Adv. & Appl. in Stat. **7**(1) (2007), 1-46

COMPARISON OF ADULT OFFENSE PREDICTION METHODS BASED ON JUVENILE OFFENSE TRAJECTORIES USING CROSS-VALIDATION

DAVID M. DAY[1], IRENE BEVC[2], THIERRY DUCHESNE[3], JEFFREY S. ROSENTHAL[4], LIANNE ROSSMAN[2] and FRANCES THEODOR[2]

[1]Department of Psychology, Ryerson University
350 Victoria Street, Toronto, Ontario, Canada M5B 2K3

[2]Program Evaluation Services
Hincks-Dellcrest Centre, Toronto, Ontario, Canada M3M 2X4

[3]Department of Mathematics and Statistics
Laval University, Québec City, Québec, Canada G1K 7P4

[4]Department of Statistics
University of Toronto, Toronto, Ontario, Canada M5S 3G3

Abstract

Considerable research has found support for the relationship between criminal offending in adolescence and criminal offending in adulthood. Estimating the strength and nature of the relationship has been facilitated by the methodological advances that have been made over the past decade. We add to this literature and describe and apply various prediction methods to examine the extent to which adult (ages 18-33 years) criminal offense trajectories can be predicted by juvenile

2000 Mathematics Subject Classification: Primary 62P25; Secondary 62-07, 62J12.

Keywords and phrases: criminal trajectories, developmental transitions, juvenile offending, persistent offending.

This research was supported with grants to the first and second authors from the Hincks-Dellcrest Centre and Ryerson University and a Natural Sciences and Engineering Research Council (NSERC) grant to the fourth author.

Received September 5, 2006

(ages 9-17 years) offense trajectories. These methods include conventional models based on latent Poisson classes (LPC) and generalised linear models (GLM) and more sophisticated Cox proportional hazards models that predict entire adult offense timelines. We also present a novel method, based on the exponential distribution, for adjusting the observed offense patterns for time-at-risk using secure custody information and a method for addressing the problem of the offense-conviction date lag. In addition, we discuss how to compare the accuracy of different prediction methods using cross-validation, thus providing a clear, unambiguous measure of prediction accuracy. We apply our methods to a data set comprising 378 male offenders in Toronto, Canada, whose criminal careers were tracked for an average of 12.1 years. Our results show that, for these data, no method can yield very accurate predictions. On the other hand, some prediction methods are able to make better use of pre-18 information to improve the precision in the predictions.

1. Introduction

The prediction of adult criminal offense patterns from adolescent criminal offense data presents many statistical challenges, from modeling criminal career patterns, to extracting usable data, to accounting for time at risk and dealing with unknown time lags between offense commissions and offense convictions. It is also difficult to assess the prediction results obtained, and to determine which of various statistical approaches is most accurate and appropriate. In this paper, we offer a number of novel statistical methodologies to deal with these various issues.

1.1. Background about criminal careers

Important advances in the past two decades have brought about greater conceptual clarification and empirical support for a dynamic and developmental approach to the study of criminal behavior. According to Loeber and LeBlanc [32], developmental criminology is concerned with within-individual changes and continuities in criminal behavior in, for example, offense mix, variety, and degree of severity. The developmental criminology perspective focuses on explicating the factors that give rise to the onset of the behavior (i.e., issues of causality) and the factors that are associated with a particular course of criminal activity over time (i.e., issues of continuity and change). In this regard, it is understood that the unfolding of an individual offender's criminal trajectory is a dynamic process that is subject to environmental and personal influences and their interaction (Andrews and Bonta [1]). This notion is in keeping with one of the main tenets of developmental psychology that the individual and environment are interdependent (Mash and Wolfe [33]). Moreover, a criminal trajectory comprises an onset, referred to as *activation*, a period of *aggravation* or a "developmental sequence of diverse forms of delinquent activities" (Loeber and LeBlanc [32, p. 382]), and termination or *desistance*. Last, this perspective is chiefly concerned with the period up until adolescence, during which time the individual experiences tremendous developmental changes that affect the onset and pattern of criminal behavior.

In a similar vein, Blumstein et al. [5] provided a conceptual and empirical framework for the criminal career paradigm. Unlike the developmental criminology notion, the criminal career approach reflects a lifespan perspective, concerned with the study of the nature and pattern of criminality across the entire life of the individual, theoretically, from the "womb to the tomb," as it were. According to Blumstein et al. [5, p. 12], a criminal career is "the longitudinal sequence of offending committed by an individual offender" that is characterized during a lifetime by three components, an onset or initiation, a termination or end, and a duration or career length (Blumstein et al. [4]).

During their careers, offenders may display changes and continuities in criminal activity on a variety of dimensions, including the rate, type, timing, and severity (Thornberry [66]). It is the pattern of transition and stability on these sorts of variables, at the level of the within-individual trajectory, across different developmental periods, as well as the underlying reasons for the observed patterns, that is of interest to researchers, theoreticians, practitioners, and policy makers. As Piquero and Mazerolle [52, p. viii] stated, a criminal career perspective "allows for an understanding of the initiation, continuation, and termination of offending behavior across the lifespan…and presents unique opportunities for developing a comprehensive understanding of criminal behavior."

While the notion of a criminal career is neither novel nor new (e.g., Shaw [59], Sutherland [65]), the current Zeitgeist has been led by various theorists and researchers who have expounded on the need for a dynamic and developmental approach to understanding crimes and criminals across the lifespan (Piquero and Mazerolle [52]). This perspective represents a significant departure from the more static theories, such as that of Gottfredson and Hirschi [24], who maintain that criminal activity throughout the lifespan is a function of a single, unchanging dimension or general propensity. In contrast, two major propositions of the dynamic life-course perspective are that past criminal behavior increases the probability of future criminal behavior and that different factors (e.g., family interactions, peer group) exert their influence at different stages of the criminal career (Nagin and Farrington [41]). Considerable research has supported these conjectures and a number of theories have been put forth to describe the processes that account for the continuities and changes in offending over time (e.g., Moffitt [37], Patterson and Yoerger [45], Sampson and Laub [56]).

The collective effect of the life course/developmental perspective has been to bring to the forefront important questions about changes and continuities in the pattern and nature of criminal behavior over time and about the dynamic processes that bring about this stability or change (Piquero et al. [51], Brame et al. [7]). These issues are of particular relevance to the chronic offender whose criminal career often begins at an early age and persists into adulthood. Chronic offenders are known to account for a large number of criminal convictions, commit serious violent offenses, and pose the greatest challenge to the criminal justice system (Piquero et al. [51]). Understanding their developmental trajectories could facilitate the development of more effective criminal justice policy regarding incarceration and treatment and rehabilitation programs.

However, the research on criminal trajectories is not without its challenges, particularly with respect to methodological strategies. As Lattimore et al. [29, p. 37] remarked, “In recent years, much attention has been devoted to developing appropriate analytical methods to model criminal careers.” A particular problem, as noted by Nagin and Tremblay [43], relates to the need for developing methods, including statistical criteria, to calibrate the adequacy of the group-based approaches. An aim of the present study is to contribute to this body of knowledge and present a novel method of addressing questions of “crime over time.”

1.2. The relationship between adolescent and adult offending

One of the most enduring questions of the developmental approach concerns the relationship between adolescent and adult offending. How are offenses committed during adolescence and adulthood linked and is there more overlap than difference? These questions have important implications for conceptualizing lifespan developmental processes. Developmental researchers in both psychology (Lerner et al. [31], Petersen and Leffert [46], Schullenberg et al. [57]) and criminology (Bottoms et al. [6], Johnson et al. [28]) concur that much can be learned about the continuities and discontinuities in the life of individuals by examining the course of behavior across the transition from adolescence to adulthood, a time when life paths become more sharply focused.

Like all developmental transitions, moving from adolescence to adulthood affords both opportunities and challenges to be negotiated by the individual. For the most part, the transition is navigated quite well. However, for some individuals, this transition is experienced as highly stressful and overwhelming (Petersen and Leffert [46]). For example, it has been suggested that individuals tend to respond to developmental transitions with a decrement in adaptation and functioning, which results in a lowered self-evaluation and heightened feelings of incompetence (Stewart [63]). These negative feelings persist until the person is able to consolidate the new roles and expectations and demonstrate a renewed sense of resilience. However, for some individuals, factors may conspire against such a normative developmental process. Two factors, in particular, that may affect the successful transition across developmental periods are the *timing* and *number* of simultaneous transitions experienced by the individual (Graber et al. [25]). In general, the premature timing and an increased number of transitions can pose difficulties for the individual, compromising his or her ability to cope with the vicissitudes of the emerging and subsequent developmental periods.

It is further suggested that involvement in serious antisocial behavior during adolescence, particularly if it begins at an early age, is protracted, and involves contact with the criminal justice system, may lead to a disruption in the normative developmental processes, bringing about a premature transition from adolescence into adulthood and a concomitant redefinition of roles and contexts (e.g., being processed as a “criminal,” making court appearances, spending a great deal of time with police, correctional, probation, and parole officers, and so forth) (Johnson et al. [28], Sullivan [64]). This non-normative process also leads to an increase in the number of transitions and non-normative stressors with which the person must contend (Petersen and Leffert [46]). The resultant effect is to impede the young person’s ability to accomplish the normative developmental tasks of adolescence, such as completing school, developing positive peer relations, and forming a healthy and integrated sense of self (Masten and Coatsworth [34]). The cumulative impact is a continued disruption in normative functioning that can interfere with the person’s ability to develop the requisite skills and capabilities to assume the socially accepted roles and expectations of adulthood. This process can result in an increased likelihood of maintaining the criminal activity into adulthood, as opportunities for completing high school and entering the labour force are diminished.

At the same time, caution must be exercised in describing these outcomes, as developmental trajectories are meant to be understood as probabilistic not deterministic (Dumas and Nilson [14]). Considerable plasticity in adaptation and adjustment allows for both continuity and discontinuity in developmental outcomes. This opens up the possibility for rehabilitative efforts to provide missed opportunities for at-risk youth and youth in contact with the justice system to facilitate their positive growth and development. Ideally, such intervention strategies are informed by a thorough understanding of development trajectories of offending behavior across developmental periods, such as childhood to adolescence and adolescence to adulthood.

1.3. Stability of offending from adolescence to adulthood

It is generally accepted within the literature that there is considerable continuity in criminal activity from one developmental period to another. As Farrington [21, p. 73] observed, “in general, the antisocial child tends to become the antisocial teenager and the antisocial adult.” As stated earlier, this homotypic continuity may be the result of a failure to achieve the normative developmental tasks of adolescence. At the same time, when examined in further detail, using different analytical tools and groups with different rates of offending, and adding covariates into the model for greater precision, the answer becomes less clear and the question of the relationship between adolescent and adult offending remains open (Piquero et al. [51], Sampson and Laub [55]).

