AUTHOR NOTES

Francois Thoedor
David M. Day, Irene Duvall, Therri Dudgeon, Jeffrey S. Rosenblum, Lianne Rossman

using Cross-Validation

based on Juvenile Offense Trajectories

Comparison of Adult Offense Prediction Methods
Abstract
In a factory, components are assembled, referred to as activation or a period of aggravation. Moreover, a criminal individual and environment are independent (Marsh & Wolfe, 2002). However, a criminal notion is in keeping with one of the main tenets of developmental psychology that the interaction and personal influences and their interaction (Andrews & Bonta, 2003), to environmental and personal influences is a dynamic process that is subject over time (the issue of continuity and change). In this regard, it is understood that the criminal development, criminological perspective focuses on explaining both the factors that give rise to the onset of the behavior (the issue of the offense mix, variety, and degree of severity). The developmental criminology perspective is concerned with within-individual changes and continuities in criminal behavior in understanding and empirical support for a dynamic and developmental approach to the study of criminal behavior. According to Loeger and Lofshult (1990), developmental criminology is an important advancement of the past two decades have brought about greater conceptual clarity.

1. Background about Criminal Careers

1.1. Introduction

With these various issues and approaches, in this paper, we offer a number of novel statistical methodologies to deal with these various issues.
While the notion of a criminal career is neither novel nor new (cf. Shaw, 1930; Silver, 1937), the current Zelizer's has been led by various theorists and researchers who have conceptualized understanding of criminal behavior.

Furthermore, heuristic allows for an understanding of the initiation, continuation, and termination of perspective, and policy makers' as Plustro and Marzecke (2001, p. 199) stated, a criminal career perspective may display changes and continuities in criminal activity.

During their careers, offenders may display changes and continuities in criminal activity, or end, and a duration of career length (Bornstein, Cohen, and Elfenbein, 1988).

In a similar vein, Bornstein, Cohen, and Visher (1986) proposed a conceptual and empirical framework for the criminal career paradigm. Unlike the developmental criminology models that address the onset and pattern of criminal behavior, will address a dynamic which time the individual experiences criminal development and termination or desistance. Last, this perspective is clearly concerned with the period up mental sequence of diverse forms of delinquent activities (Loeber & LeBLANC, 1990, p. 382).
with respect to methodological strategies. As Latimer, MacDonnell, Piquero, Lister, 
However, the research on criminal recidivism is not without its challenges, particularly 
criminal justice policy regarding incarceration and treatment and rehabilitation programs. 
standing their developmental trajectories could facilitate the development of more effective 
pose the greatest challenge to the criminal justice system (Piquero et al., 2003). Under- 
to account for a large number of criminal convictions, common sections violent offenses, and 
career often begins at an early age and persists into adulthood. Chronic offenders are known 
(2003; Brene, Bushway, Poynter, & Thorn- 
In a career (Piquero, Poynter, & Rosenthal, 2003; Brene, Bushway, Poynter, & Thorn- 
criminal behavior over time and about the dynamic processes that bring about this stability 
portant questions about changes and continuities in the pattern and nature of 
The collective effect of the course/developmental perspective has been to bring to the 
over time (e.g., Worth, 1993; Patterson & Yorek, 1993; Sampson & Laub, 1993). 
forth to describe the process that account for the continuities and changes in offending 
Considerable research has supported these conclusions and a number of theories have been 
exert their influence at different stages of the criminal career (e.g., Poynter, 1992). 
of these criminal behaviors and different facets (e.g., family interactions, peer group) 
the dynamic life-course perspective are that past criminal behavior increases the probability 
the importance dimension of general propensity. In contrast, two major propositions of 
(1990), who maintain that criminal activity throughout the lifespan is a function of a sin-
without criminal activity from the more static theories, such as that of Gottfredson and Hirschi 
and criminal risk across the lifespan (Piquero & Maasgul, 2001). Thus, perspective represents a 
expounded on the need for a dynamic and developmental approach to understanding crime.
suggested that individuals tend to respond to developmental transitions with a decrement in
as high stressful and overwhelming (Petersen & Lentert, 1999). For example, if has been
situation is balanced quite well. However, for some individuals, this transition is experienced
opportunities and challenges to be resolved by the individual. For the most part, the tran-
Like all developmental transitions, moving from adolescence to adulthood affords both
adolescence, a time when the paths become more sharply focused.

of individuals by examining the course of behavior across the transition from adolescence to
2004) concern that much can be learned about the continuities and discontinuities in the life
actors (Bottoni, Shapland, Casteleijn, Holmes, & July, 2004; Johnson, Shimron, & Cooper,
researchers in both psychology (Lentfer, Lentfer, von Eye, Osliom, Nee, Thayer-Sonn, &
portant implications for conceptualizing how are associated with developmental processes. Developmental
and adolescence linked and is there more overlap than difference? These questions have im-
ship between adolescence and adult offending. How are offenses committed during adolescence
One of the most enduring questions of the developmental approach concerns the relation-

1.2. The Relationship Between Adolescence and Adult Offending

method of addressing questions of “crime over time”
aim of the present study is to contribute to this body of knowledge and present a novel
includes statistical criteria to calibrate the adequacy of the group-based approaches. An
as noted by Naegele and Tendyka (2001), it leads to the need for developing methods,
developing appropriate analytical methods to model criminal careers” A particular problem,
and Visher (2004, p. 37) remarked, “In recent years, much attention has been devoted to

Comparison of Adult Offense Prediction Methods

6
nontypical functioning that can interfere with the person's ability to develop the requisite
self-identity (Coatsworth, 1998). The cumulative impact is a continued disruption in
school, development of positive peer relations, and forming a healthy and integrated sense of
ability to accomplish the nontypical developmental tasks of adolescence, such as completing
most courses (Pearsen & Letter, 1997). The resultant effect is to impede the young person's
increase in the number of transgressions and non-nontypical stresses with which the person
so forth (Johnson et al., 2004; Sullivan, 1996). This non-nontypical process also leads to
spending a great deal of time with police, correctional, probation, and parole officers, and
of roles and contexts (e.g., being processed as a "truant," making court appearances,
about a premature transition from adolescence into adulthood and a communal orientation
Justice systems may lead to a disruption in the nontypical developmental process, bringing
particular if it begins at an early age is pronounced, and involves contact with the criminal
also further suggested that involvement in serious antisocial behavior during adolescence,
periods.

or her ability to cope with the repercussions of the emerging and subsequent developmental
an increased number of transgressions can pose difficulties for the individual, compromising his
individual (Graham, Proctor-Curnin & Pearson, 1996). In general, the premature ending and
compensatory periods are the timing and number of simultaneous transitions experienced by the
process. Two factors, in particular, may affect the successful transition across develop-
However, for some individuals, factors may combine to impair both a nontypical developmental
To address the new roles and expectations and demonstrate a renewed sense of resilience,
ies of incompetence (Stewart, 1982). These negative feelings persist until the person is able
adaptation and functioning, which results in a lowered self-evaluation and heightened feel-
Comparison of Adult Offense Prediction Methods
and the question of the relationship between adolescence and adult offending remains open. In further detail, using different analytical tools and groups with different rates of offending, in adolescence and adulthood, as observed earlier, this phenomenon continues to be the result of a failure to observe, in general, the adolescent child tends to become the adult criminal activity from one developmental period to another. As mentioned (1996, p. 73) it is generally accepted within the literature that there is considerable continuity in

1.3. Stability of Offending from Adolescence to Adulthood

Adolescence to adulthood, at the same time, caution must be exercised in describing these outcomes as developed deterministic, as expectations for completing high school and entering the labor force are.

This process can result in increased likelihood of maintaining the criminal activity into skills and capabilities to assume the socially accepted roles and expectations of adulthood.
the transition from adolescence to adulthood was continued.

