1 Introduction

Unlike in the physical sciences, theories in criminology (as in the other social and behavioural sciences) rarely derive or test exact quantitative predictions. In a true science, exact quantitative predictions are derived and tested empirically. Most criminological theories aim to explain and predict (qualitatively rather than quantitatively) only the total prevalence of offending. In developmental and life-course criminology (see e.g. Farrington, 2005), theories also aim to specify the extent to which total prevalence is influenced by classic individual, family, and socio-economic risk factors. However, it is highly desirable to propose theories that also explain and predict important criminal career features such as the age–crime curve, the probability of recidivism after each offence, and time intervals between offences. The main aim of this Element is to make progress towards truly scientific criminology by deriving and testing exact quantitative predictions about key criminal career features from simple theories. We hope that criminologists will build on our simple theories to develop and test more complex theories (e.g. incorporating risk and protective factors) that make exact quantitative predictions.

1.1 The Moffitt Theory

As an example of a criminological theory that does not yield quantitative predictions but does attempt to explain the age–crime curve, consider the, very influential, developmental taxonomy theory of Moffitt (1993). She aimed to reconcile two apparently incongruous facts about offending, namely that it shows impressive continuity over time but that its prevalence changes dramatically at different ages, peaking in the teenage years. She proposed that there were two distinct categories of individuals, namely life-course-persistent (LCP) and adolescence-limited (AL) offenders, who differed in kind rather than in degree.

Moffitt stated that LCP offenders were influenced by many classic risk factors, including pregnancy problems of their mothers, neuropsychological deficits including impulsivity and inattention, cognitive deficits including low intelligence and attainment, poor child-rearing including low parental warmth and inconsistent discipline, disrupted families, and socio-economic deprivation. In contrast, AL offenders were motivated by the gap between their biological and social maturity (e.g. they wanted material goods and status but could not achieve them legitimately during adolescence), and they were influenced by their LCP peers (‘social mimicry’). They ceased offending in young adulthood because they could then achieve their aims legitimately when they had social skills and no cognitive deficits (e.g. they did not have low intelligence or low academic attainment).

This theory is undoubtedly one of the most important and most researched criminological theories. Moffitt (2018) and McGee and Moffitt (2019) have summarized developments in the quarter-century since the theory was propounded. Generally, the theory has held up well, but there have been concerns with new categories of offenders such as low-level chronics and adult-onset offenders and also with abstainers who refrain from offending entirely.

While this theory is excellent in many ways, it does not make or test exact quantitative predictions about such basic parameters as the prevalence of LCP and AL offenders and their frequency of offending at different ages. Also, the theory does not specify the exact quantitative impact of risk factors on types of offenders. Moffitt (1993) says that about 5 per cent, or at least less than 10 per cent, of males would be LCP offenders (at all ages combined), while a much higher fraction (peaking at over 60 per cent in her figure 3) would be AL offenders. However, she has not yet attempted to quantify her theory and show to what extent it can predict actual criminal career data (e.g. the number of crimes committed at different ages by a birth cohort).

1.2 Group-Based Trajectory Modelling

Moffitt's theory has been tested empirically using group-based trajectory modelling (GBTM). This was first used by Nagin and Land (1993), who stated explicitly that 'Analyses reported herein are also designed to test for the existence of the sorts of distinctive offending trajectories predicted by Moffitt's theory' (pp. 328–329). GBTM fits criminal career data by assuming that individuals commit crimes according to a Poisson process (i.e. randomly over time). They assume that Ln(lambda), where lambda is the individual's underlying frequency of offending, depends on time-stable characteristics of the individual as well as age and (age-squared). It was important to control for these time-stable characteristics because Nagin and Farrington (1992) previously concluded that the link between past and future offending was largely driven by persistent heterogeneity (the persistence over time of individual differences in underlying criminal potential) rather than by state dependence (the effect of past offending on future offending). Nagin and Land (1993) analysed Cambridge Study in Delinquent Development (CSDD, discussed later) data up to age thirty-two and found three categories of offenders (high-rate chronics, low-rate chronics, and AL) plus non-offenders, in fitting criminal career data. These analyses were later extended by Nagin et al. (1995).

While GBTM is a very important method, it is important to point out that it is a method of fitting criminal career data rather than a criminological theory. For example, there is no theoretical reason advanced to explain why the frequency of offending should depend on (age-squared). Skardhamar (2010, p. 299), in a critique of GBTM, stated that it is ‘a “theory-free method” (Moffitt, 2006, p. 581), one that allows us to identify unobservable groups that “emerge from data itself” (Nagin, 2005, p. 2) rather than being assumed *a priori*’. Skardhamar (2010) further argued that, even when no groups exist in reality and the data are truly continuous, GBTM will reveal several categories of offenders.

