.11	-1.17	0.24
Class 1 – Low Rate Persistent	0.00			
Class 2 – Adolescent limited	0.62	0.09	6.9	<0.0001
Class 3 – Late Starters	1.8	0.12	14.9	<0.0001
Class 4 – High Rate Persistent	2.33	0.19	12.3	<0.0001
Gender – Male	0.1	0.12	0.85	0.39
	10 year B-Spline4			
	Estimate	**Std. Error**	**z value**	**P-value**
Intercept	-0.13	0.11	-1.18	0.24
Class 1 – Low Rate Persistent	0.00			
Class 2 – Adolescent limited	0.63	0.09	6.92	<0.0001
Class 3 – Late Starters	1.83	0.12	15.04	<0.0001
Class 4 – High Rate Persistent	2.33	0.19	12.28	<0.0001
Gender – Male	0.1	0.12	0.83	0.4

Table 6.6 Predicted Probabilities of recidivism with Gender Covariate included

OI (group sizes are in brackets Quartic4/B-Spline4)	2 year		10 year	
	Quartic4	B-Spline4	Quartic4	B-Spline4
Class 1 – Low Rate Persistent (43%) (Quartic) / Low Rate Adolescent Limited (55%) (B-Spline)	0.06	0.05	0.17	0.23
Class 2 – Low Rate Adolescent Limited (42%) (Quartic) / Low Rate Persistent (31%) (B-Spline)	0.05	0.07	0.24	0.17
Class 3 – High Rate Adolescent Limited (11%/11%)	0.15	0.17	0.33	0.37
Class 4 – High Rate Persistent (4%/4%)	0.47	0.54	0.69	0.73
CCLS (group sizes are in brackets Quartic4/B-Spline4)	**2 year**		**10 year**	
	Quartic4	**B-Spline4**	**Quartic4**	**B-Spline4**
Class 1 – Low Rate Persistent (60%/60%)	0.178	0.177	0.492	0.492
Class 2 – Adolescent Limited (17%/17%)	0.314	0.314	0.644	0.644
Class 3 – Late Starters (15%/15%)	0.528	0.532	0.854	0.857
Class 4 – High Rate Persistent (8%/8%)	0.751	0.752	0.909	0.909

In examining the estimates from Table 6.6, the gender effects are all non-significant for all 8 analyses, with p-values lying between 0.05 and 0.9.

Surprisingly, the inclusion of gender has little effect, the direct effect of the gender covariate appears to be already subsumed by the trajectory groups.

6.3 Assessing Model Predictive Performance

There are a number of strategies in assessing the predictive performance of the logistic regression model. These include consideration of the Akaike Information Criterion (AIC), the null and residual deviance, confusion matrices, and Receiver Operating Characteristic (ROC) curves. The AIC measures the goodness of fit of the models to the given dataset and estimates the quality for each of the models compared to one another. It penalizes the model for added model coefficients so therefore models which minimise the AIC are preferred. Whilst residual deviance can give some indication of the goodness of fit of a binary logistic model, McCullagh and Nelder (1989) point out that this quantity cannot be used in this way, and does not have a chi-squared distribution. Confusion matrices are well known in statistical learning and are described in Altman and Bland (1994). A confusion matrix cross-tabulates the actual vs. predicted values of the model after dichotomising the predicted probabilities into two categories “predicted reconviction” and “predicted non-reconviction” using a cut-off probability of 0.5 and assists in evaluating the accuracy of the model predictive performance. The counts of the true positives, negative positives, false positives and negative positives are displayed in a 2x2 table and the accuracy of the model is calculated with:

$$\frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives\ +\ False\ Positives + False\ Negatives}.$$

Finally, another technique is the ROC (receiver operating characteristic) curve, which extends the idea of the confusion matrix. This is used for summarising and visualising the models performance via a graphical output. It plots the sensitivity (the true positive rate showing the model’s ability to predict an event correctly) on the Y-axis and (1-specifity) (false positive rate) on the X-axis for various choices of cut-off probability. When the ROC is representing a perfect predictive model, the sensitivity is equal to 1 and specificity is equal to 0, with the plotted curve touching the top left hand corner of the graph. Therefore, the steeper ROC curve the better the predictive performance of the model.

The area under the curve (AUC), sometimes called the index of accuracy or concordance index, gives the indication of the overall measure of fit of the model. It can be shown that the AUC also measures the probability that if a random pair of subjects (one true positive and one true negative) were chosen, the positive subject would have a higher predicted probability of the event in comparison the negative subject. Therefore, the AUC calculates the overall ability of the regression to classify between those offenders who reoffend and those who do not. An unpredictive model has an area of 0.5 and a perfect predictive model has an area of 1.

In Table 6.7 and Table 6.8 the AUC values have been calculated for all the logistic regression models and an example of the ROC curves are shown in Figure 6.1. The CCLS AUC values overall are slightly higher than the OI models, suggesting that the models predictive performance is doing better for CCLS dataset. The models are performing better for the 2 year follow up for both datasets and the B-Spline model AUC values are higher, suggesting that the B-Spline model is a better choice. However, for the 10 year follow up, the AUC area values are lower (between 0.6-0.7) which are relatively poor predictions.

It is not really surprising that the ten year prediction models are poorer than the two year predictions as it is trying to predict so far into the future. If a developmental view of offending is taken, then life course variables and life -events such as gaining steady employment, marriage and parenthood will reinforce desistance from criminal behaviour and affect the offending outcome (Akers, 1999, Farrington, 1986, Sampson and Laub, 2003). However, other factors such as unemployment, alcoholism, drug taking, marrying another offender and association with law-violating peers may contribute to reoffending and intervene in this ten-year period and also affect the offending outcome (Akers, 1999, Tremblay et al., 2004, Lipsey and Derzon, 1999).

Table 6.7 AUC values for basic model

Offenders Index	2 year		10 year	
	Quartic4	B-Spline4	Quartic4	B-Spline4
AUC	0.677	0.688	0.6618	0.6628
CCLS	**2 year**		**10 year**	
	Quartic4	**B-Spline4**	**Quartic4**	**B-Spline4**
AUC	0.7104	0.7116	0.6628	0.6006

Table 6.8 AUC for models with gender

Offenders Index	2 year		10 year	
	Quartic4	B-Spline4	Quartic4	B-Spline4
AUC	0.678	0.6879	0.6139	0.6022
CCLS	**2 year**		**10 year**	
	Quartic4	**B-Spline4**	**Quartic4**	**B-Spline4**
AUC	0.7144	0.7156	0.6634	0.6643

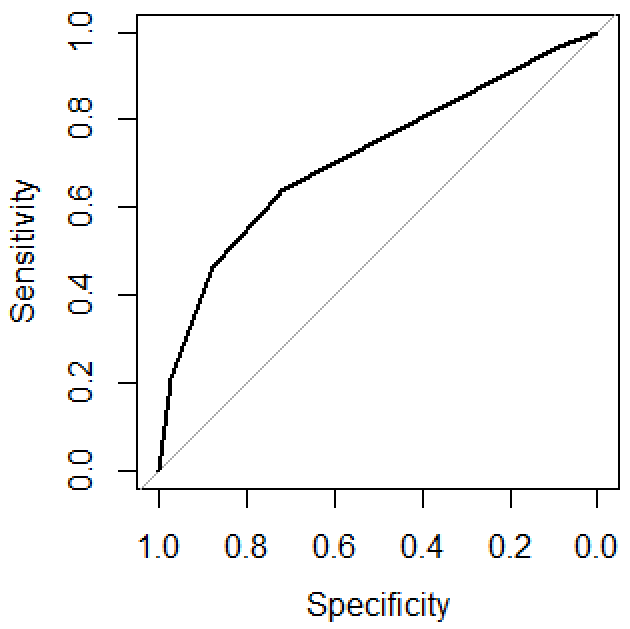


Figure 6.1 CCLS 2yr quartic model with gender covariate – Example of ROC curve. Note that the x-axis is reversed in this plot.

6.3.1 Conclusion

Trajectory membership is a powerful predictor of reconviction for both the two year and ten year follow up periods. This method is a good alternative to other reconviction methods which use other summaries of criminal history such as age of onset (Nagin and Farrington, 1992) or offending rate (Francis et al., 2007). The offending trajectory groups generated by the GBTM and BGBTM models have proven to be strong predictors of reconvictions. For both datasets, offenders belonging to the Class 4 (High Rate Chronic offenders) have reconviction probabilities ranging from 46%-75% chance of reconviction in 2 years and from 69%-91% chance of reconviction in 10 years. Those offenders in the OI dataset belonging to the 'Low Rate Persistent' group (Class 1 for the Quartic4 models and Class 2 for the B-Spline4 model) have reconviction probabilities from 6%-7% chance of reconviction in 2 years and from 16%-17% chance of reconviction for 10 years. In the CCLS dataset those offenders belonging to the 'Low Rate Persistent' group have reconviction probabilities of 17% chance of reconviction in 2 years and 49% chance of reconviction in 10 years. Although there is variation in the reconviction probabilities between the two datasets, it must be noted that the datasets are different and therefore the trajectory groups will differ and should not be directly compared. It would be sensible to expect some variation in the predicted reconviction probabilities because of this. The models perform better at predicting reconviction probabilities within 2 years than 10 years which is understandable as trying to predict so far into the future is difficult due to a number of factors which are discussed above. Using the AUC and ROC curve values to assess the predictive performance of the models shows that the 2 year B-Spline4 model is the best performing model, having the highest AUC values. This is encouraging for the use of B-Splines instead of polynomials in GBTM. Adding gender as a covariate in the model only marginally increases the models performance, suggesting that the effect of gender is already accounted for by the trajectory groups. This is possibly due to the fact there are very few females in either dataset and from the exploratory analysis in Chapter 4, it was revealed that males are the most likely to be recidivists.

The GBTM, BGBTM and logistic regression model results do not actually explain very much about the types of offenders that belong to both the datasets. It is unknown when or how many times each offender was reconvicted in the follow up periods. The dependent variable is only an indication of whether or not each offender had another conviction. The trajectory groups themselves are based solely on the number of offences committed over adolescence and early adulthood. No information about the type of offences is used or if offenders change the type of offences they commit. There could be some strong differences between the two countries if this information was used and may produce some very different offending groups if this was taken into account. The following chapters in this study begin to examine crime mix patterns and pathways using the types of offences through Latent Transition Analysis.

7 CHANGING CRIME-MIX PATTERNS OF OFFENDING

7.1 Introduction

The previous chapters focused on looking at the frequency of offending over the life course and how these frequency trajectories change with age. However, as already mentioned, this tends to ignore any changes in the patterns and types of offences being committed. Studying patterns of offending behaviour in detail is important for several reasons. For example, it allows the identification of which offence typologies appear to be precursors for other types of offences (Francis et al., 2004). Such information has both practical and theoretical implications. Knowing what types of offences criminals may commit prior to being involved in serious crime like murder, for example, can be of great importance to law enforcement agencies and policy makers as there may be scope for targeting such offenders before they move on to the more worrying pathways of criminal activity. Further, understanding possible links between various types of crime also has a theoretical force, which may enable to distinguish between different types of offenders. In short, gaining detailed knowledge of crime mix patterns – which is the main focus of this chapter – may well help with the understanding of offending behaviour and the causes behind it.

The goal of this chapter is primarily focused upon finding crime mix patterns and how they develop over the life course, within the two longitudinal datasets of criminal conviction histories. Using a latent Markov modelling approach, the criminal careers of offenders can be examined in more depth by exploring both the crime mix patterns of offenders and how these develop and change over time. By developing the methodology used in Francis et al. (2010), where the idea of lifestyle specialisation and short-term crime typologies (crime mixes) over five-year age-periods was introduced, a latent Markov model can be applied to identify different crime mix patterns and also estimate the transition probabilities from one age period to the next in the datasets.

It is worth a comment here about terminology. The terms of "latent transition analysis" and "latent Markov modelling" are both used in the literature for what is essentially the same model. The method has become more popular in the analysis of criminal career data as it shows the different phases of criminal careers and the transition between each phase. Thus Massoglia and McGloin et al have used the term Latent transition analysis (Massoglia, 2006, McGloin et al., 2009), whereas Bartolucci et al., (2007) used the term Latent Markov modelling. Which terminology is used seems to depend on the software used. Thus MPLUS and SAS users will tend to use the term Latent Transition analysis, whereas R, Latent Gold and other bespoke software will tend to use Latent Markov modelling. The term Latent Markov modelling will be used in this study except when referring to previous work.

This chapter proceeds with a discussion of crime mix patterns followed by the preparation and exploratory analysis of the datasets. This is then continued with a discussion of the methodological approaches and statistical analysis. Finally, the results of the Latent Markov Models are presented and discussed.

7.2 Crime-mix patterns

Crime mix (the variety of different offences committed over the period of time), is a criminal career dimension that has been explored by other researchers (Block et al., 2010, Piquero et al., 2007, Piquero et al., 2003). Crime mix patterns refer to the different types of offences committed by an individual within an age period and the particular offending characteristics or styles these individuals hold. The crime mix patterns can be thought of as different groups or classes of offending, where the individuals belonging to that group all share similar patterns of offending styles. The aim of this next chapter is to build upon the latent Markov modelling methodology used by Bartolucci et al., (2007) and Francis et al , (2010) to identify the crime mix patterns and to see how offenders transition between such groups as they age in both the OI for England & Wales and the CCLS Netherlands. Unlike Francis et al., (2010) (who focused on a female sample), the focus will be on both male and female offenders. The aim is therefore to try and identify if there are any common patterns of offending behaviour over the life course in the two countries.

7.3 Preparation and exploratory analysis of datasets

The preparation and alignment of the two datasets has been discussed in Chapter 3. However, for this work, further restrictions and restructuring were imposed.

The analysis of changing patterns of criminal behaviour over the life course will use eight five-year age periods from 12-51. This is a suitable length of time that represents an offender's criminal career over the life course.

The following restrictions were added to the datasets.

Similar years of birth of the two samples

The Netherlands CCLS data was restricted to those with years of birth within two years or less of the three England and Wales OI birth cohorts. Therefore, this restricted the dataset to include only those with a year of birth between 1926 and 1965 making them aged 12–26 in 1977. This condition minimises any generational or year of birth differences in the two samples.