This transition is characterized by several key factors. First, the finding of a consensus across development periods was examined as a proxy for arrest. The relationship between specific offense types (e.g., property, violent) across different development periods was examined. Second, the relationship between arrest and factors such as socioeconomic status and race, income, and education was considered. Using multiple regression, they found that inferred arrest frequency was a significant predictor of adult arrest frequency; controlling for socioeconomic status and race, arrest frequency was observed by 

Wolfgang et al. (1987) further examined the relationship between adolescence and adult offending by the developmental approach of within-individual variation of offending/ incapacities across adolescence and adulthood. They examined the patterns across development periods (Buskens et al., 2003) and found that the arrest rates of stable patterns to be similar. Moreover, these findings reflect essentially stable rates of stability in adolescence. However, these findings reflect essentially stable rates of stability in adolescence (Wolfgang et al., 1987) has been found to adolescence and adulthood for specific types of offenses, including substance use and aggression (Partin, 1990), offense before and after age 18 years. Similar findings were observed by (RVDS) already before and after age 18 years. Wolkman, Thornberry, and Ferraro (1987) found that 39% of those arrested in the Cambridge study and convicted in both adolescence and adulthood, and Buskens, Thornberry, Ferraro (1990) reported that 42% of those convicted as adolescents in the Cambridge Study. With regard to the stability of criminal activity across development periods, Partin et al. (1992)
the offense and the date of conviction (France's, Societé, Acte, 2004). Porter, 2001) and accumulate for a time lag in our official criminal records between the date of arrest and the date of offense. 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 this analysis: accounting for crime-at-risk \( (\text{Risk}) \) and accounting for time-lag (\( t \)).

To explore these issues, we consider the information about adult offense convictions before and after 15 years of age. In this paper, we compare conventional prediction methods based on latent, Parsimonious classes (LPC) and generalized latent models (GLMs), with another method based on Cox proportional hazards models. Our particular focus is on the extent to which adult offense convictions after age 15 can be predicted from adolescent offense conviction patterns. More specifically, we can predict the extent to which adolescent convictions are related to adult convictions, and how these predictions change over time, within a wide range of offense types (e.g., violent, property, etc.). Including some victimization, and for adults to show some specialization in all offense types. To be more, specialized, it is crucial 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.
and models available for our prediction problem, we consider only certain methods here. To predict the post-18 offense conviction timeline, given the variety of statistical methods available, we also consider Cox proportional hazards model, which attempts to predict the total number of adult offenses.

In the present paper, we concentrate on predicting the total number of adult offenses as a function of age, that is, a full curve of the cumulative number of offense convictions at any given age. Moreover, we could attempt to predict an entire post-18 offense timeline, for instance, between ages 16 and 20. However, we cannot model the number of convictions from a specific age (e.g., between 16 and 18), the age of first conviction, the gender of first conviction (e.g., between ages 14 and 16, or between ages 16 and 18), the age of arrest, etc., for reasons of confidentiality. Thus, the total number of adult offenses is available in our Toronto data set for analyzing longitudinal criminal trajectories. We also describe a method for assessing the validity of the prediction methods using cross-validation.

2. STATISTICAL METHODOLOGY

Comparison of Adult Offense Prediction Methods
Once the $\theta$ and $\nu$ are estimated, then the model gives a posterior probability $q^*$ that

$$(1 = \int \prod_{u=1}^{\nu} \int \prod_{u=1}^{\nu} = \alpha_{\nu} \beta)$$

subject to the constraints that $0 < q^* < q$ and the $q^*$ are then estimated by maximizing the likelihood.

Given this likelihood, the $\theta$ and $\nu$ are then estimated by minimizing the likelihood

$$\max_{\theta, \nu} \prod_{u=1}^{\nu} \int \prod_{u=1}^{\nu} = \alpha_{\nu} \beta$$

gives rise to the definition of the Poisson distribution to a pre-18 likelihood function.

Letting $C$ denote the total number of pre-18 conviction classes of individuals, this model

$$T = 1, \ldots, \nu$$

indicates in one of the $\nu$ classes, with unknown prior probabilities distributed as $\text{Poisson}(\gamma)$, where the $\gamma$ and $\nu$ are unknown. These three steps that each

pre-18 conviction counts distributed as $\text{Poisson}(\gamma)$, and post-18 conviction counts

have to define $\gamma$ latent classes. It is assumed that the individuals in class $T = 1, \ldots, \nu$ $\gamma$

distributes the criminal propensity that is unobserved, and that this propensity is used

to consider the use of latent Poisson classes (Vannin 1999). These researchers assume that each

and others (e.g., Bushway, Bram, & Parentesco, 1999; Piquero, Bram, & Howlett, 2003)

criminal trajectories across developmental transitions. For example, Parentesco et al. (2003)

to increase sophistication and precision with which to explore the nature and pattern of

recent methodological developments in the analysis of longitudinal data (e.g., Busby, 2005).

2. Latent Poisson Classes ($TPC$)

approaches that could be taken. Which seemed most appropriate for the problem at hand. There are, of course, many other

Comparison of Adult Offense Prediction Methods

12
We investigate how good an estimate we obtain by this LFC method. We discuss in §2.3 the question of how to select an optimal number of latent classes. In

\[ \sum_{f}^{\text{LFC Estimated number of post-18 convictions}} = \hat{b}_{f} \]

(1)

For this model, namely the predicted mean given by:

\[ \hat{\mu} = \frac{1}{p(\hat{f}_{y})} \prod_{u=1}^{I_{y}} \hat{b}_{u} \]

where the \( \hat{\mu} \) and the probabilities are as estimated above, where the \( \hat{f}_{y} \) and the probabilities are as estimated above.

\[ \frac{\hat{b}_{f}}{p(\hat{f}_{y})} \sum_{f}^{\text{post-18 convictions}} = \text{[sum of post-18 convictions]} \]

Post-18 convictions for \( p \) = 0, 1, 2, \ldots (is given by \( \hat{\mu} \))

Given this likelihood, the \( \hat{f}_{y} \) are then estimated by maximizing the likelihood. The final

\[ \hat{f}_{y} = \arg \max \left( \sum_{i=1}^{I} i^{(i)} / \hat{C}(\hat{g}) \prod_{u=1}^{I_{y}} \hat{b}_{u} \right) \]

then give rise to a post-18 likelihood function

Let \( \hat{g}(\hat{f}) \) denote the total number of post-18 conviction data of individual \( i \), the \( \hat{g} \)

\[ \frac{\hat{g}(\hat{f})}{\hat{g}(\hat{f})} \prod_{u=1}^{I_{y}} \hat{b}_{u} = \left( \frac{i^{(i)} / \hat{C}(\hat{g})}{\prod_{u=1}^{I_{y}} \hat{b}_{u}} \right) \]

individual \( i \) in class \( f \), given by

\[ \text{Comparison of Adult Conviction Prediction Methods} \]
In a lack of interpretability, while few covariates may fail to explain the outcome.. 
use too many covariates may lead to overfitting problems (discussed further below).

In general, it is possible to adjust the covariates for time-related, as described in 3 below. 

Therefore, we consider a more specific covariate such as the number of convictions of the offender between the ages of 16 and 18 (Adler et al), and so forth. Of the first conviction (Adler et al), the total number of convictions is represented in each conviction. Thus, we can use such covariates as the total number of pre-18 convictions are represented in each conviction. Hence, the model leads to many choices. Most obviously, what covariates should be considered?

This model leads to many choices. Most obviously, what covariates should be considered?

\[ \{ d_1 x_1 d_2 x_2 + \ldots + \Sigma x_1 x_2 \} \]

Prediction for the number of post-18 offenses for individual is given by: \( \Sigma x_1 x_2 \).