Nagin and Tremblay (2005) were careful to point out that ‘trajectory groups, like all statistical models, are not literal depictions of reality. They are meant only as a convenient statistical approximation’ (p. 882). They further stated (p. 887) that the idea that individuals actually belong to a trajectory group is a ‘misconception’ and that ‘the number of groups and the shape of each group’s trajectory are not fixed realities’ (p. 888). Indeed, Farrington et al. (2013) documented how membership of the trajectories changed in the CSDD as the follow-up age was extended from 24 to 32 to 40 to 48 and finally to 56. Nagin and Tremblay (2005, p. 898) summarized that ‘it is important for users and consumers of the analyses [to] remember that individuals do not actually belong to a trajectory group, that the number of trajectory groups in a sample is not immutable, and that individuals do not follow the group-level trajectory in lockstep’.

GBTM has advanced knowledge greatly, but it is not a criminological theory. We now turn to very simple criminological theories (criminal career models) that yield quantitative predictions that can be tested in criminal career data.

1.3 Criminal Career Models

Blumstein et al. (1985) attempted to predict observed recidivism probabilities in four cohort studies. Their key assumption was that each offender had a constant probability of persisting after each offence. They found that observed recidivism probabilities and, more generally, the distribution of offences over offenders could be predicted by a model that partitioned each sample into three subgroups: *innocents*, who had no offending record; *desisters*, who had a low recidivism probability; and *persisters*, who had a high recidivism probability. The observed aggregate recidivism probability increased after each arrest because the desisters tended to drop out and leave behind a sample composed increasingly of the persisters.

Blumstein et al. (1985) then applied their mathematical model (of innocents, desisters, and persisters) to the CSDD data. The best fit to the recidivism probabilities in the CSDD was obtained by assuming that the probability of persisting after each conviction was 0.87 for persisters and 0.57 for desisters.

The proportion of first offenders who were persisters was 0.28, while the fraction of the sample who were innocents was 0.67. Persisters and desisters differed in their a priori probabilities of persisting, not in their a posteriori number of convictions (as chronics did). This model fitted the data very accurately.

Interestingly, the number of empirically predicted chronics among the offenders (37 ‘high-risk’ offenders with four or more out of seven childhood risk factors) was similar to the predicted number of persisters (36.7) according to the model. Remarkably, the individual process of dropping out of crime by the predicted chronics in the empirical data closely matched the aggregate dropout process for persisters predicted by the model with parameters estimated from aggregate recidivism data. Therefore, the high-risk offenders might be viewed as the identified persisters. This analysis shows the important distinction between prospective empirical predictions (e.g. high-risk offenders), underlying theoretical categories (e.g. persisters), and retrospectively measured outcomes (e.g. chronics).

Barnett and Lofaso (1985) analysed the Philadelphia cohort data of Wolfgang et al. (1972). In contrast to Blumstein et al. (1985), they did not focus on the probability of persistence, but rather on the frequency of offending. They aimed to predict the individual offending frequency (the average number of offences per offender per year) rather than the number of offences committed. They assumed that offences were committed probabilistically (at random) over time, which meant that offenders committed crimes according to a stationary Poisson process (with a constant mean rate). They found that the best predictor of the future individual offending frequency (crimes per year) was the past individual offending frequency.

Barnett et al. (1987) then combined the approaches of Blumstein et al. (1985) and Barnett and Lofaso (1985). They analysed conviction data from the CSDD and aimed to predict the number of offences of each person at each age as well as time intervals between crimes. They tested several models of criminal careers containing two key parameters: (1) p = the probability that an offender terminates the criminal career after the kth conviction; for any given offender, p is assumed to be constant for all values of k, and (2) μ = the individual offending frequency per year, or the annual rate at which the offender sustains convictions while free during the active career. The individual offending frequency cannot be estimated from aggregate data simply by dividing the number of convictions at each age by the number of offenders at each age because some active offenders who have embarked on a criminal career may not be convicted at a particular age.