A conviction between ages 12-51 in one of the 11 offence categories

It is possible that some offenders may not have had a conviction between the ages of 12-51. Therefore, only offenders that had at least one conviction occasion in one of the 11 offence categories between these ages were included.

After these restrictions had been imposed upon the CCLS dataset, only 2,267 offenders were left to be included in analyses. This was significantly less than the sample for OI cohorts even after the restrictions had been imposed. Latent class analysis and latent Markov modelling tend to detect more groups if the sample size is larger, and so a final alignment condition was to make the sample sizes equal in the two datasets. A random sample of 2,267 offenders across the three cohorts was therefore taken from the OI dataset.

7.3.2 Data Exploration

A similar approach to the Francis et al. (2010) study was adopted, dividing the conviction histories into 8 five-year age groups (12–16, 17–21, 22–26, 27-31, 32-36, 37-41, 42-46, and 47-51), but analysing males and females together. This age range period has been chosen since from earlier work (Francis et al. 2010) it is believed that there would be a large amount of transiting between various crime mix groups within this chosen age period.

In total, the individuals in both dataset samples combined accumulated 30,450 convictions over the 40 year period. Each offender has a conviction in at least one of the age groups and one of the common 11 offence categories. Table 7.1 and Table 7.2 show the number of conviction occasions for each of the 11 offence categories across the eight different age periods for OI and for the CCLS datasets respectively.

There are slightly more male offenders in the CCLS data sample, which may explain why the total number of convictions is higher, as past research shows that males commit more crimes than females (Blumstein et al., 1986, Piquero et al., 2003).

Overall, the majority of offending occurred in the 17–21 year age group; however, this is not the case for all offence categories. For example, drug offenders are more prevalent in the 22–26 age period. This would agree with the findings from Massoglia (2006) who also discovered that drug offences increased after the transition into adulthood.

There are a number of other differences between the two datasets that can be noticed. The CCLS dataset has a higher proportion of offenders committing offences in each age group, and has a substantially higher number of offenders for sexual offences, blackmail, robbery, drugs and public order offences. The OI dataset has lower proportions of offenders in each age group and therefore is lower than the CCLS for all of the offence categories. Burglary and theft was the most common category in both datasets.

The drugs category was the fifth most common and blackmail the lowest category for both datasets. To obtain more information on the offenders' crime mix patterns, both the offence categories which are likely to co-occur and also the membership of individuals to crime mix patterns, needs to be identified over the 40-year period. To be able to do this a latent class approach is needed.

Table 7.1 Number of OI offenders convicted of specific offences in each five year period (N=2267)

	Age Period							
Offence Category	**12-16**	**17-21**	**22-26**	**27-31**	**32-36**	**37-41**	**42-46**	**47-51**
Murder / Violence	84	230	177	111	86	63	36	20
Firearms	6	33	24	19	16	12	5	3
Authority	3	44	30	25	15	15	4	3
Sexual Offences	19	29	27	23	12	13	5	1
Blackmail	3	0	1	0	0	1	0	0
Robbery	11	33	14	10	8	7	2	0
Burglary / Theft	568	724	421	260	147	103	47	18
Fraud / Forgery	128	253	184	149	84	53	24	14
Criminal Damage	127	262	145	96	53	42	17	8
Drugs	2	66	86	69	46	43	19	7
Public Order	3	15	6	17	8	5	8	0

Note: Offenders can contribute to more than one offence category and to more than one age period

Table 7.2 Number of CCLS offenders convicted of specific offences in each five-year period (N=2267)

	Age Period							
Offence Category	**12-16**	**17-21**	**22-26**	**27-31**	**32-36**	**37-41**	**42-46**	**47-51**
Murder / Violence	99	539	510	343	300	236	192	94
Firearms	5	108	161	89	34	4	0	0
Authority	10	145	252	102	54	51	47	8
Sexual Offences	102	191	140	52	34	25	19	7
Blackmail	7	46	33	30	27	15	4	3
Robbery	19	135	115	67	72	41	29	13
Burglary / Theft	705	1260	866	562	396	293	208	85
Fraud / Forgery	82	297	278	220	155	104	78	32
Criminal Damage	87	394	275	177	159	109	85	34
Drugs	5	185	314	213	120	36	1	0
Public Order	43	257	154	104	65	48	41	23

Note: Offenders can contribute to more than one offence category and to more than one age period

Table 7.3 and Table 7.4 display the number of convictions occasions for each offence category and age period for the OI and the CCLS, respectively. Again, the majority of conviction occasions occur within the age period 17-21 years for both countries. The CCLS dataset has more conviction occasions overall for all offence categories except for Fraud and Forgery, where the OI dataset has a higher number. The OI has more conviction occasions for firearm offences in the older age periods between 37-51 years and again for drug offences. Interestingly, the OI also has higher number of conviction occasions for several of the offence categories in the first age period 12-16 including; Murder & Violence; Criminal Damage; Burglary & Theft and Fraud & Forgery.

Table 7.3 OI Number of conviction occasions per 5-year age period (N=2267 cases)

	Age Period								
Offence Category	**12-16**	**17-21**	**22-26**	**27-31**	**32-36**	**37-41**	**42-46**	**47-51**	***Total***
Murder / Violence	112	347	302	173	133	112	49	31	*1259*
Firearms	7	46	61	25	17	12	8	4	*180*
Authority	4	56	49	34	16	22	4	6	*191*
Sexual Offences	24	39	50	37	23	49	42	2	*266*
Blackmail	3	0	1	0	0	1	0	0	*5*
Robbery	13	37	33	12	10	8	2	0	*115*
Burglary / Theft	1227	1780	1421	647	395	394	172	70	*6106*
Fraud / Forgery	161	419	435	324	222	146	64	52	*1823*
Criminal Damage	153	382	291	160	88	78	32	15	*1199*
Drugs	2	96	145	121	112	79	48	21	*624*
Public Order	3	19	3	18	8	5	8	0	*64*
Total	*1709*	*3221*	*2791*	*1551*	*1024*	*906*	*429*	*201*	*11832*

Table 7.4 CCLS Number of conviction occasions per 5-year age period (N=2267 cases)

	Age Period								
Offence Category	**12-16**	**17-21**	**22-26**	**27-31**	**32-36**	**37-41**	**42-46**	**47-51**	***Total***
Murder / Violence	104	705	659	439	399	329	237	109	2981
Firearms	5	114	190	98	35	4	0	0	446
Authority	10	156	143	118	61	59	58	11	616
Sexual Offences	109	236	172	72	42	30	28	9	698
Blackmail	7	50	34	31	30	17	5	3	177
Robbery	20	150	128	76	86	56	31	13	560
Burglary / Theft	989	2494	1689	1082	789	558	354	131	8086
Fraud / Forgery	85	330	333	263	194	124	86	40	1455
Criminal Damage	89	462	331	209	192	132	99	37	1551
Drugs	5	229	447	315	158	43	1	0	1198
Public Order	45	306	180	120	67	54	51	27	850
Total	1468	5232	4306	2823	2053	1406	950	380	18618

7.4 Methodological approach

The latent Markov modelling approach used in this chapter will now be discussed. This will start off in Section 7.4.1 by describing Latent Class Analysis for binary indicators (that is, whether each of the 11 offence types have had a conviction in each of the eight age groups for each individual). This is proceeded in Section 7.4.2 with a description of Latent Markov Modelling for Binary Indicators. Then in Section 7.4.3 Latent Markov Modelling for count data (the count of the number of convictions for each offence type in each of the eight age-periods for each individual) will be explained.

Some early work on Latent Markov Modelling on this dataset has already been published, with myself as primary author (Elliott et al., 2017). This work focused on binary indicators and looked at only three age periods, giving two possible transitions. This chapter instead focuses on Poisson count data, and therefore uses more information from the dataset than used in the above paper.

7.4.1 Latent Class Analysis

As already discussed, the criminal careers of each offender are divided into five year time periods. Starting at age 12, which is the age of criminal responsibility in the Netherlands (age 10 in England and Wales). Five-year time periods allow a reasonable amount of time for an offender to accumulate offences so that an understanding can be gained of the varied nature of their offending, while still allowing for the offender to switch behaviour. Thus, any changes in criminal activity and pathways can be assessed as offenders get older, making it possible to check for periods of specialisation, versatility or non-offending. To be able to identify crime mix patterns in the data, latent Marvov modelling is used. The latent class analysis model is firstly described, and then extended to allow transitions to be included.

7.4.2 The Latent Class Model for binary indicators

A set of binary indicator variables were constructed to indicate whether there was a conviction for each of the 11 offence categories in each specific time period for each offender. The binary indicator had the value of 1 if a conviction for that offence category occurred and 0 if not. Within each age period, we look at which of the $J = 11$ offence types have had convictions to get a response pattern. The end data file produces a prevalence matrix of offence categories by person-time period strips.

Let O_{ijt} be the observed binary response for offender i and offence category j in time period t, with $O_{ijt} = 1$ if offender i has at least one conviction for offence category j

within time period t, and 0 if not. Let $\boldsymbol{O_{it}} = (O_{i1t}, O_{i2t}, \dots, O_{iJt})$ to be the vector of responses for each time period for each offender assuming there are K clusters or classes within the dataset.

Let class k $(k = 1 \dots K)$ have probability $\pi(k)$, and $p(\mathbf{O}_{it}|k)$ is the probability of observing the indicator vector O_{it} given member ship of class k.

The likelihood is then

$$L = f(\mathbf{O}) = \prod_{it} \sum_{k} \pi(k) p(\mathbf{O}_{it}|k)$$

And, under the assumption of conditional independence, assume that

$$p(\mathbf{O}_{it}|k) = \prod_{j} p_{jk}^{\boldsymbol{O}_{ijt}} (1 - p_{jk})^{1-\boldsymbol{O}_{ijt}}$$

is a product of Bernoullis, where $p_{jk}$is the probability that there is at least one conviction for offence category j in any time period, given that the offender belongs to class k.

The unknown parameters – p_{jk} and $\pi(k)$ – are obtained by finding the true maximum of the likelihood. As there are many unknown parameters to estimate, finding the true maximum of the likelihood becomes complex. By using 100 different start sets and taking the best solution, this ensured as far as possible that the global maximum of the likelihood is reached.

To be able to determine the optimal number of latent classes K, the Bayesian Information Criterion (BIC) statistic and Bayesian Information Criterion (BIC-ICL) are used. Both of these are discussed in more detail in Chapter 5. The BIC is based upon the likelihood and addresses the problem of over fitting by the addition of a penalty term based on both the number of parameters and the sample size.

Choosing the value of K that minimises the BIC, which is defined to be:

$$BIC = -2\log L + v\log(n)$$

where v is the number of parameters and n is the number of offenders.

After reaching model convergence, it becomes possible to obtain the posterior probabilities that an offender i belongs to one of the latent classes k in a specific time period t. This can be given by q_{ikt} where;

$$q_{ikt} = \frac{\pi(k)\prod_j (p_{jk})^{O_{ijt}}(1-p_{jk})^{1-O_{ijt}}}{\sum_{k=1}^{K}\pi(k)\prod_j (p_{jk})^{O_{ijt}}(1-p_{jk})^{1-O_{ijt}}}$$

and q_{ikt} provides different estimated probabilities for each offender for each time period.

Another model assessment criterion which has been mentioned in previous chapters is the ICL-BIC. This extends the BIC to take account of the posterior probabilities and whether they are close to 1 or not. In the Latent GOLD program, the same information obtained for the states is the same as for latent classes and for models with covariates the reported classification statistics will also contain information for the states. The computation of entropy is adapted for LMM. It is acquired by adding up the states and therefore disregards any dependencies between the latent variables at differing levels (Vermunt and Magidson, 2016).

The ICL-BIC is the BIC plus twice the entropy of the model:

$$ICL - BIC = -2\log L + v\log(n) + 2\sum_{ikt}\hat{q}_{ikt} log\hat{q}_{ikt}$$

where q represents of the number of parameters in the model and n represents the number of individuals or offenders. The entropy of a model is used in statistical learning methods , and asseses how close to 1 or 0 the $\hat{q}_{ikt}$ are.

7.4.3 Latent Markov modelling for binary indicators

As described above, Latent Class Analysis makes the assumption that the time periods within an individual's criminal history are independent, and this is an unrealistic assumption – there is likely to be dependence between an offenders criminal patterning at adjacent time points.

The Latent Class model is therefore extended to a Latent Markov model (LMM). Latent Markov modelling incorporates an extra set of unknown transition matrices, each of which gives the probability of membership of a latent class membership at one time point t (or more strictly age-period *t*) given latent class membership at the previous time point $t-1, (t = 2, \dots T)$ which are estimated along with the other parameters. LMM is based on Markov chain models (Kaplan, 2008, Langeheine and Van de Pol, 2002), and captures the discrete stages of individuals' movement of latent states through time. A consequence of that is that membership of a latent class at time t depends only on the observed data at time *t* and membership at time $t-1$, and not at any earlier time points.

Unlike the above LCA models, classes in Latent Markov models are usually known as latent *states.* In these Latent Markov models, transition probability parameters are added which estimate the probability of switching between the different latent states from time $t-1$ to time t.

As before, let O_{ijt} represent the observation of the $\boldsymbol{j}th$ indicator variable of interest at time point t for an individual i. Also let $\boldsymbol{O_{it}} = \left(O_{i1t}, O_{i2t}, \dots, O_{iJt}\right)$ to be the vector of the *j* responses for each time period. Now the matrix $\boldsymbol{O_i} = (\boldsymbol{O_{i1}}, \boldsymbol{O_{i2}}, \dots, \boldsymbol{O_{it}})$ is set to be the complete set of indicator responses over the T time periods.

Let s_t denote a state of the latent variable at time point *t,* where $1 \leq s_t \leq K$. Let $T(s_0)$ be the initial state probabilities, and let $T(s_t|s_{t-1})$ represent the latent transition probabilities from time $(t-1)$ to time t and $p(\boldsymbol{O_{it}}|s_t)$ is the probability of observing the indicator vector O_{it} at time t given member ship of state s_t.

