Modeling the changes in delinquent behavior of adolescents transitioning into adulthood: Methods comparison with simulated data and Pathways to Desistance data

Siny Tsang

Hong Kong

B.S., National Taiwan University, 2007
M.A., Sam Houston State University, 2010

A Dissertation presented to the Graduate Faculty of
the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Psychology

University of Virginia
April, 2015

Committee Members:
Timo von Oertzen (Chair)
Steven M. Boker
Karen M. Schmidt
John Monahan (School of Law)
Barbara A. Spellman (School of Law)

Abstract

Various methods are available to model longitudinal data, for example, growth mixture modeling (GMM), latent class growth analysis (LCGA), k-means cluster analysis, and latent transition analysis (LTA). However, the extent to which different methods can adequately model different types of longitudinal data remains unclear. Using a set of simulated data, the current study evaluated how well the four methods perform under various simulation conditions. The extent to which the methods were able to *i*) accurately identify the number of latent classes, and *ii*) correctly assign individuals into their corresponding latent classes in the simulated data were compared with one another. Based on the simulation results, suggestions were made with respect to which method(s) were best applied to model the heterogeneity of longitudinal data. The present study further applies the methods of interest to model the changes in offending behavior as delinquent youths mature into young adults. Similarities and differences of modeling solutions from different methods are reported; recommendations are made for the exploration of inter- and intra-individual differences in longitudinal data.

List of Tables

List of Figures

Table of Contents

Acknowledgments

The completion of this dissertation marks the end of an inspiring journey, but before venturing to face new challenges, I would like to take the time to express my heartfelt appreciation to those who have made this possible.

To Marbles, thank you for always being there for me. Thank you for your unconditional love and unlimited patience, I could not have finished this dissertation without you. My dear Stella, thank you for having me in your life; you brought me so much joy in the past three years. Emma & Cooper, thank you for reminding me to have a life! Chandler, Lance, Mona, Fifi, and Oliver, thank you for being part of my graduate career. Hattie & Mariah, rest in peace, you two have been wonderful friends.

My deepest gratitude goes to my closest friends. Brian, thank you for putting up with my stubbornness. The past two years have been especially tough on you; thank you for always being such a good sport. Helen, no words can describe how grateful I am to have you standing by my side every step of the way. Lunn, this is *so sads*. Elsie, perhaps one of these days, we will be able to find that door to Mars.

To everyone who has always, and continues to, believe in me, thank you.

1. Introduction

The importance of longitudinal studies in the understanding of delinquent behavior as adolescents transition into young adults has been well documented (e.g., Monahan, Steinberg, Cauffman, & Mulvey, 2009; Mulvey et al., 2010; Piquero, 2008). Trajectories of delinquent behavior have most often been examined using latent class growth analysis, in which individuals who share similar patterns of change are identified (e.g., Monahan et al., 2009; Mulvey et al., 2010). Based on adolescents' self-report offending behavior across time, adolescents who share similar growth curve features have been identified. However, an assumption of this approach is that individuals within the same latent class follow some sort of functional form over time. It is unclear whether conventional trajectory analysis is the best method to model the changes in offending behavior as adolescents transition into adults. The objective of this study is to evaluate the utility of four different methods, growth mixture modeling (GMM), latent class growth analysis (LCGA), k-means cluster analysis, and latent transition analysis (LTA), using simulated longitudinal data sets. The heterogeneity of longitudinal data may be characterized by both inter- and intra-individual differences over time. When longitudinal data is available, issues of interest may include: Is it possible to separate the heterogeneous data into groups of more homogeneous data? For example, do subgroups of individuals who demonstrate similar trends of change over time exist in the repeated measured data? If so, how many groups, or latent classes, are there? Are these latent classes relatively similar or are they relatively different? In other words, what is the degree of separation between latent classes in the data?

Unfortunately, these features are unknown in real data; in fact, statistical modeling techniques are applied to the data with the hope to answer the above questions. Because the correct answer is not available in real data, researchers typically rely on variations of model fit indices, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), to select the optimal modeling solution. However, do different methods perform equally well under different conditions? For instance, do they perform better when there exist fewer or more latent classes in the data? Does the degree of separation between classes affect the classification accuracy of different methods? More importantly, is there a method that consistently out-performs all other methods? Or are there specific conditions in which a method yields the best or worst performance? In an attempt to answer the above questions, the current study compares the estimation results obtained using different methods under a variety of simulation conditions.

The use of simulated data, in which the number of latent classes and the degrees of separation between classes are known a priori, would allow one to examine how the relative performances of these methods compare to one another. Results from this simulation study would help researchers make informed decisions as to which modeling technique(s) should be used to model the longitudinal data available. Although it would be ideal to identify a method that out-performs other methods across different conditions, it is more likely that different aspects of data characteristics are captured by different methods, meaning that a "gold-standard" for the statistical modeling of repeatedly measured data may not exist. Nonetheless, by comparing the relative performance between different methods, it would be possible to make use of the advantages of different methods in order to obtain a more comprehensive understanding of the data.

The extent to which different techniques may yield the best solution also has practical implications for the substantive context in the current study. GMM and LCGA have been frequently used to explore the trajectories of delinquent behavior over time (e.g., Monahan et al., 2009; Mulvey et al., 2010), yet it is unclear whether similar conclusions would be drawn when different approaches are used to model heterogeneity of offending behavior as youths transition from adolescence into young adulthood. It is possible that the use of a variety of modeling techniques may uncover patterns otherwise undiscovered. Given the amount of personnel and financial effort invested into such large, longitudinal studies, it is particularly important to make the best use of the existing data. The potential discovery of inter- and/or intra-individual differences may help identify possible relations between factors that were previously unexamined.

2. Literature Review

Previous literature shows two main groups of youths — adolescence-limited (AL) and life-course persistent (LCP) — exist even among serious offending adolescents (Bartusch, Lynam, Moffitt, & Silva, 1997; Moffitt, Caspi, Dickson, Silva, & Stanton, 1996; Moffitt, Lynam, & Silva, 1994; Piquero, 2000; Piquero & Brezina, 2001; Tibbetts & Piquero, 1999). Therefore, the goal of modeling changes in delinquent behavior over time is to identify factors that can predict whether delinquent youths would be more likely to become AL or LCP offenders. The ability to predict youths' future group membership may allow interventions to be tailored to AL or LCP offender groups. Because most AL offenders are likely to desist from antisocial behavior as they mature (Moffitt, 1993), crime control or prevention policies may not be as cost-effective when targeted towards AL offenders. On the other hand, efforts in crime control may focus on the small subgroup of LCP offenders with strategies of selective incapacitation and '3 Strikes' policies (Barnett & Lofaso, 1985; Caulkins, 2001; Decker & Salert, 1988; Ezell & Cohen, 1997; Greenwood, 1983). Similarly, interventions may be designed to be offender-group specific, rather than a one-size-fits all approach. While delinquent behavior may be "normative" for AL offenders, the onset of offending behavior may become criminal "careers" for LCP offenders. If it is possible to identify youths who are "at risk" of becoming LCP, intervention strategies can be developed to target this particular subgroup of youths, which may be implemented at the onset of adolescents' antisocial behavior, but before they become long-term "career criminals." For instance, adolescents who seek excitement may be challenged in after school activities (e.g., rock climbing, regional debates), and positive social support may be introduced through mentorship programs. Being able to effectively identify youths who are most likely to become life-course persistent offenders may be informative in the design and implementation of crime control and prevention policies, which may potentially guide this small subgroup of "high risk" youths away from the pathways to criminality.

2.1 Moffitt's Developmental Taxonomy

In recent years, Moffitt's developmental taxonomy (Moffitt, 1993) has become one of the most influential theories of delinquency and antisocial behavior (DeLisi, Barnes, Beaver, & Gibson, 2009) and one of the most robust and empirically-supported theories (Walsh, 2009). It has even been described as the "most innovative approach to age-crime relations" in the literature (Tittle, 2000, p. 68).

It had long been known that the vast majority of youth who engage in offending behavior are involved in antisocial behavior during adolescence. Only a much smaller group of delinquent adolescents continues to offend throughout adulthood (Walsh, 2009; Moffitt, 1993). Moffitt refers to the former group as adolescent-limited (AL) offenders, and the latter group as life-course persistent (LCP) offenders. Although LCP offenders comprise a small proportion, approximately 5-6% of the male adolescent population (DeLisi et al., 2009; Moffitt, 1993), they account for over 50% of all violent crimes (DeLisi et al., 2009; Henry, Capsi, Moffitt, & Silva, 1996).

Life-course persistent (LCP) offenders typically begin offending in early childhood, to the onset of puberty, and continue serious antisocial behavior well into adulthood (Moffitt, 1993). Consistent with Moffitt's theory, studies have found individuals who began offending prior to the onset of puberty are the most frequent and serious offenders among different age groups (Farrington, 1996; Walsh, 2002). Compared with adolescents with later delinquent onset, adolescents with earlier onset are more likely to be convicted of more serious crimes against persons, such as assault, robbery, rape, and domestic violence (Moffitt & Walsh, 2003). Moffitt describes the continuity of antisocial behavior of LCP offenders as: "biting and hitting at age 4, shoplifting and truancy at age 10, selling drugs and stealing cars at age 16, robbery and rape at age 22, and fraud and child abuse at age 30; the underlying disposition remains the same, but its expression changes form as new social opportunities arise at different points in development (pp. 679)." This age-consistent behavior is matched by consistent cross-situational behavior, as LCP offenders "lie at home, steal from shops, cheat at school, fight in bars, and embezzle at work (pp. 679)."

Moffitt's description characterizes the LCP offenders as suffering from neuropsychological deficits, resulting in poor verbal and executive functions, low IQ, inattentiveness, and hyperactivity. These problems may have resulted from a variety of factors, including pre- and/or post-natal exposure to drugs, alcohol and/or toxic agents, birth complications, and genetic factors. Environmental risk factors, such as a criminogenic home environment, being the offspring of a single teenage mother, low socio-economic status, and abuse/neglect have been identified to exacerbate the negative person-environment interactions of neuropsychological deficits (DeLisi et al., 2009; Walsh, 2009). The cumulation of these biological and environmental negative effects may result in stable and pervasive antisocial attitudes and behavior of LCP offenders (Moffitt, 1993; Walsh, 2009).

Adolescent limited (AL) offenders show relatively little continuity in their antisocial behavior. The AL offenders comprise the vast majority of the delinquent youth, and their antisocial behavior is less serious, and more age-normative. AL offenders sometimes discontinue their offending behavior during their brief crime "careers," and are usually inconsistent in their delinquent acts across situations (e.g., they experiment with drugs but obey rules under other circumstances; Moffitt, 1993). Moffitt suggests that AL offenders are motivated to participate in antisocial behavior due to a "maturity gap" between biological maturity and social maturity. Due to better health and nutrition in recent years, the average age of biological maturity (i.e., puberty) has decreased. At the same time, the average age of social maturity has increased as a result of modernization, in that certain activities that represent adulthood, such as voting, consuming alcohol, and getting a driver's license, remain off limits for adolescents. The "maturity gap" symbolizes the difference between an adolescents' level of biological maturity and social maturity (DeLisi et al., 2009).

Caught in the "maturity gap", AL offenders are attracted to their LCP peers who are already engaging in adult-like behaviors such as smoking and drinking. Moffitt (1993) posits that AL offenders, in their attempts to reduce the "maturity gap," imitate the antisocial behavior demonstrated by their LCP peers. AL offenders "mimic" the adult-like behavior of their LCP counterparts in attempt to increase their social maturity to the level of their biological maturity (DeLisi et al., 2009). AL offenders gain reinforcement of delinquency in the form of peer approval and acceptance, as well as negative consequences such as damaged relationships with parents and provoked responses from authority (e.g., teachers and police).

As AL offenders transition into young adulthood and more adult privileges become available, the "maturity gap" is minimized, and eventually eliminated. AL offenders no longer feel motivated to participate in antisocial behavior. As AL offenders transition into young adults, certain delinquent behavior is no longer illegal, such as drinking, smoking. In addition, when AL offenders mature socially, they recognize the long-term negative consequences of illegal behavior: for instance, that having a criminal record will limit their future job opportunities, or having unprotected sex may result in unwanted pregnancy. In addition, Unlike the LCP offenders who are saddled with neurological deficits, AL offenders are "psychologically healthy." Because "healthy youths respond adaptively to changing contingencies," AL offenders eventually desist from offending (Moffitt, 1993).