With regard to the stability of criminal activity across developmental periods, Farrington [20] reported that 45% of those convicted as adolescents in the Cambridge Study in Delinquent Development (Farrington and West [22]) were reconvicted at ages 25 to 32 years. Wolfgang et al. [67] found that 39% of their sample had criminal convictions in both adolescence and adulthood, and Bushway et al. [10] indicated that 53.9% of their sample from the Rochester Youth Development Study (RYDS) offended both before and after age 18 years. Similar findings were observed by others (McCord [35]; Stattin and Magnusson [62]). The rate of continuity from adolescence to adulthood for specific types of offenses, including substance use and aggression (Farrington [18, 19]) and theft and property damage (Wolfgang et al. [67]) has been found to be similarly high. However, as these findings reflect essentially static rates of stability across developmental periods (Bushway et al. [10]), they tell us little about the patterns of within-individual covariation of offending trajectories between adolescence and adulthood engendered by the developmental approach.

Wolfgang et al. [67] further examined the relationship between adolescent and adult offending. Using multiple regression, they found that juvenile arrest frequency was a significant predictor of adult arrest frequency, controlling for socioeconomic status and race. Second, the relationship between specific offense types (e.g., property, violent) across developmental periods was examined as a Markov chain process using transitional probability matrices to determine whether the probability of a subsequent event is the same or different than the previous occurrence of the event. Once again, the finding of a consistency across the transition from adolescence to adulthood was confirmed.

At the same time, the correlations are not unity, indicating that the pattern of offending over time includes some change (i.e., not all antisocial youth become antisocial adults). For example, Piquero and Buka [50] found that while having a chronic offender status as a juvenile (chronicity was defined as more than six arrests) predicted chronic adult offender status, violent offending in adolescence was unrelated to violent offending in adulthood. The authors speculate that this may be due to the tendency for juveniles to engage in a wide range of offense types (i.e., versatility), including some violence, and for adults to show greater specialization in all offense types. To be sure, analytical tools need to be able to capture the complex patterns of stability and change in criminality across developmental periods and make full use of the longitudinal data.

In this paper, we compare conventional prediction methods based on latent Poisson classes (LPC) and generalised linear models (GLM) with another method, based on Cox proportional hazards models. Our particular focus is on the extent to which adult offense conviction patterns can be predicted from adolescent offense conviction patterns. More specifically, we consider what information about offense convictions before age 18 can be used to predict offense convictions after age 18. That is, we investigate to what extent adult offense patterns can be estimated, based on juvenile offense data. As well, two further methodological issues are addressed in the analyses, accounting for time-at-risk (Eggleston et al. [17], Piquero et al. [51]) and accounting for a time lag in our official criminal records between the date of the offense and the date of conviction (Francis et al. [23], Porter et al. [53]).

2. Statistical Methodology

In this section, we examine several statistical approaches available for investigating longitudinal trajectories of crime. These include latent Poisson classes (LPC), generalised linear models (GLM), and Cox proportional hazards regression. We also describe a method for assessing the validity of the prediction methods using cross-validation.

A wide variety of pre-18 offense conviction information is available in our Toronto data set for analyzing longitudinal criminal trajectories. These include the total number of pre-18 convictions, the number of pre-18 convictions of each of five offense types (property, violent, drug, sex, technical violation), the number of convictions (total, or of a specific type) between specified pre-18 ages (e.g., between ages 14 and 16, or between ages 16 and 18), the age of first offense, the age of first drug offense (if before age 18), and so forth.

Similarly, there is a variety of items that we may wish to predict about post-18 offense conviction behavior. We could predict the total number of adult offense conviction dates, the total number of adult offense conviction dates of a specific type, or the number of conviction dates (total, or of a specific type) between specified post-18 ages (e.g., between 18 and 20 or between 20 and 24). Likewise, we could attempt to predict an entire post-18 offense conviction timeline, that is, a full curve of the cumulative number of offense convictions as a function of age.

In the present paper, we concentrate on predicting the total number of adult offense conviction dates. However, we also consider a Cox proportional hazards model that attempts to predict entire post-18 offense conviction timelines. Given the variety of statistical methods and models available for our prediction problem, we consider only certain methods here, which seemed most appropriate for the problem at hand. There are, of course, many other approaches that could be taken.

2.1. Latent Poisson classes (LPC)

Recent methodological developments in the analysis of longitudinal data (e.g., Bushway et al. [10], Eggleston et al. [17], Nagin [40]) have lead to increased sophistication and precision with which to explore the nature and pattern of criminal trajectories across developmental transitions. For example, Paternoster et al. [44] and others (e.g., Bushway et al. [9], Piquero et al. [49]) consider the use of latent Poisson classes (Nagin [40]). These researchers assume that each individual has a criminal propensity that is unobserved, and that this propensity is used to define J latent classes. It is assumed that the individuals in class $j\,(j = 1, \ldots, J)$ have pre-18 total conviction counts distributed as Poisson (δ_j), and post-18 total conviction counts distributed as Poisson (λ_j), where the δ_j and λ_j are unknown. They then suppose that each individual is in one of the J classes, with unknown prior probabilities π_j, $j = 1, \ldots, J$.

Letting C_i denote the total number of pre-18 conviction dates of individual i, this model gives rise (using the definition of the Poisson distribution) to a pre-18 likelihood function

$$L_{pre} = \prod_{i=1}^{n}\left[\sum_{j=1}^{J} \pi_j e^{-\delta_j}(\delta_j)^{C_i}/(C_i)!\right].$$

Given this likelihood, the δ_j and the π_j are then estimated by maximizing the likelihood L_{pre} (subject to the constraints that $\delta_j \geq 0$, $\pi_j \geq 0$, and $\sum_{j=1}^{J} \pi_j = 1$).

Once the δ_j and π_j are estimated, then the model gives a posterior probability q_{ij} that individual i is in class j, given by

$$q_{ij} = \frac{\pi_j e^{-\delta_j}(\delta_j)^{C_i}/(C_i)!}{\sum_{k=1}^{J} \pi_k e^{-\delta_k}(\delta_k)^{C_i}/(C_i)!} = \frac{\pi_j e^{-\delta_j}(\delta_j)^{C_i}}{\sum_{k=1}^{J} \pi_k e^{-\delta_k}(\delta_k)^{C_i}}.$$

Letting D_i denote the total number of post-18 conviction dates of individual i, the q_{ij}, then give rise to a post-18 likelihood function

$$L_{post} = \prod_{i=1}^{n}\left[\sum_{j=1}^{J} q_{ij} e^{-\lambda_j}(\lambda_j)^{D_i}/(D_i)!\right].$$

Given this likelihood, the λ_j are then estimated by maximizing the likelihood L_{post}. The final prediction of this model, then, is that the probability that individual i will have precisely d post-18 convictions (for $d = 0, 1, 2, \ldots$) is given by

$$P[i \text{ has } d \text{ post-18 convictions}] = \sum_{j=1}^{J} q_{ij} e^{-\lambda_j}(\lambda_j)^{d}/d!,$$

where the λ_j and the probabilities q_{ij} are as estimated above.

For their data, Paternoster et al. [44] show that this model with 3 latent classes (i.e., $J = 3$) gives a good fit for the *frequency* of adult conviction counts over the population (at least on their Cambridge data).

That is, they accurately predict what *fraction* of adults in their sample will have zero adult offense convictions, or one, or two, and so forth. However, this is quite different from the question of whether this model provides good predictions for *individual* offense counts (i.e., the number of adult offenses that each *individual* will commit), which is our focus here. To predict individual offense counts, we use the usual point estimate for this model, namely the predicted *mean*, given by

$$\text{LPC Estimated number of post-18 convictions} = \sum_{j=1}^{J} q_{ij}\lambda_j. \tag{1}$$

We discuss in Subsection 2.5 the question of how to select an optimal number of J latent classes. In Section 5 we investigate how good an estimate we obtain by this LPC method.

2.2. Generalised linear models (GLM)

We next consider a Poisson regression model. For each individual i, we write Y_i for the total number of post-18 offenses and write $x_{i1}, \ldots, x_{ip}$ for the pre-18 covariate information. We then assume that the relationship between Y_i and the covariates is given by a Poisson regression model. Specifically, given the covariates $x_{i1}, \ldots, x_{ip}$ the Y_i independently follow Poisson (μ_i) distributions, where $\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})$, or equivalently $\log \mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$.

Estimation of the regression coefficients $\beta_0, \ldots, \beta_p$ is then done using a maximum likelihood procedure (readily available in most statistical packages including R, S-Plus, SAS, and GLIM). Once estimates of the regression coefficients $\hat{\beta}_0, \ldots, \hat{\beta}_p$ have been obtained, a prediction for the number of post-18 offenses for individual i is given by $\hat{\mu}_i = \exp\{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}\}$.

This model leads to many choices. Most obviously, what covariates should be considered? We have access to considerable information about the pre-18 convictions, such as the individual's age at conviction and which offense types are represented in each conviction. Thus, we can use such covariates as the total number of pre-18 conviction dates (TotConv), the age of the first conviction (Age1st), the number of pre-18 conviction dates involving charges of type drug (TotDrug), the total number of conviction dates between the ages of 14 and 16 (Tot1416), and so on. We can also consider more specific covariates such as the number of convictions of type Violent between the ages of 16 and 18 (Violent1618), and so forth. (It is also possible to adjust the covariates for time-at-risk, as described in Section 3 below.) In general, using too many covariates may lead to overfitting problems (discussed further below) and to a lack of interpretability, while too few covariates may fail to exploit the detail in the available data.

For ease of understanding and interpretability, for the most part, we chose our covariates using our own judgement about what quantities appropriately summarised the pre-18 conviction information. However, we also conducted a more systematic search for the best model, using the backward elimination procedure. We started with a model with every potential covariate present, and then removed the least significant covariate, provided its significance level did not exceed a pre-selected retention level (we tried with levels 10%, 5% and 1% to obtain a few candidate models). We then fitted the remaining model and eliminated the least significant covariate. We repeated these elimination steps until we were left with a model whose covariates all had an associated significance level whose value was less than that of the retention level.

The backward elimination procedure described above was run using the GENMOD procedure in SAS. We also ran a backward elimination procedure using the stepAIC function available in the statistical package R. This latter algorithm functions in the same manner as the classical backward elimination procedure except that the algorithm does not stop when every variable in the model has a significance level inferior to the preset retention level, but rather when the Akaike Information Criterion (AIC) of the model starts to increase.

Another factor to consider is whether (and how) to *stratify* the population into distinct subgroups, on the basis of their pre-18 offense data. In general, too much stratification leads to groups that are too small to detect patterns, while too little stratification might force, into one statistical model, offender types with vastly different post-18 offense patterns. We decided to consider stratifications of the population into various numbers (typically between one and eight) of different subgroups, on the basis of total number of pre-18 offense conviction dates. Given the tremendous number of different GLM models available to us, we needed to determine which ones led to the best predictions. For this we used a cross-validation criterion, as discussed in Section 5 below.