The LMM model is therefore;

$$P(\boldsymbol{O}_i) = \sum_{s_1=1}^{K}\sum_{s_2=1}^{K}\ldots\sum_{s_T=1}^{K} T(s_0)\prod_{t=2}^{T} T(s_t|s_{t-1})\prod_{t=1}^{T} p(\boldsymbol{O}_{it}|s_t).$$

with the likelihood equal to

$$L = \prod_{i=}^{n} P()$$

again, assuming conditional independence, and, as before set

$$p(\boldsymbol{O_{it}}|s_t) = \prod_{j} p_{js}^{\boldsymbol{O}_{ijt}}(1-p_{js})^{1-\boldsymbol{O}_{ijt}}$$

where $p_{js}$is the probability that there is at least one conviction for offence category j in any time period, given that the offender belongs to state s. Note that the p_{js} do not vary over time – the latent states remain static and do not change definition – it is the offender who is able to switch between these static latent states and provides change in offending patterns.

The Latent Markov model, therefore, needs to estimate three sets of probabilities; the initial state probabilities *T(s*$_0$*)*, the *T-1* sets of transition probabilities $T(s_t|s_{t-1})$ and the p_{js}.

7.4.4 Latent Markov modelling for Poisson counts

It is now possible to modify the above development for binary data, replacing it to be a series of counts. Let O_{ijt} represent the count of the number of conviction occasions for the $\boldsymbol{j}th$ crime type variable of interest, at time point t, for an individual i.

Then as before, $\boldsymbol{O_{it}} = \left(O_{i1t}, O_{i2t}, \dots, O_{iJt}\right)$ is the vector of the *j* responses for each time period. Similarly 0_{i}, is the matrix of all counts over all crime types and time periods and is defined as before. Similarly, the probability of observing $\boldsymbol{O_i}$ can be written as

$$P(\boldsymbol{O}_i) = \sum_{s_1=1}^{K} \sum_{s_2=1}^{K} \dots \sum_{s_T=1}^{K} T(s_0) \prod_{t=2}^{T} T(s_t|s_{t-1}) \prod_{t=1}^{T} p(\boldsymbol{O}_{it}|s_t).$$

The only change is the definition of $p(\boldsymbol{O}_{it}|s_t)$, which becomes a product of Poisson probabilities:

$$p(\boldsymbol{O_{it}}|s_t) = \prod_j \frac{e^{-\lambda_{js}} \lambda^{O_{ijt}}}{O_{ijt}!}.$$

So for a count data LMM, a set of profile mean rates is estimated, one for each latent state, for each of the crime types. These again can be used to define the latent states. It is possible to look at both the relative sizes of the λ's within a latent state, and the differential sizes of the λ's across latent states. As before, the model requires estimation of both the initial class sizes and the transition matrices giving movement between states.

7.4.5 Missing data and Latent Markov modelling

Finally, missing data needs to be discussed. A full information maximum likelihood approach (Enders and Bandalos, 2001) is taken in the estimation of parameters in the presence of partially observed count sequences. Typically, these will occur at the ends of sequences. Thus, for the OI data, those in the 1968 birth cohort will only be followed up to age 40, and the counts in the last two age periods will be missing.

Full information maximum likelihood modifies the likelihood so that the calculation for $P(\boldsymbol{O}_i)$ includes only those age periods which are observed. This means that the assumption of the missing process is missing at random (Graham, 2009). This seems reasonable in these datasets as the missingness is caused by design and not by an offender refusal.

7.4.6 Software issues for latent Markov models

There are four main software packages that can fit Latent Markov models. PROC LTA (Latent Transition Analysis) is an add-on to SAS written by members of the Methodology Center at Penn State University (Lanza et al., 2007). It is however restricted in the type of data that can be fitted, and can only deal with binary observations.

The MIXTURE comment in MPLUS (Muthén and Muthén, 1998-2012) is also able to fit latent Markov models (known as Latent Transition Analysis in the package). Data of various types is allowed, but there are specification problems if users want to move beyond two or three time periods. The clumsiness of the code specification needed also make this possibility unrealistic for the data.

The real choice to make comes between the two remaining solutions – the library LMest in R (Bartolucci and Pandolfi, 2016) – described as Latent Markov models with and without covariates - and the use of Latent Gold 5.1 (Vermunt and Magidson, 2016). Both are recent additions to their respective packages, with the library LMest being developed in 2014 and the Latent Markov facility on Latent Gold being added around the same time. The routine est_LM_basic is the basic user function. While it is able to fit an arbitrary number of time periods, it is unable to deal with count data. In addition, the user interface is less straightforward to use. Therefore, Latent Gold is the chosen software package for these models.

7.5 Results

The results of fitting the latent Markov model to each dataset are reported below in turn. Firstly, a sequence of models were fitted to determine the optimal number of states for the LMM analysis. Table 7.5 shows the log-likelihood and associated BIC and ICL-BIC values for various latent Markov models for the OI and CCLS datasets specifying from one state up to eight states.

The BIC values in both datasets continue to decline as the number of states increases, and do not reach a minimum. However the ICL-BIC values do provide information on the optimal number of states. The lowest ICL-BIC value was achieved for the 5 state model for the OI dataset – however a 3 state model minimised the ICL-BIC for the CCLS dataset. As it was necessary for comparative purposes to specify the same number of groups for both countries, the 5 state model was chosen as the optimal solution for both datasets. This was similar to previous analyses performed on this dataset (Elliott et al., 2017), where a 5 state model was also chosen as the most parsimonious model when using the indicators as binary variables instead of counts of conviction occasions.

Table 7.5 Log- Likelihood BIC and ICL-BIC statistics for various latent Markov models.

Latent Markov Models	Log-likelihood	-2 log likelihood	BIC	ICL-BIC
Offenders Index				
1 states	-43262.2	86524.31	86609.3	86609.3
2 states	-43262.2	86524.31	86609.3	65079.02
3 states	-31691.5	63382.96	63591.57	66027.41
4 states	-29532.2	59064.3	59427.43	64847.33
5 states	-28465.9	56931.76	57480.32	***63935.99***
6 states	-27685.4	55370.86	56135.76	67291.71
7 states	-27203.1	54406.19	55418.32	65977.51
8 states	-26688.3	53376.63	54666.91	65270.46
CCLS dataset				
1 states	-57581.1	115162.3	115247.3	115247.3
2 states	-47318.4	94636.75	94845.36	98757.09
3 states	-45786.1	91572.12	91935.25	***101312.9***
4 states	-45046.8	90093.64	90642.2	101481.3
5 states	-44449	88897.99	89662.89	102098.8
6 states	-44049.3	88098.54	89110.67	103297
7 states	-43708	87416.04	88706.31	104353.2
8 states	-43425.5	86851.01	88450.34	104570.8

Note: The model that minimises the ICL-BIC is given in bold italic.

Table 7.6 Offenders Index estimated mean rates over five years of an offender in each state for each offence category

Offenders Index	State				
	1	2	3	4	5
Murder / Violence	0.0134	0.059	0.7117	0.7783	0.0624
Firearms	0.0006	0.0043	0.103	0.1614	0.0583
Authority	0.0015	0	0.1187	0.1708	0.031
Sexual Offences	0.0031	0.0079	0.007	0.0882	0.5939
Blackmail	0.0001	0.0012	0.0011	0	0
Robbery	0.0004	0.0134	0.0383	0.0835	0.0421
Burglary / Theft	0.0075	0.9695	0.402	9.0233	0.8663
Fraud / Forgery	0.0051	0.1628	0.1038	1.2354	2.9419
Criminal Damage	0.0038	0.1203	0.5533	0.9774	0.0204
Drugs	0.0043	0.0017	0.4907	0.3053	0.0498
Public Order	0.0011	0.0028	0.0362	0.0178	0.0068

Note: Figures in lighter shading have rates greater than or equal to 0.1 but less than 0.5; those in darker shading are greater or equal to 0.5

Table 7.7 Interpretation of the Offenders Index five latent states based on the estimated mean rates

Offenders Index	LM States
State 1	Non-offending / Low offending
State 2	Low Burglary & Theft offending
State 3	Violence, Drugs and Criminal Damage offending
State 4	Versatile and more serious offending
State 5	Fraud Forgery & Sexual Offences

Table 7.8 CCLS estimated mean rates over five years of an offender in each state for each offence category

CCLS	State				
	1	2	3	4	5
Murder / Violence	0.0255	0.2796	0.0304	1.0033	0.0977
Firearms	0.0005	0.0301	0	0.1398	0.097
Authority	0.0045	0.0632	0.0031	0.1945	0.033
Sexual Offences	0.0033	0.1204	0.0416	0.0817	0.0067
Blackmail	0.0004	0.0067	0.0021	0.0813	0.0156
Robbery	0	0.0132	0	0.319	0.0174
Burglary / Theft	0.0156	0.3846	0.4741	2.2344	1.6258
Fraud / Forgery	0.0083	0.0854	0.0405	0.3867	0.3104
Criminal Damage	0.0066	0.1763	0.0334	0.4733	0.0376
Drugs	0.0031	0.0302	0.0001	0.177	0.6499
Public Order	0.0041	0.0848	0.0137	0.2908	0.0216

Note: Figures in lighter shading have rates greater than or equal to 0.1 but less than 0.5; those in darker shading are greater or equal to 0.5

Table 7.9 Interpretation of CCLS five latent states based on the estimated mean rates

CCLS	LM States
State 1	Non-offending / Low offending
State 2	Low offending with Burglary & theft, Violence, Sexual Offence and Criminal Damage
State 3	Low Burglary & Theft offending
State 4	Versatile and more serious offending
State 5	Drugs and Burglary & Theft offending

7.5.1 Interpreting the latent states

Table 7.6 and Table 7.8 show the profiles of the five latent states the OI and CCLS datasets respectively. Formally, the tables contain the five-yearly mean rates for an offender in one of the latent states s for each offence category j. It is immediately apparent that some latent states have higher mean rates for particular offences than for others.

To understand these tables, the following example is given. The estimated mean rate that an OI offender in state 2 having the offence ‘Burglary & Theft’ is 0.9695; suggesting that most of the offenders assigned to state 2 will have on average at least one conviction in the ‘Burglary & Theft’ category. The next highest mean rate for state 2 was for ‘Fraud & Forgery’, with a mean rate of 0.16 for this offence. The rest of the offence categories in state 2 had mean rates of 0.12 or lower. Therefore, state 2 is considered a specialist ‘Burglary & Theft’ offending group. Even though there are a number of offenders who have specialised in ‘Burglary & Theft’ offending, there are still many offenders for whom ‘Burglary & Theft’ offences are just one in a wide range of different types of convictions within the five-year period.

Thus, using these mean rates, the five latent states in Table 7.7 and Table 7.9 were given a label that best summarised the crime mix of offences in each state. These labels should only be used as an indication of the crime mix but these are as a short-hand indication of the profile of the latent state.

For the Offenders Index dataset, two of the latent states can be considered to be specialised (state 2 - ‘Low Burglary and Theft’ and state 5 - ‘Fraud & Forgery and Sexual Offences’). State 2 – ‘Low Burglary and Theft’ has very low mean rates for all offence categories except for ‘Burglary & Theft’ which has a mean rate of 0.97, which is at least 0.8 higher than all other offence categories. This suggests that offenders belonging to this state are most likely to have a ‘Burglary & Theft’ conviction over all other offence categories.

Interestingly state 5 is a ‘Fraud & Forgery and Sexual Offences’ offending group, this is a rather specialised state, containing higher mean rates for some of the more rarer types of offences and would probably be more common with older offenders. The offence category for ‘Fraud & Forgery’ has the highest mean rate at almost 3. Also, state 5 is the only latent state that has the highest mean rate for ‘Sexual Offences’ out of all the other latent states. Therefore, offenders belonging to state 5 are most likely to have convictions for ‘Fraud & Forgery’ and/or ‘Sexual Offences’.

However, there are also states that are more versatile (state 3 – ‘Violence, Drugs and Criminal Damage’ and state 4 – ‘Versatile and more serious offending’) having higher mean rates in several offence categories. Particularly, state 4 - ‘Versatile and more serious offending’ has the highest mean rates with over 9 for ‘Burglary & Theft’. Offenders belonging to these states are therefore more likely to have several convictions over a wider range of offence categories and be involved in more serious crimes.

The CCLS dataset have two states that can be considered specialist (state 3 – ‘Burglary and Theft’ and state 5 – ‘Drugs and Burglary & Theft’). State 3 ‘Burglary and Theft’, overall has very low mean rates for all offence categories except for ‘Burglary & Theft’ which has a mean rate of almost 0.5. This is at least 0.43 higher than all other offence categories, suggesting that offenders belonging to this state all have very low rates of conviction for most offence categories but with a higher chance of receiving a conviction for ‘Burglary & Theft’ in a 5 year period. State 5 – ‘Drugs and Burglary & Theft’, has a mean rate of 0.65 for ‘Drugs’, which is the highest mean rate out of all other states for this offence. It also has a mean rate of 1.6 for ‘Burglary & Theft’ suggesting that offenders in this state are likely to convicted of ‘Drugs’ and/or ‘Burglary & Theft’ offences at a higher rate than for other offence categories.

The CCLS also has versatile states, a low offending (state 2 – ‘Low Burglary & Theft, Violence, Sexual and Criminal Damage offending’) and a higher more serious offending group (state 4 – ‘Versatile and more serious offending’). The remaining state – state 1 – is a very low offending and non-offending group with very low rates of conviction for all offences.

7.5.2 Transition probabilities

It is expected from extensive research on the age-crime curve discussed in Chapter 2 that the majority of offenders would be actively offending during the second age period 17–21 years (Farrington, 1986; Hirschi and Gottfredson, 1983; Moffitt, 1993). Therefore, it is predicted most offenders will be at some point in the 'non-offending/low offending' group, most likely at age 12–16 years and again in one of the later age periods. However, some offenders will start their criminal careers at a much earlier stage and it is predicted that these offenders will continue offending as they grow older and become involved in more serious and varying offences (Mazerolle et al., 2000; Moffitt, 1993; Piquero et al., 2007; Piqueroet al.,1999). More recently, it has been posited that late bloomers or adult onset offenders exist (Krohn et al., 2013). These issues will be examined through the investigation of the estimated transition matrices in each of the jurisdictions.

Table 7.11 and Table 7.13 show the estimated transition probabilities for a time heterogeneous Markov model (that is, the transition probabilities vary over time) for the OI and CCLS datasets. They describe how the offending behaviour changes from each age period starting with age 12–16 years right through to age 47-51 years. The Latent Markov model also estimates initial state probabilities (class sizes) for the age period 12-16, and using the above transition probabilities it is possible to also estimate the state probabilities for all other age groups (see Table 7.10 and 7.12).