The classification of different types of delinquent youths may "serve as a powerful organizing function, with important implications for theory and research on the causes of crime (Moffitt, 1993, p.675)." The ability to classify adolescents based on the patterns of their offending profiles may allow customized crime control and prevention strategies. Because antisocial behavior may be considered to be "normal" among AL adolescents, and most AL offenders tend to desist from crime as they mature, emphasis on crime control and prevention may be placed on factors specific to the period of adolescent development.

Crime theories and corresponding crime control policies regarding AL specific offending should take into account factors that contribute to desistance from crime. In contrast, the onset of antisocial behavior may be the beginning of LCP youths' criminal career. Theories of crime and prevention policies should locate potential causal factors early in their childhood, and efforts should be made to develop intervention strategies to steer these youth away from pathways of criminality.

2.2 Modeling the development of adolescent delinquency

The etiology of delinquency among adolescents has also been examined in samples of adolescents in western Germany (Boers, Reinecke, Seddig, & Mariotti, 2010), using a combined latent autoregressive Markov model (Heise, 1969) and a latent growth curve model (Bollen & Curran, 2006; Meredith & Tisak, 1990), described as a *conditional time varying covariate latent curve model*. The relations between distal factors (e.g., traditional value orientations, hedonistic value orientations), proximal factors (e.g., violent peer groups, pro-violent norms), and the development of delinquency were explored among adolescents enrolled in the German prospective panel study Crime in the Modern City. Delinquency was operationalized using annual self-reported frequencies of offenses, including theft, shoplifting, robbery, assault, and vandalism.

Trajectory analysis of these data showed a six latent class solution of delinquent/non-delinquent adolescents–non-offenders, low-rate offenders, adolescence-limited offenders, early desisters, late-onset offenders, and persistent offenders. The persistent offenders, consistent with Moffitt's description of LCP offenders (Moffitt, 1993), self-reported the most involvement in delinquent behavior (see Lacourse, Dupere, & Loeber, 2008; Odgers et al., 2007; Thornberry, 2005). In contrast to the continuity of offending behavior among LCP offenders, Boers et al. found that the LCP offenders in their sample showed a reduction in delinquent behavior during late adolescence. Besides early onset offenders (Boers et al., 2010; Farrington, 1996; Odgers et al., 2007), LCP offenders also consist of late onset offenders (Boers et al., 2010), suggesting that LCP offenders may consist of a heterogeneous group of delinquent adolescents, thus increasing the difficulty of predicting this particular group of offenders.

Longitudinal studies of delinquent behavior during the transition from adolescence to adulthood have also explored factors that contribute to the continuation and cessation of criminal activity. For example, based on Steinberg and Cauffman's (1996) psychosocial maturation model, research suggests more antisocial behavior is reported by youths with lower social maturity (Steinberg & Cauffman, 2000), and that deficiencies in psychosocial maturity are associated with increased continuation of offending behavior (Monahan et al., 2009). Using a longitudinal study of serious offending adolescents, Monahan, Steinberg, Cauffman, and Mulvey (2013) found relation between the development of psychosocial maturity and desistance from offending behavior. When compared with youths who desisted from criminal activities, adolescents who demonstrated continuing antisocial behavior were found to exhibit diminished psychological maturity development. Early desisting youths were also found to exhibit greater psychosocial maturity than late desisting youths. These findings suggest that individual differences in the changes in antisocial behavior may, in part, be due to differences in the development of self-regulation processes, such as psychosocial maturity.

The consistent findings of increased antisocial behavior during adolescence and decreased criminal activities in adulthood in developmental criminology have raised concerns as to whether the age-crime curve phenomena may be due to demographic variants. Specifically, Males and Brown (2014a; 2014b) posit that pattern of the age-crime curve is an illusion of age differences in poverty, rather than characteristics of adolescence-limited offending behavior. They argue that, controlling for poverty rates, the discrepancies between youth and adult arrest no longer exist. However, these results were obtained based on cross-sectional data and do not necessarily represent the relation between the changes in criminal activities and economic status from adolescence to adulthood (Shulman, Steinberg, & Piquero, 2014). Furthermore, analysis using data from the National Longitudinal Study of Youth showed that the patterns of the age-crime curve remain similar even after controlling for social economic status (Shulman, Steinberg, & Piquero, 2013). These results highlight the need for longitudinal studies to fully understand the changes in offending behavior, as well as the association with other factors, across time.

As an increasing number of longitudinal studies examining the relation between continuity and change in offending across the life course (e.g., the National Youth Survey, NYS; the National Longitudinal Study of Adolescent Health, Add Health, the National Longitudinal Survey of Youth-1997, NLSY97, and the Pathways to Desistance) are developed and become available, concerns have been raised regarding the methodological and statistical challenges of the design and analysis of longitudinal data (Eggleston, Laub, & Sampson, 2004). Because of the longitudinal overlapping cohort design adopted in many of the longitudinal studies, individuals were at different ages at the baseline of measurement. The concern is that, at the beginning of the observation, individuals may be at different points in their age-crime curves, thus follow different age-crime paths across subsequent observation time points (Lauritsen, 1998). It is possible that the same data analyzed by age-at-assessment or by time-of-assessment may result in dramatically different conclusions (Piquero, Monahan, Glasheen, Schubert, & Mulvey, 2013).

Evidence from three prior investigations suggest conclusions drawn from applying age-based and time-based modeling strategies may not necessarily be the same, despite using the same longitudinal data. Osgood, O'Malley, Bachman, and Johnston (1989) examined the relations between time and age trends in crime among high school seniors (age range = 17 to 23) from 1976 to 1986. Analysis of age trends, using four waves of follow-up data, showed a decrease in most illegal behavior over time. However, the findings were not replicated in the time trends analysis. Analysis of time trends were suggested to be "more erratic and thus summary conclusions were deemed premature" (Piquero et al., 2013, p. 41).

Changes in delinquent behavior throughout adolescence and into early adulthood were examined using data from the NYS, a longitudinal study of over 1,600 adolescents (age range: 11 to 17 at initial data collection; Lauritsen, 1998). "[T]reating the NYS panel data as an accelerated longitudinal design with multiple age cohorts (pp. 129)," trends of general delinquency and serious offending were found to decline for the 13-, 15-, and 17-year-old cohorts, but not the 11-year-old cohort. Such findings indicate that, when analyzed using age cohorts, the expected age-crime curves were not found. Results further suggested that a single developmental trajectory, such as the age-crime curve, may not be adequate in describing the changes in delinquency from ages 11 to 21 for adolescents in the sample.

In a recent study, the issue of whether different modeling techniques (age-based or time-based) matter was further investigated using a sample of serious juvenile offenders in the Pathways to Desistance study (Piquero et al., 2013). Both age-based and time-based trajectory analysis resulted in a five-group solution, with modest consensus between the two solutions. Especially at the extreme ends of the offending distribution (i.e., very high rates or very low rates of antisocial behavior), similar groups of individuals were identified in both modeling techniques. On the other hand, individuals who engaged in a moderate amount of antisocial behavior were less consistently identified between the two methods. Furthermore, the delinquency trajectories identified using age-based and time-based assessments were not identical (Piquero et al., 2013).

These studies highlight the methodological and statistical challenges in the study of antisocial behavior across time. Given that different conclusions may be drawn depending on the choice of modeling approach, it is important that model selection for such trajectory analyses should be driven by the research questions. For instance, age-based analyses should be used in studies that examine the relation between antisocial behavior and age-related developmental changes. In contrast, time-based approaches should be adopted to investigate the differential effects of life events among samples of delinquent youths.

In addition to the age-based and time-based discrepancies described above, the robustness and sensitivities of the trajectory attributes have been questioned. Using a sample of 500 delinquent boys and their official offending records from ages 7 to 70, Eggleston et al. (2004) demonstrated that offending trajectory patterns produced from a semi-parametric group-based approach (Nagin & Tremblay, 1999) may be substantially affected by the length of follow-up, inclusion of incarceration information, and the exclusion of mortality information. Such differences may impact subsequent analyses based on these identified trajectory groups. As the shape and population of each trajectory group may vary across different analysis conditions (e.g., change in the number of observation times, inclusion or exclusion of certain variables), the results of subsequent analyses examining the relation between trajectory group memberships and other covariates (e.g., parental monitoring, impulsivity, substance abuse) may potentially be affected.

2.3 Modeling Longitudinal Data

In this following section, an overview of four methods frequently used to model longitudinal data is provided.

2.3.1 Growth Mixture Modeling (GMM)

The growth mixture modeling (GMM) is a semi-parametric, group-based modeling approach used for identifying different trajectories of behaviors across groups (Nagin & Tremblay, 1999). Similar to hierarchical and latent curve models, the association between an individual's age and behavior is estimated using a polynomial expression:

$$y_{it}^{*j} = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \varepsilon$$

where y_{it}^{*j} is a latent variable characterizing the behavior (e.g., delinquency) of individual i at time t given membership in group j, Age_{it} is individual i's age at time t, Age_{it}^2 is the square of individual i's age at time t, and ε is the error, assumed to be normally distributed with zero mean and constant variance σ^2. As the shape of the trajectory, (β_0^j, β_1^j, and β_2^j), is not constrained to be the same across the j groups, this approach allows the flexibility of estimating different developmental trajectories across groups. In addition to the estimated proportion of the population in each trajectory group, the posterior probability of group membership for individuals in the sample can also be estimated. That is, the probability with which each individual belongs to each group can be computed, based on the estimated model coefficients, upon which individuals can be classified to the trajectory group that best correspond to their observed behavior.

The Bayesian Information Criterion (BIC) is suggested as the basis for model selection in GMM (Nagin & Tremblay, 1999; D'Unger, Land, McCall, & Nagin, 1998). When there is limited prior information on the correct model, such as under circumstances of exploratory studies, Kass and Raftery 1995 and Raftery 1995 recommend selecting the model with the smallest BIC, in which BIC is computed as:

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

where L is the maximized likelihood of the model, n is the sample size, and k is the number of parameters in the model. Because the BIC rewards parsimony, the optimal model selected based on the BIC will tend to have fewer groups.

This growth mixture model method has been used to develop trajectories of delinquency (e.g., Muthen & Shedden, 1999; Piquero et al., 2013; Roeder, Lynch, & Nagin, 1999; White, Bates, & Buyske, 2001). Assuming that the population consists of a mixture of distinct groups characterized by their developmental trajectories, this approach can test whether theoretical developmental trajectory can be identified in the population.

2.3.2 Latent Class Growth Analysis (LCGA)

Latent Class Growth Analysis (LCGA) can be considered to be a special type of GMM (Jung & Wickrama, 2008). LCGA assumes that individual heterogeneity is completely captured by the mean growth trajectories of the k latent classes. That is, the variance-covariance estimates for the growth factors within each trajectory are constrained to be zero. LCGA assumes that all growth trajectories within a class are homogeneous, and that the within-class deviations from the mean trajectory are assumed to be random error.

2.3.3 Cluster Analysis

The goal of grouping n individuals into K groups, with no preassigned system of classification, can be achieved using cluster analysis (Azzalini & Scarpa, 2012). Typically, for the ith individual, there are p variables, $\tilde{x}_i = (x_{i1}, ..., x_{ip})^\top$. Individuals who are more similar are assigned into the same group, whereas those who are dissimilar are assigned into different groups. The dissimilarity, $d(i, i')$, between individuals i and i' is defined by their distance, based on the dissimilarities in each of the observed p variables, such that $d_j(x_{ij}, x_{i'j})$ for $j = 1, ..., p$,

$$d_j(x, x) \geq 0, \quad d_j(x', x) = d_j(x, x')$$

and

$$d(x, x') = 0$$

if and only if $x = x'$.

For quantitative variables, such as self-reported offending, the squared Euclidean distance can be used as a distance measure (Aldenderfer & Blashfield, 1984; Hair & Black, 2000),

$$d_j(x, x') = (x - x')^2.$$

The simplest option to obtain dissimilarities between two individuals for all variables $d(i, i')$ is to sum the dissimilarity $d_j(x, x')$ over all *j*,

$$d(i, i') = \sum_{j=1}^{p} d_j(x_{ij}, x_{i'j}).$$

The values of $d(i, i')$ can be arranged in a *n* × *n dissimilarity matrix D*, with zero diagonal and non-negative off-diagonal elements. D is used as the basis for most of the clustering methods.