Barnett et al. (1987) found that models assuming that all offenders had the same p and μ did not fit the data and therefore assumed that there were two categories of offenders: ‘frequents’ and ‘occasionals’. Each category had its own value of p and μ, which were assumed to be constant over time. They found that the model that best fitted the data had the following parameters: μ_F (conviction rate of frequents per year) = 1.14, μ_o (conviction rate of occasionals per year) = 0.41, p_F (termination probability of frequents after each conviction) = 0.10, p_o (termination probability of occasionals after each conviction) = 0.33, and α (fraction of frequents compared to occasionals) = 0.43. Thus, 43 per cent of the offenders were frequents, and this group had a higher individual offending frequency and a lower probability of terminating their criminal careers after each conviction. Barnett et al. (1987) did not suggest that there were in reality only two categories of offenders, but rather that it was possible to fit the conviction data (the number of convictions of each offender at each age) accurately using a simple model that assumed only two categories.

Barnett et al. (1987) basically showed that a very simple criminological theory, focussing only on the frequency of offending and the probability of termination after each conviction, could produce accurate quantitative predictions of the number of offences of each person at each age and of time intervals between crimes. Furthermore, Barnett et al. (1989) carried out a test of the predictive validity of this model using the CSDD data. The model was developed on conviction data between the tenth and twenty-fifth birthdays and tested on conviction data between the twenty-fifth and thirtieth birthdays. The aim was to predict the number of reoffenders, the identities of reoffenders, the number of reconvictions, the age at the first reconviction, and the time intervals between reconvictions in this follow-up period. Generally, the model performed well.

These very simple quantitative theories are the starting point for our Element. Surprisingly, since the 1980s, there have been very few attempts to develop and test simple theories of the type developed by Blumstein and his colleagues (see Farrington et al., 2016). An exception is the book, *Explaining Criminal Careers*, by MacLeod et al. (2012). Some key features of this book are described in Section 2.

1.4 The ‘Great Debate’ in Criminology

Rocque et al. (2016) pointed out that the ‘great debate’ in criminology focussed on the explanation of the age–crime curve, which is clearly a crucial criminological phenomenon. This debate was between Gottfredson and Hirschi on one side and Blumstein and his collaborators on the other side.

In the landmark report of the US National Academy of Sciences Panel on Criminal Career research, Blumstein et al. (1986) emphasized the need to distinguish different features of criminal careers. Farrington (1992, p. 521) summarized these key features: 'A criminal career has a beginning (onset), an end (desistance), and a career length in between (duration). Only a certain proportion of the population (prevalence) has a criminal career and commits offences. During their careers, offenders commit offences at a certain rate (frequency) while they are at risk of offending in the community (i.e. not incarcerated or hospitalized)'.

Hirschi and Gottfredson (1983) argued that the age–crime curve was 'invariant' regardless of sex, race, country, time period, or crime type. They stated that crime rates decreased with age (after the peak) because of 'inexorable ageing' and decreases in biological factors such as energy, physical strength, and testosterone (in males). They further argued that criminal career research and longitudinal studies were not needed because the correlates of offending were the same at all ages. Gottfredson and Hirschi (1986) contended that all criminal career features reflected the single underlying construct of 'criminal propensity'; when this was high, the onset of offending was early, the desistance of offending was late, the duration of offending was high, and the frequency of offending was high. Therefore, prevalence and frequency both reflected criminal propensity; the causes of offending were the same at all ages; and the causes of onset and desistance were the same.

Blumstein et al. (1988a, 1988b) contended that these arguments were incorrect. For example, in the CSDD, they reported that the predictors of conviction (onset) were generally different from the predictors of reconviction (persistence). Earlier, Farrington (1986) showed that the age–crime curve was not invariant but varied over time, place, sex, and crime type and that it reflected prevalence rather than frequency. Later, Farrington and Hawkins (1991) in the CSDD and Loeber et al. (1991) in the Pittsburgh Youth Study (PYS, discussed later) showed in more detail that different criminal career features had different predictors.

Rocque et al. (2016, p. 4) concluded that:

> More recent research on age and crime has failed to unequivocally adjudicate these two positions, but it seems as if the 'criminal career' camp has garnered more support. In other words, more recent research on age and crime has shown that there is a benefit to longitudinal methodologies, that something is to be gained by examining different parts of the criminal career, and that the relationship between age and crime is not entirely invariant.

We believe that Blumstein and his collaborators are correct. Nevertheless, most criminological theories are still concordant with the Gottfredson–Hirschi approach in only trying to explain influences on the prevalence of offending, not influences on the onset, persistence, frequency, desistance, or duration.

Furthermore, most criminological theories are very complex. We believe that a simple theory that explains and predicts a wide range of results is preferable. This point is also made in Agent-Based Modelling, which aims to develop the simplest possible theory and model for a simulation that will provide a realistic set of outcomes (see Weisburd et al., 2017). We hope that our Element will stimulate more adequate, more scientific, and more quantitative theories that aim to explain criminal career features.