As before, let s_t denote a state of the latent variable at time point *t,* where $1 \leq s_t \leq K$. Let $T(s_0)$ be the initial state probability for state s, and let $T(s_t|s_{t-1})$ represent the latent transition probabilities from state s_{t-1} at time $(t-1)$ to state $s_t$at time *t*. Then $T(s_{t-1})$ is defined to be the state probability for state s_{t-1} at time *(t-1).*

The state probabilities $T(s_t)$ for state s_t are therefore given by the following recursive expression (in matrix form):

$$T(s_t) = T(s_{t-1})^T \; T(s_t|s_{t-1}).$$

Both the initial state probabilities and the estimated state probabilities for later age periods are presented in Table 7.10, Table 7.11 and 7.12 for the OI and the CCLS datasets respectively.

Table 7.10 OI estimated state membership probabilities from one age period to the next

	State	1	2	3	4	5
Age Period	12-16 years	0.6068	0.3674	0.0046	0.0193	0.0019
	17-21 years	0.4479	0.3898	0.0997	0.0414	0.0213
	22-26 years	0.5833	0.2473	0.1086	0.0305	0.0304
	27-31 years	0.7326	0.1343	0.0872	0.0179	0.0282
	32-36 years	0.8519	0.0615	0.0572	0.0100	0.0196
	37-41 years	0.9267	0.0244	0.0311	0.0057	0.0120
	42-46 years	0.9657	0.0089	0.0147	0.0034	0.0073
	47-51 years	0.9835	0.0032	0.0065	0.0021	0.0046

Note: 1. Figures in lighter shading are greater than or equal to 0.1; those in darker shading are greater or equal to 0.5.
2. The 12-16 probabilities are estimated directly from the model as T(s_j).
3. The 17-21 through to 47-51 age period probabilities are calculated from the T(S_j) and the transition probabilities.

Table 7.11 OI transition probabilities from one age period to the next

(a) Transition probabilities from 12-16 to 17-21 years

		17-21 years				
	State	1	2	3	4	5
12-16 years	1	0.488	0.3949	0.0903	0.0011	0.0257
	2	0.4078	0.3957	0.1091	0.0774	0.01
	3	0.282	0.0183	0.6813	0.0035	0.0149
	4	0.0158	0.2416	0.0782	0.6283	0.0361
	5	0.1935	0.0168	0.0681	0.0582	0.6634

(b) Transition probabilities from 17-21 to 22-26 years

		22-26 years				
	State	1	2	3	4	5
17-21 years	1	0.7204	0.1842	0.069	0.0014	0.025
	2	0.5515	0.388	0.033	0.0139	0.0136
	3	0.3717	0.0206	0.5893	0.0061	0.0123
	4	0.033	0.263	0.0989	0.569	0.0361
	5	0.3391	0.0288	0.0935	0.015	0.5236

(c) Transition probabilities from 22-26 to 27-31 years

		27-31 years				
	State	1	2	3	4	5
22-26 years	1	0.8659	0.07	0.0429	0.0015	0.0198
	2	0.6445	0.3288	0.0086	0.0022	0.0159
	3	0.4694	0.0223	0.4884	0.0102	0.0097
	4	0.0669	0.2774	0.1211	0.4995	0.035
	5	0.4996	0.0416	0.108	0.0033	0.3476

(d) Transition probabilities from 27-31 to 32-36 years

		32-36 years				
	State	1	2	3	4	5
27-31 years	1	0.9366	0.0239	0.024	0.0014	0.0141
	2	0.7153	0.2646	0.0021	0.0003	0.0177
	3	0.5666	0.023	0.3869	0.0162	0.0073
	4	0.1295	0.279	0.1414	0.4179	0.0323
	5	0.6388	0.0521	0.1083	0.0006	0.2003

(e) Transition probabilities from 32-36 to 37- 41 years

		37-41 years				
	State	1	2	3	4	5
32-36 years	1	0.9685	0.0078	0.0128	0.0012	0.0096
	2	0.7729	0.2073	0.0005	0	0.0192
	3	0.6543	0.0227	0.2931	0.0246	0.0053
	4	0.2328	0.2608	0.1535	0.3251	0.0277
	5	0.7385	0.0589	0.0981	0.0001	0.1043

(f) Transition probabilities from 37-41 to 42-46 years

		42-46 years				
	State	1	2	3	4	5
37-41 years	1	0.9833	0.0025	0.0067	0.0011	0.0064
	2	0.8199	0.1595	0.0001	0	0.0205
	3	0.7256	0.0215	0.2133	0.0358	0.0037
	4	0.3786	0.2205	0.1507	0.2287	0.0215
	5	0.8026	0.0627	0.0836	0	0.0511

(g) Transition probabilities from 42-46 to 47-51 years

		47-51 years				
	State	1	2	3	4	5
42-46 years	1	0.9905	0.0008	0.0035	0.0009	0.0043
	2	0.8575	0.1209	0	0	0.0215
	3	0.7773	0.0197	0.15	0.0505	0.0025
	4	0.5459	0.1653	0.1312	0.1427	0.0148
	5	0.8426	0.0645	0.0688	0	0.0242

Table 7.12 CCLS estimated state membership probabilities from one age period to the next

	State	1	2	3	4	5
Age Period	12-16 years	0.2729	0.0308	0.6684	0.0269	0.0009
	17-21 years	0.0305	0.4978	0.1217	0.2084	0.1415
	22-26 years	0.2930	0.3778	0.0008	0.1632	0.1652
	27-31 years	0.5135	0.2509	0.0040	0.1087	0.1228
	32-36 years	0.6692	0.1734	0.0133	0.0794	0.0645
	37-41 years	0.7686	0.1249	0.0288	0.0550	0.0227
	42-46 years	0.8274	0.0849	0.0551	0.0261	0.0063
	47-51 years	0.8467	0.0418	0.1462	0.0116	0.0014

Note: 1. Figures in lighter shading are greater than or equal to 0.1; those in darker shading are greater or equal to 0.5.
2. The 12-16 probabilities are estimated directly from the model as T(s_j).
3. The 17-21 through to 47-51 age period probabilities are calculated from the T(S_j) and the transition probabilities.

Table 7.13 CCLS transition probabilities from age period to the next

(a) Transition probabilities from 12-16 to 17-21 years

		17-21 years				
	State	1	2	3	4	5
	1	0	0.427	0.4458	0	0.1272
	2	0.5162	0.4837	0	0.0001	0
12-16 years	3	0.0216	0.5381	0	0.2882	0.1521
	4	0.0019	0.2487	0.0003	0.5858	0.1633
	5	0.0834	0.0467	0.0005	0.015	0.8544

(b) Transition probabilities from 17-21 to 22-26 years

		22-26 years				
	State	1	2	3	4	5
	1	0.7518	0	0	0	0.2482
	2	0.4974	0.5024	0	0.0002	0
17-21 years	3	0.0081	0.5391	0	0.2882	0.1647
	4	0.0044	0.2606	0.0015	0.5912	0.1423
	5	0.1453	0.0547	0.0032	0.0343	0.7625

(c) Transition probabilities from 22-26 to 27-31 years

		27-31 years				
	State	1	2	3	4	5
	1	1	0	0	0	0
	2	0.4785	0.5209	0	0.0006	0
22-26 years	3	0.003	0.5349	0.0002	0.2853	0.1766
	4	0.0099	0.2702	0.0069	0.5903	0.1228
	5	0.2309	0.0585	0.0175	0.0719	0.6211

(d) Transition probabilities from 27-31 to 32-36 years

32-36 years

	State	1	2	3	4	5
27-31 years	1	1	0	0	0	0
	2	0.4593	0.539	0	0.0017	0
	3	0.0011	0.5268	0.0036	0.2804	0.188
	4	0.0216	0.2721	0.0308	0.5726	0.1029
	5	0.3104	0.0529	0.0812	0.1275	0.4279

(e) Transition probabilities from 32-36 to 37- 41 years

		37-41 years				
	State	1	2	3	4	5
32-36 years	1	1	0	0	0	0
	2	0.4392	0.5556	0.0005	0.0047	0
	3	0.0004	0.4837	0.0723	0.257	0.1866
	4	0.0429	0.249	0.1251	0.5046	0.0783
	5	0.3062	0.0351	0.2766	0.1658	0.2163

(f) Transition probabilities from 37-41 to 42-46 years

		42-46 years				
	State	1	2	3	4	5
37-41 years	1	1	0	0	0	0
	2	0.4086	0.5572	0.0214	0.0127	0
	3	0.0001	0.1932	0.6237	0.1024	0.0806
	4	0.0643	0.1719	0.3834	0.3354	0.0449
	5	0.1897	0.0146	0.5917	0.1354	0.0687

(g) Transition probabilities from 42-46 to 47-51 years

		47-51 years				
	State	1	2	3	4	5
42-46 years	1	1	0	0	0	0
	2	0.2033	0.2988	0.4795	0.0184	0
	3	0	0.014	0.9724	0.0074	0.0063
	4	0.0588	0.0724	0.717	0.1361	0.0157
	5	0.0772	0.004	0.8318	0.0727	0.0143

7.5.3 Transition probabilities in the OI England and Wales Dataset

To interpret the movements between states, both the transition probabilities and the state probabilities are examined.

The initial state values at age 12–16 years (the first row of Table 7.10) give the estimated sizes in terms of probability of the five states in the 12–16 age groups. The first state – the ‘Non-offending/Low offending’ group – is the largest group at age 12–16 for the OI dataset (p=0.6068). Examination of the first transition matrix In Table 7.11 gives the transition probabilities from 12-16 to the next age group (17-21). It can be seen that 49% remain in this ‘non-offending/Low offending’ group (p=0.488) and 39% move to state 2 the ‘Low Burglary & Theft’ specialist group (p=0.3949). The second transition from 17-21 to 22-26 (Table 7.11 b) – shows that the majority of offenders in state 1 at age 17-21 remain in state 1 (72%) and only 18% move to state 2. For the fourth age group (Table 7.10) it can be seen that there are more offenders in state 1 than at the initial state sizes at age 12-16. The size of state 1 continues to increase at each subsequent age period as more offenders from other active offending states transit into state 1.

Offenders who begin in state 2 at ages 12-16 – ‘Low Burglary and Theft’ - are in the second largest state initially. The majority of those in state 2 tend to move into state 1 with a slightly smaller proportion remaining in state 2 and a few moving into state 3 – ‘Violence, Drugs and Criminal Damage’. Through subsequent transition periods, more offenders move into state 1 and less remain in state 2 and even less move to other states. State 2 remains the second largest group at each age period, decreasing in size each time till age 37-41 years, where state 3 becomes the second largest group.

The majority of offenders beginning in state 3 - ‘Violence, Drugs and Criminal Damage’, tend to mostly stay in state 3 but the proportion decreases with the age period. Those offenders who transit out of state 3, tend to move to state 1 (essentially non-offending), and this increases with each age period.

The offenders who begin in state 4 – ‘Versatile and more serious offending’, tend to stay in state 4. However, this also decreases with each age period. The offenders who do move from state 4 tend to move to state 2 and then to state 3. Only by age

32-36 years do offenders move from state 4 into state 1 and by the final transition over half of the offenders remaining in state 4 move into state 1; although numbers are very small.

State 5 – 'Fraud & Forgery and Sexual Offences', like states 3 and 4, has the majority of its offenders remaining in state 5 for the first two transitions. Those offenders who do move from state 5 tend to move into state 1. By the third transition, more offenders start to move into state 3 and this continues for the following transition. By the age period 42-46 years, not many offenders continue in state 5 and most move into state 1.

7.5.4 Transition Probabilities in the CCLS Netherlands Dataset

Unlike the OI dataset, the first state – the 'Non-offending/Low offending' group – is not the largest group at age 12–16 for the CCLS (first row of Table 7.12) and only contains 27% of offenders. Instead the largest group is actually state 3 - 'Low Burglary and Theft offending' - which contains 67% of offenders. State 3 is very similar to state 1 but with a higher mean rate for burglary and theft offending. In other words, the two largest states initially are the lowest offending or non-offending groups.

Examination of the first transition matrix in Table 7.13 shows that, as expected, these offenders in state 1 transition into one of the other offending groups for the age period 17–21 years. In fact, zero offenders remain in state 1, the majority of offenders move into state 3 the 'Low Burglary & Theft offending' group (45%), and a similar proportion of offenders (43%), move into state 2 the 'Low versatile offending Burglary & Theft, Violence, Sexual and Criminal Damage' group. The second transition (Table 7.13 b) shows the majority of offenders move from state 1 to state 5 'Drugs, Burglary & Theft' group. For the third transition onwards all offenders in state 1 remain in state 1 for every age period. Similarly to state 1 in the OI dataset, the size of state 1 increases at each subsequent time period.

Offenders who begin in state 2 - ‘Low versatile offending Burglary & Theft, Violence, Sexual and Criminal Damage’, at ages 12-16 are a very small state group initially. Around half of those in state 2 tend to move into state 1, with almost half remaining in state 2. This is a similar pattern throughout the remaining transitions with slightly less offenders moving into state 1 each time and a slight increase with those offenders remaining in state 2. However, for the final transition at ages 44-46 years, nearly half of offenders in state 2 (48%) move into state 3 - ‘Burglary & Theft’. The final size of state 2, again, is very small.

For the offenders who begin in state 3 – ‘Burglary & Theft’ at ages 12-16, are in the largest group initially (67% of all offenders). No offenders remain in state 3 for the first three transitions. Offenders in state 3 tend to follow a similar path for each transition till ages 42-46 years. Offenders are split with around half tending to transit into state 2 and the majority of the remaining offenders move into state 4 - ‘Versatile and more serious offending’ . These offenders who move into state 4 are showing evidence of escalation by continuing with offending and committing more serious offences. However, those that move into state 2 are possibly reducing their offending. By ages 42-46 years 62% of offenders actually remain within state 3, but the actual number of offenders is relatively small. By the final transition, 72% of offenders stay within in state 3 and there are more offenders from other states moving into state 3 for the last age period.

The offenders belonging to state 4 - ‘Versatile and more serious offending’ – are one of the smallest states at ages 12-16 years, containing only 2.7% of offenders. This state does increase in size for the second age period to 21% and then decreases with each subsequent transition. The majority of offenders in state 4 tend to stay and continue with offending, however, with each transition the proportion of offenders remaining decreases. Those offenders who transit out of state 4 is normally into state 2 or state 5 - ‘Drugs and Burglary & Theft offending’. By ages 42-46 this pattern changes and offenders tend to move into state 3 (38%) and less into state 2 (17%). In the final transition, only a small proportion of offenders remain in state 4 and the majority move into state 3 (72%).