Finally, we note that as a more general GLM, we could consider an *overdispersed* Poisson regression model, which assumes that the Y_i's are independent with $E[Y_i|x_{i1}, ..., x_{ip}] = \mu_i$ and $\mathrm{Var}[Y_i|x_{i1}, ..., x_{ip}] = \phi\mu_i$. (The Poisson case corresponds to $\phi = 1$.) In this model, the parameter ϕ is the dispersion parameter (for further discussion of this issue, see Francis et al. [23]). When $\phi > 1$ the model is overdispersed (i.e., the variability in the Y_i's is greater than Poisson variability). For our data, we observe that all such models have significant overdispersion ($\phi \approx 5$), which is not surprising since extra Poisson variability seems to be the norm when modeling offense counts (e.g., Paternoster et al. [44]). Now, this overdispersion may affect the choice of model through the backward and AIC elimination procedures. However, since it affects only variances and not means, it will *not* change our estimates for a particular model.

2.3. Cox proportional hazards regression

As a final issue that we consider, it would be of interest to determine whether the entire post-18 offense trajectory (timeline), rather than just the post-18 offense count, could be predicted from the pre-18 information. One possible approach is Cox Proportional-Intensity (i.e., non-parametric time-inhomogeneous Poisson) Regression Models, as suggested by Day et al. [13].

Cox Proportional-Intensity models estimate the cumulative intensity rate, say, $\Lambda_i(t)$, corresponding to the expected number of offense dates for individual i between ages 18 and t, for $t \geq 18$. The estimate is of the form

$$\hat{\Lambda}_i(t) = \hat{\Lambda}_0(t)\exp(\beta'\mathbf{x}_i),$$

where $\hat{\Lambda}_0(t)$ is a *baseline* cumulative intensity rate, which is estimated non-parametrically from the data, and is the same for all individuals; and $\mathbf{x}_i$ is a list of pre-18 offense covariates for individual *i*; and β is a vector of regression coefficients, to be estimated parametrically.

Thus, in the end, the Cox model provides a complete trajectory $\hat{\Lambda}_i(t)$ for each individual *i*, predicting their number of offenses by each age *t*. This is more ambitious than the LPC and GLM, which attempt to predict only the total number of adult offenses. However, there is some relation: with the Cox model, the predicted total number of adult offenses is then given by $\hat{\Lambda}_i(\infty) - \hat{\Lambda}_i(18)$. It is then possible to compare such predictions with the true number of observed adult offenses for each individual. We would expect such predictions to be *worse* than those of GLM, since GLM is specifically designed to predict the total number of adult offenses, while the Cox model is attempting to predict entire trajectories. We investigate this question further below. As with GLM, Cox models allow for many choices in terms of what covariates are considered, whether and how to stratify the population, and so forth. We consider a variety of different options herein.

2.4. Overfitting and information criteria

We address here the issue of assessing the fit of prediction methods. Evaluating prediction methods is a subtle issue, since it is always possible to find a model that fits the available data extremely well, but at the expense of *complicating* the model greatly. In the most extreme case, a model could use so much pre-18 offense information as to *uniquely identify* every individual in the study. One might then obtain an overly specific rule, such as, “If you have two property offenses precisely at ages 15.8913 and 16.7759, and just one drug offense at precisely age 17.4326, then you will commit precisely seven offenses as an adult.” enough such rules, derived from an exhaustive analysis of all the pre- and post-18 offense data, might precisely describe the available adult-offense data; but there would be no reason to think that the rules so generated would generalise in any way to new young offenders.

This is the problem of *overfitting* a model. With enough covariates, one can describe observed data quite well, but the resulting conclusions will be entirely determined by the details of one's available data, with no ability to generalise. So the question becomes how to determine how many covariates are too many? That is, how can we determine which covariates are useful to obtain a good prediction model and which are too specific to our actual data and should be discarded?

It has been proposed that *Information Criteria* can be used to control overfitting. The *Bayesian Information Criterion* (e.g., Brame et al. [8], d'Unger et al. [16], Nagin and Land [42], Schwarz [58]) is defined as

$$\mathrm{BIC} = -2\log(L) + n\log(k),$$

where L is the likelihood function, n is the number of individuals being studied, and k is the number of parameters in the model. Similarly, the *Akiake Information Criterion* (e.g., Sakamoto et al. [54]) is defined as

$$\mathrm{AIC} = -2\log(L) + 2k.$$

In either case, it is argued that *minimising* the Information Criterion leads to the best model. The intuition is that including more parameters can lead to a better fit, and hence increase L, but at a penalty of also increasing k; and minimising the BIC or AIC is an attempt to balance these two effects and thus avoid overfitting.

While these information criteria do have some theoretical justification, they are only indirect measures (or approximations) of the overfitting problem. Furthermore, their application in criminology is somewhat inconsistent. For example, Eggleston et al. [17, p. 506] note that:

"Although the Bayesian Information Criterion has been emphasized as the primary criterion to assess the optimal number of groups, the model selection process is often more complex and thus, group selection remains somewhat subjective. As Nagin and Land note in their original article on this subject, the groupings may be seen as only an approximation of a postulated underlying continuous dimension of hidden heterogeneity in offending propensity [42, p. 357]. Since these groupings are abstractions or approximations and not a true reflection of reality, researchers tend to use the BIC as one criterion for choosing the number of groups, but not the sole criterion. For instance, Brame et al. [8] find a six group model to be the optimal model based on the BIC for their childhood aggression analysis and yet they describe the four-group model because the results from this more parsimonious solution are qualitatively similar."

The basic problem is that the Information Criteria approach is an *indirect* attempt to compare prediction methods and deal with the overfitting problem. A more direct method is cross-validation, as we now discuss.

2.5. A fair comparison: cross-validation

Given the multitude of methods, models, covariates, stratifications, and so forth, that are available for predicting post-18 offense patterns from pre-18 offense data, it is important to have some method of *comparing* different prediction methods, to determine which appear to be most accurate. To properly assess the validity of a prediction model, it is necessary to distinguish between those data that are used to *fit* the model and those data that are subsequently used to *test* the model. Ideally, one would have *two* large samples of data. A model would be developed using the first sample and then assessed as to how accurately the model predicts the observations in the second sample. If such predictions are accurate, then the model is likely a good one. However, if such predictions are highly inaccurate, then perhaps the model involved overfitting or other problems that allowed it to model the *first* set of data well, but not to make accurate predictions on the second (i.e., fresh) set of data.

Of course, in practice, it is difficult to obtain one large sample of data to fit a model, and usually a second "test" sample is not available. However, one way around this problem is suggested by *cross-validation* (e.g., Hjorth [27]). The idea of cross-validation is that one individual, i, is temporarily *excluded* from the data. The proposed model is then fit using all the *other* individuals. Subsequently, the accuracy of the fit to predict the behavior of individual i and how much error results is computed. The cross-validation prediction error is then the average of these errors, averaged over all individuals i.

To be more precise, for a particular prediction method and model, write Pred(i) for the predicted total number of adult offenses of individual i, after fitting the model by temporarily *excluding* individual i. And, write Obs(i) for the actual total number of adult offenses of individual i. Then the *cross-validation error* of the model is given by

$$\text{cross-validation error} = \frac{1}{n}\sum_{i=1}^{n} |\,\text{Pred}(i) - \text{Obs}(i)\,|. \tag{2}$$

When comparing two models, the one that has a *smaller* cross-validation error should be considered superior. In this way, we can compare, not only different modeling paradigms (latent Poisson classes versus GLM, or GLM versus Cox models), but also different covariate choices (e.g., distinguish between different offense types or not, consider the age of offenses or not, etc.), and also different stratifications (e.g., divide the sample into high-rate and low-rate offenders, or drug and non-drug offenders, or not). In principle, cross-validation can be used to compare the accuracy of any two prediction methods. And, since it directly measures the prediction accuracy on individuals who were *not* used to fit the data, it is not fooled by overfitting (i.e., an overfit model will lead to a large cross-validation error).

We close with two remarks. First, it is true that cross-validation is somewhat computer intensive, since the entire model must be re-fit n times, once for each choice of individual i to be excluded. However, we have not found this overly burdensome. With our data of $N = 378$ individuals, on a standard personal computer, performing a complete cross-validation analysis typically takes less than one minute for GLM, and 10-30 minutes for Cox models. Second, the statistic given by equation (2) is an L1 cross-validation measure. It is also possible to define an L2 *cross-validation error*, by

$$\text{L2 error} = \frac{1}{n}\sum_{i=1}^{n} (\text{Pred}(i) - \text{Obs}(i))^2.$$

Either L1 or L2 leads to fair comparison of different methods. However, the L2 error is overly sensitive to the existence of a few very inaccurate predictions, so we consider only the L1 error herein.

2.6. Adjusting for time-in-study

Our data collection ceased at a particular point in time and not all offenders were at the same age at that time. Furthermore, some individuals had left the country, died, were deported, or were otherwise unable to be traced for the entire study period. Thus, for each individual i, we have an age L_i at which they left the study and a corresponding number of years $T_i = L_i - 18$ after age 18, during which they could potentially commit adult offenses for the study.

It is possible to take account of this time period T_i, as follows. When fitting models, we divide the post-18 offense counts by T_i to obtain post-18 offense frequencies and fit our model to this modified data. Then, when making final adult-offense predictions, we multiply our predicted post-18 offense frequency by T_i to obtain a predicted post-18 offense count, which can then be compared to the observed post-18 offense count.

For the generalised linear models, this adjustment for time in study (T_i) can be accomplished particularly simply, by setting $\log \mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \log T_i$, where now $\log T_i$ is referred to as an offset term. For Cox models, the situation is even simpler: predicting the complete timeline means that adjustment for time in study is accomplished automatically. In our comparisons below, for each model we consider both options: adjusting for time in study or not. We shall see that usually adjusting for time-in-study leads to significant improvements. We next turn to a more complicated form of adjustment of variables.

3. Adjusting for Time-at-risk

It is highly unlikely that an individual will be charged with a new set of offenses that were committed while in secure custody. It is only while *not* in secure custody that an individual is "at risk" of offending. Piquero et al. [51] use the term "street time" to refer to this notion. Arguably, our prediction models should take this factor into account by adjusting the covariate and predictor variables to measure the rate of offending while at risk. For example, if an individual commits five crimes between the ages of 16 and 18 and was in secure custody for a total of one full year between the ages of 16 and 18, then their effective offense rate during that time period was twice as large as it would have been had they been at risk the entire time.

It seems plausible that we will obtain more accurate predictions of adult conviction patterns if we take such factors into account. Indeed, Eggleston et al. [17], see also Piquero et al. [47], Piquero et al. [51] argue that prediction models that fail to take into account the time during which the offender is incarcerated lead to inaccurate estimations of criminal trajectories. More specifically, ignoring this information could result in an underestimate of an individual offender's criminal propensity.

Unfortunately, the data in the present study do not allow us to make such adjustments directly. On the positive side, we do have accurate timelines for each individual's dates in secure custody. On the negative side, we do not have access to the actual dates corresponding to when the offenses were committed, only the dates corresponding to when the convictions were recorded. As noted by Porter et al. [53], "criminal records for individual offenders give the date of adjudication, not the date of crime commission" (p. 658). The time lag between the date of offense and the date of conviction is unknown, making it impossible to directly combine the data about conviction dates with the data about time-at-risk.