One estimate of the regression coefficients \( \Sigma x_1 x_2 \) have been obtained, a bivariate procedure (readily available in most statistical packages including R, S-Plus, SAS).

Estimation of the regression coefficients \( \Sigma x_1 x_2 \) is then done using a maximum likehood procedure (readily available in most statistical packages including R, S-Plus, SAS).

\[ \{ d_1 x_1 d_2 x_2 + \ldots + \Sigma x_1 x_2 \} \]

Poisson distribution, where \( \Sigma x_1 x_2 \) is the independent follow regression model. Specifically, given the covariates \( \Sigma x_1 x_2 \), the \( \Sigma x_1 x_2 \) independent follow regression model. We then assume that the relationship between \( \Sigma x_1 x_2 \) and the covariates is given by a Poisson total number of post-18 offenses, and write \( d_1 x_1 \) for the pre-18 covariance information.

We next consider a Poisson regression model. For each individual \( i \), we write \( Y_i \) for the

2.2. Generalised Linear Models (GLM)
one and eight of different subgroups, on the basis of their number of pre-18 office conviction
decided to consider estimations of the population into various numbers (Appendix B) between
into one statistical model, another types with vastly different post-18 offence patterns. We
in groups that are too small to detect patterns, while too high the stratification might force,
suppose, on the basis of their pre-18 offence data. In general, too much stratification leads
Another reason to consider is whether (and how) to stratify the population into distinct
rather when the Akaike Information Criterion (AIC) of the model starts to increase.
every variable in the model has a significance level inferior to the preset reception level, but
the classical backward elimination procedure expected that the algorithm does not stop when
available in the statistical package R. This latter algorithm functions in the same manner as
execute in SAS. We also ran a backward elimination procedure using the stepAIC function
The backward elimination procedure described above was run using the GENMOD pro-
that of the reception level
model whose covariates all had an associated significance level whose value was less than
least significant covariate, we removed those elimination steps until we were left with a
obtain a few candidate models). We then fitted the remaining model and eliminated the
level did not exceed a pre-selected reception level (we used with levels 10%, 5%, and 1% to
level present, and then removed the least significant covariate, provided its significance
using the backward elimination procedure. We started with a model with every potential
information. However, we also conducted a more systematic search for the best model,
unlike our own independent about what quantities appropriately summarized the pre-18 counts
For ease of understanding and interpretability, for the most part, we chose our covariates
available data.
Corresponding to the expected number of offense dates for individual $i$, between ages 18 and

$C_{Proportional-Hazards}$ models estimate the cumulative intensity rate, say, $A(t)$.


The non-parametric type-identifying properties (Proportion) Ressmann Models (e.g., Ressmann et al., 2001). Now, $\phi$ is non-parametric type-identifying properties. For our data, we observe that all such models have

significant overdispersion ($\phi$), which is not surprising since extra-Poisson variability is greater than Poisson variability ($\phi > 1$). When $\phi < 1$ the model is overdispersed (i.e., the variability in the $A_i$'s is greater than the Poisson model). The parameter $\phi$ is the dispersion parameter (for further discussion of this issue, see Frances

Finally, we note that in a more general GEM, we could consider an overdispersed Poisson regression model, which assumes that the $A_i$'s are independent with the $A_i$'s are independent with $A_i$. In this model, $\phi = \phi_i$, the Poisson case corresponds to $\phi = 1$, $\phi = \phi_i$. In this model, $n_i = \sum_{i=1}^{n} A_i$. This is a $\phi$-

Competition of Adult Offense Prediction Methods

16
2.4. Overfitting and Information Criteria

We consider a variety of different options here, when covariates are considered, whether and how to strata the population, and so forth, this question further below. As with GLM, Cox models allow for many choices in terms of offences, while the Cox model is attractive in that it can include more general models, where the latter is specifically designed to predict the total number of observed adult offences for each individual. We would expect such predictions to be more by \( Y(t) - \int_0^t Y_s(s) \, ds \) than those of GLM, since GLM is specifically designed to predict the total number of such predictions with the time varying of some parameter with the Cox model, the predicted total number of adult offences is then given with \( \ell \), which attempts to predict only the total number of adult offences. However, there is little additional information with the Cox model provides a complete model for each individual.

Thus, in the end, the Cox model provides a complete model for each individual for individual \( i \) and \( \beta \) is a vector of regression coefficients, to be estimated parametrically, from the data, and is the same for all individuals, and \( \mathbf{x}_i \) is a list of pre-18 offence covariates, where \( Y_0(t) \) is a baseline cumulative intensity rate, which is estimated non-parametrically,

\[
\lambda(t) = \exp(\mathbf{x}_i \beta)
\]

for \( t \geq 18 \), the estimate is of the form
The intuition is that including more parameters can lead to a better fit, and hence increase

\[ AIC = -2 \log L + 2k \]

\[ BIC = -2 \log L + \log n \cdot k \]

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 Akaike Information Criterion (AIC) is defined as

\[ \text{AIC} = -2 \log L + 2k \]

\[ \text{BIC} = -2 \log L + \log n \cdot k \]

It has been proposed that Information Criteria can be used to control overfitting. The actual data and should be discarded.

If the rules so generated would generalize in any way to new young offenders, to think that the rules so generated would generalize in any way to new young offenders, to think that the rules so generated would generalize in any way to new young offenders, to think that the rules so generated would generalize in any way to new young offenders, to think that the rules so generated would generalize in any way to new young offenders.
2.5. A Pair Comparison: Cross-Validation

As we now discuss, cross-validation is the process of evaluating prediction methods and determining the most accurate method of comparing different prediction models. To determine which models are most accurate, it is necessary to have some method of cross-validation that allows for the selection of the best model from pre-100 scores. Given the multitude of methods, models, and results derived, it is important to critically evaluate the methods used.

Generally, similar to the K-fold Cross-Validation method, the random selection of folds for the models is critical. For example, consider a model selection based on the number of folds and the average model performance. The K-fold Cross-Validation method is a useful technique for evaluating the performance of different models. Although the Bayesian Information Criterion has been emphasized in the past, its use may be constrained by the complexity of the model selection process.

In summary, it is crucial to consider the consistency and the number of folds, as well as the number of iterations, when applying cross-validation. While these information criteria do have some theoretical justification, they are only intended to balance the two effects and avoid overfitting. A higher number of folds and iterations can provide more consistent results, but this may also increase the computational burden.
or not, etc., and also different strata/histo/... (e.g., divide the sample into high-risk and
choices (e.g., distinguish between different offense types or not), consider the age of offenses
(heart or Poisson classes versus GLM) or GLM versus Cox models), but also different covariates
considered supported. In this way, we can compare not only different models/predictions
When comparing two models, the one that has a smaller cross-validation error should be

\[ \text{cross-validation error} = \sum_{i=1}^{n} \frac{\hat{y}_i - y_i}{u} \]

Individually, then the cross-validation error of the model is given by

Excluding individual i, and write \( q \) for the actual error, for the actual error, for the model by temporarily
predicted local number of actual offenses of individual i, after putting the model by temporarily
predicted for the model by temporarily
To be more precise, for a particular prediction method and model, write \( \hat{y}_i \) for the
is then the average of these errors, averaged over all individuals j.
individuals j, and how much error results is computed. The cross-validation prediction error
all the other individuals. Subsequently, the accuracy of the fit to predict the behavior of
individual i, is temporarily excluded from the data. The proposed model is then shown
suggested by cross-validation (e.g., Horvitz, 1994). The idea of cross-validation is that one
usually a second "less" sample is not available. However, one way around this problem is
Of course, in practice, it is difficult to obtain one large sample of data for all models, and
set of data well, but not to make accurate predictions on the second (or, fresh) set of data,
performs the model involved averaging of other problems that allowed it to model the
then the model is highly a good one. However, if such predictions are highly inaccurate, then
the model predicts the observations in the second sample. If such predictions are accurate,
model would be developed using the first sample and then assessed as to how accurately
For the study, years \( T = \frac{1}{18} \) after age 18, during which they could potentially commit all offenses.