One of the most commonly used non-hierarchical clustering method is K-means, designed for continuous variables. This method identifies aggregating points, known as $centroids$, around which to construct clusters, and attributes observations to the nearest centroid. Note that the centroids are not irrevocably fixed, instead, they are subject to sequential updating as the algorithm proceeds (Azzalini & Scarpa, 2012).

Based on some *a priori* criterion, assume that the observations can be divided into K groups. The total dissimilarity, the sum over all the elements of D, can be decomposed as

$$\begin{aligned} \sum_{i,i'} d(i, i') &= \sum_{k=1}^{K} \left(\sum_{i'}^{G(i')=k} d(i, i') + \sum_{i'}^{G(i')\neq k} d(i, i') \right) \\ &= D_{within} + D_{between} \end{aligned}$$

where $G(i)$ indicates the group to which the ith individual is assigned. In order to cluster individuals with high similarity and separate those with high dissimilarity, the goal is to minimize D_{within}, or maximize $D_{between}$.

K-means is a heuristic to minimize

$$D_{within} = 2 \sum_{k=1}^{K} \sum_{G(i)=k} \parallel \tilde{x}_i - m_k \parallel^2$$

where m_k is the mean vector of the individuals of the kth group. The algorithm proceeds in an iterative manner, clustering individuals around the centroids, which are subject to iterative updating, until convergence. The iterative four-step process is illustrated below:

1. With K number of groups specified *a priori*, the initial positions of centroids m_k are set with random starting values.

2. For i = 1,...,n, $\tilde{x_i}$ is assigned to group k for such that $\parallel \tilde{x_i} - m_k \parallel$ is minimum.

3. For k = 1,...,K, m_k is set as the centroid of the new cluster.

4. Steps 2 and 3 are repeated until $m_1, ..., m_k$ stabilize.

To avoid the initial determination of K groups, clustering can also be achieved using hierarchical methods. For example, an agglomerative approach begins from a state in which K = n, that is, each individual belongs to his/her separate group. The method then proceeds iteratively by aggregating previously formed groups with low dissimilarity. Because K groups are formed by aggregating two groups, this results in a hierarchical structure in which the subdivision in $K + 1$ groups belongs to the subdivision in K groups. This process continues until K = 1, when all the individuals belong to the same group. The optimal multiple-cluster solutions can be identified via multiple visual approach (Aldenderfer & Blashfield, 1984), such as inspection of the dendrogram and Euclidean distances plot (Hoeve et al., 2008).

Cluster analyses have been used in identifying psychological profiles of serious juvenile offenders (e.g., Espelage et al., 2003), as well as factors associated with delinquency, such as parenting styles (Hoeve et al., 2008) and verbal aggression (Donovan & Brassard, 2011).

2.3.4 Latent Transition Analysis (LTA)

Latent transition analysis (LTA), a type of latent Markov model (Everitt, 2006), can be used to model longitudinal data by estimating the event transitions across time (Collins

& Lanza, 2010). In addition to examining the number and nature of latent classes within observations, LTA further allows the estimation of change, and the nature of change, between latent classes across time. To distinguish the latent classes in LTA with those in latent class analysis (LCA), latent classes in LTA are subsequently referred to as *latent statuses*. The term $status$ is used to indicate temporary states in LTA latent statuses, and that individuals may move in and out of these states (Collins & Lanza, 2010).

Suppose there are $j = 1, ..., J$ observed variables, each measured at $t = 1, ..., T$ times. Each observed variable j has $r_j = 1, ..., R_j$ response categories. A contingency table of size $W = T \cdot \prod_{j=1}^{J} R_j$ can be formed by cross-tabulating the J variables at T times. Each of the W cells correspond to a vector of response patterns, $y = (r_{1,1}, ..., r_{J,T})$, to the J variables at each of the T times. Let Y be the random variable which response patterns have been chosen. Accordingly, each response pattern y can be associated with probability $P(Y = y)$, where the sum over all y's, $\sum P(Y = y) = 1$.

Let L represent the overall categorical latent variable, with S latent statuses. Thus, L_T represents the categorical latent variable at Time T, with possible values $s_T = 1, ..., S$. Three sets of parameters are estimated in the LTA, (i) latent status prevalences (δ_{s_t}), (ii) item-response probabilities ($\rho_{j,r_{j,t}|s_t}$), and (iii) transition probabilities ($\tau_{s_{t+1}|s_t}$). The prevalence of latent status s at Time t is denoted by δ_{s_t}, representing the probability of membership in latent status s at Time t. As the latent statuses are mutually exclusive and exhaustive at each time point,

$$\sum_{s_t=1}^{S} \delta_{s_t} = 1.$$

The probability of response $r_{j,t}$ in observed variable j, conditional on membership in latent status s_t at Time t, is denoted as $\rho_{j,r_{j,t}|s_t}$. For each combination of latent status s and observed variable j, there are R_j item-response probabilities. Because each individual provides only one response alternative to variable j at Time t, for all j, t,

$$\sum_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|s_t} = 1.$$

The probability of a transition to latent status s at Time $t+1$, conditional on membership in latent status s at Time t, is represented as $\tau_{s_{t+1}|s_t}$. Among the individuals in latent status s_t at Time t, each individual is only in one latent latent status s_{t+1} at Time $T+1$, where s_{t+1} may or may not be the same latent status as s_t. Because the latent status membership at each Time t is mutually exclusive and exhaustive,

$$\sum_{s_{t+1}=1}^{S} \tau_{s_{t+1}|s_t} = 1.$$

Let an indicator function be $I(y_{j,t} = r_{j,t}) = 1$ when the response to variable $j = r_j$ at Time t, and $I(y_{j,t} = r_{j,t}) = 0$ otherwise. The probability of observing a particular response pattern at Time t can be expressed as:

$$P(Y=y) = \sum_{s_1=1}^{S} ... \sum_{s_T=1}^{S} \delta_{s_1} \tau_{s_2|s_1} ... \tau_{s_T|s_{T-1}} \prod_{t=1}^{T} \prod_{j=1}^{J} \prod_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|s_t}^{I(y_{j,t}=r_{j,t})}.$$

Based on the latent status prevalences from Time 1 and transition probabilities, the latent status prevalences for Times 2 through Time T can be computed by:

$$\delta_{s_t} = \sum_{s_{t-1}=1}^{S} \delta_{s_{t-1}} \rho_{s_t|s_{t-1}}$$

for all $t \geq 2$.

LTA has been applied in the literature to examine alcohol use during the transition to adulthood (Auerbach & Collins, 2006), the onset of substance use (Collins, 2002; Patrick et al., 2009), and adolescent depression (Collins & Lanza, 2010), among others.

2.4 The Current Study

The ultimate goal of modeling longitudinal data is to identify the underlying patterns of change across time, and it is possible that different methods are better at modeling different types of longitudinal data (e.g., different number of latent classes, different distances between latent classes). The purpose of the current study is to evaluate the utility of four different methods: *i*) growth mixture modeling (GMM), *ii*) latent class growth analysis (LCGA), *iii*) k-means clustering, and *iv*) latent transition analysis (LTA)[1] in modeling the heterogeneity of changes in self-report offending behavior as adolescents transition into young adults. Using a set of simulated data, generated to mirror the patterns of offending behavior among serious offending adolescents, this simulation study aims to answer two main questions:

(a) Can the methods accurately identify the number of latent classes in each simulated data set?

(b) Can the methods correctly classify individuals into their corresponding latent classes in each simulated data set?

Findings in the simulation analyses provide information on the performance of difference methods in modeling longitudinal data with different number of latent classes with varying distances between classes.

The methods are also applied to the data from the Pathways to Desistance ("Pathways") project. The objective of the application study is to examine how well each of the methods were in modeling the heterogeneity of changes in real self-report offending be-

---

[1] Analyses using LTA timed out after about 120 hours of computation, due to the overly-large contingency table required for the estimation. As such, model estimations using LTA was found to be infeasible and were subsequently excluded in the evaluation. Additional details were described in the Results section and suggestions to potentially reduce the computation requirements were illustrated in the Discussion section.

havior data. Such application work illustrated how the methods perform in real data, when the true number of latent classes is unknown.

3. Methods

3.1 Real-world Data

3.1.1 Data

The current study utilized data the 1,170 adolescent boys enrolled in the Pathways to Desistance ("Pathways") project, a large longitudinal study of serious adolescent offenders from Maricopa County, Arizona, and Philadelphia County, Pennsylvania. The purpose of Pathways was to examine the mechanisms that influence the desistance of antisocial activity within a group of serious adolescent offenders who are making the transition from adolescence into early adulthood (see Mulvey et al., 2004). Enrollment criteria required potential participants to be less than 18 years old at the time of the study index offense and found guilty of a serious offense (overwhelmingly felony offenses, with a few exceptions for less serious property offenses, misdemeanor sexual assault, or misdemeanor weapons offenses). Enrollment of males was limited to 15% drug offenders to maintain a heterogeneous sample of serious offenders. The enrolled sample represented approximately one in three adolescents adjudicated on the enumerated charges in these locales during the recruitment period (November 2000 to January 2003).

Participants were individuals who were at least 14 years old and less than 18 years old at the time of the study index interview (M = 16.2, SD = 1.1). The sample was ethnically diverse with 20% Caucasian, 41% African American, 33.5% Hispanic youth, and 5% youth of "other ethnicity." These individuals had, on average, 3.2 (SD = 2.2) petitions prior to the baseline interview. For 350 individuals (25.8%), the study index petition was their first petition to court. The study index petition represented a felony assault or felony weapon charge for 39% of enrolled youth, followed by a drug felony (18%), burglary (15%), major property felony (10%), felonies not mentioned (7%), murder/rape/arson (7%), or other less serious charges (4%).

Baseline interviews were conducted at (or around the time of) recruitment. Follow-up interviews were conducted every six months for the next three years (6-, 12-, 18-, 24-, 30-, and 36th month) and every twelve months for the subsequent four years (48-, 60-, 72-, 84th month). Including the baseline interviews, a total of 11 waves of data was available.

3.1.2 Self-Report Offending (SRO)

Twenty-two items were adapted to measure adolescents' amount of involvement in antisocial and illegal activities. Participants were asked to indicate whether they engaged in the 22 different antisocial acts in their lifetime (at baseline interview) and within the past six or twelve months (at subsequent interviews). The 22 antisocial acts include drug (e.g., sold marijuana), aggressive (e.g., set fire, destroyed/damaged property), and income offending (e.g., used check/credit card illegally, shoplifted). In order to examine the amount of self-reported offending behavior within a consistent timeframe across time, the bi-annual SRO recorded in the first 36 months were collapsed into an annual SRO dichotomous variable, which was either "yes" for having reported an offense, or "no" otherwise. For instance, participants who reported having engaged in an antisocial act in either, or both, of the 6- and 12-month follow-up interviews would be coded as having participated in an offending act in the first year follow-up. On the other hand, those who did not report offending behavior in both of the 6- and 12-month follow-up interviews would be coded as not having participated in an offending act in the first year follow-up.

A total SRO score was computed by summing an individual's response on all 22 SRO items, indicating the amount of self-reported offending behavior in their lifetime (at baseline interview) or within the prior year (at subsequent interviews). Participants with higher SRO scores self-reported having engaged in more offending behavior, whereas those with lower SRO scores self-reported having participated in less offending behavior.

Since participants in "Pathways" were serious delinquent juveniles, many of them spent time in custody (e.g., in jail or prison) at some point during the seven years in which data was collected. When a participant is in custody, fewer opportunities are available for him to engage in offending behavior; this means that a decrease in self-reported offending behavior may be due to the lack of offending opportunity, rather than a true reduction in criminal activities. In order to take into account the availability of offending opportunities, SRO scores were multiplied by the inverse of the proportion of time participants were not in custody (i.e., have opportunities to take part in antisocial activities). Nonetheless, the probabilities of participating in illegal behavior, albeit small, still exist when individuals were in custody (e.g., beat up someone, set fire). When a participant reported having engaged in offending behavior even though he was in custody for the entire year of assessment, meaning the proportion of time on the streets was zero, a very small number (0.1) was added to the proportion to avoid division by zero, which would be undefined.