The fifth state - ‘Drugs and Burglary & Theft’, at ages 12-16 years is the smallest state and contains only 0.1% of offenders to begin with. For the first 4 transitions, the majority of offenders in state 5 remain although this proportion does decrease each time. The offenders that do move from state 5 tend to move into state 1, and this proportion increases with each transition till ages 32-26, where some offenders have started to move into state 4. By ages 37-41 years, the offenders in state 5 tend to move into state 3 (28%) with some still moving into state 1 and state 4 and even less remaining in state 5. In the sixth transition into ages 42-46 years, nearly all offenders in state 5 move, with the majority into state 3. However, for the final age period, nearly all offenders in state 5 remain in state 5, with a few offenders moving into state 3.

7.5.5 Estimating the Most Likely Conviction Pathways

From the latent Markov models, the individual posterior probabilities of state membership for all time points together were examined to see what offending group each individual offender belonged to for each time period. These are essentially posterior probabilities of conviction pathways. This allows for examination of which transition patterns (pathways through the offending groups) were the most popular. There were a possible 390,625 pathways for offenders to take through the five offending groups. The OI offenders took 284 of these pathways and the CCLS offenders took 290 of these pathways.

The most common pathway for OI offenders was to start in state 1 the ‘Non offending

/ Low offending' group and remain in this group throughout all 7 transition periods. From exploratory analysis of the data discussed in Chapter 4, many of the OI offenders only have one recorded conviction which meant they would not necessarily transition into another offending group unless it was for a more serious or rare offence.

The most common pathway for the CCLS offenders was to start in state 3 - 'Low burglary & Theft offending' and to join state 2 - Low versatile offending Burglary & Theft, Violence, Sexual and Criminal Damage' in the next age period and then to move to state 1 - 'Non-offending/Low offending', for the remaining age periods. These pathways are somewhat different, however there is not much difference between state 1 and state 3 in the CCLS as they both have very low mean rates for all offence categories, except that state 3 has a higher mean rate of 0.47 for 'Burglary & Theft'. Both these pathways by the third age period (22-26 years) have offenders in state 1 'Non-offending/Low offending' group for the remainder of the age periods.

The second most common pathway is different for both countries. For the OI 14% (316) offenders begin in state 2 'Low Burglary and Theft offending' group and then move into state 1 for the remaining age periods.

For the CCLS 9% (200) offenders start in state 1 'Non-offending / Low offending' group then move into state 2 'Low versatile offending Burglary & Theft, Violence, Sexual and Criminal Damage' group and continue in this state for the third age period finally ending up in state 1 by the fourth age period. These offenders could be considered as following the age–crime curve. They start in the 'Non-offending/Low offending' group and move into the 'Low Burglary & Theft' group for age 17–21 years and then move back to the 'Non-offending/Low offending' group after adolescence / early adulthood. This also agrees with the Adolescent Limited (AL) offender typology suggested by Moffitt (1993) and explains why a significant proportion of offenders have taken this pathway.

The third most popular pathway for each country is considerably different. The third OI pathway shows offenders beginning in the first age period in state 1 - ‘Non-offending / Low offending’, and then moving into state 2 - ‘Low Burglary & Theft’ during 17–21 years and then back to state 1 for the last age period. Again, these offenders could be considered following the age-crime curve as described above.

The third CCLS pathway shows offenders starting in state 3 - ‘Low Burglary & Theft’ and then moving into state 2 - ‘Low versatile offending Burglary & Theft, Violence, Sexual and Criminal Damage’ during 17-21 years and continuing in this state for the next age period. Then this pathway changes to state 1 - ‘Non-offending / Low offending’ for the remaining age-periods.

Table 7.14 The three most common pathways for the Offenders Index and CCLS datasets

Dataset	**Transition Pattern** **12-16 → 17-12 → 22-26 → 27-31 → 32-36 → 37-41 → 42-46 → 47-51**	**Frequency**	**%**
Offenders Index	State 1→ State 1→ State 1→ State 1→ State 1→ State 1→ State 1→ State 1	408	18
	State 2→ State 1→ State 1→ State 1→ State 1→ State 1→ State 1→ State 1	316	14
	State 1→ State 2→ State 1→ State 1→ State 1→ State 1→ State 1→ State 1	239	11
Netherlands	State 3→ State 2→ State 1→ State 1→ State 1→ State 1→ State 1→ State 1	661	29
	State 1→ State 2→ State 2→ State 1→ State 1→ State 1→ State 1→ State 1	200	9
	State 3→ State 2→ State 2→ State 1→ State 1→ State 1→ State 1→ State 1	181	8

7.6 Conclusion

The purpose of this chapter was to examine the OI and CCLS conviction datasets, to identify any crime mix patterns and how offenders may switch between these as they age. The aim was to investigate if there were any common patterns of offending behaviour and if the probabilities of transition vary between the two countries. By examining the two longitudinal datasets of criminal conviction histories, containing information of the types of offences, several distinct crime mix patterns were found.

There are both similarities and differences between the two countries. One such similarity is the identification of a non-offending/low-offending group in each dataset which became the largest state by the final age period at 47-51 years. Evidence was also found of early onset and late onset offenders. Most offending and transitions occurred at 17-21 years and 22-26 years for both countries.

One difference between the two countries is that the makeup of the latent states differed. This meant the conviction pathways offenders took were also different between the two countries.

Another difference noted was discovered in the transition probabilities. In some of the CCLS transitions, there were very clearly defined pathways as the probabilities for some of the states were equal to 0 or 1. This meant 100% of all offenders belonging to those specific states either moved or remained for those particular age periods.

Evidence was found of specialised and versatile crime mix patterns over time for both datasets. Some offenders tended to stay in the same crime mix offending group over each five-year age period; whereas other offenders would switch groups over time. The make-up of the crime mix groups varied for each dataset along with the transition probabilities, showing differences between the two countries.

The results indicate that the models chosen to examine the datasets were suitable for revealing the patterns of offending. Using the ICL-BIC, it was decided that the five state model was the best fit for the datasets using the 11 offence categories observed at eight 5-year age periods.

The five latent states were easily interpretable and the sizes were estimated at each age period. From clustering the 11 offending categories into 5 crime mix offending groups, it was found that certain offences clustered together. For offenders in England and Wales, ‘Fraud and Forgery’ co-occurred with ‘Sexual Offences’, and another group consisting of ‘Murder and Violence’ and ‘Drugs’ and ‘Criminal Damage’ was also found. For offenders in the Netherlands, ‘Drugs’ and ‘Burglary and theft’ co-occurred and another group consisting of ‘Murder and Violence’, ‘Burglary and Theft’, ‘Sexual Offences’ and ‘Criminal Damage’ was also discovered.

Both the datasets have different offence compositions of their versatile offending groups. Some of these crime mix groups are common with other latent class results within criminology. Certain crime mix groups can be interpreted as specialised; both datasets have a specialised burglary and theft group but with differing mean rates for this offence. There are also very versatile crime mix groups identified with a number of different offence categories co-occurring.

Seven transition matrices for each sample were estimated, the first one from ages 12–16 to 17–21 years right through to 42-46 to 47-51 years. It was discovered that most offenders tend to begin offending in the 17–21 or 22-26 year age periods, usually followed by desistance for the next age periods. This is typical of the criminological literature on the age–crime curve and the ‘AL’ typology suggested by Moffitt (1993).

Both datasets have different common routes or pathways through the crime mix groups. Although there are 390,625 possible pathways that offenders can take through the crime mix groups, there are a few that display evidence of specialisation by continuing to offend in the same crime mix group for the next age periods. However, there are also some offenders who tend to stay in state 4 - ‘Versatile and more serious’, for many subsequent age periods which can be interpreted as specialised - as they are remaining in the same state - or versatile as they belong in an offending group which consists of many offences.

Onset age plays an important part in the crime mix pattern of an offender. As mentioned previously, most offenders start their criminal behaviour during late adolescence; however, it has been discovered that there is evidence for a small group of offenders who have a late onset age and continue offending throughout the later age periods. There is also evidence to show that early onset can lead to escalation; continuing to offend at a more versatile manner and commit more serious offences as they grow older.

The results show that specialisation can occur over the shorter term but can also be versatile over the longer term by changing crime mix offending groups, and agree with the results of Sullivan et al. (2006). Other research by (McGloin et al., 2007) also concur with this finding and found that offenders can display short term offending patterns consistent with specialisation but illustrate patterns of versatility over longer periods. This indeed is one of the benefits of using five-year age periods to assess criminal histories rather than summarising the full lifetime of convictions.

The changes in the patterns of offending behaviour could be due to the changes in local life circumstances. Events such as gaining employment, getting married or drug and alcohol abuse, according to (McGloin et al., 2007), all influence offending patterns of offence specialisation or versatility. The results contradict life course theories, which tend to assume that offenders become more specialised as they age (Lussier et al., 2017).

There are caveats to this work that need mentioning. Differences between jurisdictions may be caused by different recording practices and different criminal justice systems in the two countries, or, alternatively, may represent real differences in offending behaviour caused by the distinctive cultures, education and social systems in the two countries. These are difficult to disentangle without different research using different methodology. To take one example, the larger proportions of those convicted for violence offences in the Netherlands could be caused by real social differences, yet could also be due to their differing diversionary policies (with more non-violent offences in the Netherlands perhaps not being prosecuted but diverted into other disposals). This important question of what drives these differences will for the moment have to remain unanswered.

The model could be further improved by adjusting the transition matrices for the later age periods where there is less data available. One possibility here is to fix the later transition matrices to be equal. Another consideration would be to analyse male and female offenders separately, however, there are very few convictions for females; particularly in the later age periods.

8 DISCUSSION AND CONCLUSION

8.1 Introduction

The purpose of this thesis was to provide approaches to identify the changing crime mix patterns and pathways from two official conviction datasets collected from two European countries. By considering both the frequency and types of offences, a more in depth examination of the changing crime mix patterns between and within individuals over the life course was possible.

The thesis has contributed to the growing research on criminal careers and the advances in statistical methods and quantitative analysis in criminology. B-Splines have been introduced as an alternative method to using cubic polynomials in trajectory modelling, and Latent Markov Modelling has been examined as a technique for estimating both crime mix patterns and how individuals transition between the different crime mix groups over the life course. Using these statistical methods contributes to a new analysis of criminal career development.

This chapter will draw together the main findings from the results presented in the previous chapters. It will outline the key themes that have materialised from the analyses and discuss possible policy implications and the impacts they may have. Finally, the chapter will close with a discussion of the original contributions presented from this thesis and potential areas for further research.

8.2 Summary of results

The previous chapters have explored and compared the crime mix patterns and pathways of criminal careers from two official conviction datasets; The England & Wales Offender's Index and The Netherlands Criminal Careers and Life Course

Study. The results of the analyses have certainly emphasised the similarities and differences between the two countries. However, the findings have also highlighted the various patterns and pathways between individual offenders within each dataset.

In chapters 3 and 4, the two conviction datasets were introduced, providing a detailed account of how they were constructed. Considerable care was taken in aligning the two datasets so that the samples were comparable as far as possible. The descriptive findings from exploring the aligned datasets were then presented. A number of criminal career dimensions were defined and examined. These aided in describing the background of the offending behaviour patterns in each country. It was found that most offenders had their first conviction occasion between the ages of 12-20 years, which was considered the beginning of their criminal career. However, a number of offenders started their criminal careers at a later age and are considered ‘late starters’. The average length or duration of a criminal career differed for each dataset. Excluding the offenders with only one conviction occasion, offenders in the OI had an average duration of 12.8 years and the CCLS offenders had an average of 21.6 years. The majority of offenders had only 1-2 conviction occasions, however a small minority of offenders were revealed in both datasets who were responsible for many of the conviction occasions. As mentioned previously, this supports the findings by many criminology researchers. Male offenders were more likely to participate in offending and be reoffenders, in comparison to female offenders. However, it was found overall that the majority of the offenders in both datasets reoffended in the period under study.

The first set of analyses performed upon the datasets in Chapter 5, set out to discover if different groups of offenders following separate offending trajectories could be identified. The chapter began with a detailed discussion of three of the main methodologies that can be utilised for modelling trajectories. The justification for the final model chosen was presented and a Group-Based Trajectory model along with an extended version of the GBTM, the B-Spline Group-Based Trajectory model, were performed upon the datasets. Four distinct offending trajectory groups were discovered in each of the dataset samples. A four-class solution was chosen as the best fitting model after careful consideration. The model's information criterion was examined along with the consideration of other factors such as the aim of the research, past experience, and interpretability of the groups. Both datasets had trajectories that represented the age-crime curve that relate to the adolescent-limited typology of Moffitt's taxonomy. These adolescent limited trajectories were quadratic in shape, however not all the trajectories were polynomial in shape.

Another aim was to try and discover any trajectories that showed offenders following a recidivating pathway. Observable in both datasets were trajectory groups that were non-polynomial in shape. A high rate persistent trajectory group was discovered that showed offenders continuing to offend as they aged. However, as expected this was the smallest offending group in both datasets and supports the research that it is a small number of offenders who are responsible for the majority of crimes committed.

Following on from the results produced in Chapter 5, the posterior probabilities estimated from the models were then used to aid in predicting reconviction over two different time periods. Chapter 6 presented the results of using the trajectory group membership as a predictor in future reoffending. Using logistic regression, the likelihood of reconviction in 2 years and 10 years was assessed. This chapter tested if offending trajectories could be useful as predictors for future offending as well as discovering if offenders belonging to a declining trajectory were less likely to reoffend than those on an upward trajectory. Results showed that using trajectory group membership was a strong predictor for reconviction. Offenders belonging to the 'High rate persistent' group had the highest probabilities of reconviction especially compared to offenders belonging to a declining trajectory. The logistic regression models were better at estimating the reconviction probabilities in the two year period than the 10 year period.