While Porter et al. acknowledge this issue as a problem, they did not address it in their study of recidivism and psychopathy as a function of age at offense and simply substituted the date of conviction for the date of offense. Francis et al. [23] also cite this issue as a difficulty with official criminal records. However, they were able to determine that, based on figures from the U. K. Crown Court for the period between 1991-2001, the time lag between offense committal and an offender's trial ranged from 12-17 weeks, though they note that this time-lag may not relate to all cases, only the most serious. In this section, we develop a model that takes into account the random lag time Δ between offense commission and offense conviction, in an effort to more accurately adjust our variables for the individual's time-at-risk.

3.1. A model for time-at-risk

We assume that an individual i has a rate of offending at age t which is given by a *product* of their propensity to commit crimes, $p_i(t)$, times an "availability to commit" indicator variable $W_i(t)$ which equals 0 while individual i is in secure custody, otherwise equals 1. That is,

$$i\text{'s offense commission rate at age } t = p_i(t)W_i(t).$$

In our data, we have access to the dates at which an individual is in secure custody, so we can easily compute $W_i(t)$. Unfortunately, we do not know the dates at which crimes are committed, which makes it very difficult to directly estimate $p_i(t)$.

Our solution is to transform this problem to one involving conviction dates, about which we have more data. That is, rather than estimate the propensities $p_i(t)$ to commit crimes, we instead estimate the propensities $\lambda_i(t)$ to obtain offense *convictions*. We assume that

$$i\text{'s offense conviction rate at age } t = \lambda_i(t)Z_i(t),$$

where now $Z_i(t)$ represents individual i's "availability for conviction" at age t. We take $Z_i(t)$ to be the *probability* that individual i was at risk (i.e., not in secure custody) at the actual offense age, $t - \Delta$.

Since the time lags Δ are unknown, we model them as random variables, each having an exponential distribution with mean $T = 90$ days, a figure which is consistent with that reported by Francis et al. [23]. That is, we assume that $\Delta \sim \text{Exponential}(T)$, with density function $e^{-s/T}/T$ for $s > 0$. (Thus, for a given offense, the lag time could be much smaller than 90 days or much larger. In particular, our model takes account of the fact that the lag time Δ is unknown and hence treated as random.)

The "availability for conviction" factor $Z_i(t)$ is then the *expected value* of $W_i(t - \Delta)$, where $\Delta \sim \text{Exponential}(T)$. Thus,

$$Z_i(t) = E[W_i(t - \Delta)] = \int_0^{\infty} W_i(t - s)\frac{e^{-s/T}}{T}\,ds.$$

Note that $Z_i(t)$ is always between 0 and 1. If $Z_i(t) \approx 0$, this means that the individual was in secure custody for the vast majority of the time preceding age t, and thus was virtually unavailable to have a new conviction at age t. If $Z_i(t) \approx 1$, this means that the individual was at large for the vast majority of the time proceeding age t and thus was almost completely available to have a conviction at age t.

3.2. Computing the availability for conviction

We compute $Z_i(t)$ explicitly as follows. Let individual i have time periods of secure custody given by $(a_1, s_1), (a_2, s_2), \ldots, (a_K, s_K)$, where $a_1 \leq s_1 \leq a_2 \leq \cdots \leq s_K$. (That is, the a_i are the times of arrival in secure custody, and the s_i are the times of release from secure custody.) Then we compute that

$$Z_i(t) = 1 - \sum_{j=1}^{K} (e^{-\max(0,\, t-s_j)/T} - e^{-\max(0,\, t-a_j)/T}). \tag{3}$$

To see why equation (3) is true, note that if $t \leq a_1$, then $Z_i(t) = 1$ (of course). If for some J we have $s_J \leq t \leq a_{J+1}$ (or $t \geq s_K$ in the case $J = K$), then

$$Z_i(t) = 1 - \sum_{j=1}^{J} (e^{-(t-s_j)/T} - e^{-(t-a_j)/T}).$$

Finally, if for some J we have $a_J \leq t \leq s_J$, then

$$Z_i(t) = e^{-(t-a_J)/T} - \sum_{j=1}^{J-1} (e^{-(t-s_j)/T} - e^{-(t-a_j)/T})$$

(where $\sum_{j=1}^{0}$ is taken as 0 in the case $J = 1$). Since $e^0 - e^0 = 0$, these formulas are all, collectively, equivalent to the single formula equation (3) above.[1]

[1] In theory, we always have $Z_i(t) > 0$. However, due to roundoff error, the computer may occasionally compute erroneously that $Z_i(t) \leq 0$, which can cause problems. To avoid this, in our computations we simply replace $Z_i(t)$ by 10^{-6} whenever $Z_i(t) < 10^{-6}$.

3.3. Aggregate conviction availability factors

In the analyses below, it will be necessary to consider individual offense counts over fixed age ranges. This will be used both to define aggregate juvenile offense counts to use as covariates, and to compare predicted adult offense counts to observed counts. To correct these offense counts for time-at-risk, it is necessary to divide them by corresponding aggregate conviction availability factors over the same time periods.

The aggregate conviction availability, for individual i over the ages from A to B, is given by

$$R_i(A, B) = \int_A^B Z_i(t)dt,$$

the integral of their conviction availability factor $Z_i(t)$ over the age range from A to B. We compute from equation (3) that

$$R_i(A, B) = (B - A) - \sum_{j=1}^{K} (I(A, B, s_j) - I(A, B, a_j)),$$

where

$$I(A, B, u) = \int_A^B e^{-\max(0,\, t-u)/T} dt$$

$$= \begin{cases} B - A, & A \le B \le u, \\ Te^{u/T}[e^{-A/T} - e^{-B/T}], & u \le A \le B, \\ (u - A) + T[1 - e^{-(B-u)/T}], & A \le u \le B. \end{cases}$$

This gives a precise formula for computing $R_i(A, B)$, the aggregate conviction availability for individual i between ages A and B.

3.4. Adjusting the pre-18 and post-18 variables

From the above analysis, we can (optionally) use the $Z_i(t)$ and $R_i(A, B)$ values to modify pre-18 and post-18 conviction variables, as follows. For pre-18 aggregate variables, we can simply replace Prop1416 with $\text{Prop1416}/R_i(14, 16)$, and similarly for the other variables. For post-18 prediction, we can divide the total adult conviction count by $R_i(18, D_i)$, where D_i is the age at which individual i departed from the study.

We believe that such adjustments provide a logical, sound method of taking into account the unknown (random) lag time between offense dates and conviction dates, thus allowing the conviction data to be coordinated with the secure-custody data. In particular, we feel that this adjustment is a theoretical improvement over simply ignoring this lag time, or, for example, simply assuming that it is always equal to 90 days. In Section 5, we consider the extent to which such adjustments do or do not actually improve the accuracy of our predictions.

4. Data

The Toronto sample comprises 378 male young offenders who had been sentenced in late adolescence or early adulthood, sometime between 1986 and 1995, to one of two open custody facilities for youth operated by the Hincks-Dellcrest Centre. The Hincks-Dellcrest Centre is a children's mental health center in Toronto, Canada. This group represents a 50% random sample of the approximately 800 young males who had been sentenced to one of these youth homes during this period. The study sample was between 16.1 and 24.4 years of age $(M = 17.6,\ SD = .9)$ at the time of admission into the group home. The average sentence length was of 124.6 days $(SD = 109.8,$ range $= 1 - 1087$ days).

The criminal data were derived from *all* of their distinct convictions, temporally sequenced, that were committed up until March 17, 2001, the end of the follow-up period. Official criminal records, comprising Phase I (committed while the youth was 12 to 15 years of age), Phase II (committed while the youth was 16 to 17 years of age), and adult offenses were obtained from four sources: (a) the (Ontario) Ministry of Community and Social Services (MCSS); (b) the (Ontario) Ministry of Correctional Services (MCS); (c) the Canadian Police Information Centre (CPIC); and (d) Predisposition Reports (PDR) maintained in the clinical files by the Hincks-Dellcrest Centre. Access to the young offender records, which are confidential in Canada under the Young Offenders Act (YOA; 1984), were obtained through a court order. The court order consisted of a number of provisions designed to ensure the anonymity of the records. Steps were taken to ensure that the identifiable information in the records was kept confidential.

Four data sources were used to ensure a high degree of completeness and accuracy for the sequenced, longitudinal conviction data, which is essential for research that requires an accurate temporal sequencing of criminal convictions (Arnold and Kay [2]; Smith et al. [60]). Although the use of official criminal records has been called into question (Dunford and Elliott [15]), studies have reported a high degree of concordance between self-report delinquency and official records (Moffitt et al. [38]). As well, official records are appropriate for our purposes because they provide the requisite precision with regard to the timing and sequence of offending (Smith et al. [60]).

The criminal trajectories for the Toronto sample were tracked for an average of 12.1 years (range = 4.9-22.8, $SD = 3.0$), from late childhood/early adolescence into adulthood, with 73% of the sample being followed for 10 years or more. Their mean age at first conviction was 15.5 years (range = 8.9-21.3, $SD = 1.8$) and the sample was, on average, 27.5 years (range = 22.2-33.5, $SD = 2.6$) at the time of the most recent follow-up. Over the course of the study period, the sample amassed a total of 5,165 convictions. These included 2,387 convictions for a property offense (referring to the most serious offense at each conviction), 1,330 violent offenses, 296 drug offenses, and 1,128 "technical" offenses, which included a range of offenses, including failure to appear in court and Highway Traffic Act violations. The lengthy criminal records of many of these individuals (up to 50 successive convictions) allow for a rich, detailed analysis of their offense patterns. The length of the follow-up and the nature of the study sample is comparable to that reported in an investigation of criminal career lengths by Piquero et al. [48]. These researchers followed up a sample of 377 male offenders from the California Youth Authority (CYA) institutions for an average of 13 years, 2 months. The age of their first offense was reported to be 11.93 years (range = 5-18 years) and they were, on average, 31 years at the end of the follow-up period. Piquero et al. concluded that the length of the study period allowed them to capture "much of the time spent in criminal careers among a serious offender population, at least until the end of adolescence and through adulthood" (p. 431).

4.1. Coding procedures

For each individual, the criminal records were coded for a range of variables for each conviction arising from a new set of charges (Day [12]), including all of the criminal charges, the sentence date, length, and type (e.g., open or closed custody), and the severity of the offenses. The severity ratings were taken from the MCS Statistical Reporting System User Manual (1995). When coding offenses for a new set of charges, the most serious offense was counted. This procedure is commonly used in this type of research (e.g., Lattimore et al. [30]; Stander et al. [61]). However, in addition, for each new conviction, the complete range of criminal charges was coded into a single variable, "OffenType," which takes into account all of the charges incurred by the individual, that led to a given conviction (as much as is available on the offender's "rap sheets"), not just the most serious offense. The variable denotes, for example, for each new conviction, whether an offender is a "pure" type (e.g., property, violent, or drug offender) or a "versatile" offender and what type of versatility he expresses (e.g., violent and drug offender, property, violent and drug offender). Using the OffenType variable as the unit of analysis provides a more complete picture of a given offender's criminal tendencies. As well, *all* of the offenses incurred at each conviction were recorded, not just those resulting in a conviction. This coding practice served to avoid a potential bias introduced by plea bargaining. Last, as stated previously, a common problem encountered in much longitudinal crime research is controlling for "time-at-risk," that is, the time the offender is at risk to offend due to being "on the street" (Blumstein et al. [5]). Our data are sufficiently detailed to allow for an accurate estimation of this variable and represents an important improvement over previous studies. These coded variables for each of the temporally sequenced convictions, then, provided the criminal data for our longitudinal analyses.