In this Element, we use two key parameters – the frequency of offending and the probability of reoffending – to define categories of offenders and investigate the extent to which this simple theory fits criminal career data in two longitudinal studies: the British CSDD (Section 3) and the American PYS (Section 5). The accuracy of predictions is quite remarkable. We then go beyond these simple theories to investigate which childhood risk factors predict categories of offenders in the CSDD (Section 4) and the PYS (Section 6). These analyses suggest how the simple theories might be extended to explain and predict a wide variety of criminal career data. We hope that our analyses will encourage criminologists to formulate and test truly scientific theories that lead to quantitative predictions about how key risk factors influence key criminal career features such as the number of offenders and offences in a cohort at each age. In turn, we hope that more accurate quantitative scientific theories of criminal behaviour will lead to more effective prevention and intervention strategies.

2 The Offenders Index (OI) and the *Risk/Rate* Model

2.1 The MacLeod et al. Analyses

MacLeod et al. (2012) proposed a quantitative theory of criminal careers based on a detailed analysis of official conviction data extracted from the UK Home Office Offenders Index (OI). The mathematical models derived from this theory were shown to fit both longitudinal and cross-sectional conviction data very well. MacLeod et al. also identified the theoretical offender categories from psychological and behavioural data from the Offender Assessment System (OASys) developed and used by the prison and probation services of England and Wales. In this Element, we test this theory using independently collected conviction and assessment data from the CSDD and the PYS. ‘Offences’ always refer to offences leading to convictions.

The OI was created in 1963 and contains records obtained from courts in England and Wales for each court appearance resulting in a conviction for one or more ‘standard list’ offences. The ‘standard list’ includes all offences that may be tried in the Crown Court (more serious indictable and ‘either-way’ offences), as well as the more serious of the offences that can only be tried in the

Magistrates' Courts. The most common types of offence are theft, violence, vandalism, fraud, and drug use. The definition of 'standard list' has changed during the period covered by the OI, with offences being added to or removed from the list, but the MacLeod et al. analyses were based on the definition used in the early 1990s. Cohort samples, comprising all court appearance records for individuals born in one of four weeks during the cohort years of 1953, 1958, 1963, 1968, and 1973, were extracted in 1992/1993, 1999/2000, and 2006 (see Ministry of Justice Statistics Bulletin, 2010). The records of the different convictions for each individual were linked together to form individual OI criminal career histories. The 1992/1993 extracts were used as the basis for the MacLeod et al. analyses, with the 1953 cohort updated to 1999 at age 46. The 1953 cohort is directly comparable to the CSDD cohort, as most of the latter's males were born in 1953.

In the MacLeod et al. analysis, for each individual in the OI cohort, court appearances were labelled with a sequence number, 1 for the first court appearance, 2 for the second, and so on. Conviction records were restricted to principal convictions, coding only the most serious offence dealt with in the court appearance, and a histogram of the count of individuals with n or more principal convictions was constructed. Plotting this histogram, with a linear x-axis (conviction number n) and a logarithmic y-axis (number of court appearances), it is clear, from Figure 1 (MacLeod et al. 2012, figure 2.3, p. 29), that the data points for $n > 6$ lie on a straight line with slope Log (p_1).

It is also clear that the residuals from that line for $n < 6$ also fall on a straight line with a much steeper slope Log (p_2). From this graphical analysis, the 'dual *risk* recidivism model' was derived Eq. (1) (MacLeod et al. 2012, Equation 2.4, p. 30)

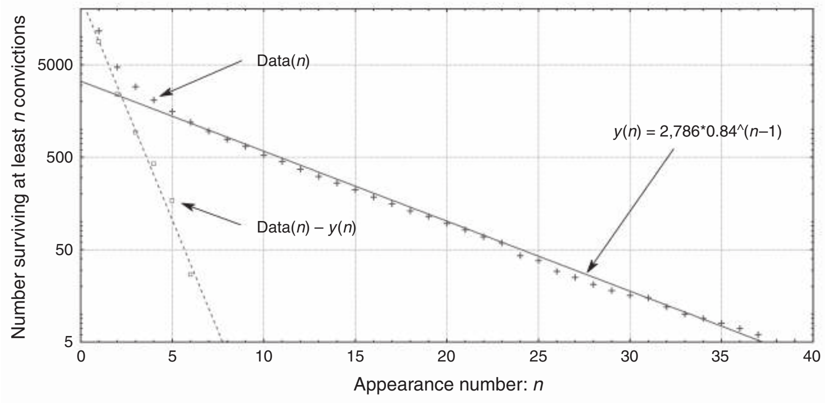

Figure 1 Number of individuals with n or more convictions in the OI 1953 cohort

$$\mathbf{Y}(n) = \mathrm{A}*\left(\mathrm{a}*\mathrm{p}_1^{(n-1)} + (1-\mathrm{a})*\mathrm{p}_2^{(n-1)}\right), \tag{1}$$

where:

Y(*n*) is the number of individuals with at least *n* convictions,
A is the total number of individuals in the cohort with at least one conviction,
a is the proportion of individuals in the *high-risk* (of reconviction) category,
$\mathbf{p_1}$ Is the *high-risk* probability of reconviction, and
$\mathbf{p_2}$ is the *low-risk* probability of reconviction.

The model parameters were estimated for the five longitudinal OI cohort samples and for a 1997 OI cross-sectional sentencing sample. All samples had the same dual *risk* structure, and parameter values were found to be consistent, allowing for the reducing follow-up periods. This model also provided a convincing explanation for the increasing probability of reconviction with conviction number in the early stages of criminal careers. This increasing probability was due to the more rapidly reducing number of *low-risk* individuals as conviction number increased.

A similar analysis of inter-conviction times, using the numbers of offenders surviving conviction-free up to time *t* from the previous conviction, was used to identify the 'dual-*rate* survival time model', Eq. (2) (MGF, 2012, Equation 2.8, p. 35). Survival times were used to smooth out random variations in the inter-conviction time data.

$$S(t) = \mathrm{S}_{0^*}\left(b^*\mathrm{e}^{-\lambda_1 * t} + (1-b)*e^{-\lambda_2 * \mathrm{t}}\right), \tag{2}$$

where:

$S(t)$ is the number of offenders surviving conviction-free, up to time **t** from the previous conviction,
S_0 is the total number of inter-conviction times in the data,
λ_1, λ_2 are the mean numbers of convictions per year for the *high-rate* and *low-rate* categories of offenders, respectively, and
b is the proportion of inter-conviction times attributed to the *high-rate* category of offenders.

The parameter values were estimated for the above-mentioned OI cohort samples. The same dual-rate structure was found, and the parameter estimates were again consistent across cohorts allowing for the length of respective follow-up periods. For the 1953 birth cohort, with the longest follow-up period, the parameters were: $a = 0.237$, $p_1 = 0.840$, $p_2 = 0.313$, $b = 0.565$,

$\lambda_1 = 0.859$, and $\lambda_2 = 0.212$. All the OI cohorts included both male and female offenders.

In the MacLeod et al. analyses, due to the lack of detailed information, immigration and emigration were assumed to balance out or at least, along with death, were assumed to have limited impact on the underlying processes. Also, the time offenders spent in prison, although clearly important at the individual level, was not taken into account in the analysis for several reasons. First, only sentence length had been recorded on the OI, so there was no allowance for remission, parole, partial suspension of sentence, or time spent on remand. Second, inter-conviction time distributions for individuals with custodial sentences were not significantly different from those of similar offenders with supervisory disposals. Lastly, half the total sentence length (to allow for remission) for all the cohorts represents less than 2 per cent of the sum of active career lengths. Again, it was assumed that time in custody had only a limited impact on the underlying distributions. It should be pointed out that the average time served in England and Wales is much less than in the USA. For example, the average time served for robbery in England and Wales in 1997 was 19.3 months, whereas the average time served for robbery in the USA in 1996 was 37.4 months (Farrington et al., 2004).

Because Eqs. (1) and (2) fitted the data of five independent cohort samples and one cross-sectional sample very well, this suggests that these are not only good models of the data but also a good representation of the processes generating the data. It is, however, important that the parameters of the models have real-world meaning and that the processes described are plausible in the context of the individuals and events creating the data. Arbitrary equations (e.g. quadratic, cubic, or higher-order polynomials) can fit continuously varying data as closely as we please.[1] For the equations to be theoretically useful, a plausible theory is required that generates the same equations and parameters that relate to measurable quantities.

2.2 The MacLeod et al. Theory

The underlying theory developed in MacLeod et al. (2012) is a categorized theory of criminal convictions. A basic legal premise of criminal convictions is that individuals are responsible for their actions and that they decide to commit or not to commit a criminal act. There are exceptions to criminal responsibility, principally if a person is underage or of unsound mind at the time of an offence. In this theory, it is proposed that individuals are more or less inclined to break

[1] Quote by physicist Dyson Freeman: ‘Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk’.