3.2 Simulation Setup

The simulation was designed to evaluate different methods' ability to (i) identify the correct number of latent classes, (ii) with minimum misclassification of individuals. Several variations of data generation were used, varying in the number of latent classes and the distance between latent classes. The data structures of the simulation are presented in Table 3.1.

The simulated datasets were generated to mirror the structure of the Pathways dataset, in order to ensure that the simulated data sets reflect the patterns of self-reported offending behavior as adolescents are transitioning into young adults. Because the structure of the simulated data sets is similar to that of the real data, it safeguards against potential bias on the part of the researcher in evaluating the utility of the different methods.

Table 3.1: Conditions for data simulation.

Parameter	Levels	Values
Occasions	1	8
Items per time point	1	22
Number of latent classes	3	3, 5, 7
Distance between classes	3	3, 5, 7
Conditions	9	
Iterations per cell	1,000	
Total iterations	9,000	

Participants in the "Pathways" dataset were followed for a period of seven years after the initial enrollment, therefore, the simulated data sets consist of eight occasions of measurement. To simulate the 22 (Yes/No) self-report offending (SRO) items used in the Pathways dataset to assess offending behavior among adolescents, SRO scores (from 0 to 22) were generated for each of the eight measurement time points. As described in the previous section, a total SRO score was computed for each individual at each time point, reflecting an individual's involvement in antisocial behavior in their lifetime (at baseline interview) or in the past year (at subsequent interviews). Because the total SRO scores were used as an indication of engagement in offending behavior, as a first step, SRO total scores were simulated at each measurement time point rather than individual SRO item responses.

Data simulated in the current study was manipulated in *a*) the number of latent classes (Class), and *b*) the average distance between classes (Distance). With respect to the number of latent classes, data was generated to simulate a three-, five-, and seven-class solution. Using the Pathways dataset, a five-class solution was previously identified (Piquero et al., 2013); thus, a five-class solution was generated to mirror the ideal number of groups of serious offending adolescents. To evaluate the utility of different methods in identifying fewer or greater number of latent classes, data was also generated to simulate three- and seven-class solutions. Within each Class condition, data was generated for a sample size of n = 1,002, 1,000, and 1,001, for the three-, five-, and seven-class solution, respectively.

The slight difference in the sample sizes was designed such that the number of participants in each latent class was generated to be the same. Each latent class consisted of 334, 200, and 147 individuals in the three-, five-, and seven-class solution, respectively. This manipulation ensured that, within each Class condition, an individual would have the same probability of being in any given latent class k. In the current simulation, classification of individuals into different latent classes was based on individuals' total SRO scores across time. Although it is possible to classify individuals according to individuals' SRO response patterns (on the 22 SRO items) across time, it is important to evaluate how well different methods can model the changes in total SRO scores across time prior to assess the utility of methods at the item-level.

In addition to the Class condition, the average distance between classes (Distance) was also manipulated in the current simulation. The average distance between classes was defined as the average SRO total score difference, across all measurement time points, between the latent classes within each Class condition. Data were generated to simulate small (3 SRO units), medium (5 SRO units), and large (7 SRO units) distance between classes, within each Class condition. The Distance condition was designed to evaluate how well different methods are able to capture the different latent classes when the classes are relatively similar (i.e., small distance) and different (i.e., large distance).

Combining the Class and Distance manipulations, there are a total of nine simulation conditions: *a*) Class: 3, Distance: 3; *b*) Class: 3, Distance: 5; *c*) Class: 3, Distance: 7; *d*) Class: 5, Distance: 3; *e*) Class: 5, Distance: 5; *f*) Class: 5, Distance: 7; *g*) Class: 7, Distance: 3; *h*) Class: 7, Distance: 5; and *i*) Class: 7, Distance: 7. For each simulation condition, 1,000 iterations were performed, generating 1,000 unique sets of data within each simulation condition. In each simulated dataset, the total SRO scores for each participant was generated for eight occasions, simulating the eight measurement time points. The simulated SRO scores were randomly drawn from a normal distribution around a set of eight centroids, defined as the mean SRO total scores. In the following paragraphs, details regarding how the SRO scores were generated for each individual within each iteration will be described. First, the methods from which the centroids, or mean SRO total scores, were selected for each time point, within each iteration will be described. Then, the way to which SRO data was generated for each individual given the centroids, within each iteration will be illustrated.

At each time point, the centroid of each latent class was defined as the average SRO scores among individuals in the corresponding latent class. In order to select the centroids, or means, of each latent class at each time point, an eight-digit binary number was used for random sampling. Each time point was represented with a single digit binary number, 0 or 1, which determined the location of the centroid (c) of latent class k. If the binary number was 0, c_k was set at 11, reflecting an average of 11 SRO for latent class k. If the binary number was 1, c_k was set at 11 $\pm$ *d*, where *d* was the distance between latent classes, which was set at 3, 5, or 7. For instance, to simulate another latent class $k+1$ with a small distance (3 SRO units) from the latent class k, c_{k+1} was set at 11 + 3 = 14. As such, the average SRO scores for latent classes k and $k+1$ were 11 and 14, respectively. Because the total SRO scores may range from 0 to 22, the "default" average SRO was set as the mid-point of the possible range (i.e., 22/2 = 11), indicating a moderate amount of self-reported offending. Setting the "default" average SRO allowed a maximum possible range of mean SRO scores in the simulation; at the maximum Distance condition, the average SRO scores may range from 4 to 18 (i.e., 11 $\pm$ 7) SRO units.

Given there were eight occasions of measurement, one can think of this as an eight-dimensional space. The centroids-selection process described above is a way to ensure a minimum distance that is equal to the conditions of distance between latent classes, within the 8-dimensional space. To measure the distance between centroids of different latent classes in the 8-dimensional space, a city-block metric was used in the current study. Based on the eight-digit binary number, the combination of 0's and 1's would indicate the centroids (c) of each time point (t) for a latent class (k). To create a data set with n number of latent classes, a random generation of the eight-digit binary number was repeated for a minimum of n iterations. For each data set, the pattern of the eight-digit binary number generated for every n-th iteration was compared with those of all previous iterations. Patterns that were not previously generated were selected for subsequent data generation, whereas those that were duplicates of previous iterations were discarded. The process was continued until a set of unique patterns of eight-digit binary number was generated for each dataset.

To illustrate the procedure of centroid determination for data simulation, an example is presented in Table 3.2. Suppose the centroids for a three-latent class solution was to be generated, the eight-digit binary number was randomly generated for n times. The 1st and 2nd iterations were both retained because their patterns were different. However, the 3rd iteration was discarded because the pattern was was the same as that in the 1st iteration. Because the 4th iteration did not repeat any of the previous patterns, the 4th iteration was retained. Accordingly, the centroids of the 1st, 2nd, and 3rd latent class in this three-latent class example was determined by the 1st, 2nd, and 4th iterations in Table 3.2, respectively. Each column in Table 3.2 represented one measurement time point in the simulation. Suppose data was to be simulated for a Class: 3, Distance: 3 condition, a mean SRO total score of 11 was set for the latent class if the value at time t was 0, whereas a mean SRO total score of 8 or 14 if the value at time t was 1.

For each of the nine simulation conditions, a new set of eight-digit binary numbers was randomly generated for each simulated data set. The centroid determination procedure described above generates up to 2^8 = 256 different possible patterns. Using the 256 different centroid patterns, it is possible to generate $\frac{256!}{(256-3)!(3!)} = 2,763,520$ different combinations of SRO patterns across time with three latent classes. Accordingly, the possible combinations of SRO patterns increase with increased number of latent classes. Because the centroid determination is generated by random sampling, this helps ensures that the simulated data sets are representative of a large variety of data patterns.

Table 3.2: Example centroid determination for data simulation.

	Time points (t)							
Iterations	1	2	3	4	5	6	7	8
1st	0	1	1	1	0	1	1	0
2nd	1	1	0	0	1	1	1	0
3rd	0	1	1	1	0	1	1	0
4th	0	1	1	1	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n-th	1	0	0	0	1	0	0	0

At time point t, a value of 0 indicates a mean SRO total score of 11, whereas a value of 1 indicates a mean SRO total score of 8 or 14. Because the pattern generated at the 3rd iteration is identical to that at the 1st iteration, the pattern of the 3rd iteration is discarded. Accordingly, the 4th iteration is used instead.

Using the methods described above, the centroids for each latent class in the eight-dimensional space were generated for each iteration, in each simulation condition. The SRO total scores for each individual belonging to each latent class was generated based on the centroids determined for each latent class, at each time point, for each iteration. For each individual, the SRO total score at each time point was drawn from a random normal distribution, with the mean, rounded to the nearest integer, at the centroid of the corresponding time point. To control for the variance of each latent class within and between data sets, the standard deviation of each latent class at each time point was constrained to 3 SRO units. Mirroring the range of the actual SRO total scores, the minimum and maximum possible values for the simulated SRO data was set to be 0 and 22, respectively. Data simulated to be less than 0 SRO units were recoded as 0, whereas those generated to be more than 22 SRO units were recoded to be 22. Each SRO total score represents the total number of "yes" responses to the 22-item SRO measure. To simulate potential item response patterns for the LTA method, the number of items corresponding to the total SRO score was randomly assigned a "yes" response. For example, if the simulated SRO total score was 8, 8 items was randomly selected to assign a "yes" response (i.e., a score of 1), whereas the remaining 14 items will be assigned a "no" response (i.e., a score of 0).

In the real-world application of SRO assessments, the amount of time individuals not in custody–not in jail/prison–is taken into account, because the amount of time individuals are not in custody (i.e., out on the streets) indicates the availability of opportunities to participate in antisocial behavior, whereas there exist fewer opportunities to engage in offending acts when individuals are in custody. Two individuals who self-reported the same amount of offending acts, with the same amount of time out on the street, would be considered to be similar, thus belong in the same latent class. However, if one were on the street most of the time whereas the other were in custody most of the time, these two individuals would not belong in the same latent class, despite having self-reported the same amount of antisocial behavior. Because the variation in amount of time out of the street would potentially alter individuals' class assignments, the current study simulated a fixed amount of time out on the street for all individuals. Specifically, the proportion of time on street between each measurement time points was simulated to be 0.7 for all individuals. This fixed value of proportion of time-on-street ensured the latent class in which individuals were based only on the number of self-reported offenses, making sure that the influence of this variable is considered correctly in the classification methods.

To ensure that similar classification results would be obtained when the proportion of time-on-street varies across individuals, an additional 100 datasets were generated for the Class: 3, Distance: 5 condition. For each individual, the proportion of time on street at each measurement time point was randomly drawn from values between 0.6 and 0.8. By restricting the simulated proportion of time on street, the latent class into which individuals were simulated remained the same; that is, the inclusion of individuals' proportion of time on street did not change the latent classes to which individuals belong. To take into account the variability of the proportion of time on streets, the simulated SRO scores in this 100 datasets were multiplied by the inverse of the proportion of time individuals were not in custody (i.e., on the streets and thus have opportunities to offend; see section 3.1.2 for how individuals' proportion of time-on-street was taken into account in the application study) .

The simulated datasets were generated with a sample size of 1,001, 1,000, and 1,002, respectively for the three-, five-, and seven-class conditions. One thousand iterations of datasets were generated per condition, resulting in a total of 9,000 simulated datasets. The repeated iterations constituted a large enough sample of datasets for a reasonably precise evaluation of the different methods. Although the precision of estimation would increase with more iterations, 1,000 iterations per condition provides a useful evaluation of the methods. In summary, a 3 (Class) x 3 (Distance) design was adopted for the simulation conditions, with 1,000 simulated datasets per condition.

3.3 Evaluation of Methods

For all simulation conditions, two to eight latent classes were fitted to the simulated SRO data using each method. In theory, the minimum number of latent classes that can be fitted to the data is one latent class, in which all individuals are determined to belong to one homogenous latent class. In the present study, because data were generated such that there existed at least three latent classes (Class: 3), the smallest number of latent classes to be fitted to the data was set to be one less than the minimum class condition, i.e., two latent classes. On the other hand, the maximum number of latent classes that can be fitted to the data is equal to the sample size in the simulated data, in which each latent class consists of one individual in the data. Having one latent class for each individual in the sample seems to be against the principle of clustering–to identify participants who demonstrate similar behavioral patterns. As such, the present study limited the maximum number of latent classes to be fitted to the simulated data to be one more than the maximum class condition, i.e., eight latent classes. To evaluate the effectiveness of the clustering methods in different simulation conditions, the correct choice of number of latent classes and the proportion of misclassification were examined. Details of these evaluation criteria will be described in the following sections.