5. Results and Comparisons

We applied each of our statistical methodologies to our data, with various choices of subgroup divisions, and calculated the cross-validation error in each case.

5.1. Baseline methods: mean and median

Since we are comparing various prediction methods, we shall also find it useful to define two very simplistic prediction methods to be used for baseline comparisons. One simplistic method is to compute the *mean* number of post-18 offenses of all the individuals in the sample and then boldly predict that all individuals will have this same number of post-18 offenses. For example, if the mean post-18 offense count in the sample is 4, then this prediction method would simply predict 4 as the post-18 offense count for each individual. Similarly, another simplistic method is to compute the *median* number of post-18 offenses of all the individuals in the sample and then predict that all individuals will have this same number of post-18 offenses. For example, if the median post-18 offense count in the sample is 7.2, then this prediction method would simply predict 7.2 as the post-18 offense count for each individual.

Both these methods are obviously of rather limited value since they do not take into account at all the differences between individuals, but rather predict exactly the same post-18 offense count for everybody. In other words, these methods make use of *zero* covariates. Nevertheless, we view these methods as *baseline* prediction methods in the hopes that any *good* prediction method would easily surpass them.

Applying cross-validation to the Mean predictor, we obtain a cross-validation error of 5.6727. Applying cross-validation to the Median predictor, we obtain a cross-validation error of 5.3836. We shall see that it is, indeed, true that our other prediction methods are superior to these baseline methods. However the margin of victory is not as overwhelming as one might expect.

5.2. Corrections and stratifications

We next tried correcting these baseline estimators for time-in-study, as discussed in Subsection 2.6. This reduced the CV errors to 5.279439 for the Mean, and 5.106577 for the Median, which is a significant improvement. We subsequently attempted to divide the population into various subgroups, ranging from one to eight, based on their total number of pre-18 offense dates, again adjusted for time-in-study. This led to total CV errors for the Mean of 5.279439, 4.920975, 4.749055, 4.718747, 4.695471, 4.731291, 4.721266, and 4.734808, respectively. For the Median, the total CV errors were 5.106577, 4.789283, 4.692084, 4.626262, 4.674750, 4.681342, 4.703925, and 4.677931, respectively. Thus, we see that in this case, subdividing the population produces further improvements, with four or five subgroups being optimal.

We also considered adjusting for time-at-risk by dividing and multiplying, not by T_i, but by $R_i(18, L_i)$, with R_i as in Subsection 3.3 (and where $L_i = T_i + 18$ is the age at which individual i left the study). Here $R_i(18, L_i)$ may be thought of as the value of T_i when adjusted for adult time-at-risk. However, this seemed to lead to poorer estimates. For example, for the Median with no subgroup divisions, it gives a CV error of 5.970111, which is significantly more than the 5.106577 obtained when adjusting by T_i. Thus, we did not consider the $R_i(18, L_i)$ for any further adjustment.

5.3. Poisson latent classes (PLC)

For the Poisson Latent Classes models of Subsection 2.1, the only choice is the number of J latent classes. We have considered $J = 1, 2, 3, 4, 5, 6, 7$ different classes. For each choice of J, the model provides estimates of adult offense counts. For each choice of J, we computed the CV error for the PLC model as follows. First, we estimated the π_j and δ_j values using all of the pre-18 data. Then, for each individual i in turn, we estimated the λ_j values using the post-18 data for the entire population *excluding* individual i. We then plugged in equation (1) to obtain a predicted value $\text{Pred}(i)$ for individual i. Finally, we computed the CV error by averaging the resulting prediction errors over all individuals i, as in equation (2).

With just 1 latent class $(J = 1)$, we obtain a CV error of 5.6727. Fitting one latent class to the entire population yields the parameters $\pi_1 = 1$, $\delta_1 = 5.2804$, and $\lambda_1 = 7.9391$. With 2 latent classes $(J = 2)$, we obtain a CV error of 5.4887, corresponding to the parameters $\pi_1 = 0.6923$, $\pi_2 = 0.3077$, $\delta_1 = 3.3189$, and $\delta_2 = 9.6943$, with $\lambda_1 = 4.0229$ and $\lambda_2 = 16.3398$. With 3 latent classes $(J = 3)$, we obtain a CV error of

6.2868, corresponding to the parameters $\pi_1 = 0.3231$, $\pi_2 = 0.5617$, $\pi_3 = 0.1153$, $\delta_1 = 2.1085$, $\delta_2 = 5.4883$, and $\delta_3 = 13.1493$ with $\lambda_1 = 2.6878$, $\lambda_2 = 14.5301$, and $\lambda_3 = 6.3599$. For larger values of J, the CV error fluctuates somewhat, but, due to overfitting, it is never as low as the CV error corresponding to $J = 2$. Indeed, no choice of J leads to particularly good predictions and, in fact, the resulting CV errors are no lower than that of the baseline Median predictor (see Table 1).

We then tried adjusting the LPC for time-in-study, as discussed in Subsection 2.6. In this case, with just one latent class the CV is reduced to 5.279452 (just like the Mean predictor). With two latent classes $(J = 2)$ it is reduced quite a bit further, to 4.750726. It achieves its smallest value, 4.714156, with four latent classes $(J = 4)$, before climbing for larger values of J.

5.4. Generalised linear models (GLM)

As discussed in Subsection 2.2, different Poisson regression models are available depending on what pre-18 covariates one chooses to consider. Consequently, we started with simple models, and built up to more complicated models. Initially, we used just the single covariate consisting of each individual's total number of pre-18 conviction dates (TotConv). This model (combined with the $\log(T_i)$ offset to account for time in study) gave a CV error of 4.821457, which is already a significant improvement over the previous methods. (Note that the $\log(T_i)$ offset is crucial. Without it the CV error is 5.339813.)

The corresponding model for our (complete) data is then given by

$$\mu_i = T_i \exp(-0.58234 + 0.07311 \text{ TotConv}).$$

For example, an individual who was in the study until age 30 (so $T_i = 12$), and who had 14 juvenile conviction dates (TotConv = 14) would have an expected number of post-18 conviction dates of

$$\mu_i = 12 \exp(-0.58234 + 0.07311 \times 14) = 18.65486.$$

Including the age at first offense (Age1st), together with TotConv, leads to a CV error of 4.842783 that, perhaps surprisingly, is no improvement over using just TotConv alone. (Using Age1st on its own gives a CV error of 5.058746.) By contrast, including, in addition to TotConv, the total number of pre-18 conviction dates which include a charge of type Violent (TotViolent), type Drug (TotDrug), and type Sex Offense (TotSexoff), leads to a reduced CV error of 4.759688. The corresponding model for our data is then given by

$$\mu_i = T_i \exp(-0.54896 + 0.07277 \text{ TotConv} - 0.01863 \text{ TotViolent} + 0.12253 \text{ TotDrug} - 0.18634 \text{ TotSexoff}). \tag{4}$$

In each model considered, a positive [negative] coefficient implies a positive [negative] correlation, and means that an increase in the corresponding covariate, with all other covariates keeping the same value, will produce a higher [lower] number of expected post-18 offenses. For instance, in the above model, if TotConv increases by 1 and the model's other covariates all remain unchanged, then the expected number of post-18 offenses will be multiplied by $\exp(0.07277) \approx 1.075$, that is, an increase of about 7.5%.

Removing TotConv from the model and, instead, including the total number of conviction dates corresponding to all five offense types (TotProp, Tot Violent, TotDrug, TotSexoff, TotTech) , gives a very similar CV error, 4.748343. The corresponding model for our data is then given by

$$\mu_i = T_i \exp(-0.52537 + 0.05868 \text{ TotProp} + 0.02388 \text{ TotViolent} + 0.18665 \text{ TotDrug} - 0.12526 \text{ TotSexoff} + 0.05039 \text{ TotTech}).$$

In this case, all the regression coefficients are positive, with the exception of that for TotSexoff. This corresponds to the well-known fact that juvenile sex offenses often correlate negatively with adult criminal behavior. On the other hand, it also shows that the negative coefficient for TotViolent in equation (4) was simply an artifact of the fact that TotConv was also included as a covariate there. That is, if TotViolent increases while TotConv stays the same, then this means that the number of convictions of some *other* offense type [Property or Technical] must have correspondingly decreased, which in this case causes an overall decrease (on average) in adult convictions.

In a different direction, including just the total number of conviction dates in the age ranges 0-14 (Tot014), 14-16 (Tot1416), and 16-18 (Tot1618), leads to a CV error of 4.734296, a small further improvement. The corresponding model for our data is then given by

$$\mu_i = T_i \exp(-0.64411 + 0.03311\ \text{Tot014} + 0.04869\ \text{Tot1416} + 0.10442\ \text{Tot1618}).$$

These covariates can then be combined in different ways. Including the offense type totals (TotProp, TotViolent, TotDrug, TotSexoff, and TotTech) together with the age range totals (Tot014, Tot1416, and Tot1618) leads to a CV error of 4.721162, slightly better.

Alternately, using TotConv, TotViolent, TotDrug, and TotSexoff together with Tot014, Tot1416, and Tot1618, gives a CV error of 4.699465, another slight improvement. The corresponding model for our data is then given by

$$\mu_i = T_i \exp(-0.60086 + 0.09991\ \text{TotConv} - 0.02074\ \text{TotViolent} + 0.10132\ \text{TotDrug} - 0.18164\ \text{TotSexoff} - 0.05787\ \text{Tot014} - 0.04672\ \text{Tot1416});$$

in particular, the variable Tot1618 does not contribute. Adding Age1st to that list, that is, using the covariates TotConv, TotViolent, TotDrug, TotSexoff, Tot014, Tot1416, Tot1618, and Age1st, gives a CV error of 4.697884, a value that is practically unchanged.

We also considered adjusting the covariates for time-at-risk, as in Subsection 3.4. However, these adjustments do not appear to reduce the CV error. For example, using just the at-risk-adjusted total number of pre-18 conviction dates (TotConvAdj) gives a CV error of 5.06137, which is significantly worse than the 4.821457 error from using the corresponding unadjusted covariate, TotConv. Similarly, using the time-adjusted version of the best set of covariates above, namely TotConvAdj, TotViolentAdj, TotDrugAdj, TotSexoffAdj, Tot014Adj, Tot1416Adj, Tot1618Adj, and Age1st, leads to a CV error of 5.17611, considerably *worse* than the value of 4.697884 achieved without adjusting for time-at-risk.