If we have an age \( t \geq \frac{1}{18} \) at which they left the study and a corresponding number of days of time spent or were otherwise unable to be traced for the entire study period, then, for each individual, we have the same age at that time. Furthermore, some individuals had left the community, died, were deported, or were otherwise unable to be traced for the entire study period. Thus, for each data collection ceased at a particular point in time and not all offenders were at risk.  

2.6. Adjusting for Time-in-Study

To correct for this, let \( \text{Adjusted} \) "time-in-study" for each offender be the period of time during which they could commit offenses. This is done by subtracting the number of days they spent or were otherwise unable to be traced from the total study period. For each individual, the adjusted "time-in-study" is calculated as follows:

\[
\text{Adjusted Time-in-Study} = T - \text{Days of Inactivity}
\]

where \( T \) is the total study period and \( \text{Days of Inactivity} \) is the number of days the individual was not at risk.

Possible to define an L2 cross-validation error by:

\[
\text{L2 Error} = \sum_{i=1}^{N} \left( y_i - \hat{y}_i \right)^2
\]

where \( y_i \) is the observed data and \( \hat{y}_i \) is the predicted value. It is also possible to define an L2 cross-validation measure. E is also a Cox model. Second, the statistic given by \( E \) is an L2 cross-validation measure. It is also

validation analysis usually takes less than one minute for each method and 10-30 minutes for each of the 170 offenders on a standard personal computer, performing a complete cross-validation analysis.

We close with two remarks. First, it is true that cross-validation is somewhat cumbersome, i.e., it involves calculating a large error or "overfitting" the model. Second, an overfit model will lead to a large cross-validation error.

Finally, we have noted that 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 equivalent to comparing 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 necessarily the best method for comparing the accuracy of different models.
at risk the entire time.

offense rate during that time period was twice as large as it would have been had they been secure custody for a total of one full year between the ages of 16 and 18, then their effective
crime rate would be 50% higher. As a result, we use the term “secure custody” to refer to this
example.

Example 1: If an individual commits five crimes between the ages of 16 and 18 and was in
the correctional and predictor variables to measure the rate of offending while at risk. For
offense committed while in secure custody. If it is only while not in secure custody that an individual
is at risk of offending, then the term “secure custody” refers to the time in which that individual
was held in secure custody. If so, then an individual will be charged with a new set of offenses that were

3. ADJUSTING FOR TIME-AT-RISK

In previous sections, we next turn to a more complicated form of adjustment for variables.
in study of not. We still see that usually adjusting for time-in-study leads to significant
cantly. In our comparisons below, for each model we consider both approaches: adjusting for time
in study is accomplished automatically. The complete time frame means that adjusting for time in study is accomplished automatically, and
is referred to as an offset term. For Cox models, the situation is easier. Simpler prediction
is based on a logistic regression model. For each of the variables in the model, the predicted
fractional part is simply by selecting the logistic regression model. For each of the variables in the model, the predicted
adjusted fraction of time is simply

\[ \hat{f}(t) = \frac{1}{1 + \exp[- \beta_0 - \sum \hat{\beta}_j \hat{X}_j(t)]} \]

For the generalized linear model, this adjusted fraction of time in study \( t \) can be computed,
which can then be compared to the observed post-18 offense counts.

Our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
our predicted post-18 offense frequency \( \hat{f}(t) \) to obtain a predicted post-18 offense count,
a model that takes into account the random time lag between offense commission and the time lag may not relate to all cases, only the most serious. In this section, we develop.

offense committed and an offender’s earliest recorded from 1–17 weeks, though they note that figures from the U.K. Crown Court for the period between 1991–2001, the time lag between difficulty with official criminal records. However, they were able to determine that based on the date of conviction for the date of offense. Figures of al. (2004) also cite this issue as a study of recidivism and psychological as a function of age at offense and simply substitute

While Porter et al. acknowledge this issue as a problem, they did not address it in their

directly combine the data about conviction dates with the data about time-lag.

between the date of offense and the date of conviction is unknown, making it impossible to give the date of adjudication, not the date of crime commission” (p. 628). The time lag

were recorded. As noted by Porter et al. (2004), “criminal records for individual offenders to which the offenses were committed, only the dates corresponding to when the convictions secure custody. On the negative side, we do not have access to the actual dates corresponding

directly. On the positive side, we do have accurate imputations for each individual’s dates in

In turn, the data in the present study do not allow us to make such inferences

Propensity.

Because this information could result in an underestimation of an individual offender’s criminal

is incorrectly lead to inaccurate estimations of criminal recidivism. More specifically, that prediction models that fail to take into account the time during which the offender

Piquero, Duminuco, Brantmeier, Horgan, Meloy, & Nagin, 2004; Piquero et al., 2003) and also

patterns if we take such factors into account. Indeed, Recidivistica et al. (2004): see also

It seems plausible that we will obtain more accurate predictions of adult conviction

Comparison of Adult Offense Prediction Methods

23
reported by Franks et al. (2004). That is, we assume that an exponential distribution with mean $\lambda = 90$ days, a figure which is consistent with that since the crime rates are unknown, we model them as random variables with the T

\[ \lambda - t, \]

offense at age $t$.  

In the population that individual $i$ was at risk (i.e., not in secure custody) at the actual (1), we take $Z_i(t)$ to be the probability that individual $i$ was at risk, let $\Lambda(i)$ represent individual $i$'s availability for conviction at age $t$. We assume that

\[ (i)^{\Lambda(i)} \alpha = t. \]

Our solution is to transform this problem to one involving convolution at a rate at which crimes are committed, which makes it very difficult to directly estimate $\Lambda(i)$.

Unfortunately, we do not know the rate at which crimes are committed, which an individual is in secure custody, so we in our data, we have access to the data at which an individual is in secure custody, so we

\[ (i)^{\Lambda(i)} \alpha = t. \]

That is,

variable $\Lambda(i)$ which equals 0 while individual $i$ is in secure custody, otherwise equals 1.

of their propensity to commit crimes, $p(i)^{\Lambda(i)}$ times an "availability to commit" indicator. We assume that an individual $i$ has a rate of offending at age $t$ which is given by a product

3.1 A Model for Time-at-Risk

Time-at-risk.

offense conviction, in an effort to more accurately adjust our variables for the individual's
\[
\left(\frac{L}{L'(\nu-\gamma)}-\nu-\gamma\right)\sum_{r}^{1=f} - 1 = (\nu)^Z
\]

In the case \( y = f \), then for some \( f \) we have:

\[
(\nu)^Z \quad \text{if} \quad f \quad \text{is true,} \quad \text{then} \quad f \leq \text{true}
\]

To see why (3) is true, note that if (3) is true, then (4) is true. Hence, we have:

\[
\left(\frac{L}{L'(\nu-\gamma)}-\nu-\gamma\right)\sum_{r}^{1=f} - 1 = (\nu)^Z
\]

We compute the availability for condition 3.2. Computing the availability for condition

available to have a conviction at age \( t \),

at least for the vast majority of the time, the procedure at age \( t \) was almost completely unavailable to have a new conviction at age \( t \). If this means that the individual was unavailable in a secure custody for the vast majority of the time, the procedure at age \( t \) and thus was virtually secure that, between 0 and 1, if this means that the individual has a maximum between 0 and 1 or 0, this means that the individual

\[
sp \frac{L}{L'/L'\gamma} (s-t)^Z \int_{0}^{\infty} = \left[(\nabla-t)^Z\right] d = (\nu)^Z
\]

where, exponential, \( L \), thus:

\((\nabla-t)^Z\) is then the expected value of \( M \), hence unknown and hence created as random.