3.3.1 Proportion of Correctly Estimated Number of Latent Classes

For each simulation condition, two to eight latent classes were fitted to the simulated data. The sample-size adjusted Bayesian information criterion (BIC) was used to determine the optimum number of latent classes in the simulated data. The BIC is defined as

$$BIC = -2 \ln L + k \ln n$$

where L is the maximized likelihood of the model, k is the number of estimated parameters in the model, and n is the sample size of the data. Taking into account the number of parameters estimated in the model, the BIC imposes a penalty for every parameter increased, and increases the penalty with increased sample size (Schwarz, 1987).

Within each simulation condition, the proportion of data sets (out of 1,000 iterations) for which the correct number of latent classes were chosen by each clustering method was computed. In addition, the patterns in which different methods selected the incorrect number of latent classes as the optimum solution were of interest. Under what conditions were the estimated optimum number of classes fewer (i.e., under-estimate) or more (i.e., over-estimate) than the correct one? Were the likelihoods of under- or over-estimation similar across different clustering methods? Accordingly, the proportions of data sets for which varying number of optimum latent classes, two to eight classes, were identified by each method were computed for all simulation conditions.

Next, a flexible approach to examine the correct choice of latent classes was adopted. The optimum solution was considered to be *close enough* to the correct number of latent classes if the number chosen by the clustering method was within a range of one latent class too few or too many. Under-estimation was defined as situations in which the optimum number of latent classes were two or more classes fewer than the correct number; over-estimation was identified when the optimum number of classes were two or more classes more than the correct one. Based on this categorization, the proportions of data sets for which the optimum number of latent classes were less than, equal to (i.e., close enough), and more than the correct number were calculated for all simulation conditions.

3.3.2 Proportion of Misclassification

Even when the correct number of latent classes were chosen, it is possible that individuals were assigned into the incorrect class. In the present simulation study, all individuals were assigned into their corresponding latent classes (e.g., $k_1,\ k_2,\ k_3$), based on their simulated SRO scores across time. Based on the clustering results, the proportion of misclassification was defined as the percentage of individuals who were not assigned into their corresponding latent class ($k_1,\ k_2,\ k_3$). For situations in which over-estimation occurred, the proportion of misclassification also included the percentage of individuals who were assigned into latent classes that did not exist in the corresponding original simulated data sets.

Proportion of misclassification was evaluated when *a*) the number of latent classes fitted to the simulated data was equal to the correct number of classes, regardless of whether the solution was found to be the optimum; and *b*) the optimal latent class solution was equal to the correct number of classes. When the number of latent classes fitted to the simulated data was equal to the simulated Class conditions, classification accuracy among the clustering techniques may be evaluated under the same conditions. An overall proportion of misclassification, from the 1,000 iterations, was computed for all simulation conditions. For ease of interpretation, the proportion of correct classification was reported, where higher values representing more accurate classification of individuals in the simulated data sets.

Nonetheless, as described in the previous section, the correct number of latent classes may not always be selected as the optimal solution. Classification accuracy in these situations may not be as informative; rather, one may be more concerned with the proportion of misclassification when the correct number of latent classes were determined to be the optimal solution. An overall proportion of misclassification, from the iterations in which the optimal latent class solution was identical to the correct number of classes, was also computed for all simulation conditions. For ease of interpretation, the proportion of correct classification was reported, where higher values representing more accurate classification of individuals in the simulated data sets.

3.4 Application

Similar to the simulation analyses, two to eight latent classes were fitted to the "Pathways" data to model the heterogeneity of SRO across time, using each clustering method. The BIC fit index was used to identify the optimum number of latent classes in the "Pathways" data. Because the correct number of latent classes and the correct classification of individuals were unknown in this real-world data, classification results were discussed in the context of the changes in SRO scores over time.

3.5 Implementation

Data for the simulation study was generated using the statistical software R (R Core Team, 2013). In order to avoid result differences due to program differences, analyses in the simulation and application studies were conducted using the same statistical programs. GMM and LCGA models were estimated using the "lcmm" package (Proust-Lima, Philipps, Diakite, & Liquet, 2014) in R, k-means models estimated using the "kmlcov" package (Benghezal, 2013) in R. LTA models were not estimated in R, but were estimated using Mplus Version 7.11 (Muthen & Muthen, 1998-2012).

4. Results

In the present study, model estimation using LTA was found to be not feasible. LTA uses a contingency table, or cross-tabulation, to represent the changes in item responses across time (Collins & Lanza, 2010). With 22 dichotomous ("Yes" and "No") SRO items across eight measurement time points, the contingency table has 2^{22*8} = 2^{176} = 9.578 x 10^{52} cells. Due to the size of this contingency table, computation was found to be too heavy for the operating system; operations timed out after approximately 120 hours of computation. In the following, only the results obtained using GMM, LCGA, and k-means are presented. Implications and suggestions with respect to LTA will be described in the discussion section.

4.1 Simulation Analyses

Self-reporting offending (SRO) data was simulated for 1,000 iterations for each of the nine simulation conditions. Within each iteration, data for a sample size of *n* = 1,002, 1,000, and 1,001 was generated for the three-, five-, and seven-class condition, respectively. For each individual in the simulated data, a total SRO score (range: 0 to 22) were generated for each of the eight measurement time points. The sample size differs for different latent class conditions to ensure that the each latent class consists of the same number of participants, within each latent class condition. In other words, each latent class consists of 334, 200, and 143 participants in the three-, five-, and seven-class condition, respectively.

The distance between latent classes was defined as the average distance (in SRO units) between latent classes across time. The average distance between latent classes was generated to be 3 (small), 5 (medium), and 7 (large) SRO units. Latent classes with small distance have relatively similar total SRO scores across time; whereas latent classes with large distance have relatively different total SRO scores across time. Across the eight measurement time points, the average difference in total SRO scores between classes is 3, 5, and 7 SRO units, respectively in the small, medium, and large distance conditions.

With nine simulation conditions, a total of 9,000 data sets were simulated. Figure 4.1 illustrates an example of the data simulated for the five latent class, with an average of 5 SRO units between latent class, condition. Each line represents the average SRO scores for 200 participants simulated for each latent class in the example. For instance, participants in class 3 (dashed line with squares) were simulated to have high SRO scores for the first six measurement time points, then decreased towards the last few measurement time points. Participants in class 4 (dashed line with crosses), on the other hand, were simulated to have more variability in their SRO scores.

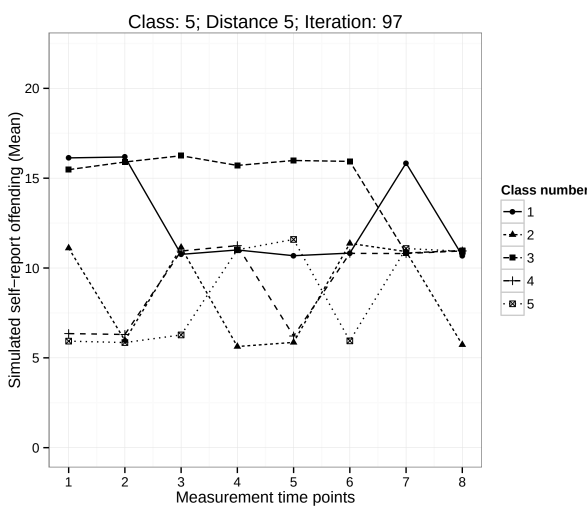

Figure 4.1: An example of the data simulated for the five latent class, with an average of 5 SRO units difference between classes. Each line in the graph represents the average simulated SRO data for each latent class.

The methods k-means, latent class growth analysis (LCGA), and growth mixture modeling (GMM) were used to determine the number of latent classes based on the simulated SRO scores in each of the simulated data set. Although the number of latent classes that can be fitted to the simulated data may range from one (i.e., all individuals belong to the same latent class) to the number of individuals in the sample (i.e., each individual constitutes his/her own latent class), the present study limited the number of latent classes fitted to the simulated data between two (one less than the smallest number of latent class condition) and eight (one more than the largest number of latent class condition). Because the “correct” number of latent classes is known a priori for the simulated data, this range of latent class number (2 to 8 classes) allows the possibility to evaluate how often each method can accurately identify, under-, or over-estimate, the number of latent classes best fit for the simulated data.

To determine the number of latent class solution best fitted for each simulated data, the BIC fit index was used. The model with the smallest BIC value was considered to be the best fit for the data. For cases in which there were ties in the smallest BIC values, parsimony was taken into account such that the model with the fewer latent class solution was considered to be the best fit for the simulated data.

Proportion of correctly identified number of classes

In each simulation condition, the percentage of iterations (out of 1,000) each latent class solution was considered to be the best fit for the simulated data is summarized for each method in Table 4.1 and Figure 4.2. Across all simulated conditions, k-means tended to determine that a larger number of latent classes was the best fit for the simulated data. Using k-means on data simulated for three latent classes, the three-class solution is unlikely to be determined as the best solution. The probabilities of identifying there were three latent classes in the simulated data ranged from 0.3% to 4%. Instead, k-means estimated that more latent classes was best fitted for the data, suggesting k-means tend to over-estimate the number of latent classes for the simulated data. This pattern of over-estimation was consistent, though less prominent, for the five-class solution. The probabilities of determining there existed five latent classes in the simulated data ranged between 7.8% and 13.3%. Similar to the results for data simulated for three latent classes, k-means determined that more latent classes was the best fit for data simulated for five latent classes. For data simulated for seven latent classes, k-means seemed to be less likely to over estimate the number of latent classes that best fitted the data. When seven latent classes existed in the generated data, k-means correctly estimated seven latent classes in the simulated data 23.3% to 27.8% of the time.

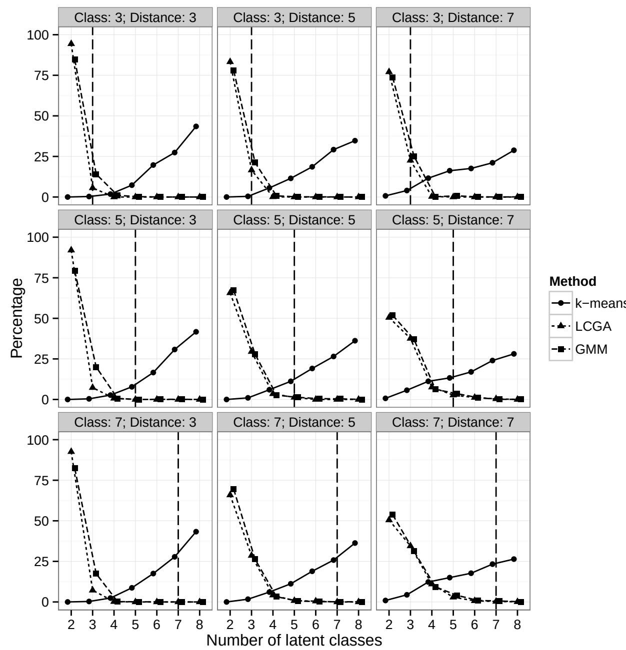

Figure 4.2: Percentage each latent class solution was estimated to be best fitted to the simulated data using different methods for each condition.

Using k-means, it appeared an increasing proportion of simulated data was determined to have more latent classes across all nine simulated conditions. That is, k-means tended to consider solutions with larger number of latent classes as the best fit for the simulated data, regardless of the number of latent classes a priori or the average SRO units difference between classes.

The pattern in which the number of latent classes was considered to be the best fit for the data in each simulated condition was different from that of k-means when using LCGA and GMM. Although the assumption that the trajectories of all individuals in the same class are homogeneous exists only for LCGA but not GMM, individuals with similar SRO patterns across time are classified into latent classes in both LCGA and GMM (Nagin,

Table 4.1: Percentage each latent class solution was estimated to be the best fit to the simulated data using different methods for each condition.