In a different direction, we considered stratifying the population into different subgroups based on their total number of pre-18 conviction dates. When using only the covariate TotConv and varying the number of groups from one to eight, the corresponding total CV errors are, respectively, 4.821457, 4.782300, 4.733757, 4.764169, 4.778888, 4.774936, 4.791977, and 4.796312. We thus see that stratifying the sample into three groups is optimal in this case, reducing the CV error from 4.821457 to 4.733757. However, the improvement is not that large, presumably because the TotConv covariate already takes into account the number of pre-18 conviction dates for each individual.

When using only the covariate Age1st, then, with from one to eight groups, the corresponding CV errors are, respectively, 5.058746, 4.911553, 4.810295, 4.821087, 4.842361, 4.820236, 4.912434, and 4.833099. This indicates that using three subgroups is again optimal and this time the improvement is somewhat greater, since Age1st is a different quantity from the stratification criterion (TotConv).

When using all five offense-type covariates (TotProp, TotViolent, TotDrug, TotSexoff, and TotTech) and varying the number of groups from one to eight, the corresponding total CV errors are, respectively, 4.748343, 4.756716, 4.762293, 4.884235, 4.932118, 5.020130, 5.017440 and 4.995730. Thus, in this case, stratifying the population does not decrease the CV error at all.

Similarly, when using the best selection of covariates above (TotConv, TotViolent, TotDrug, TotSexoff, Tot014, Tot1416, Tot1618, and Age1st) and varying the number of groups from one to eight, the corresponding total CV errors are, respectively, 4.697884, 4.768348, 4.778368, 4.947666, 5.100318, 5.121288, 5.305890, and 5.121633. Once again, stratifying the population does not decrease the CV error at all. In summary, using just TotConv gives a fairly reasonable CV improvement, which can be further improved by using information about offense types and/or conviction ages. However, adjusting for time-at-risk and stratifying into subgroups offers very little further improvement.

Further trial and error leads to other models, as well. It happens that, for our data, we can do somewhat better by concentrating largely on covariates corresponding to the 16-18 age range. Specifically, if we use just the three covariates, Tot1618, Sexoff1618, and Age1st, we obtain a CV error of 4.652834, which, again, is a slight improvement over our previous results. This may correspond to a "true" discovery or this very slight improvement could be just an artifact of our data.

Finally, we did a more systematic search to determine the best GLM models available. For this search, we sorted the pre-18 convictions by age range (0-14, 14-16, and 16-18) and by offense types (property, violent, drug, sex offense, and technical). We thus allowed ourselves the use of such covariates as the number of conviction dates including a property offense type between the ages of 14 and 16 (Prop1416), and so on. We considered those covariates both with and without adjustment for time-at-risk. We also considered stratification into different numbers of subgroups (1 through 12). We used backward elimination procedures, as discussed in Subsection 2.2.

Among these many GLM models, it turned out that the lowest possible CV error was 4.606149. This best predictive power was obtained by stratifying into two equal-sized subgroups, determined by the number of pre-18 offenses, adjusted for time-at-risk (as in Subsection 3.4) and summed over all offense types. Those individuals whose value of this quantity is below the median are designated as low-rate offenders, while those above the median are designated as high-rate offenders.

For this stratification, the model then uses just two covariates, Tot1618 (total number of offense conviction dates between ages 16 and 18) and Sexoff1618Adj (number of sex offenses between the ages of 16 and 18, adjusted for time-at-risk). These covariates were selected using the backward elimination method discussed earlier, using a significance level for retention of 5%; the model does not change if we instead use a level of 1%. The model then predicts that for low-rate offenders, their total number of post-18 conviction dates is estimated by

$$\mu_i = T_i \exp(-0.5006 - 0.2882 \text{ Sex1618Adj} + 0.0322 \text{ Tot1618}).$$

For high-rate offenders, the corresponding estimate is

$$\mu_i = T_i \exp(-0.34405 - 0.34543 \text{ Sex1618Adj} + 0.08993 \text{ Tot1618}).$$

This new CV error of 4.606149 represents some improvement over our previous bests of 4.697884 and 4.652834, though the improvement was only moderate. Furthermore, the model is somewhat “unnatural” in that one covariate is adjusted for time-at-risk, while the other is not. Overall, we suspect that this “best” model is largely an artificial result in that the more models one tries, the more likely one is to find a model that happens to provide a good fit for the data just by chance, sometimes referred to as the “data mining effect.” Of course, it is possible that we have uncovered a general fact that for optimal prediction, the number of sex offenses should be adjusted for time-at-risk, while the total number of offenses should not. However, it seems more likely that this result was merely an artifact of the particulars of our data and would not be repeated with a new set of data, which we are in the process of developing.

5.5. Cox proportional hazards regression

We next turn to Cox proportional hazards models, which attempt to predict the entire adult offense trajectory $\hat{\Lambda}_i(t)$. As discussed in Subsection 2.3, for comparison purposes, we predict the total number of adult offenses by $\hat{\Lambda}_i(\infty) - \hat{\Lambda}_i(18)$. To apply cross-validation, we predict each $\hat{\Lambda}_i(t)$ using only data for the *other* individuals, that is, with individual i *excluded* from the data set.

When using only the covariate TotConv, we obtain a cross-validation error of 4.879009. This is only slightly larger than the corresponding CV error for GLM of 4.821457, which suggests that, in predicting the entire adult offense timeline, we only lose a small amount of accuracy, compared to predicting total adult offenses, directly. The corresponding Cox model is given by

$$\hat{\Lambda}_i(t) = \Lambda_0(t)\exp(0.0698 \ \text{TotConv}),$$

where $\Lambda_0(t)$ is a baseline hazard function. That is, for each individual i, $\hat{\Lambda}_i(t)$ represents their estimated number of offense convictions between the ages of 18 and t. These predicted adult offense trajectories are different for each individual i, but we can get a sense of the figures by plotting the trajectories for different individuals. Figure 1 shows the predicted total number of adult offenses, as a function of age, for typical individuals at the 90th percentile, median, and 10th percentile of the range of individual predictor values.

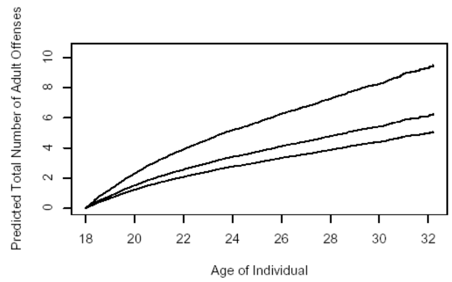

Figure 1. Estimates of post-18 offenses as a function of age, using TotConv only, for typical individuals at the 90th percentile (top), median (middle), and 10th percentile (bottom).

When using covariates corresponding to all five offense-type covariates (TotProp, TotViolent, TotDrug, TotSexoff, and TotTech), the CV error reduces slightly to 4.857658. Instead, using the total number of offenses in each of the three age ranges (Tot014, Tot1416, and Tot1618) reduces the CV error still further to 4.807447.

Combining all these eight covariates (TotProp, TotViolent, TotDrug, TotSexoff, TotTech, Tot014, Tot1416, and Tot1618) gives a CV error of 4.785165, slightly lower and fairly comparable with the corresponding GLM figure of 4.721162. Figure 2 shows the predicted total number of adult offenses using these eight covariates for typical individuals at the 90th percentile, median, and 10th percentile of the range of individual offense patterns. The corresponding model is now given by

$$\begin{aligned}\hat{\Lambda}_i(t) = \Lambda_0(t)\exp(&0.00497\ \text{TotProp} - 0.03264\ \text{TotViolent}\\ &+ 0.09370\ \text{TotDrug} - 0.20284\ \text{TotSexoff} - 0.01595\ \text{TotTech}\\ &+ 0.04483\ \text{Tot014} + 0.04940\ \text{Tot1416} + 0.11218\ \text{Tot1618}).\end{aligned}$$

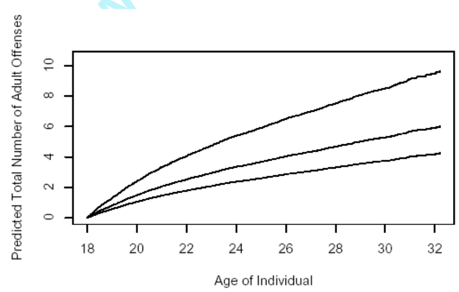

Figure 2. Estimates of post-18 offenses as a function of age, using eight covariates, for typical individuals at the 90th percentile (top), median (middle), and 10th percentile (bottom).

Again, we considered stratifying the population into different numbers of subgroups, according to total number of pre-18 conviction dates. When using only the covariate TotConv, the optimal number of subgroups was three, reducing the CV error slightly from 4.879009 to 4.832820. When using the five offense type covariates, two subgroups provided a very small improvement from 4.857658 to 4.857006, while three subgroups were slightly worse. When using the three age-range covariates, three subgroups reduced the CV error from 4.807447 to 4.784290. When using all eight covariates, stratification into subgroups only increased the CV error. So, overall, stratification into subgroups for the Cox models provided at best slight reductions in the CV error.

5.6. Summary

The results of our cross-validation comparisons are summarized in Table 1.

Table 1. Values of the cross-validation criterion for various models, for the Toronto data. Here “Method” is the statistical method used (either “LPC” for the latent Poisson classes model described in Subsection 2.1, or “GLM” for the generalised linear model described in Subsection 2.2, or “Cox” for the Cox proportional hazards regression model described in Subsection 2.3); “#Groups” is the number of subgroups into which the population was stratified, based on pre-18 convictions; “Covariates” summarises which pre-18 covariates were used; “TIS” tells whether or not the covariates were adjusted for time-in-study (as in Subsection 2.6); “TAR” tells whether or not the covariates were adjusted for time-at-risk (as in Subsection 3.4); “CV error” is the value of the cross-validation error statistic given by equation (2); and “CV fraction” is the value of this statistic expressed in terms of % of the CV statistic obtained with the baseline Mean estimator

Method	#Groups	Covariates	TIS	TAR	CV error	CV fraction
Mean	1	None	No	No	5.6727	100%
Mean	1	None	Yes	No	5.2794	93.07%
Mean	5	None	Yes	No	4.6955	82.77%
Median	1	None	No	No	5.3836	94.90%
Median	1	None	Yes	No	5.1066	90.02%
Median	4	None	Yes	No	4.6263	81.55%