To gain more detailed knowledge of crime mix patterns, the types of offences being committed needed to be incorporated into analyses. This allowed for the examination of any crime mix patterns changes that may have occurred over the life course. In Chapter 7, a more in depth analysis of the criminal careers of offenders was undertaken through a latent Markov modelling approach. This permitted for the identification of different crime mix patterns and how offenders transitioned between these over time. The results revealed a number of distinct crime patterns evident in both datasets. It was decided that a 5-state solution was the best fitting model and the latent states were interpretable. A ‘Non-offending/ Low offending’ group was common in both datasets and this became the largest state by the last age period, with most offenders transitioning into this group. As expected, the majority of offending occurred at 17-21 years and 22-26 years which was made evident by most transitions into the other four offending groups occurring within these age periods. However, there appears be early onset and late onset offenders which follow different patterns and pathways to the majority of other offenders.

Differences in the crime mix patterns were found between the two datasets. The make up of the latent states or crime mix offending groups varied, meaning that the transitions and possible conviction pathways would also vary. However, both specialised and versatile crime mix patterns were evident in the datasets. There were offenders that stayed within the same crime mix group throughout all age periods, so were therefore considered specialised. However, a small group of offenders would also stay within the same crime mix group from one age period to the next but this would be a versatile crime mix group. These offenders can be considered specialised as they stick with the crime mix group but will engage in a wide variety of offences.

Versatile crime mix patterns were less common but still discovered in the datasets.

These would see offenders switching between the different crime mix groups through the different age periods. A small proportion of offenders could be interpreted as escalating as they got older. These offenders would start off in a less serious crime mix group before transitioning into a more serious crime mix group as they aged. They also tended to have an early onset age.

The crime conviction pathways through the crime mix offending groups, tended to show that most offenders would be on a declining/desisting pathway. Most transitions into one of the four active offending groups occurred at the 2nd and 3rd age periods before moving into the 'non-offending/low offending' group for the remainder of age-periods. Almost all pathways that the offenders took ended with a transition into the 'non-offending/low offending' group by the final age period even for the high rate persistent offenders.

8.3 Implications for Policy

The research on criminal careers is highly valuable in influencing criminal justice policies. Understanding the various patterns of criminal behaviour and the causes, can immensely help when trying to develop effective crime reduction initiatives. The research undertook in this thesis and the results produced have implications for policy.

The discovery of various criminal patterns and pathways show that there is a need for a number of different crime prevention strategies are needed for different types of offenders. Criminal justice policies to reduce reoffending might be effective for one group of offenders but not for others, therefore policy makers should consider implementing a variety of crime reduction intiatives.

The findings show that the majority of offending occurs at late adolescence and early adulthood. Therefore it is important to target persistent offenders early in their criminal careers. Unfortunately, by the time they are identified as a persistent offender, it is normally too late. This is due to factors such as getting older and maturing, these offenders are already declining in their offending and desisting from crime. The general static theories of crime such as Gottfredson and Hirschi (1990) encourage the idea of crime preventative initiatives that are targeted at children before they reach adolescence. This therefore would help reduce the criminal propensity to offend.

Although, many offenders have an early onset age, the results also discovered a group of offenders who started offending at a later age and other offenders who continued to offend into later adulthood. As mentioned earlier, many prior studies on criminal careers has focused upon juvenile offenders, ignoring any criminal patterns by older offenders. Therefore policy makers should also be considering crime preventative strategies that are specifically targeted at adult offenders.

A preventative focus needs to be taken with the small group identified as life course persistent offenders. Preventative strategies should be targeted at early ages to reduce the effects of potential negative life circumstances which trigger their criminal behaviour and set them on a path of chronic offending. With regard to incarceration, it is possibly more beneficial to society if the identified life course persistent offenders were imprisoned. These offenders have begun committing crime already, therefore they are not influenced by rehabilitative techniques or positive changes in life circumstances such as job opportunities or marriage. This then prevents LCP offenders from committing further crime by incarcerating them, and is considered effective by reducing the number of offences.

Selective incarceration can only be effective if the small group of chronic persistent offenders are reliable identified early on. From the results in Chapter 5, the OI and

CCLS trajectories displayed a ‘High-rate persistent’ group which was also very similar to the adolescent limited trajectory groups between the ages of 12-17 years. Therefore, it is very important to be able to distinguish between these offending groups, otherwise the effectiveness of imprisonment policies will be decreased. Placing the other offending groups in prison would be considered a waste of resources as these offenders are on a declining trajectory after adolescence. The number of prevented offences per prisoner is then reduced, if already desisting offenders are imprisoned.

A number of criminal pathways were identified that showed offenders on escalating pathways. This highlights a need to do further research identifying risk factors which is necessary for public policy.

To effectively evaluate offender differences, there is a need to incorporate other data on individuals such as socio-economic status, family background, education and other personal characteristics.

8.4 Original Contributions of this Thesis

The research undertaken in this thesis has provided original contributions to the literature. Some of these contributions will now be discussed and can be summarised into three main areas; research design, methodology and cross national comparisons.

The research design undertaken for this study has been explained in previous chapters. Many aspects had to be carefully considered when comparing datasets from two different countries. From investigating the literature, it appears that no other studies have attempted to align two official conviction datasets from separate countries in the same way as this thesis. There are many challenges to overcome to ensure the data is valid and comparable. Very few studies directly compare official conviction data as it requires extensive recoding of the data, which can be very time consuming and complex. A number of strategies were utilised to carefully align the two datasets and to help overcome some of the problems encountered with having different datasets. In particular, the offence categories were examined in depth as this is where many of the differences were encountered between the two countries. It meant many categories were excluded to create common offence categories which could be compared and analysed. The research design offers a methodology for creating categories of offending that can be used in other jurisdictions. Although the dataset samples cannot be perfectly aligned, the research design has proved effective for the analyses undertaken and has shown that a great deal can be achieved using existing datasets.

This thesis also contributes the use of advanced statistical methods in criminology. In Chapter 5, an extension to the group-based trajectory model was demonstrated by incorporating the use of B-Splines. This proved to be a useful method for overcoming the potential upward ticks produced in the plotted trajectories and provided a much more flexible approach to modelling trajectories. In Chapter 7, latent Markov modelling was introduced as a method to examine crime mix patterns and pathways in the datasets. It proposed the use of Poisson distributed count data for examining both the crime mix and frequency of conviction occasions in a single analysis. This extends and advances the research undertaken in the previous published paper by Elliott et al. (2017), which only used binary indicators in the LMM. The method has great potential for examining a number of criminal career dimensions and this thesis encourages of the use of LMM on other data sources. Finally, this thesis has contributed to cross national research. As previously noted, cross national studies are rarely undertaken due to the difficulties they present. However, cross national research provides insightful and valuable knowledge on criminal behaviour that may challenge some of the existing theories of crime. Earlier, it was mentioned that an overwhelming amount of criminal career research has been carried out in the USA. This raises the important question of how relevant this research is for other countries, like those in Europe. This study has contributed to the criminal career research in European countries, and unlike many USA studies, conviction data instead of arrest data has been used. The research in this study has demonstrated that different jurisdictions are distinct in their criminal career behaviour. Despite there being numerous differences, it does not suggest that the offenders themselves are completely different. Interventions, the social environment, criminal justice policies and laws all contribute to these differences. Cross national studies that are based on conviction data (which is therefore based on different jurisdictions) should be carefully considered and interpreted. Researchers need to be wary in assuming results developed in one country can apply to offenders in another. Policies and interventions in one country may not necessarily work elsewhere and this thesis has explicitly shown that there are offending patterns unique to both England & Wales and the Netherlands. Criminal justice practitioners should therefore tailor crime reduction initiatives to ensure they are beneficial and cost effective.

8.5 Suggested Further Research

Although this thesis has attempted to address some of the limitations apparent in the current research on criminal careers, there are still areas that are in need of further investigation. Some of these areas will be discussed below.

It is recommended that analysis should be undertaken on criminal conviction datasets that include more recent offenders. The conviction datasets used in this thesis are of older offenders and include convictions from over 50 years ago, consequently the analyses presented here are somewhat historical. Changes in the law, criminal justice system, police service and the environment all have an impact on the patterns of criminal activity. Therefore, the patterns which have been discovered here should be tested to see if they are relevant to more recent offenders and to see if offending is stable. By comparing with more recent conviction data, any potential generational effects can also be examined.

The CCLS dataset is considerably smaller than the OI, and the Netherlands would benefit from a larger and more recent sample of offenders. A larger sample would allow for the robustness of the dataset to be assessed. The OI has potential for this already, it would involve dividing the data samples and fitting models to them separately. The assessment of the model fits can then be compared.

There is also potential for further statistical methodology development. More research is needed on latent Markov modelling. It would beneficial to extend the use of the Poisson distribution for counts in the negative binomial distribution, this would allow for overdispersion that occurs when using count data. Furthermore, the model needs to incorporate the possibility of fixing the parameter in one of the latent states to define this as a non-offending group. This would allow offenders to move in and out of this non-offending group for periods of intermittency.
The results discovered conviction pathways that appeared to show offenders moving into more serious or less serious crime mix offending groups. This suggested that offenders were escalating or de-escalating in their offending behaviour. Measure of crime seriousness into the latent states should be incorporated so that movement between the states can be defined as escalation or de-escalation. This would contribute to the literature on escalation that has been investigated previously using the OI conviction data in Francis and Liu (2015).

Finally, there is scope for examining long term recidivism in more depth. It would be beneficial to discover if certain actuarial risk factors fade over time or still stay important. For example, would early onset still be important if an offender has not reoffended in the first 5 years following release or conviction? A dynamic survival model with time-varying risk effects could examine the effects of long term recidivism. New methods of assessing risk would be highly valuable to policy makers and law enforcement agencies.

It is hoped that this thesis has demonstrated the value of comparative research and encourages the use of more sophisticated methodologies to assess patterns and pathways of criminal careers on other datasets.

BIBLIOGRAPHY

Altman D, Bland J. 1994. Statistics Notes: Diagnostic tests 1: sensitivity and specificity *BMJ,* 308 :1552

Agnew, R. 2006. *Pressured into crime: An overview of general strain theory.,* Los Angeles, Roxbury.

Aitkin, M., Vu, D. & Francis, B. 2015. A new Bayesian approach for determining the number of components in a finite mixture. *METRON,* 73**,** 155-176.

Akers, R. 1999. *Criminological Theories: Introduction and Evaluation*, Taylor & Francis Group.

Akers, R. & Sellers, C. 2004. *Criminological theories: Introduction, evaluation, and application.,* Los Angeles, Calif, Roxbury Pub.

Andersson, F., Levander, S., Svensson, R. & Levander, M. 2012. Sex differences in offending trajectories in a Swedish cohort. *Criminal Behaviour and Mental Health,* 22**,** 108-121.

Armstrong, T. & Britt, C. 2004. The Effect of Offender Characteristics on Offense Specialization and Escalation. *Justice Quarterly,* 21**,** 843-876.

Baker, T., Falco Metcalfe, C. & Jennings, W. 2013. What Are the Odds?: Predicting Specialization in Offending Over the Life Course. *Criminal Justice and Behavior,* 40**,** 909 - 932.

Bartolucci, F. 2013. *Latent Markov models : applications in social science and economics,* Boca Raton FL, CRC Press.

Bartolucci, F. & Pandolfi, S. 2016. LMest: Latent Markov Models with and without Covariates. [Online], R package version 2.3. Available: https://CRAN.R-project.org/package=LMest.

Bartolucci, F., Pennoni, F. & Francis, B. 2007. A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society Series a-*

Statistics in Society, 170**,** 115-132.

Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* 67**,** 1-48.

Bauer, D. & Curran, P. 2003. Overextraction of Latent Trajectory Classes: Much Ado About Nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods,* 8**,** 384-393.

Berk, A. 2008. *Statistical learning from a regression perspective,* New York, Springer.

Bersani, B., Nieuwbeerta, P. & Laub, J. 2009. Predicting Trajectories of Offending over the Life Course: Findings from a Dutch Conviction Cohort. *Journal of Research in Crime and Delinquency,* 46**,** 468 - 494.

Block, C., Blokland, A., Van Der Werff, C., Van Os, R. & Nieuwbeerta, P. 2010. Long-Term Patterns of Offending in Women. *Feminist Criminology,* 5**,** 73-101.

Blokland, A. 2005. *Crime over the life span; trajectories of criminal behavior in Dutch offenders*, Netherlands Institute for the study of Crime and Law Enforcement, Faculty of Law, Leiden University.

Blokland, A., Nagin, D. & Nieuwbeerta, P. 2005. Life Span Offending Trajectories of a Dutch Conviction Cohort. *Criminology,* 43**,** 919-954.

Blokland, A. & Nieuwbeerta, P. 2005. The effects of life circumstances on longitudinal trajectories of offending. *Criminology,* 43**,** 1203-1240.

Blokland, A. & Nieuwbeerta, P. 2010. Considering Criminal Continuity Testing for Heterogeneity and State Dependence in the Association of Past to Future Offending. *Australian and New Zealand Journal of Criminology,* 43**,** 526-556.

Blokland, A. & Van Os, R. 2010. Life Span Offending Trajectories of Convicted Dutch Women. *International Criminal Justice Review,* 20**,** 169-187.

Blumstein, A., Cohen, J. & Farrington, D. 1988. Criminal Career Research - Its Value for Criminology. *Criminology,* 26**,** 1-35.

Blumstein, A., Cohen, J., Roth, J. A. & Visher, A. 1986. *Criminal Careers and "Career*

Criminals", Washington, D.C., National Academy Press.

Brame, R., Mulvey, E. & Piquero, A. 2001. On the development of different kinds of criminal activity. *Sociological Methods & Research,* 29**,** 319-341.

Brame, R., Paternoster, R. & Bushway, S. 2004. Criminal Offending Frequency and Offense Switching. *Journal of Contemporary Criminal Justice,* 20**,** 201-214.

Brame, R. & Piquero, A. 2003. Selective Attrition and the Age-Crime Relationship. *Journal of Quantitative Criminology,* 19**,** 107-127.

Britt, C. 1996. The Measurement of Specialization and Escalation in the Criminal Career: An Alternative Modeling Strategy. *Journal of Quantitative Criminology,* 12**,** 193-222.

Buker, H. 2011. Formation of self-control: Gottfredson and Hirschi's general theory of crime and beyond. *Aggression and Violent Behavior,* 16**,** 265-276.

Bushway, D., Sweeten, G. & Nieuwbeerta, P. 2009. Measuring Long Term Individual Trajectories of Offending Using Multiple Methods. *Journal of Quantitative Criminology,* 25**,** 259-286.