Condition			Latent class solution						
Class	Distance	Method	2	3	4	5	6	7	8
3	3	k-means	0	0.3	1.8	7.3	19.7	27.4	43.5
		LCGA	94.4	5.6	0	0	0	0	0
		GMM	84.7	14.1	1.0	0.2	0	0	0
3	5	k-means	0	0.4	5.6	11.5	18.6	29.2	34.7
		LCGA	83.3	16.6	0.1	0	0	0	0
		GMM	77.9	21.2	0.7	0.1	0.1	0	0
3	7	k-means	0.7	4.0	11.6	16.2	17.6	21.1	28.8
		LCGA	77.1	22.6	0.3	0	0	0	0
		GMM	73.8	25.1	0.2	0.7	0.1	0	0.1
5	3	k-means	0	0.4	2.7	7.8	16.6	30.8	41.7
		LCGA	92.0	7.3	0.7	0	0	0	0
		GMM	79.2	20	0.6	0	0.1	0.1	0
5	5	k-means	0	1.0	6.0	11.2	19.1	26.5	36.2
		LCGA	65.7	29.5	3.5	1.3	0	0	0
		GMM	67.2	27.9	2.7	1.3	0.5	0.4	0
5	7	k-means	0.7	5.7	11.2	13.3	17	24	28.1
		LCGA	50.5	37.6	7.6	2.7	1.2	0.4	0
		GMM	51.9	37.0	6.2	3.6	1.1	0.1	0.1
7	3	k-means	0	0.3	2.4	8.7	17.5	27.8	43.3
		LCGA	92.6	7.2	0.2	0	0	0	0
		GMM	82.3	17.4	0.2	0.1	0	0	0
7	5	k-means	0	1.7	6.1	11.2	18.9	25.8	36.3
		LCGA	65.9	28.6	4.2	0.8	0.5	0	0
		GMM	69.5	26.3	3.2	0.6	0.2	0.1	0.1
7	7	k-means	0.9	4.4	12.3	15.0	17.7	23.3	26.4
		LCGA	50.5	34.5	11.0	2.9	0.7	0.3	0.1
		GMM	53.8	31.2	9.3	4.1	0.9	0.6	0.1

Note. Latent class solution: Percentage of latent class solution estimated to be the best fit to the simulated data. Class: Number of latent classes in the simulated condition. Distance: Average distance (in SRO units) between latent classes in the simulated condition.

1999). As such, the results estimated with the LCGA and GMM are relatively similar (see Table 4.1 and Figure 4.2).

Both LCGA and GMM tended to determine that fewer latent classes was the best fit for the simulated data. Using LCGA and GMM on data simulated for three latent classes, the three-class solution was unlikely to be determined as the best solution. The probabilities of identifying there were three latent classes in the simulated data ranged from 5.6% to 25.1%. Both LCGA and GMM estimated that two latent classes were the best fit for the data, suggesting both methods tend to under-estimate the number of latent classes for the simulated data. Unlike the pattern of over-estimation using k-means, LCGA and GMM were very unlikely to determine that a larger number of latent classes were the best fit for data simulated with three latent classes. It appeared the pattern of under-estimation was consistent for data simulated for five latent classes. The probabilities of correctly determining there existed five latent classes in these simulated conditions ranged between 0% and 3.6%. Similar to the results for data simulated for three latent classes, both LCGA and GMM determined that fewer latent classes was the best fit for data simulated for five latent classes. For data simulated for seven latent classes, the trend to under-estimate the number of latent classes remained consistent, though slightly less prominent. When there existed more number of latent classes a priori, both LCGA and GMM were very unlikely (less than 1% of the time) to estimated there were seven latent classes in the simulated data.

The same patterns of correctly identified latent class number were observed in the 100 datasets simulated with varying proportion of time on streets (Table 4.2). Consistent with the above findings, k-means was very unlikely to identify there existed three latent classes. Among the 100 iterations, the three-class solution was never found to be the best solution. Rather, k-means was much more likely to estimate that the datasets were made up of more latent classes. On the contrary, both LCGA and GMM determined that the datasets consisted of no more than three latent classes. Although the datasets were generated for the three-class condition, both LCGA and GMM was more likely to find the two-class solution to be a better fit for the data (76 - 78% of the time). These findings show the classification results are consistent even when the proportion of time-on-street was generated to vary across individuals at each measurement time points, suggesting the inclusion of control variables (e.g., proportion of time on streets) did not affect the accuracy of classification across different methods.

Table 4.2: Percentage each latent class solution was estimated to be the best fit to the simulated data with varying proportion amount of time out using different methods for the three class condition (Distance = 5).

	Latent class solution						
Method	2 classes	3 classes	4 classes	5 classes	6 classes	7 classes	8 classes
k-means	0	0	4.0	9.0	20.0	20.0	47.0
LCGA	78.0	22.0	0	0	0	0	0
GMM	76.0	24.0	0	0	0	0	0

Note. Latent class solution: Percentage of latent class solution estimated to be the best fit to the simulated data. A total of 100 iterations was performed for each method.

An alternative way to examine the extent to which k-means, LCGA, and GMM identify the solution best fit for the simulated data was to summarize the latent class solutions in terms of less than, equal to, or more than the simulated class condition (Table 4.3 & Figure 4.3). This method is more flexible in identifying the number of latent classes best fitted for the data; the identified solution was considered to be close enough to the simulated class condition if the number of latent class was within a range of one latent class too few or too many. Solutions that were two or more classes less than the simulated class condition were considered to be under-estimated, whereas solutions that were two or more classes more than the simulated class condition were considered to be over-estimated. Using the five latent class condition as an example, solutions between four to six latent classes were considered to be equal to the simulated class condition; solutions less than four latent classes or more than six latent classes were respectively considered to be less than or more than the simulated class condition.

Table 4.3: Percentage in which the latent class solution estimated to be the best fit for the data was less than, equal to, or more than the simulated class condition.

Condition	Method	Less than	Equal	More than
Class: 3; Distance: 3	k-means	-	2.1	97.9
	LCGA	-	100.0	0
	GMM	-	99.8	0.2
Class: 3; Distance: 5	k-means	-	6.0	94.0
	LCGA	-	100.0	0
	GMM	-	99.8	0.2
Class: 3; Distance: 7	k-means	-	16.3	83.7
	LCGA	-	100.0	0
	GMM	-	99.1	0.9
Class: 5; Distance: 3	k-means	0.4	27.1	72.5
	LCGA	99.3	0.7	0.1
	GMM	99.2	0.7	0.1
Class: 5; Distance: 5	k-means	1.0	36.3	62.7
	LCGA	95.2	4.8	0.4
	GMM	95.1	4.5	0.4
Class: 5; Distance: 7	k-means	6.4	41.5	52.1
	LCGA	88.1	11.5	0.4
	GMM	88.9	10.9	0.2
Class: 7; Distance: 3	k-means	11.4	88.6	-
	LCGA	100.0	0	-
	GMM	100.0	0	-
Class: 7; Distance: 5	k-means	19.0	81.0	-
	LCGA	99.5	0.5	-
	GMM	99.6	0.4	-
Class: 7; Distance: 7	k-means	32.6	67.4	-
	LCGA	98.9	1.1	-
	GMM	98.4	1.6	-

Note. Less than: Best class solution identified is at least two classes less than the simulated condition. Equal: Best class solution identified is one class more/less than the simulated condition. More than: Best class solution identified is at least two classes more than the simulated condition. Class: Number of latent classes in the simulated condition. Distance: Average distance (in SRO units) between latent classes in the simulated condition.

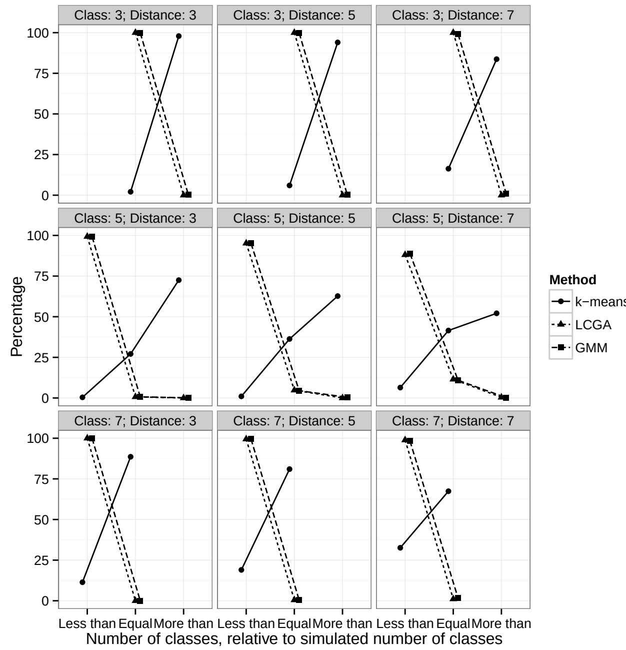

Figure 4.3: Percentage in which the latent class solution estimated to be the best fit for the data was less than, equal to, or more than, the simulated class condition.

When there existed few classes in the data (i.e., number of latent classes = 3), k-means tended to over-estimate the number of classes more than 80% of the time, whereas both LCGA and GMM were able to estimate the correct number of classes almost all the time ($> 98\%$). For data in which there was a moderate number of classes (i.e., number of latent classes = 5), both LCGA and GMM tended to under-estimate the number of classes more than 88% of the time. On the other hand, k-means under-estimated the number of latent classes less than 10% of the time. Consistent with the trend, though to a lesser extent, when there were few classes, k-means tended to over-estimate the number of classes. When there existed more classes in the data (i.e., number of latent classes = 7), k-means estimated the correct number of classes much more often than either LCGA or GMM. As with the previous under-estimation trend using LCGA and GMM, both methods tend to estimate fewer number of classes almost all the time ($> 98\%$) when there were more classes in the simulated data. Such findings were replicated with the 100 datasets simulated with varying proportion of time on streets (Table 4.4). These results suggested k-means tend to over-estimate the number of classes in the data, unless there existed more classes a priori. LCGA and GMM, on the contrary, were more likely to estimate there were fewer number of classes in the data, which is consistent with previous studies showing that large samples were required to obtain accurate estimates in GMMs (e.g., Kim, 2012; Peugh & Fan, 2012).

Table 4.4: Percentage in which the three-class solution estimated to be the best fit for the simulated data with varying amount of time on the streets was less than, equal to, or more than the simulated three class condition (Distance = 5).

Condition	Method	Less than	Equal	More than
Class: 3; Distance: 5	k-means	-	4.0	96.0
	LCGA	-	100.0	0
	GMM	-	100.0	0

Note. Less than: Best class solution identified is at least two classes less than the simulated condition. Equal: Best class solution identified is one class more/less than the simulated condition. More than: Best class solution identified is at least two classes more than the simulated condition. A total of 100 iterations was performed for each method.

Proportion of misclassification

The extent to which individuals in the simulated data were classified into the "correct" class were examined for each simulation condition, using k-means, LCGA, and GMM. Because the latent class to which each simulated individual was known a priori, it is possible to compare the latent class assigned by the methods against the correct class assignment. In addition, the classification results were compared against the probabilities of correct classification by random chance. There existed a 33.4%, 20%, and 14.3% probability, by chance, of correctly classifying an individual into the correct class for the three-, five-, and seven-class solutions, respectively.

First, the proportion of misclassification was examined when the latent class solution fitted to the data was the same as the simulated class condition, regardless of whether the solution was determined to be the best fit for the data. Averaging the proportion of correct classification for all 1,000 data sets in each simulation condition, the results were summarized in Table 4.5 and Figure 4.4.

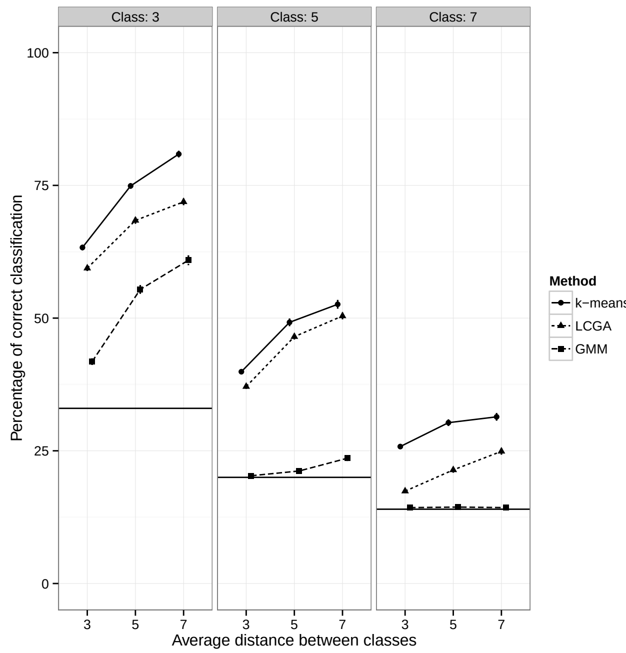

Figure 4.4: Percentage of correct classification using different methods, when the latent class solution fitted to the data was equal to the simulated class condition. (Horizontal line indicates chance performance.)