The availability for conviction factor (\( t \)) is then the expected value of \( Z \), time is unknown and hence created as random.
\[
\begin{align*}
\{ [L/(n-B)-L - I], L + (V - n) & \in \mathbb{R}_{n+1} \\
\{ [L/B - L - I], V - B & \in \mathbb{R}_{n+1} \}
\end{align*}
\]
\[
\int_{0}^{V} P_{L/(n-1)}^{f} \quad (n, B, V, I)
\]

where

\[
((f_0, B, V, I) - (f_s, B, V, I)) \sum_{t=0}^{f} (V - B) = (B, V, I)^{f_0}
\]

the integrated over the range from \( V \) to \( B \).

The aggregate convolution availability for individual \( I \) over the range from \( V \) to \( B \) is given

\[
\int_{B}^{V} P_{L/(n-1)}^{f} \quad (B, V, I)
\]

\[
P_{L/(n-1)}^{f}
\]

The aggregate availability factors over the same time periods.

These offense counts for time-at-risk, it is necessary to divide them by corresponding aggregate counts, and to compute predicted adult offense counts in observed counts. To correct correlations, and to compute predicted adult offense counts, in observed counts to use as age ranges. This will be used both to define aggregate juvenile offenses counts to use as age ranges. In the analyses below, this will be necessary to consider individual offense counts over each.

3.3. AGGREGATE CONVOLUTION AVAILABILITY FACTORS

Collectively equivalent to the single formula (3) above,

\[
\begin{align*}
\int_{0}^{f} \sum_{t=0}^{f} (L/(n-1) - L/(f_{t} - 1)) - L/(f_{t} - 1) = (f_{t})^{Z}
\end{align*}
\]

Finally, if for some \( f \) we have \( f \geq f_{t} \), then

\[
(f_{t} - L/(f_{t} - 1)) \sum_{t=0}^{f} - L/(f_{t} - 1) = (f_{t})^{Z}
\]
4. DATA

   With such adjustments do or do not actually improve the accuracy of our predictions.
   For example, simply assuming that it is always equal to 90 days. In 3, we consider the extent to
   that this adjustment is a theoretical improvement over simply ignoring this lag time, or, for
   the conviction data to be considered with the secure custody data.
   In particular, we feel the unknown (random) lag time between offense dates and conviction dates, thus allowing
   the model to account

   We believe that such adjustments provide a logical, sound method of taking into account

   D, is the age at which individual i departed from the study.
   For post-15 prediction, we can divide the total adult conviction count by $\frac{1}{2} I_{(15, D)}$, where
   simply replace $P_{15} \rightarrow P_{14.16}$ with $P_{15} \rightarrow P_{1-18}$ and similarly for the other variables.
   From the above analyses, we can (optionally) use the $Z \rightarrow (i)$ and $P_{15} \rightarrow (i)$ values to modify

3.4. Adjusting the Pre-15 and Post-15 Variables

   For individual $i$ between ages $A$ and $B$.

   This gives a precise formula for computing $P_{15} \rightarrow (i), P_{1-18} \rightarrow (i)$ and the aggregate conviction availability

   Comparison of Adult Offense Prediction Methods
As well, official records are appropriate for our purposes because they provide the required information about the chronological and behavioral development of offenders. Although the use of official criminal records has been called into question in some instances, longitudinal criminal conviction data, which is essential for research that requires an accurate temporal sequencing of criminal convictions (Wright & Heffernan, 1999; Wright & Hindes-Delcast, 1997), has been used to ensure a high degree of completeness and accuracy for the samples under study.

The records were kept confidential. Steps were taken to ensure that the identifiable information in the records was revealed to the court. The court order consisted of a number of provisions designed to ensure the confidentiality of the records. Steps were taken to ensure that the identifiable information in the records was not revealed to the public.

The criminal records were derived from all official criminal records, comprising Phase I, which included information on all criminal convictions that were committed while the youth was 16 to 17 years old. The average sentence length was 1-1087 days. The average sentence length was 1-1087 days. The average sentence length was 1-1087 days. The average sentence length was 1-1087 days. The average sentence length was 1-1087 days. The average sentence length was 1-1087 days. The average sentence length was 1-1087 days. The average sentence length was 1-1087 days.
4.1. Coding Procedures

At the end of adolescence and through adulthood (p. 431), much of the time spent in criminal careers among a serious offender population at least period. Pigrams et al. concluded that the length of the study period allowed them to capture years (range = 6-18 years) and they were, on average, 31 years at the end of the follow-up for an average of 7.2 years, 2 months. The age of their first offense was reported to be 11.92.

up a sample of 277 male offenders from the California Youth Authority (CYA) institutions and the purpose of the study sample is comparable to that reported in an investigation of available for a rich, detailed analysis of their offense patterns. The length of the follow-up is illustrated by the criminal records of many of the individuals (up to 50 successive convictions).

range of offenses, including failure to appear in court and Hicks' (1979) and Williams (1988) which included a
1,339 violent offenses, 2,951 drug offenses, and 1,178 "technical" offenses, which included a

The criminal records of the sample amassed a total of 2,162 convictions. Those included 2,387

The criminal records for the Toronto sample were tracked for an average of 12.1 years.

Comparison of Adult Offense Prediction Methods
Longitudinal analyses.

For each of the temporally sequenced convictions, then, provided the criminal data for our
and represents an important improvement over previous studies. These coded variables
1986). Our data are sufficiently detailed to allow for an accurate estimation of the variable
that is, the crime time the offender is at risk to offend due to being „on the street‟ (Bosma et al.,
problem encountered in much longitudinal crime research is controlling for „time-at-risk,‟
... to avoid a potential bias introduced by plea bargaining. Last, as stated previously, a common
... to record, not just those resulting in a conviction. The coding practice selected
... picture of a given offender‟s criminal tendencies. As well, all of the offenses incurred at each
date offender. Using the OJJDP Type variable as the unit of analysis provides a more complete
and what type of seriously the expressive (e.g., violent and drug offender; property, violent and
... in „true‟ Type (e.g., property, violent, or drug offender) or „versatile‟ offender
... serious offense. The variable denote, for example, for each new conviction, whether an
... to a given conviction (as much as is available on the offender‟s „typical‟ offenses), not just the
„Offense,‟ which takes into account all of the charges incurred by the individual, that led
... conviction, the complete range of criminal charges was coded into a single variable,
... with a variable, 1994: Standard; Participation, Hill et al., 1986). However, in addition, for each
... used in these Type of research (e.g., Lattimore, 1991). Moreover, this
... offense was
... offenses. The severity ratings were taken from the NICS Statistical Reporting System User
... the sentence data, length, and type (e.g., open or closed custody) and the severity of the
that is, the crime time the offender is at risk to offend due to being „on the street‟ (Bosma et al.,
problem encountered in much longitudinal crime research is controlling for „time-at-risk,‟
... picture of a given offender‟s criminal tendencies. As well, all of the offenses incurred at each
date offender. Using the OJJDP Type variable as the unit of analysis provides a more complete
and what type of seriously the expressive (e.g., violent and drug offender; property, violent and
... in „true‟ Type (e.g., property, violent, or drug offender) or „versatile‟ offender
... serious offense. The variable denote, for example, for each new conviction, whether an
... to a given conviction (as much as is available on the offender‟s „typical‟ offenses), not just the
„Offense,‟ which takes into account all of the charges incurred by the individual, that led
... conviction, the complete range of criminal charges was coded into a single variable,
3.6.7.2. Applying cross-validation to the Mean predictor, we obtain a cross-validation error of

good prediction method would easily surpass them.