Bushway, S., Thornberry, T. & Krohn, M. 2003. Desistance as a developmental process: A comparison of static and dynamic approaches. *Journal of Quantitative Criminology,* 19**,** 129-153.

Cauffman, E. 2008. Understanding the female offender. *Future of children,* 18**,** 119-142.

Chaiken, J. & Chaiken, M. 1982. *Varieties of Criminal Behavior,* Santa Monica, CA, Rand Corp.

Coleman, C. 1996. *Understanding crime data : haunted by the dark figure,* Buckingham, Open University Press.

Collins, L. & Lanza, S. 2010. *Latent Class and Latent Transition Analysis : With Applications in the Social Behavioral, and Health Sciences*, Wiley.

Colman, R., Kim, D., Mitchell-Herzfeld, S. & Shady, T. 2009. Delinquent Girls Grown Up: Young Adult Offending Patterns and Their Relation to Early Legal, Individual, and Family Risk. *Journal of Youth and Adolescence,* 38**,** 355-366.

Copas, J. & Marshall, P., 2002. The offender group reconviction scale: a statistical reconviction score for use by probation officers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(1),159–171.

Cox, D. 1966. Some procedures connected with the logistic qualitative response curve. *In:* David, F. N. (ed.) *Research Papers in Probability and Statistics (Festschrift for J. Nyeman).* London: Wiley.

Cullen, F. & Agnew, R. 1999. Criminological Theory: Past to Present Essential Readings. Los Angeles, CA: Roxbury Publishing Co.

D'unger, A., Land, K., Mccall, P. & Nagin, D. 1998. How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses. *American Journal of Sociology,* 103**,** 1593-1630.

Delisi, M. 2005. *Career Criminals in Society,* London, SAGE.

Delisi, M. & Vaughn, M. 2007. The Gottfredson-Hirschi Critiques Revisited: Reconciling Self-Control Theory, Criminal Careers, and Career Criminals. *International Journal of Offender Therapy and Comparative Criminology,* 52**,** 520-537.

Eggleston, E., Laub, J. & Sampson, R. 2004. Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology,* 20**,** 1-26.

Elliott, A., Francis, B., Soothill, K. & Blokland, A. 2017. Changing crime-mix patterns of offending over the life course: a comparative study in England & Wales and the Netherlands. *In:* Blokland, A. & Van Der Geest, V. (eds.) *The Routledge International Handbook of Life-Course Criminology.* London: Routledge.

Enders, C. & Bandalos, D. 2001. The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal,* 8**,** 430-457.

Epperson, J. 1987. On the Runge Example. *The American Mathematical Monthly,* 94**,** 329-341.

Ezell, M. & Cohen, L. 2005. *Desisting from Crime: Continuity and Change in Long-term Crime Patterns of Serious Chronic Offenders,* Oxford, Oxford University Press.

Farrington, D. 1986. Age and Crime. *Crime and Justice-a Review of Research,* 7**,** 189-250.

Farrington, D. 1992. Criminal Career Research In The United Kingdom. *British Journal of Criminology,* 32**,** 521-536.

Farrington, D. 2001. What Has Been Learned from Self-Reports About Criminal Careers And the Causes of Offending? Home Office Report. London.

Farrington, D. 2003. Developmental and Life-Course Criminology: Key Theoretical and Empirical Issues - The Sutherland Award Address *Criminology,* 41**,** 221-225.

Farrington, D. 2015. Cross-national comparative research on criminal careers, risk factors, crime and punishment. *European Journal of Criminology,* 12**,** 386-399.

Farrington, D., Coid, W., Harnett, M., Jolliffe, D., Soteriou, N., Turner, E. & West, D. 2006. Criminal careers up to age 50 and life success up to age 48: new findings from the Cambridge Study in Deliquent Development. *In:* Home Office Research, D. a. S. D. (ed.) 2nd ed. London: Home Office.

Farrington, D., Loeber, R., Elliott, D., Hawkins, J., Kandel, D., Klein, M., Mccord, J., Rowe, D. & Tremblay, R. 1990. Advancing Knowledge About The Onset Of Delinquency And Crime. *Advances in Clinical Child Psychology,* 13**,** 283-342.

Farrington, D., Snyder, H. & Finnegan, T. 1988. Specialization in Juvenile-Court Careers. *Criminology,* 26**,** 461-487.

Farrington, D. & West, D. 1990. The Cambridge Study in Delinquent Development: A Long-Term Follow-Up of 411 London Males. *In:* Kerner, H.-J. & Kaiser, G.

(eds.) *Kriminalität: Persönlichkeit, Lebensgeschichte und Verhalten.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Farrington, D. & Zara, G. 2015. *Criminal Recidivism : Explanation, prediction and prevention*, Cullompton : Taylor and Francis.

Fitzmaurice, G., Laird, N. & Ware, J. 2004. *Applied Longitudinal Analysis,* Hoboken, Wiley.

Francis, B. & Crosland, P. 2002. The Police National Computer and the Offenders Index: can they be combined for research purposes? Full Report. *Home Office Online Report.* London.

Francis, B., Elliott, A. & Weldon, M. 2016. Smoothing Group-Based Trajectory Models Through B-Splines. *Journal of Developmental and Life-Course Criminology,* 2**,** 113-133.

Francis, B. & Humphreys, L. 2016. Investigating the Relationship Between the Diversity Index and Frequency of Offending. *Journal of Developmental and Life-Course Criminology,* 2**,** 397-416.

Francis, B. & Liu, J. 2015. Modelling escalation in crime seriousness: a latent variable approach. *METRON,* 73**,** 277-297.

Francis, B., Liu, J. & Soothill, K. 2010. Criminal Lifestyle Specialization: Female Offending in England and Wales. *International Criminal Justice Review,* 20**,** 188-204.

Francis, B., Soothill, K. & Fligelstone, R. 2004. Identifying Patterns and Pathways of Offending Behaviour: A New Approach to Typologies of Crime. *European Journal of Criminology,* 1**,** 48-87.

Francis, B., Soothill, K. & Humphreys, L. 2007. Development of a reoffending measure using the Police National Computer database.

Gibbons, D. 1975. Offender Typologies-Two Decades Later. *British Journal of Criminology,* 15**,** 140-156.

Gibbons, R., Hedeker, D. & Dutoit, S. 2010. Advances in Analysis of Longitudinal

Data. *Annual Review of Clinical Psychology, Vol 6,* 6**,** 79-107.

Gibson, C. 2010. 'Gottfredson, Michael R., and Travis Hirschi: Self-control theory',. *In:* Ft Cullen, A. P. W. (ed.) *Encyclopedia of criminological theory.* Thousand Oaks, CA.: SAGE Publications, Inc.

Gittens, E. 2011. *Criminal Careers of a sub-set of offenders in Barbados.* University of Liverpool.

Glueck, S. & Glueck, E. 1930. *500 criminal careers,* New York, A.A. Knopf.

Gottfredson, M. & Hirschi, T. 1986. The True Value of Lambda Would Appear to be Zero - An Essay on Career Criminals, Crimal Careers, Selective Incapacitation, Cohort Studies, and Related Topics. *Criminology,* 24**,** 213-234.

Gottfredson, M. & Hirschi, T. 1988. Science, Public Policy, and The Career Paradigm. *Criminology,* 26**,** 37-55.

Gottfredson, M. & Hirschi, T. 1990. *A General Theory of Crime,* Stanford, Calif., Stanford University Press.

Graham, J., Collins, L., Wugalter, S., Chung, N. & Hansen, W. 1991. Modeling Transitions in Latent Stage - Sequential Processes - A Substance Use Prevention Example. *Journal of Consulting and Clinical Psychology,* 59**,** 48-57.

Graham, J. W. 2009. Missing data analysis: Making it work in the real world. *Annual review of psychology,* 60**,** 549-576.

Greenberg, D. 1991. Modeling Criminal Careers. *Criminology,* 29**,** 17-46.

Hastie, T. & Tibshirani, R. 1991. *Generalized additive models,* Boca Raton, Fla, Chapman & Hall.

He, X. & Ng, P. 1999. COBS: qualitatively constrained smoothing via linear programming. *Computational Statistics,* 14**,** 315-337.

Home Office 1998a. The Offenders Index Codebook. London: Home Office.

Home Office 1998b. The Offenders Index: A user's guide. London: Home Office.

Horney, J. & Marshall, I. 1991. Measuring Lambda Through Self-Reports. *Criminology,* 29**,** 471-495.

Howard, P. & Kershaw, C. 2000. Using Criminal Career Data in Evaluation. *In:* Mair, G. & Tarling, R. (eds.) *The British Society of Criminology Conference.* Liverpool: British Society of Criminology.

Jones, B. & Nagin, D. 2007. Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. *Sociological Methods & Research,* 35**,** 542-571.

Jones, B., Nagin, D. & Roeder, K. 2001. A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological methods & research,* 29**,** 374-393.

Jung, T. & Wickrama, K. 2008. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass,* 2**,** 302-317.

Junger-Tas, J. & Block, R. 1988. *Juvenile delinquency in the Netherlands,* Amstelveen, The Netherlands, Kugler Publications.

Kaplan, D. 2008. An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology,* 44**,** 457-467.

Keele, L. 2007. Splines. *Semiparametric Regression for the Social Sciences.* John Wiley & Sons, Ltd.

Kempf, K. 1987. Specialization and the Criminal Career. *Criminology,* 25**,** 399-420.

Krohn, M., Gibson, C. & Thornberry, T. 2013. Under the Protective Bud the Bloom Awaits: A Review of Theory and Research on Adult-Onset and Late-Blooming Offenders. *In:* Gibson, C. L. & Krohn, M. D. (eds.) *Handbook of Life-Course Criminology: Emerging Trends and Directions for Future Research.* New York: Springer New York.

Laird, N. & Ware, J. 1982. Random-Effects Models for Longitudinal Data. *Biometrics,* 38**,** 963-974.

Langeheine, R. & Van De Pol, F. 2002. Latent markov chains. *Applied latent class analysis***,** 304-341.

Lanza, S., Lemmon, D., Schafer, J. & Collins, L. 2007. PROC LCA & PROC LTA User's Guide Version 1.1.3 beta. Pennsylvania The Pennsylvania State University, The Methodology Center.

Laub, J. 2004. The Life Course of Criminology in the United States: The American Society of Criminology 2003 Presidential Address. *Criminology,* 42**,** 1-26.

Lipsey, M. & Derzon, J. 1999. Predictors of Violent or Serious Delinquency in Adolescence and Early Adulthood: A Synthesis of Longitudinal Research. *In:* Farrington, R. L. D. P. (ed.) *Serious & violent juvenile offenders: Risk factors and successful interventions.* Thousand Oaks, California: SAGE Publications, Inc.

Liu, S. 2015. Is the Shape of the Age-Crime Curve Invariant by Sex? Evidence from a National Sample with Flexible Non-parametric Modeling. *Journal of Quantitative Criminology,* 31**,** 93-123.

Loeber, R. & Farrington, D. 2012. *From Juvenile Delinquency to Adult Crime: Criminal Careers, Justice Policy and Prevention*, OUP USA.

Loeber, R. & Snyder, H. 1990. Rate of Offending in Juvenile Careers - Findings of Constancy and Change in Lambda. *Criminology,* 28**,** 97-109.

Lussier, P., Leblanc, M. & Proulx, J. 2005. The generality of criminal behavior: A confirmatory factor analysis of the criminal activity of sex offenders in adulthood. *Journal of Criminal Justice,* 33**,** 177-189.

Lussier, P., Mccuish, E., Deslauriers-Varin, N. & Corrado, R. 2017. *Crime specialization as a dynamic process? In:* Blokland, A. & Van Der Geest, V. (eds.) *The Routledge International Handbook of Life-Course Criminology.* London: Routledge.

Lynam, D., Piquero, A. & Moffitt, T. 2004. Specialization and the Propensity to

Violence. *Journal of Contemporary Criminal Justice,* 20**,** 215-228.

Macleod, J., Grove, P. & Farrington, D. 2012. *Explaining criminal careers : implications for justice policy,* Oxford, Oxford University Press.

Marshall, J. 2006. Juvenile offending trajectories: A South Australian study. *In:* Research, O. O. C. S. A. (ed.). South Australia: Office of Crime and Statistics.

Massoglia, M. 2006. Desistance or Displacement? The Changing Patterns of Offending from Adolescence to Young Adulthood. *Journal of Quantitative Criminology,* 22, 215–239.

Mccullagh, P. & Nelder, J. 1989. *Generalized Linear Models, Second Edition,* London, Taylor & Francis.

Mcgloin, J., Sullivan, C. & Piquero, A. 2009. Aggregating to Versatility?: Transitions among Offender Types in the Short Term. *British Journal of Criminology,* 49**,** 243-264.

Mcgloin, J., Sullivan, C., Piquero, A. & Pratt, T. 2007. Local Life Circumstances and Offending Specialization/Versatility. *Journal of Research in Crime & Delinquency,* 44**,** 321-346.

Mclachlan, G. & Peel, D. 2000. *Finite mixture models,* Chichester Wiley.

Menard, S. & Elliott, D. 1990. Self-reported offending, maturational reform, and the Easterlin hypothesis. *Journal of Quantitative Criminology,* 6**,** 237-267.

Moffitt, T. 1993. Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: a Developmental Taxonomy. *Psychological Review,* 100**,** 674-701.

Moffitt, T., Caspi, A., Dickson, N., Silva, P. & Stanton, W. 1996. Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and Psychopathology,* 8**,** 399-424.

Monahan, K. & Piquero, A. 2009. Investigating the longitudinal relation between offending frequency and offending variety. *Criminal Justice and Behavior,* 36, 653–673.

Morgan, B. & Beaujean, A. 2014. An Investigation of Growth Mixture Models for

Studying the Flynn Effect. *Journal of Intelligence,* 2, 156-179.

Muthén, B. & Shedden, K. 1999. Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics,* 55**,** 463-469.

Muthén, L. & Muthén, B. 1998-2012. Mplus User's Guide, Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Nagin, D. 1999. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4(2), 139–157.

Nagin, D. 2005. *Group-based modeling of development,* Cambridge. Harvard University Press.

Nagin, D. & Farrington, D. 1992. The Onset and Persistence of Offending. *Criminology,* 30**,** 501-523.