For data simulated with three latent classes, the proportion of correct classification increases as distance between classes increases, regardless of which method was used. k-means consistently out-performed both LCGA and GMM; on average, k-means tended to classify a higher proportion of individuals into the correct class. Nonetheless, all three methods out-performed the probability of correct classification by chance. Similar results were observed for the 100 datasets generated with varying proportion of time on streets for the three-class, five-distance condition. When the proportion of time on streets varied across individuals for different measurement time points, k-means, LCGA, and GMM correctly assigned individuals into their corresponding latent classes 74.4$\%$ (SE = 1.40), 67.9$\%$ (SE = 1.50), and 55$\%$ (SE = 2.10) of the time, respectively. These findings showed the extent to which methods were accurate in classifying individuals into different latent classes was not affected by the changes in proportion of time on streets.

Table 4.5: Proportion of misclassification using different methods, when the latent class solution fitted to the data was equal to the simulated class condition.

Condition		Random chance		k-means		LCGA		GMM	
Class	Distance	% Correct	S.E.	% Correct	$S.E.$	% Correct	$S.E.$	% Correct	$S.E.$
	3			63.3	0.40	59.4	0.40	41.8	0.50
3	5	33.33	0.015	74.9	0.40	68.4	0.50	55.4	0.70
	7			80.9	0.50	71.9	0.50	60.9	0.80
	3			39.9	0.30	37.1	0.30	20.3	0.10
5	5	20.00	0.013	49.2	0.60	46.5	0.40	21.2	0.20
	7			52.6	0.70	50.4	0.50	23.6	0.40
	3			25.8	0.30	17.4	0.20	14.3	0
7	5	14.29	0.011	30.3	0.50	21.4	0.40	14.4	0
	7			31.4	0.60	24.9	0.50	14.3	0

Note. Random chance: The probability of classifying an individual into the correct (i.e., assigned through simulation) latent class by chance. Class: Number of latent classes in the simulated condition. Distance: Average distance (in SRO units) between latent classes in the simulated condition.

When there existed five latent classes in the simulated data, the proportion of correct classification dropped drastically, although the trend of increased proportion of correct classification with distance between classes remained. The patterns of increased rate of correct classification with distance between classes were very similar between k-means and LCGA. Although k-means consistently out-performed LCGA, the standard errors for k-means also increased with as the distance between classes increased. Nonetheless, less than half of the individuals, on average, were correctly classified into their correct class when the classes were moderately close or close to one another. Even when the classes were relatively different from each other, only approximately 50$\%$ of the individuals, on average, were classified correctly. Compared with k-means and LCGA, which both out-performed the 20$\%$ random chance of correct classification, GMM was barely better than the results by chance. The proportion of correct classification using GMM increased slightly as the distance between classes increased; however, the standard errors also increased as there was more distance between classes. These results suggested that it is possible that the proportion of correct classification using GMM may be no better than random chance.

Compared with the two previous latent class conditions, the proportion of correct classification decreased further for data simulated with seven latent classes. Regardless of which methods were used, less than half of the participants were assigned into their correct class. k-means out-performed both LCGA and GMM, with slight increase in correct classification as the distance between classes increased. However, there also existed an increase in standard errors as the classes were more different from one another. This pattern was similar using LCGA, although the overall proportion of correct classification was smaller than that using k-means. In addition, the probabilities that individuals were put in the correct class using LCGA were not better than random chance. When GMM was used, the probability of correct classification was not better than random chance, regardless of the distance between latent classes.

Next, the proportion of misclassification was examined when the best-fitted latent class solutions was equal to the simulated class conditions. As shown in the previous sections, the simulated class conditions were not necessarily determined to be the best fit for the data, meaning that the number of data sets in which the simulated class conditions were the same as the best fitted solution may vary across simulation conditions and methods. The proportion of correct classification in each simulation condition was thus computed by taking the average of the data sets in which the best fitted number of classes was the same as the simulated condition. The results were summarized in Table 4.6 and Figure 4.5.

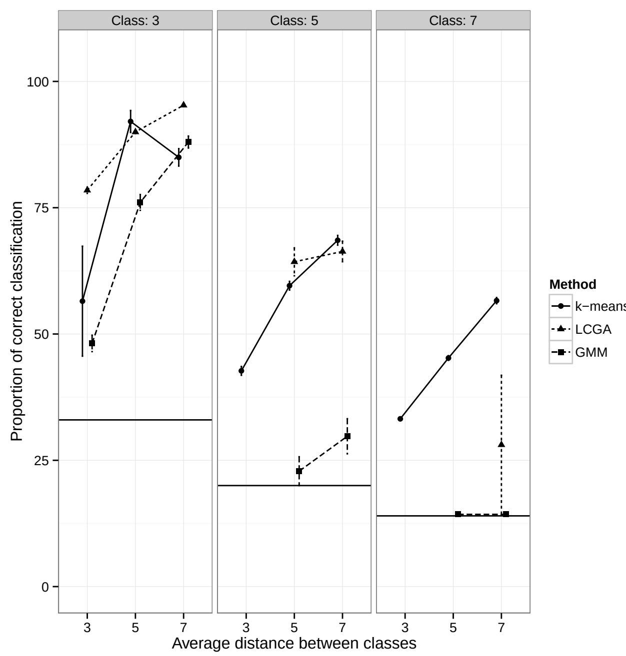

Figure 4.5: Percentage of correct classification using different methods, when the latent class solution estimated to be best-fitted to the data was equal to the simulated class condition.

In the three-class conditions, the proportion of individuals classified into the "correct" class increased as the distance between classes increased for both LCGA and GMM. Compared with GMM, the accuracy rate was higher for LCGA overall. Both methods consistently out-performed the 33.4% probability of correct classification by chance. More importantly, when the classes are relatively different from each other (e.g., 5 or 7 SRO units apart), both methods were able to put individuals into their corresponding a priori classes more than three-quarters of the time. Results from k-means suggested that the use of this clustering method was better than the probability of random correct classification. It appeared that the performance of k-means was best when the classes were the latent classes were further apart, although it did not out-perform LCGA nor GMM. As for conditions in which the latent classes were relatively close to each other (e.g., 3 or 5 SRO units apart), it may be premature to interpret k-means' performance due to the small number of data sets in which the three-class solution was determined to be the best fit for the data.

Using the100 datasets generated with varying proportion of time on streets for the three-class, five-distance condition, the results were partially replicated. As shown in Table 4.2, k-means, LCGA and GMM determined the three-class solution to be the best fit for the data 0$\%$, 22$\%$, and 24$\%$ of the the time, respectively. Considering only the datasets in which the three-class solution was determined to be the best solution, the classification accuracy rates of LCGA and GMM were 88.2$\%$ (SE = 0.98) and 72.3$\%$ (SE = 4.95), respectively. Consistent with the previous findings, results suggested that LCGA out-performed GMM. Given that k-means did not identify the three-class solution to be the best fit for the data, it was not possible to evaluate its classification accuracy rate at present.

When data was simulated for five and seven latent classes, the proportion of correct classification increased with increased distance between classes for k-means. In both conditions, the accuracy rates of classification using k-means was much better than those by random chance. Comparing across the three-, five-, and seven-class conditions, it is noticeable that there existed an overall decrease in classification accuracy as the number of classes in the simulated data increased. When the same number of individuals belong to an increasing number of classes, each class constitutes fewer individuals; it therefore becomes increasingly difficult to correctly assign them into their corresponding classes.

Table 4.6: Percentage of misclassification using different methods, when the best-fitted latent class solution was equal to the simulated class condition.

Condition		Random chance		k-means			LCGA			GMM		
Class	Distance	% Correct	$S.E.$	N	% Correct	$S.E.$	N	% Correct	$S.E.$	N	% Correct	$S.E.$
	3			3	56.50	10.89	56	78.46	0.63	141	48.14	1.63
3	5	33.33	0.015	4	92.07	2.18	166	89.99	0.30	212	76.05	1.57
	7			40	84.97	1.73	226	95.29	0.18	251	88.00	1.19
	3			78	42.71	0.88	0	-	-	0	-	-
5	5	20.00	0.013	112	59.56	0.87	13	64.29	2.77	13	22.85	2.85
	7			133	68.55	0.97	27	66.34	2.04	36	29.76	3.47
	3			278	33.21	0.27	0	-	-	0	-	-
7	5	14.29	0.011	258	45.25	0.46	0	-	-	1	14.30	-
	7			233	56.61	0.62	3	28.07	13.77	6	14.30	0

Note. Random chance: The probability of classifying an individual into the correct (i.e., assigned through simulation) latent class by chance. Class: Number of latent classes in the simulated condition. Distance: Average distance (in SRO units) between latent classes in the simulated condition. N: Number of iterations (out of 1,000) in which the latent class solution estimated to be the best fit to the data was equal to the number of latent classes in the simulated condition.

Turning to results using LCGA, a large decrease was observed moving from the three-, to five-, to seven-class conditions. Although the use of LCGA resulted in much better classification accuracy than the probability by random chance in the five-class conditions, the classification accuracy was similar to that using k-means. However, k-means clearly out-performed LCGA when there existed more classes in the data. As for the classification accuracy using GMM in the five- and seven-class conditions, GMM's performance was no better than random chance. These results were consistent with previous observations that large samples are required to yield accurate estimates using GMM (e.g., Kim, 2012; Peugh & Fan, 2012).

4.2 Application Analyses

The application section of this study utilized the data obtained from the Pathways to Desistance ("Pathways") study (see Mulvey et al., 2004). As described in the methods section, self-report Offending (SRO) data was collected from 1,170 serious offending boys enrolled in "Pathways." SRO information was collected at the time of enrollment, every six months for the following three years, and annually for the subsequent four years. At each time point, individuals were asked to whether they engaged in 22 different offending acts in their lifetime (at baseline interview) and within the past six months or one year (at subsequent interviews). In order to maintain consistency in the time intervals between measurement points, the bi-yearly SRO records were combined into annual SRO records. If an individual indicated he participated in the said offending behavior in both, or either one of the six-month interviews (within the same year), the entry for the offending behavior for the corresponding year was coded as "yes", reflecting his self-reported offending act in the past year. Otherwise, the entry for the offending behavior for the corresponding year was coded as "no", indicating he did not participate in said offending acts in the past year. As such, the data consisted of a total of eight measurement time points of SRO. For each time point, a total SRO score was computed by summing an individual's response on all 22 SRO items, reflecting the amount of self-reported offending acts within the prior year.

All participants' SRO scores across the measurement time points were illustrated in Figure 4.6. Each line in Figure 4.6 represented the SRO data for each participant. Despite the overlapping lines, the amount of offending behavior reported by the individuals appears to decrease over time, suggesting an overall reduction in self-reported offending behavior as these individuals became older. However, the pattern of decreased self-report offending behavior may not be consistent for all individuals. As illustrated in Figure 4.7, the patterns of self-report offending behavior vary for the ten random samples of participants. The offending behavior reported by some individuals decreased across time, whereas others reported a diverse amount of offending acts at different measurement time points.

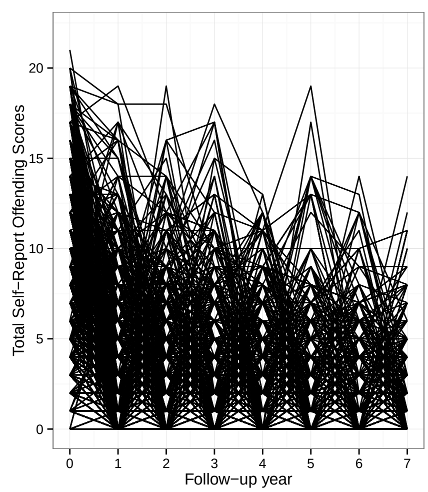

Figure 4.6: Self-report offending scores fro all participants in the "Pathways" data.