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

5.1. Baseline Methods: Mean and Median

We applied each of our statistical methodologies to our data, with various choices of

RESULTS AND COMPARISONS

Comparison of Adult Offense Prediction Methods
Any further discussion.

The 2.10677 obtained when analyzing by \( L_i \)’s, Thus, we did not consider the \( R_i \)’s. This error of 2.97071, which is significantly more than with no subgroups divided, it seems to lead to poorer estimates. For example, for the 2-4 risk study, however, this seems to lead to poorer estimates. For example, for the 2-4 risk study, the median 2-4 risk. Here, \( R_i \)’s may be thought of as the value of \( L_i \) when analyzed for adult

We also considered analyzing for time-at-risk by dividing and multiplying, not by \( L_i \), but

the population produces further improvements, with four or the subgroups being optimal.

4.681342, 4.703932, and 4.677331, respectively. Thus, we see that in this case, subdividing

the median, the total CV errors were 2.10677, 4.789232, 4.692984, 4.66262, 4.67470, 4.929727, 4.74958, 4.772417, 4.931291, 4.712826, and 4.731803, respectively. For

again subdivided for time-in-study. This led to total CV errors for the mean of 2.279439

subgroups, ranging from one to eight, based on their total number of pre-15 age class;

and a significant improvement. We subsequently attempted to divide the population into various

This reduced the CV errors to 2.279439 for the mean, and 2.10677 for the median, which is

We next tried correcting these baseline estimates for time-in-study, as discussed in 2.6.

5.2. Corrections and Stratifications

as one might expect.

superior to these baseline methods. However, the margin of victory is not as overwhelming

erro of 2.3836. We shall see that if it is indeed, line that our other prediction methods are
value. A few hidden classes (r = f) before computing for larger values of r. Indeed, no choice of r leads to particularly good predictions and, in fact, the resulting CV error is never as low as the CV error corresponding to r = 2. For larger values of r, the CV error increases with r. We then used adjusting the IPC for time-studies as discussed in 2.6. In this case, with errors no larger than those of the baseline median predictor (see Table 1),

"..."
\[ f(t) = \alpha + \beta t + \gamma t^2 + \delta t^3 \]

The corresponding model for our data is then given by

\[ Y = \exp(-0.51896 + 0.07277 L + 0.01863 L \text{days}) \]

where \( L \) is the log of the CY error of 4.79383. The corresponding model for our data is then given by

\[ Y = \exp(-0.51896 + 0.07277 L + 0.01863 L \text{days}) \]

Reduced CY error of 4.79383. The corresponding model for our data is then given by

\[ Y = \exp(-0.51896 + 0.07277 L + 0.01863 L \text{days}) \]

6.4. Generalised Linear Models (GLM)
In a different direction, incrementing just the total number of conviction dates in the age
average in adult convictions, mean[ must have correspondingly decreased, which in this case causes an overall decrease (on
then this means that the number of convictions of some other offense type[ property of Tech-
include as a covariate there, that is if TOLVorden increases while TOLVorden stays the same,
coeficient for TOLVorden in (4) was simply an artefact of the fact that TOLVorden was also
sex[. This corresponds to the well-known fact that juvenile sex offenses often correlate
In this case, all the regression coefficients are positive, with the exception of that for TOL-

\[ L' = \exp(-0.37237 + 0.02889 L'_{\text{total}}) \]

then given by

TOLVorden' gives a very similar CV error, 4.7483. The corresponding model for our data is
then does correspond to all the offense types (TOLVorden' TOLVorden, TOLVorden', TOLVorden'
Removing TOLVorden' from the model and instead, incrementing the total number of convic-

\[ \exp(0.07277) \approx 1.072 \] that is, an increase of about 7.2%.

all remain unaltered, then the expected number of post-18 offenses will be multiplied by
instance, in the above model, if TOLVorden increases by 1 and the model's other covariates
become the same value, will produce a higher [lower] number of expected post-18 offenses.
Relation, and means that an increase in the corresponding covariance with all other covariates
In each model considered, a positive [negative] coefficient implies a positive [negative] cor-

\textbf{Comparison of Adult Offense Prediction Methods}
of 4.697884 achieved without adjusting for time-at-risk.

TOTAL16324, and A\^2 error, leads to a CV error of 3.7446, 
considerably more than the value
named T-session, T-total, T-navigation, T-reason, T-total, T-total, T-total.
are, T-session, Similarly, using the time-adjusted version of the best set of covariates above,
significantly worse than the 4.982457 error from using the corresponding unadjusted covari-
we also compared adjusting the covariates for time-at-risk, and in 3.4. However, these ad-
and A\^2 error gives a CV error of 4.697884, a value that is practically unchanged.
by adjusting the covariates T-session, T-total, T-navigation, T-reason, T-total, T-total, T-total.
in particular, the variable T-total does not contribute. Adding A\^2 error to that list, that is,

\[ L = \exp(-0.009886 + 0.0469784 \times 3) \]

Corresponding model for our data is then given by

TOTAL14, and TOTAL16, gives a CV error of 4.694659, another slight improvement. The
Alternatively, using T-session, T-total, T-navigation, T-reason, T-total, T-total, T-total.
TOTAL14, TOTAL14, and TOTAL16 (leads to a CV error of 4.712162, slightly better.
(TotalProp), T-total, T-navigation, T-total, T-total, T-total, together with the age range factors.
These covariates can then be combined in different ways. Including the choice of age

\[ L = \exp(-0.009886 + 0.0469784 \times 3) \]

a small further improvement. The corresponding model for our data is then given by
Comparison of Adult Offline Prediction Methods

Population does not decrease the CV error at all. In summary, using just Toccor gives
4.77839, 4.9733, 5.1003, 5.2128, 3.90790, 4.3.12163. Once again, standardizing the
from one to another, the corresponding total CV errors are, respectively, 4.69781, 4.76348,

Similarly, when using the best selection of covariates above (Tocor, Toccor, Toccor,

CV error at all.

3.07410, and 4.99730. Thus, in this case, standardizing the population does not decrease the
CV errors, respectively, 4.74834, 4.7594, 4.76223, 4.84223, 4.93211, 4.95030.
and Toccor) and varying the number of groups from one to another, the corresponding total

When using all the offline-empire covariates (top, rfp, Tocor, Tocor, Tocor, Tocor,

from the standardization criterion (Tocor),

and the time the improvement is somewhat greater, since A is is a different quantity
4.82029, 4.91721, and 4.83099. This indicates that using three subgroups is again opti-
responding CV errors are, respectively, 3.05796, 4.91177, 4.8105, 4.82169, 4.84296,

When using only the covariate A is, then, with from one to each groups, the cor-

pre-ls correction dates for each individual
large, presumably because the Tocor correlate already takes into account the number of

reduce the CV error from 4.82147 to 4.73872. However, the improvement is not that
4.79632, we then see that standardizing the sample into three groups is optimal in this case,

are, respectively, 4.82147, 4.78231, 4.77333, 4.74169, 4.77888, 4.77936, 4.79177, and

and varying the number of groups from one to another, the corresponding total CV errors

Only using the correlation dates. When using only the correlate Top-

In a different direction, we considered standardizing the population into different subgroups

Comparison of Adult Offline Prediction Methods
High-rate offenders.