LPC	1	TotConv only	No	No	5.6727	100%
LPC	2	TotConv only	No	No	5.4887	96.76%
LPC	3	TotConv only	No	No	6.2868	110.8%
LPC	1	TotConv only	Yes	No	5.2795	93.07%
LPC	4	TotConv only	Yes	No	4.7142	83.10%
GLM	1	TotConv only	No	No	5.3398	94.13%
GLM	1	TotConv only	Yes	No	4.8215	84.99%
GLM	1	Age1st only	Yes	No	5.0587	89.18%
GLM	1	By offense type	Yes	No	4.7483	83.71%
GLM	1	TotConv & certain types	Yes	No	4.7597	83.91%
GLM	1	By age range	Yes	No	4.7343	83.46%
GLM	1	By type & age range	Yes	No	4.7212	83.23%
GLM	1	TotConv & certain types & age range	Yes	No	4.6995	82.84%
GLM	1	Type & age range & Age1st	Yes	No	4.6979	82.81%
GLM	1	TotConv only	Yes	Yes	5.0614	89.22%
GLM	1	Type & age range & Age1st	Yes	Yes	5.1761	91.25%
GLM	3	TotConv only	Yes	No	4.7338	83.45%
GLM	2-8	By offense type	Yes	No	≥ 4.7567	≥ 83.85%
GLM	2-8	Type & age range & age1st	Yes	No	≥ 4.7683	≥ 84.06%
GLM	1	Age 16-18 & Age1st	Yes	No	4.6528	82.02%
GLM	2	Optimal choice	Yes	Mixed	4.6061	81.20%
Cox	1	TotConvOnly	Yes	No	4.8790	86.01%
Cox	1	By offense type	Yes	No	4.8577	85.63%
Cox	1	By age range	Yes	No	4.8074	84.75%
Cox	1	By type & age range	Yes	No	4.7852	84.35%
Cox	3	TotConv Only	Yes	No	4.8328	85.19%
Cox	2	By offense type	Yes	No	4.8570	85.62%
Cox	3	By age range	Yes	No	4.7843	84.34%

As indicated in Table 1, without adjusting for time-in-study, all of the methods perform fairly poorly. This includes the latent Poisson classes (LPC) models, which may be effective at modeling population characteristics (Paternoster et al. [44]), but not very useful for predicting individuals' adult offense patterns without adjusting for time-in-study. However, once time-in-study adjustments are made, even the baseline

Median method is quite competitive, provided the individuals are first carefully stratified by total number of pre-18 offenses. The LPC with adjustment for time-in-study also give comparable results.

The various generalised linear models (GLM) use information such as the number of offenses of different types over different pre-18 age ranges, which are not taken into consideration by the baseline and LPC models. Even using just the total number of pre-18 conviction dates (TotConv), the models reduce the CV error to about 85% of baseline. Taking into account offense types or age ranges further reduces this to about 83-84%. Taking both into account reduces the CV error to below 83%. Concentrating on the 16-18 age range reduces this to near 82%, and doing a search for an optimal model (which uses both adjusted and unadjusted covariates) gets down to just above 81%. On the other hand, more direct use of adjusted covariates, or of stratifications into subgroups, does not further improve the predictions.

As for the Cox models, the findings largely mirror the GLM results. The use of just the TotConv covariate provides reasonable results and the use of more detailed covariates (especially the age-range ones) reduces the CV error somewhat. Stratification into subgroups provides only slight improvements. Overall, the CV errors for Cox are slightly higher than the corresponding GLM ones, but this is not surprising given that the Cox models attempt to predict the entire post-18 offense trajectory, rather than just the total number of adult offenses. The most striking conclusion from Table 1 is that none of the prediction methods performs particularly well. For example, none of them gets below 80% of the baseline CV error. We discuss this issue further below.

6. Discussion and Conclusions

This paper reviewed and analysed some problems associated with predicting adult (i.e., post-age-18) criminal offenses from adolescent (i.e., pre-age-18) criminal offenses. We have seen that many prediction methods are available for predicting post-18 offenses from pre-18 offenses. Some, like latent Poisson classes, make use only of pre-18 total offense counts and predict only post-18 total offense counts. Others, like general linear models, can make use of the ages at as well as types of pre-18 offenses. Some models, like the Cox Proportional-Intensity model, can even attempt to predict full post-18 offense timelines. We have also presented a novel method that uses an exponential distribution model to adjust the offense data to take into account the time-at-risk of individuals.

We have presented the cross-validation error statistic and argued that it provides a precise, fair method for comparing the accuracy of different offense prediction methods, including different choices of stratification and covariate adjustment. We hope that, in the future, the cross-validation technique is applied to compare other prediction methods on other criminal data sets.

In the present study, we applied these concepts to a sample of 378 young offenders from Toronto. Our results indicate that, for these data, it is possible to reduce the CV error to just over 80% of baseline. This can be done in several ways, each of which requires adjusting for time-in-study, which appears to be a critical factor. The models then require the use of further pre-18 offense data, either stratification into fine counts of number of pre-18 convictions, and/or covariates based on the types and ages of pre-18 offenses.

6.1. Poisson variability

It is worth asking why none of the prediction methods considered performs particularly well. For example, no method results in a CV error that is less than 80% of the baseline CV error. This can be partially explained through *Poisson variability*. That is, even if we could predict precisely the adult conviction propensity of each individual, there is still a degree of randomness regarding when that individual would have the specific opportunity to commit a crime and be apprehended, charged, and convicted for the offense. Specifically, each individual's number of adult conviction dates can be modeled as a random variable Y_i, having the distribution given by Poisson (μ_i). Most of our work has concentrated on estimating the values of the μ_i. But even if we knew μ_i precisely, there would still be some randomness in the observed value Y_i.

To estimate how much this Poisson variability contributes to the CV error values as shown in Table 1, we assume that each μ_i is, in fact, equal to the observed value, say n_i. We then compute the expected value

$$ee = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}|\, Y_i - n_i \,|,$$

where each Y_i has the distribution given by Poisson (n_i). This computation was carried out using a simple R program, and resulted in a value of $ee \doteq 1.935285$. In other words, even if we could predict the mean values μ_i perfectly, we would still expect a CV error of about 1.935285, or 34.1% of baseline. Hence, *over one-third of the baseline CV error can never be eliminated*, no matter how precise a statistical prediction method is employed and no matter how well the pre-18 conviction patterns predict the post-18 criminal propensities. Furthermore, as noted at the end of Subsection 2.2, the associated Poisson distributions appear to be overdispersed, which may further increase the amount of the CV error due to Poisson variability. This observation places the figures in Table 1 into some perspective. On the other hand, it still only partially explains the relatively poor results seen there.

6.2. Limitations and further work

It appears to be the case that, among a juvenile criminal population, the pre-18 offense behavior cannot completely predict the post-18 criminal activity. In addition to the Poisson variability discussed above, there may be other reasons why none of our statistical prediction methods performs particularly well for our data. These include:

• Out of all of the timeline data for all offense types, and so forth, we had to select those covariates that appeared most promising for prediction. This included grouping the adolescent offenses by offense types and by offender age ranges. It is possible that we have not done this wisely, and that alternative choices of covariates would lead to better predictions.

• The effect of the covariates on the mean in the generalised linear model was assumed to be multiplicative (linear on the log mean). Perhaps the relationship between these factors and the mean is more complex, and nonlinear effects could improve the predictive power. It may be possible to apply the generalised additive models of Hastie and Tibshirani [26] to find nonparametric estimates of the transformation required on each covariate.

- We restricted ourselves primarily to three different statistical prediction methods, LPC, GLM, and Cox. It is possible that some other method, not yet explored, would lead to better estimates.

- The number of individuals in our study $(N = 378)$, while not small, is not sufficiently large to allow all statistical effects to manifest themselves. We believe that with a larger data set, the more sophisticated statistical methodology considered here (including such factors as complicated covariates, adjusting for time-at-risk, etc.) would become more useful for reducing prediction error.

- As mentioned in Section 3, we modeled the lag time between offense and conviction as a random variable. While we consider our method to be innovative, it is nevertheless an approximation and may explain why our adjustments for time-at-risk have not significantly improved our estimates. If we could find data on the actual conviction lags, then we could more directly use the time-at-risk timeline information in our estimates.

- Our predictions were made using only the pre-18 criminal conviction data. Other pre-18 information and observations may be available, such as psychiatric diagnoses (Bevc et al. [3]) and severity of offenses. It is possible that such additional information, combined with adolescent criminality data, would lead to better predictions.

- Perhaps most importantly, the individuals in our study were quite homogeneous in that they all were young offenders of somewhat similar criminal backgrounds housed in similar custodial settings. We had no comparison groups of non-offenders, extremely slight offenders (with no time in custody), extremely violent offenders (held in more secure facilities), and so forth. As a result, even baseline prediction methods would work reasonably well for this group. With a more heterogeneous sample of offenders, we believe that more sophisticated statistical analyses would prove highly useful in separating out different types of individuals and provide better predictions of future offense behavior. Indeed, this is a conclusion that Piquero et al. [48], see also Piquero et al. [49] arrive at in their study of serious male offenders.

We believe there is considerable scope for applying the prediction methods presented here, as well as additional statistical methods, to other criminal data sets in an effort to further explore the question of which prediction methods work best and why. Furthermore, this paper focused on estimating the *total* number of adult conviction dates. However, it is also possible to directly apply the various methods to predicting the number of adult offenses of different *types*. We believe that would be a very natural extension of our work.

In addition, as discussed in Subsection 2.3, it is possible to consider models that predict entire adult offense trajectories, rather than simply total offense counts. In this paper we considered such models solely from the point of view of their prediction of total adult offense dates. A more detailed evaluation would instead consider the extent to which they have successfully predicted offenses at various adult *ages*.

Finally, there are, of course, many other questions besides the prediction problem that can be asked about data such as that considered here. For example, how do an individual's offense types change as a function of time? To what extent do criminals "specialise" in one particular type of crime as they age? What is the relationship between sentence given, and sentence served, and how is this relationship affected by previous conviction history? Indeed, longitudinal adolescent criminal conviction data is full of mysteries waiting to be studied and unravelled.

References

[1] D. A. Andrews and J. Bonta, The Psychology of Criminal Conduct, 3rd ed., Anderson, Cincinnati, OH, 2003.

[2] B. L. Arnold and F. M. Kay, Early transition stages and heterogeneity in criminal careers among young offenders, Canadian Review of Sociology and Anthropology 36 (1999), 157-177.

[3] I. Bevc, T. Duchesne, J. S. Rosenthal, L. Rossman, F. Theodor and E. Sowa, Young offenders' diagnoses as predictors of subsequent adult criminal behaviour, Poster Presented at the 111th Convention of the American Psychological Association, Toronto, Canada, 2003.

[4] A. Blumstein, J. Cohen and D. P. Farrington, Longitudinal and criminal career research: Further clarifications, Criminology 26 (1988), 57-74.

[5] A. Blumstein, J. Cohen, J. Roth and C. A. Visher, Criminal Careers and "Career Criminals", Vol. 1, National Academy Press, Washington, DC, 1986.

[6] A. Bottoms, J. Shapland, A. Costello, D. Holmes and G. Muir, Towards desistance: Theoretical underpinnings for an empirical study, The Howard J. 43 (2004), 368-389.

[7] R. Brame, S. D. Bushway, R. Paternoster and T. P. Thornberry, Temporal linkages in violent and nonviolent activity, J. Quantitative Criminology 21 (2005), 149-174.

[8] R. Brame, E. P. Mulvey and A. R. Piquero, On the development of different kinds of criminal activity, Sociol. Methods Res. 29 (2001), 319-341.