Nagin, D., Farrington, D. & Moffitt, T. 1995. Life-Course Trajectories of Different Types of Offenders. *Criminology,* 33**,** 111-139.

Nagin, D., Jones, B., Lima Passos, V. & Tremblay, R. 2016. Group-based multi-trajectory modeling. *Statistical Methods in Medical Research***.**

Nagin, D. & Land, K. 1993. Age, criminal careers, and population heterogeneity: Specification and estimation of a non-parametric, mixed poisson model. . *Criminology,* 31**,** 327-362.

Nagin, D. & Paternoster, R. 1991. On the relationship of past to future participation in delinquency. *Criminology,* 29**,** 163-189.

Nagin, D. & Paternoster, R. 2000. Population Heterogeneity and State Dependence: State of the Evidence and Directions for Future Research. *Journal of Quantitative Criminology,* 16**,** 117-144.

Nagin, D. & Piquero, A. 2010. Using the Group-Based Trajectory Model to Study Crime Over the Life Course. *Journal of Criminal Justice Education,* 21**,** 105-116.

Nieuwbeerta, P., Blokland, A., Piquero, A. & Sweeten, G. 2011. A Life-Course Analysis of Offense Specialization Across Age: Introducing a New Method for

Studying Individual Specialization Over the Life Course. *Crime & Delinquency,* 57**,** 3-28.

Nylund, K., Asparouhov, T. & Muthén, B. 2007. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal,* 14**,** 535-569.

Osborne, M., Presnell, B. & Turlach, B. 1998. Knot Selection for Regression Splines via the LASSO. *Comput Sci Stat,* 30**,** 44-49.

Osgood, D. & Schreck, C. 2007. A New Method for Studying the Extent, Stability, and Predictors of Individual Specialization in Violence. *Criminology,* 45**,** 273–312.

Paternoster, R., Brame, R., Piquero, A., Mazerolle, P. & Dean, C. W. 1998. The forward specialization coefficient: Distributional properties and subgroup differences. *Journal of Quantitative Criminology,* 14, 133–154.

Paternoster, R., Dean, C., Piquero, A., Mazerolle, P. & Brame, R. 1997. Generality, continuity, and change in offending. *Journal of Quantitative Criminology,* 13**,** 231-266.

Patterson, G. 1993. Orderly change in a stable world: the antisocial trait as a chimera. *Journal of consulting and clinical psychology,* 61**,** 911-919.

Piquero, A. 2008. Taking Stock of Developmental Trajectories of Criminal Activity over the Life Course. *In:* Liberman, A. M. (ed.) *The Long View of Crime: A Synthesis of Longitudinal Research.* New York: Springer New York.

Piquero, A., Blumstein, A., Brame, R., Haapanen, R., Mulvey, E. & Nagin, D. 2001. Assessing the Impact of Exposure Time and Incapacitation on Longitudinal Trajectories of Criminal Offending. *Journal of Adolescent Research,* 16**,** 54-74.

Piquero, A., Farrington, D. & Blumstein, A. 2003. The Criminal Career Paradigm. *Crime and Justice: a Review of Research,* 30**,** 359-506.

Piquero, A., Farrington, D. & Blumstein, A. 2007. *Key Issues in Criminal Career*

Research : New Analyses of the Cambridge Study in Delinquent Development, Cambridge, University Press.

Piquero, A., Hawkins, D. & Kazemian, L. 2012. Criminal Career Patterns. *In:* Loeber, R. & Farrington, D. P. (eds.) *From Juvenile Delinquency to Adult Crime: Criminal Careers, Justice Policy, and Prevention.* United States: Oxford University Press.

Piquero, A., Paternoster, R., Mazerolle, P., Brame, R. & Dean, C. 1999. Onset Age and Offense Specialization. *Journal of Research in Crime and Delinquency,* 36**,** 275-299.

Piquero, A., Schubert, C. & Brame, R. 2014. Comparing Official and Self-report Records of Offending across Gender and Race/Ethnicity in a Longitudinal Study of Serious Youthful Offenders. *Journal of Research in Crime and Delinquency,* 51**,** 526-556.

Proust-Lima, C., Philipps, V., Diakite, A. & Liquet, B. 2017. lcmm: Extended Mixed Models Using Latent Classes and Latent Processes. R package version: 1.7.8, https://cran.r-project.org/package=lcmm.

Raftery, A. 1995. Bayesian model selection in social research. *Sociological methodology***,** 25, 111-163.

Ram, N. & Grimm, K. 2009. Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. *International journal of behavioral development,* 33**,** 565-576.

Rodriguez, G. 2001. Smoothing and Non-Parametric Regression. Available: http://data.princeton.edu/eco572/smoothing.pdf [Accessed 18/11/2015].

Sampson, R. & Laub, J. 1991. Crime and Deviance Over the Life Course: The Salience of Adult Social Bonds. *American Sociological Review,* 55**,** 609-627.

Sampson, R. & Laub, J. 1993. *Crime in the making: pathways and turning points through life. ,* Cambridge, Havard University Press.

Sampson, R. & Laub, J. 2003. Life-Course Desisters? Trajectories of Crime Among

Delinquent Boys Followed to Age 70. *Criminology,* 41, 555-592.

Schwarz, G. 1978. Estimating the Dimension of a Model. *The Annals of Statistics,* 6, 461-464.

Silverman, B. W. 1985. Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting. *Journal of the Royal Statistical Society. Series B (Methodological),* 47**,** 1-52.

Silverthorn, P. & Frick, P. 1999. *Developmental pathways to antisocial behavior: The delayed-onset pathway in girls*. *Development and Psychopathology,* 11, 101–126.

Simmons, J., Legg, C. & Hosking, R. 2003. National Crime Recording Standard (NCRS): an analysis of the impact on recorded crime. Part One: The national picture. London. Home Office.

Skardhamar, T. 2009. Reconsidering the Theory on Adolescent-Limited and Life-Course Persistent Anti-Social Behaviour. *British Journal of Criminology,* 49**,** 863-878.

Soothill, K., Ackerley, E. & Francis, B. 2008. Criminal convictions among children and young adults: Changes over time. *Criminology & Criminal Justice,* 8, 297-315.

Sullivan, C., Mcgloin, J., Pratt, T. & Piquero, A. 2006. Rethinking the "Norm" of Offender Generality: Investigating Specialization in the Short-Term. *Criminology,* 44**,** 199-233.

Sullivan, C., Mcgloin, J., Ray, J. & Caudy, M. 2009. Detecting Specialization in Offending: Comparing Analytic Approaches. *Journal of Quantitative Criminology,* 25**,** 419-441.

Sweeten, G. 2014. Group-based trajectory models. In Bruinsma, G., & Weisburd, D. (Eds.) *Encyclopedia of criminology and criminal justice*, 1991-2003. New York: Springer.

Tarling, R. 1993. *Analysing offending : Data, models and interpretations*, London, HMSO.

Thornberry, T. 2005. Explaining multiple patterns of offending across the life and course and across generations. *Annals of the American Academy of Political and Social Science,* 602**,** 156-195.

Thornberry, T. & Krohn, M. 2003. Comparison of Self-report and Official Data for Measuring Crime. *In:* And, J. V. P. & Petrie., C. V. (eds.) *Measurement Problems in Criminal Justice Research: Workshop Summary.* Washington, DC: National Academy Press.

Tollenaar, N. & P. G. M. Van Der Heijden, 2012. Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 565–584.

Tracy, P., Wolfgang, M. & Figlio, R. 1990. *Delinquency careers in two birth cohorts,* New York, Plenum Press.

Tremblay, R., Nagin, D., Séguin, J., Zoccolillo, M., Zelazo, P., Boivin, M., Pérusse, D. & Japel, C. 2004. Physical Aggression During Early Childhood: Trajectories and Predictors. *Pediatrics,* 114**,** 43-50.

Vaughn, M., Delisi, M., Beaver, K. & Howard, M. 2008. Toward a Quantitative Typology of Burglars: A Latent Profile Analysis of Career Offenders. *Journal of Forensic Sciences,* 53**,** 1387-1392.

Verbeke, G. & Lesaffre, E. 1996. A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. *Journal of the American Statistical Association,* 91**,** 217-221.

Vermunt, J. & Madigson, J. 2005. *Latent Gold 4.0 User Manual,* Belmont, Massachusetts, Statistical Innovations Inc.

Vermunt, J. & Magidson, J. 2016. Upgrade Manual for Latent GOLD 5.1. Basic, Advanced, and Syntax. Available: https://www.statisticalinnovations.com/wp-content/uploads/LGtecnical.pdf [Accessed 20/7/2017].

Ward, A., Day, D., Bevc, I., Sun, Y., Rosenthal, J. & Duchesne, T. 2010. Criminal

Trajectories and Risk Factors in a Canadian Sample of Offenders. *Criminal Justice and Behavior,* 37.

West, D. & Farrington, D. 1973. Who Becomes Delinquent: Second Report of the Cambridge Study in Delinquent Development. London: Heinemann.

Wickham, H. 2014. Tidy Data. *Journal of Statistical Software,* 1, 1-23.

Wolfgang, M., Robert Figlio & Sellin, T. 1972. *Delinquency in a Birth Cohort.,* Chicago, University of Chicago Press.

Yessine, A. & Bonta, J. 2008. Pathways to serious offending. Canada.Public Safety Canada. Available: http://www.publicsafety.gc.ca/res/cor/rep/_fl/ptso-2008-eng.pdf [Accessed 25/10/2015].

Yessine, A. & Bonta, J. 2009. The Offending Trajectories of Youthful Aboriginal Offenders. *Canadian Journal of Criminology and Criminal Justice* 51**,** 435-472.

APPENDIX

Offenders Index 38 Offence Categories

1. 'Serious violence'
    a. Murder
    b. Attempted Murder
    c. Threats, conspiracy, or incitement to murder
    d. Manslaughter
2. 'Violence'
    a. Wounding and Endangering life
    b. Malicious wounding
    c. Assault
    d. Hijacking
    e. Rioting
    f. Unlawful Assembly
    g. Causing an affray
    h. Intimidating a juror or witness or person assisting in the investigation of offences
    i. Harming or threatening to harm a witness, juror etc
    j. Common assault and battery
3. 'Firearms/dangerous weapon (possession etc)'
    a. Possession of firearms etc with intent to endanger life or injure property
    b. Possession of offensive weapons without lawful authority or reasonable excuse
    c. Possessing firearm or ammunition without certificate
4. 'Resist arrest etc'

a. Assault with intent to commit felony or resist apprehension
b. Assault with intent to resist apprehension or assault a person assisting a constable
c. Assaulting, resisting or obstructing a person assisting a constable
d. Absconding from lawful custody

5. 'Kidnapping and false imprisonment'
6. 'Sexual Offences against 16+' - (Categories 6-9 ‘Sexual Offences’ include the following sub categories;
   a. Rape and attempted rape
   b. Buggery and attempted buggery
   c. Sexual Assault and attempted
   d. Procuration
   e. Indecent exposure
   f. Obscene publications
   g. Gross Indecency with Children
   h. Taking, permitting to be taken or making, distributing or publishing indecent photographs or pseudo-photographs of children)
7. 'Sexual Offences under 16'
8. 'Sexual Offences consensual'
9. 'Prostitution'
10. 'Burglary (dwelling)'
11. 'Burglary (aggravated etc)'
12. 'Burglary (other)'
13. 'Going equipped'
14. 'Robbery'
15. 'Blackmail'
16. 'Vehicle taking (aggravated etc)'
17. 'Theft'

18. 'Theft from person'
19. 'Theft from employee'
20. 'Theft (in a dwelling)'
21. 'Theft (machines/meters/electricity)'
22. 'Theft from vehicles'
23. 'Theft of vehicles'
24. 'Attempted theft of/from vehicle'
25. 'Shoplifting'
26. 'Fraud and forgery'
27. 'Receiving and handling'
28. 'Criminal damage'
29. 'Drugs (possession etc only)'
30. 'Drugs (possession/supply)'
31. 'Drugs (import/export/production)'
32. 'Absconding/bail/breach offences'
33. 'Public order'
34. 'Perjury/attempting to pervert course of justice'
35. 'Dangerous Driving'
36. 'Immigration'
37. 'Child cruelty etc'
38. 'Other'

CCLS 28 Offence Categories

1. ‘Murder’
2. ‘Culpose Death’
3. ‘Threatening’
4. ‘Assault’

5. ‘Guns’
6. ‘Authority’
7. ‘Sexual abuse (child)’
8. ‘Sexual Assault’
9. ‘Rape’
10. ‘Flashing’
11. ‘Simple Theft’
12. ‘Aggravated Theft (Burglary)’
13. ‘Violent Theft’
14. ‘Extortion’
15. ‘Forgery’
16. ‘Embezzlement’
17. ‘Swindling’
18. ‘Vandalism’
19. ‘Drugs’
20. ‘Public Order’
21. ‘Fencing’
22. ‘Discrimination’
23. ‘Offences causing general danger’
24. ‘Other criminal law’
25. ‘Other non-criminal law’
26. ‘Unknown’
27. ‘Misdemeanour’
28. ‘Traffic offences’

11 Offence Categories

Offence Category	OI	CCLS
Murder/Violence	• Murder • Threats • Manslaughter • Violence • Assault • Malicious Wounding	• Murder • Culpose Death • Threatening • Assault
Firearms	• Firearms	• Guns
Authority	• Resisting Arrest/Authority • Absconding	• Resisting authority
Sexual Offences	• Sexual 16+ • Sexual under 16 • Sexual Consensual • Prostitution	• Sexual Abuse (child) • Sexual Assault • Rape • Indecent Exposure (flashing)
Robbery	• Robbery	• Robbery
Blackmail	• Blackmail	• Extortion
Burglary and Theft	• Burglary (dwelling) • Burglary (aggravated) • Burglary (other) • Going equipped • Theft • Theft from person • Theft (in dwelling) • Theft (machines /meters /electricity) • Shoplifting • Vehicle taking • Theft from vehicles	• Simple Theft • Aggravated Theft • Violent Theft

	• Theft of vehicles, • Attempted theft of/from vehicle	
Fraud and Forgery	• Fraud and Forgery • Receiving and Handling	• Forgery • Embezzlement • Swindling
Criminal Damage	• Criminal Damage	• Vandalism
Drugs	• Drugs (import /export/ production)	• Drugs
Public Order	• Public Order	• Public Order