Median are designated as low-rate offenders, while those above the median are designated as

summed over all offense types. Those individuals whose value of the quantity is below the

groups, determined by the number of pre-15 offenses, adjusted for crime at-risk (less in 3.4) and

4.696:9. This best predictive power was obtained by stratifying into two equal-sized sub-

Among these many CML models, it remains one that the lowest possible CV error was

as discussed in 2.2.

different numbers of subgroups (1 through 17), we used backward elimination procedures,

both with and without adjustment for crime at-risk. We also considered stratification into

ype between the ages of 14 and 16 (Prop146), and so on. We considered those covariates

the use of such covariates as the number of conviction days including a property offense

offense types (property, violent, drug, sex offenses, and lewdness). We thus allowed ourselves

this search, we sorted the pre-15 convictions by age ranges (0-14, 14-16, and 16-18) and by

Finally, we did a more systematic search to determine the best CML models available. For

Just as resistant or our data,

results. This may correspond to “true” discovery of this very slight improvement could be

we obtain a CV error of 4.69284, which already is a slight improvement over our previous

age range. Specifically, if we use just the three covariates, Total18, SexAt16, and Age18,

can also improve better by concentrating largely on covariates corresponding to the 16-18

Further trial and error leads to other models, as well. If happens that, for our data, we

Futhe into subgroups offers very little further improvement.

about offense types and/or conviction ages. However, adjusting for crime at-risk and strat-

a fairly reasonable CV improvement, which can be further improved by using information

Comparison of Adult Offense Prediction Methods

38
We next turn to Cox proportional hazards models, which attempt to predict the entire

5.5. Cox Proportional Hazards Regression

In the process of developing the particular of our data, and would not be repeated with a new set of data, which we are

offenses should not. However, it seems more likely that this result was merely an artifact of the number of sex offenses should be adjusted for time-at-risk, while the total number of offenses is not. Overall, we suspect that this "best" model is largely an artificial result in that other is not.

a good fit for the data just by chance, sometimes referred to as the "data mining effect." 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. Furthermore, the model is somewhat "unrealistic" in that one covariate is adjusted for time-at-risk, while the other is not.

This new CV error of 4.6284% represents some improvement over our previous results.
Combining all three age categories (TorTop, TorTol, TorDec, TorDecExt, Tor-

Total 168 reduces the CV error slightly further to 4.8747.

When using covariates corresponding to all five offenses/five categories of

[INSERT FIGURE 1 ABOUT HERE]

and 10th percentile of the range of individual predictor values.

and compare the LOS of each for typical individuals at the 90th percentile, median,

plotting the results for different individuals. Figure 1 shows the predicted total number

offense trajectories are different for each individual; but we can get a sense of the shapes by

estimated number of offense convictions between the ages of 18 and 26. These predicted adult

where \( \dot{V}_{(t)} \) is a baseline hazard function. That is, for each individual \( i \), \( \dot{V}_{(t)} \) represents their

model is given by

of accuracy, compared to predicting total adult offenses, directly. The corresponding Cox

suggests that, in predicting the ending adult offense lifetime, we only lose a small amount

This is only slightly larger than the corresponding CV error for GLM of 4.8747, which

When using only the covariate TorTop, we obtain a cross-validation error of 4.87909.

data set:

\[ \dot{V}_{(t)} = \exp(0.99008 \text{TorTop} + 0.0084) \]

Comparison of Adult Offense Prediction Methods
The results of our cross-validation cohorts are summarized in Table 1.

6.6. Summary

Shifts reductions in the CV error

The CV error, so overall, satisfaction into subsamples for the Cox models provided a best to 4.781290. When using all eight covariates, satisfaction into subsamples only increased using the three age-range covariates, three subsamples reduced the CV error from 4.80447 to 4.87796, while these subsamples was slightly worse. When improvement from 4.827820 to 4.877960, while these subsamples provided a very small the optimal number of subsamples was three, reducing the CV error slightly from 4.877960 to combine to total number of pre-trained covariates. When using only the covariate Toolkit,

Again, we considered stratifying the population into different numbers of subsamples ac-

Comparison of Adult Offense Prediction Methods
to predict the entire post-18 offense trajectory, rather than just the total number of adult
the corresponding GLM ones, but this is not surprising given that the Cox models attempt
provide only slight improvements. Overall, the CV errors for Cox are slightly higher than
especially the age-range ones) reduces the CV error somewhat. Standardization into subgroups
the Torcorm covariance provides reasonable results and the use of more detailed covariates
As for the Cox models, the findings largely mirror the GLM results. The use of just
predictions.

of adult offenders, or of standardizing into subgroups, does not further improve the
unadjusted covariates (keeps down to just above 81%). On the other hand, more direct use
this to near 82%, and done a search for an optimal model (which uses both adjusted and
account reduces the CV error to below 82%. Covariating on the 16–18 age range reduces
account offenders' ages of age ranges further reduces this to about 82–84%. Taking both into
dates (Torcorm), the models reduce the CV error to about 85% of baseline. Taking into
account of different types of age range populations reduce the error to about 85% of baseline. Therefore, the various generalized linear models (GLM) use information such as the number of
The LPC with adjustment for time-in-study also give comparable results.
the LPC with adjustment for time-in-study are first carefully stratified by ordinal number of pre-18 offenses,
which, provided the individuals are first carefully stratified by local number of pre-18 offenses;
the baseline. However, predictions with individual's alone do not perform well for
models, which may be effective
As indicated in Table I, without adjustment for time-in-study, all of the methods perform
Comparison of Adult Offense Prediction Methods
administrative for crime-in-study, which appears to be a critical factor. The models then require to just over 80% of baseline. This can be done in several ways, each of which requires Toronto. Our results indicate that, for these data, it is possible to reduce the CV error.

In the present study, we applied these concepts to a sample of 275 young offenders from

data sets.

cross-validation techniques is applied to compare other prediction methods on other criminal
different choices of structuration and covariance adjustment. We hope that, in the future, the
case, our method for comparing the accuracy of different offense prediction methods, including

We have presented the cross-validation error statistic and argued that it provides a pre-

data to take into account the time-at-risk of individuals.

presented a novel method that uses an exponential distribution model to adjust the offense

intensity model, can even attempt to predict full post-18 offense timeline. We have also

of the ages at as well as types of pre-18 offenses. Some models, like the Cox Proportional-

predict only post-18 total offense counts. Others, like general linear models, can make use

predict only post-18 total offense counts. Some, like latent Poisson classes, make use only of pre-18 total offense counts and

seen that many prediction methods are available for predicting post-18 offenses from pre-18

post-18 (≤18) criminal offenses from adolescent (i.e., pre-18-≤18) criminal offenses. We have

This paper reviewed and analyzed some problems associated with predicting adult (≤18)

6. DISCUSSION AND CONCLUSIONS

error, we discuss this issue further below.

perform particularly well. For example, none of them falls below 80% of the baseline CV

offenses. The most striking conclusion from Table 1 is that none of the prediction methods

Comparison of Adult Offense Prediction Methods

43
6.1. Poisson Variability

Poisson Variability is a measure of the variability in the number of occurrences of a particular event. It is often used in the context of crime prediction to assess the variability of crime rates within a region. The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event.

The Poisson distribution is defined by the probability mass function:

\[ P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \]

where \( P(X = k) \) is the probability of observing \( k \) events, \( \lambda \) is the expected number of events, and \( k! \) is the factorial of \( k \).

In the context of crime prediction, the Poisson distribution can be used to model the number of crimes in a given area. The parameter \( \lambda \) can be estimated based on historical crime data. If the variability in the number of crimes is not accounted for, the model's predictions can be inaccurate.

In the example provided, the calculation of Poisson Variability is shown, where the expected value \( E \) is calculated as:

\[ E = \sum_{u} u \cdot \frac{1}{u} = \sum_{u} 1 = n \]

Then, the Poisson Variability is calculated as:

\[ \text{Variance} = n \cdot (n - 1) \]

This calculation helps in understanding the variability in the number of crimes, which is crucial for improving the accuracy of crime prediction models.
We restricted ourselves primarily to three different statistical prediction methods, i.e.,

1990

It may be possible to apply the generalized additive models of Hastie and Tibshirani and the mean is more complex, and nonlinear effects could improve the predictive power.