[9] S. Bushway, R. Brame and R. Paternoster, Assessing stability and change in criminal offending: A comparison of random effects, semiparametric, and random effects modeling strategies, J. Quantitative Criminology 15 (1999), 23-61.

[10] S. D. Bushway, T. P. Thornberry and M. D. Krohn, Desistance as a developmental process: A comparison of static and dynamic approaches, J. Quantitative Criminology 19 (2003), 129-153.

[11] C. C. Cottle, R. J. Lee and K. Heilburn, The prediction of criminal recidivism in juveniles: A meta-analysis, Criminal Justice and Behavior 28 (2001), 367-394.

[12] D. M. Day, Risk for court contact and predictors of an early age for a first court contact among a sample of high risk youths: A survival analysis approach, Canad. J. Criminology 40 (1998), 421-446.

[13] D. M. Day, I. Bevc, J. S. Rosenthal, T. Duchesne, L. Rossman and F. Theodor, Predicting adult offenders' criminal trajectories from their juvenile criminal trajectories, Poster Presented at the 111th Convention of the American Psychological Association, Toronto, Canada, 2003.

[14] J. E. Dumas and W. J. Nilson, Abnormal Child and Adolescent Psychology, Allyn and Bacon, NY, 2003.

[15] F. W. Dunford and D. S. Elliott, Identifying career offenders with self-reported data, J. Research in Crime and Delinquency 21 (1984), 57-86.

[16] A. V. d'Unger, K. C. Land, P. L. McCall and D. S. Nagin, How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses, Amer. J. Soc. 103 (1998), 1593-1630.

[17] E. P. Eggleston, J. H. Laub and R. J. Sampson, Methodological sensitivities to latent class analysis of long-term criminal trajectories, J. Quantitative Criminology 20 (2004), 1-26.

[18] D. P. Farrington, Self-reported and official offending from adolescence to adulthood, Cross-national Research in Self-reported Crime and Delinquency, M. W. Klein, ed., Kluwer, Dordrecht, Netherlands, 1989, pp. 399-423.

[19] D. P. Farrington, Age, period, cohort, and offending, Policy and Theory in Criminal Justice: Contributions in Honor of Leslie T. Wilkins, D. M. Gottfredson and R. V. Clarke, eds., Avebury, Aldershot, 1990, pp. 51-75.

[20] D. P. Farrington, Criminal career research in the United Kingdom, British J. Criminology 32 (1992), 521-536.

[21] D. P. Farrington, The explanation and prevention of youthful offending, Delinquent and Crime: Current Theories, J. D. Hawkins, ed., Cambridge University Press, Cambridge, UK, 1996, pp. 68-148.

[22] D. P. Farrington and D. J. West, The Cambridge study in delinquent development: A long-term follow-up of 411 London males, Criminality: Personality, Behavior, Life History, H. J. Kerner and G. Kaiser, eds., Springer-Verlag, Berlin, 1990, pp. 115-138.

[23] B. Francis, K. Soothill and E. Ackerley, Multiple cohort data, delinquent generations, and criminal careers, J. Contemporary Criminal Justice 20 (2004), 103-126.

[24] M. R. Gottfredson and T. Hirschi, A General Theory of Crime, Stanford University Press, Stanford, CA, 1990.

[25] J. A. Graber, J. Brooks-Gunn and A. C. Petersen, Adolescent transitions in context, Transitions Through Adolescence: Interpersonal Domains and Context, J. A. Graber, J. Brooks-Gunn and A. C. Petersen, eds., Erlbaum, Mahwah, NJ, 1996, pp. 369-383.

[26] T. J. Hastie and R. J. Tibshirani, Generalized Additive Models, Chapman & Hall, London, UK, 1990.

[27] J. S. U. Hjorth, Computer Intensive Statistical Methods Validation, Model Selection, and Bootstrap, Chapman & Hall, London, UK, 1994.

[28] L. M. Johnson, R. L. Simons and D. D. Conger, Criminal justice system involvement and continuity of youth crime, Youth & Society 36 (2004), 3-29.

[29] P. K. Lattimore, J. M. MacDonald, A. R. Piquero, R. L. Linster and C. A. Visher, Studying the characteristics of arrest frequency among paroled youthful offenders, J. Research in Crime Delinquency 41 (2004), 37-57.

[30] P. K. Lattimore, C. A. Visher and R. L. Linster, Specialization in juvenile careers: Markov results for a California cohort, J. Quantitative Criminology 10 (1994), 291-316.

[31] R. M. Lerner, J. V. Lerner, A. von Eye, C. W. Ostrom, K. Nitz, R. Talwar-Soni and J. Tubman, Continuity and discontinuity across the transition of early adolescence: A developmental contextual perspective, Transitions Through Adolescence: Interpersonal Domains and Context, J. A. Graber, J. Brooks-Gunn and A. C. Petersen, eds., Erlbaum, Mahwah, NJ, 1996, pp. 3-22.

[32] R. Loeber and M. LeBlanc, Toward a developmental criminology, Crime and Justice: A Review of Research, M. Tonry and N. Morris, eds., University of Chicago Press, Chicago, IL, 1990, pp. 375-471.

[33] E. J. Mash and D. A. Wolfe, Abnormal Child Psychology, 3rd ed., Thompson Wadsworth, Belmont, CA, 2005.

[34] A. S. Masten and J. D. Coatsworth, The development of competence in favorable and unfavorable environments: Lessons from research on successful children, American Psychologist 53 (1998), 205-220.

[35] J. McCord, Family relationships, juvenile delinquency, and adult criminality, Criminology 29 (1991), 397-417.

[36] Ministry of the Solicitor General and Correctional Services, Statistical Reporting System User Manual, Statistical Services, Correctional Services Division, North Bay, Ontario, Canada, 1995.

[37] T. E. Moffitt, "Life-course persistent" and "adolescence-limited" antisocial behavior: A developmental taxonomy, Psychological Review 100 (1993), 674-701.

[38] T. E. Moffitt, A. Caspi, N. Dickson, P. Silva and W. Stanton, Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years, Development and Psychopathology 8 (1996), 339-424.

[39] R. H. Myers, Classical and Modern Regression with Applications, 2nd ed., Duxbury, Pacific Grove, CA, 1990.

[40] D. S. Nagin, Analyzing developmental trajectories: A semi-parametric group-based approach, Psychological Methods 4 (1999), 139-177.

[41] D. S. Nagin and D. P. Farrington, The onset and persistence of offending, Criminology 30 (1992), 501-523.

[42] D. S. Nagin and K. C. Land, Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model, Criminology 31 (1993), 327-362.

[43] D. S. Nagin and R. E. Tremblay, Analyzing developmental trajectories of distinct but related behaviors: A group-based method, Psychological Methods 6 (2001), 16-34.

[44] R. Paternoster, R. Brame and D. P. Farrington, On the relationship between adolescent and adult conviction frequencies, J. Quantitative Criminology 17 (2001), 201-225.

[45] G. R. Patterson and K. Yoerger, Developmental models for delinquent behavior, Mental Disorder and Crime, S. Hodgins, ed., Sage, Newbury Park, CA, 1993, pp. 140-172.

[46] A. C. Petersen and N. Leffert, What is so special about adolescence? Psychosocial Disturbances in Young People, M. Rutter, ed., Cambridge University Press, Cambridge, UK, 1995, pp. 3-36.

[47] A. R. Piquero, A. Blumstein, R. Brame, R. Haapanen, E. P. Mulvey and D. S. Nagin, Crime in emerging adulthood, Criminology 40 (2001), 137-170.

[48] A. R. Piquero, R. Brame and D. Lynam, Studying criminal career length through early adulthood among serious offenders, Crime and Delinquency 50 (2004), 412-435.

[49] A. R. Piquero, R. Brame and T. E. Moffitt, Extending the study of continuity and change: Gender difference in the linkage between adolescent and adult offending, J. Quantitative Criminology 21 (2005), 219-243.

[50] A. R. Piquero and S. L. Buka, Linking juvenile and adult patterns of criminal activity in the Providence cohort of the National Collaborative Perinatal Project, J. Criminal Justice 30 (2002), 259-272.

[51] A. R. Piquero, D. P. Farrington and A. Blumstein, The criminal career paradigm: Background and recent developments, Crime and Justice: A Review of Research, M. Tonry, ed., Vol. 30, pp. 359-506, University of Chicago Press, Chicago, IL, 2003.

[52] A. R. Piquero and P. Mazerolle, Life-course Criminology: Contemporary and Classic Readings, Wadsworth, Belmont, CA, 2001.

[53] S. Porter, A. Birt and D. P. Boer, Investigation of the criminal and conditional release profiles of Canadian federal offenders as a function of psychopathy and age, Law and Human Behavior 25 (2001), 647-661.

[54] Y. Sakamoto, M. Ishiguro and G. Kitagawa, Akaike Information Criterion Statistics, D. Reidel Publishing Co., 1986.

[55] R. J. Sampson and J. H. Laub, Crime and deviance in the life course, Annual Review of Sociology 18 (1992), 63-84.

[56] R. J. Sampson and J. H. Laub, Crime in the Making: Pathways and Turning Points Through Life, Harvard University Press, Cambridge, MA, 1993.

[57] J. E. Schullenberg, A. L. Bryant and P. M. O'Malley, Taking hold of some kind of life: How developmental tasks relate to trajectories of well-being during the transition to adulthood, Development and Psychopathology 16 (2004), 1119-1140.

[58] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978), 461-464.

[59] C. R. Shaw, The Jack-Roller: A Delinquent Boy's Own Story, University of Chicago Press, Chicago, IL, 1930.

[60] D. R. Smith, W. R. Smith and E. Norma, Delinquent career-lines: A conceptual link between theory and juvenile offenses, The Sociological Quarterly 25 (1984), 155-172.

[61] J. Stander, D. P. Farrington, G. Hill and P. M. E. Altham, Markov chain analysis and specialization in criminal careers, British J. Criminology 29 (1989), 317-335.

[62] H. Stattin and D. Magnusson, Stability and change in criminal behavior up to age 30, British J. Criminology 31 (1991), 327-346.

[63] A. J. Stewart, The course of individual adaptation to life changes, J. Personality and Social Psychology 42 (1982), 1100-1113.

[64] M. L. Sullivan, Developmental transitions in poor youth: Delinquency and crime, Transitions Through Adolescence: Interpersonal Domains and Context, J. A. Graber, J. Brooks-Gunn and A. C. Petersen, eds., Erlbaum, Mahwah, NJ, 1996, pp. 141-164.

[65] E. Sutherland, The Professional Thief, University of Chicago Press, Chicago, IL, 1937.

[66] T. P. Thornberry, Developmental Theories of Crime and Delinquency, Transaction, New Brunswick, NJ, 1997.

[67] M. E. Wolfgang, T. P. Thornberry and R. M. Figlio, From Boy to Man-From Delinquency to Crime: Follow-up to the Philadelphia Birth Cohort, University of Chicago Press, Chicago, 1987.

■