The effect of the covariances on the mean in the generalized linear model was assumed to produce predictions.

have not done this work, and that alternative choices of covariances would lead to better

methods perform particularly well for our data. These include:

variability discussed above, there may be other reasons why none of our statistical prediction

behaviour cannot completely predict the post-18 criminal activity. In addition to the Possion

It appears to be the case that, among a juvenile criminal population, the Pre-18 offence

6.2. Limitations and Further Work

other hand, it still only partially explains the relatively poor results seen there.

This observation places the figures in Table 1 into some perspective. On the

be overdispersed, which may further increase the amount of the CV error due to Possion

Furthermore, as noted at the end of 7.2, the associated Possion distributions appear to

matter how well the Pre-18 conviction patterns predict the post-18 criminal predictions.

be diminished, no matter how precise a statistical prediction method is employed and no

Comparison of Adult Offense Prediction Methods
that more sophisticated statistical analyses would prove highly useful in separating out

somalib well for this group. With a more heterogeneous sample of offenders, we believe

result, as a result, even baseline prediction methods would work rea-

offenders (with no time in custody), extremely violent offenders (held in more secure

other custody settings). We had no comparison groups of non-offenders, extremely slight

they all were young offenders of somewhat similar criminal backgrounds housed in sim-

Perhaps most importantly, the individuals in our study were quite homogeneous in that

with anecdotal criminality data, would lead to better predictions.

2003) and severity of offenses. It is possible that such additional information, combined

information and observations may be available, such as psychiatric diagnoses (Devo et al.,

Our predictions were made using only the pre-18 criminal conviction data. Other pre-18

more directly use the time-at-risk/lifestyle information in our estimates.

improved our estimates. If we could find data on the actual conviction rates, we could

imputation and may explain why our predictions for time-at-risk have not significantly

varied. While we consider our method to be innovative, it is nevertheless an approx-

As mentioned in 3, we modeled the lag time between offense and conviction as a random

useful for reducing prediction error.

factors as correlated covariates influencing for time-at-risk, etc.) would become more

data set, the more sophisticated statistical methodologies considered here (including such

large to allow all statistical effects to manifest themselves. We believe that with a larger

The number of individuals in our study (N = 378), while not small, is not sufficiently

better estimates.

Cox, and GML, and GML. It is possible that some other method, not yet explored, would lead to

Comparison of Adult Offense Prediction Methods

46
Longitudinal addresses criminal conviction data is full of mysteries waiting to be studied and sentence served, and how is the relationship affected by previous conviction history? Indeed, particular types of crime, as they age? What is the relationship between sentence given, and offence types change as a function of time? To what extent do criminals, especially in one

Finally, there are, of course, many other questions besides the prediction problem that predicted offences at various adult ages.

more detailed evaluation would instead consider the extent to which they have successfully such models solely from the point of view of their prediction of adult adult crime. A offence trajectories, rather than simply local offence counts. In this paper we considered In addition, as discussed in 2.2, it is possible to consider models that predict entire adult of different types. We believe that would be a very natural extension of our work, also possible to directly apply the various methods to predict the number of adult offences this paper focused on estimating the total number of adult conviction data. However, it is further explore the question of which prediction methods work best and why. Furthermore, here as well as additional statistical methods, to other criminal data sets in an effort to We believe there is considerable scope for applying the prediction methods presented.

Indeed, this is a condition that Piquero et al. (2009; see also Piquero et al., 2005) arrive

different types of individuals and provide better predictions of future offense behaviour.
REFERENCES
class of delinquent / criminal careers: Results from mixed Poisson regression


Poster presented at the 11th Conference of the American Psychological Association.

Predicting adult offenders' criminal trajectories from their juvenile criminal trajectories.

Day, D., Mil, K. R., & Haining, R. (2004). The prediction of criminal recidivism in


Collet, C., Lee, R. J., & Helsham, K. (2001). The prediction of criminal recidivism in

Criminal Psychology, 19, 129-133.

Comparison of adult offence prediction methods.
and Hall


and Criminal Justice: Journal of Contemporary Criminal Justice, 20, 103-126.


and uninformative environment: Lessons from research on successful children. American
University of Chicago Press.

C. Peterson (Eds.), Transitions through adolescence: Interpersonal domains and context
Tthman, T. (1996). Continuity and discontinuity across the transition of early adoles-
Ioter, R. M., Ioter, T. W., von Eye, Oston, C. W., Niz, K., Tlamer-Sohn, R.,
and of research in Crime Prevention 4: 4, 27-42.
Studying the characteristics of arrest frequency among paroled youthful offenders, four-
and continuity of youth crime. Youth Socioy 36, 33-33.


Briefly introduce the Reader, V. F., & Hare, G. F. (1986). A guide to information criteria and statistics.

Gender differences in the productivity of Canadian Federal offenders as a function of psychopathy and age. Law

Reform, 2, 497-501.


Background and recent developments. In M. Tonry (Ed.), Crime and justice: A review


Criminal Justice, 30, 259-273.

in the Providence cohort of the National Collaborative Perinatal Project. Journal of


and change: Gender difference in the link between adolescent and adult offending


(Ed.)). Physical social disruptions in young people (pp. 3-36). Cambridge: UK: Cambridge

Persson, A. C. & Lefever, N. (1999). What is so special about adolescence? In M. Rutter


Comparison of Adult Offense Prediction Methods

Interpersonal domains and contact (pp. 111-164). Minneapolis, MN: Federation.


This between theory and juvenile offenses. The Sociological Quarterly 24, 155-172.


of Chicago Press.


to adolescence. Development and Psychopathology 16, 1119-1140.

Here how developmental tasks relate to narratives of well-being during the transition

Sellers, T. E. Byrne, A. L., & O'Malley, P. M. (2004). Taking hold of some kind of

through life. Cambridge, MA: Harvard University Press.

In the making: Pathways and turning points

of Sociology 18, 63-48.


of Sociology 18, 63-48.


D. Reidel Publishing Co.

baseline mean estimation.

The value of this statistic expressed in terms of the CV statistic obtained with the

Table 1. Values of the cross-validation criterion for various models for the Toronto

<table>
<thead>
<tr>
<th>%</th>
<th>Yes</th>
<th>No</th>
<th>By age range</th>
<th>By offense type</th>
<th>Tocconi Only</th>
<th>By age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>2</td>
<td>Cox</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>82.6%</td>
<td>Yes</td>
<td>No</td>
<td>3</td>
<td>Cox</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>81.9%</td>
<td>Yes</td>
<td>No</td>
<td>1</td>
<td>Cox</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>81.7%</td>
<td>Yes</td>
<td>No</td>
<td>1</td>
<td>Cox</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>82.0%</td>
<td>Yes</td>
<td>No</td>
<td>1</td>
<td>Cox</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>80.9%</td>
<td>Yes</td>
<td>No</td>
<td>≤ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>88.6%</td>
<td>Yes</td>
<td>No</td>
<td>≤ 3-16</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≤ 3-16</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>80.1%</td>
<td>Yes</td>
<td>No</td>
<td>≤ 3-16</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>83.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
<tr>
<td>84.3%</td>
<td>Yes</td>
<td>No</td>
<td>≥ 16-18</td>
<td>Age range</td>
<td>2-8</td>
<td>8-18</td>
</tr>
</tbody>
</table>
Figure 1: Estimates of post-IQ offenses as a function of age, using Torciano's Generalized Offender Typology. 

Graph: 
- X-axis: Age of Individual 
- Y-axis: Predicted Total Number of Adult Offenses 
- Three lines representing different percentiles (below, median, and above)
Figure 2: Estimates of post-18 offenses as a function of age, using cell correlations for typical individuals at the 90th percentile (top), median (middle), and 10th percentile (bottom).