In addition to individuals' SRO data, the amount of time individuals were not in custody within each measurement time interval was available in the "Pathways" data. This value, also known as the proportion of time on streets, refers to the amount of time there existed opportunities for participants to engage in offending behavior. This information should be taken into account when examining individuals' offending behavior; as the amount of time on streets increases, there exist more opportunities for individuals to

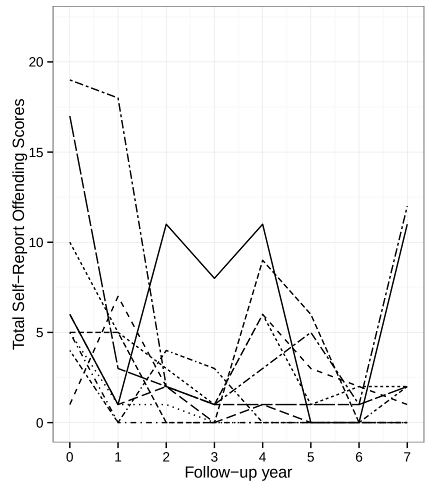

Figure 4.7: A random sample of 10 participants.

participate in offending acts. For instance, suppose two individuals self-reported the same amount of offending behavior within a time period (e.g., 5 SRO), the one who was in custody 80% of the time (i.e., with less time on streets) would be regarded as more delinquent than the one who was in custody only 20% of the time (i.e., with more time on streets). Thus, the subsequent analyses took into account the proportion of time individuals were not in custody.

When modeling the changes in the simulated SRO data, simulation analyses showed conclusions drawn from using k-means, LCGA, and GMM may differ depending on the number of latent classes and the distance between classes in the data set. With respect to the proportion of correctly identified number of classes and the proportion of misclassification, none of the methods consistently out-performed the rest across the simulation conditions. Unlike the simulated data in which the number of latent classes and individuals' "correct" class assignments were know, such information was not known a priori for the "Pathways" data. Thus, all three methods were applied to determine the number of latent classes in the "Pathways" data.

As with the simulation analyses, the number of latent classes fitted to the "Pathways" data ranged from two to eight. To determine the number of latent class solution best fitted for each simulated data, the BIC fit index was used. The model with the smallest BIC value was considered to be the best fit for the data. For cases in which there were ties in the smallest BIC values, parsimony was taken into account such that the model with the fewer latent class solution was considered to be the best fit for the simulated data.

Results of the cluster solutions (2 to 8 clusters) obtained from k-means, LCGA, and GMM were presented in Table 4.7. Using k-means, model fitness increased with increased number of classes, with the eight-class solution found to be the best-fitting solution. When LCGA and GMM were applied to the "Pathways" data, solutions with fewer latent classes–four and six–were found to be the best fit for the data, respectively. However, latent class assignments for the "Pathways" participants, computed based on the posterior class-membership probabilities obtained from GMM using the "lcmm" package (Proust-Lima et al., 2014) in R (R Core Team, 2013), showed there existed only three latent classes (see Table 4.7, 3rd column). This classification results indicated that even though the six-class solution was found to be the best-fitting solution, three of the classes were empty, meaning that individuals had very low probabilities of being assigned to those three classes. Such classification results may have occurred due to inaccurate estimations by the current assumption that the "Pathways" data is normally distributed, when in fact, the data should be modeled as a zero-inflated poisson distribution. In order to make sure this issue was not program specific, the same GMM analyses were repeated with Mplus (Muthen & Muthen, 1998-2012). Although the BICs were remarkably different between the two programs ("lcmm" and Mplus), the pattern of the results were somewhat similar (see below). It can be speculated that the difference may be due to different pre-processing procedures of severely non-normally distributed data, such as the current "Pathways" data. GMM results from Mplus (Table 4.7, 4th column) found the eight-class solution to be the best-fitting solution. Upon examining the classification results, one class was found to consist of less than 1% of the "Pathways" individuals, suggesting that this solution, though best-fitting, has minimum practical implications. The second best-fitting solution, with six latent classes, was also found to include one class that consisted of less than 1% of the participants. As such, the third best solution, with four latent classes, was determined to be the best fit for the "Pathways" data obtained using GMM in Mplus. Because it is more reasonable that a latent class solution in which there exist individuals in each latent class is a better fit for the data than one in which empty classes exist (i.e., classes that consist of no participants), GMM results obtained using Mplus (Muthen & Muthen, 1998-2012) was reported in lieu of those reported using the "lcmm" package (Proust-Lima et al., 2014) in the subsequent sections.

Unlike the simulation analyses, the "correct" number of classes was unknown in the "Pathways" data. Without knowledge of the "correct" number of classes in the data, whether k-means, LCGA, or GMM yielded the most accurate clustering results could not be determined at this point. As such, classification results based on the best-fitting solution obtained using different methods were explored. Subsequently, the extent to which individuals with similar SRO patterns across time were classified into the same classes by different methods was examined.

4.2.1 GMM Classification Results

Applying GMM to the "Pathways" data[1], the solution found to be the best fit for the data consisted of four latent classes. The mean SRO total scores across time for individuals assigned into different classes were presented in Table 4.8 and Figure 4.8. Results from the GMM solution suggested the majority of the "Pathways" participants (84.27%) belong to the same latent class ($Class_{GMM}$ 1). These individual self-reported a moderate amount of offending behavior at the initial interviews; in the following year, the amount of antisocial behavior in which they reported having participated decreased dramatically. In the next

[1]Despite the difference in the BICs, both the relative percentages and the patterns of self-report offending were very similar with the analyses using the "lcmm" package (Proust-Lima et al., 2014) in R (R Core Team, 2013) and Mplus (Muthen & Muthen, 1998-2012).

Table 4.7: Bayesian information criterion (BIC) results of varying latent class solutions, using k-means, LCGA, and GMM.

Solution	Fit criteria: BIC			
	k-means	LCGA	GMM†	
			lcmm()	Mplus
2 classes	50166.21	50947.39	50583.74	57047.02
3 classes	49335.49	50968.52	50604.86	56646.42
4 classes	48605.32	**49944.21**	50625.99	56458.87
5 classes	48168.64	49965.34	50647.12	56511.89
6 classes	47882.40	49986.46	**50229.31**	56279.00
7 classes	47699.81	50007.59	50250.44	56546.19
8 classes	**47545.05**	50289.00	50271.56	**56175.54**

†*Estimation results using both the "lcmm" package (Proust-Lima et al., 2014) in R (R Core Team, 2013) and Mplus (Muthen & Muthen, 1998-2012) were presented here, because three classes were found to be empty upon posterior classification with the results obtained using lcmm(). Note. Smallest BIC values in* ***bold****, indicating the best-fitting solution for the data.*

seven years, they self-reported a relatively consistent amount of minimum (approximately one) antisocial behavior. By the final interview at year seven, these individuals reported, on average, having participated in less than one offending behavior in the prior year. Based on the early decrease in offending behavior, these individuals may be known as the "early desisters," reflecting their early desistance from offending behavior.

The second and third latent class ($Class_{GMM}$ 2 & $Class_{GMM}$ 3) identified by GMM consisted 10.49% and 3.67% of the "Pathways" participants, respectively. These individuals self-reported a relatively large amount of offending behavior (approximately 11 SRO) at the initial interviews (see Table 4.8 & Figure 4.8). However, the amount of offending behavior in which they self-reported having participated decreased consistently over the next seven years. By the final interview, the number of offending acts they reported was about one-third of that in the initial interview. Although these participants were classified into two separate classes, note that the overall patterns of these two groups were very similar, with the exception that the decline in offending behavior was slightly slower for $Class_{GMM}$ 3 participants than $Class_{GMM}$ 2 participants in the first two years. These two groups of individuals may be identified as "slow desisters," indicating their gradual reduction in their involvement of antisocial behavior.

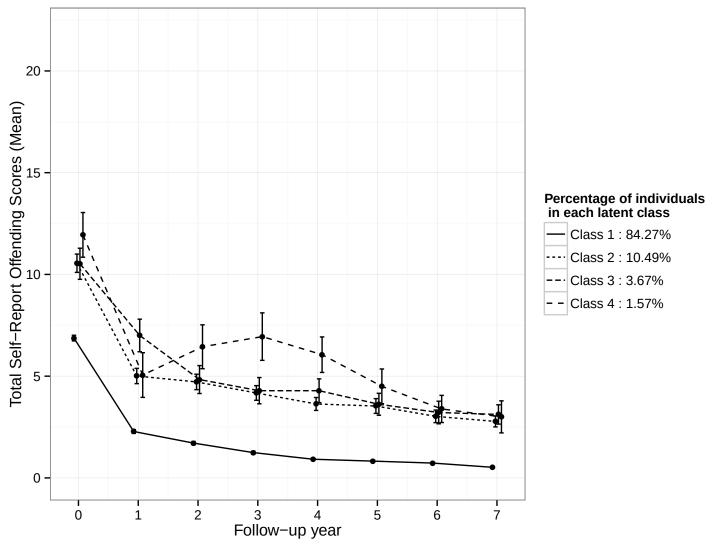

Figure 4.8: Classification results using GMM in Mplus.

The smallest group of individuals (1.57%) identified using GMM followed a different pattern of changes in SRO over time. This group of individuals ($Class_{GMM}$ 4) reported a slightly elevated amount of offending behavior (about 12 SRO) as compared to those in $Class_{GMM}$ 2 & $Class_{GMM}$ 3 , the changes in their participation in antisocial acts appeared to be curvilinear. After the initial drastic decrease in offending behavior in the first year of follow-up, the self-reported amount of antisocial behavior increased slightly for the next two years, then gradually decreased after the third year. The behavior reported by this small group of individuals seemed to suggest continuous offending behavior across time. Even though there existed an eventual reduction in offending behavior, these individuals maintained an average of about three antisocial acts by the final interview. Given their continuous involvement in antisocial behavior, this small group of individuals may be identified as "continuous offenders."

Table 4.8: Average self-report offending scores for each latent class in each measurement time points.

Method	Class	Proportion in each class (%)	Follow-up year 0	1	2	3	4	5	6	7
GMM†	1	84.27	6.87	2.28	1.71	1.24	0.92	0.82	0.73	0.52
	2	10.49	10.55	5.01	4.72	4.17	3.63	3.53	3.02	2.77
	3	3.67	10.52	7.00	4.83	4.29	4.29	3.62	3.21	3.12
	4	1.57	11.94	5.06	6.44	6.94	6.06	4.50	3.39	3.00
LCGA	1	86.98	6.81	2.14	1.75	1.37	1.08	1.02	0.90	0.75
	2	9.27	11.89	7.35	4.75	3.85	2.91	2.38	1.93	1.65
	3	2.45	10.21	5.21	5.54	5.61	5.61	5.25	4.21	2.68
	4	1.31	15.07	8.40	8.87	5.07	4.93	2.47	3.20	1.60
k-means	1	51.84	5.62	1.09	0.82	0.59	0.60	0.54	0.50	0.41
	2	3.67	11.36	7.57	4.50	3.55	1.88	1.50	1.10	1.29
	3	5.51	8.86	3.87	3.52	3.78	3.89	4.02	3.73	3.08
	4	3.32	13.05	7.76	5.18	4.50	4.39	3.63	2.24	2.05
	5	1.66	11.74	6.05	6.37	6.63	4.16	4.79	3.26	2.00
	6	22.81	7.92	3.61	2.80	2.07	1.53	1.38	1.13	0.96
	7	9.62	10.34	4.29	4.02	3.13	1.85	1.68	1.77	1.17
	8	1.57	14.17	8.22	8.11	4.83	4.61	2.28	2.83	1.89

†GMM results from Mplus are presented here. Note. Follow-up year: 0 refers to the initial interview. Total SRO scores range from 0 to 22; higher scores indicate larger amount of offending behavior reported by the individuals. GMM: Growth mixture modeling; LCGA: Latent class growth analysis.

4.2.2 LCGA Classification Results

When LCGA was applied to the "Pathways", a four-class solution was found to be the best fit. The mean SRO total scores across time for individuals assigned into different classes were presented in Table 4.8 and Figure 4.9. Most of the participants (86.98%) were classified into the same latent class ($Class_{LCGA}$ 1). The changes in offending behavior reported by these participants were similar to those of the "early desisters" ($Class_{GMM}$ 1); these individuals started with a moderate amount of offending behavior, reported a large reduction in antisocial acts in the following year, and maintained a minimum amount of illegal activities in the subsequent years. In fact most of the $Class_{LCGA}$ 1 participants were also assigned into $Class_{GMM}$ 1 (Table 4.9), suggesting that both methods were relatively consistent in identifying this group of "early desisters" in the "Pathways" data.

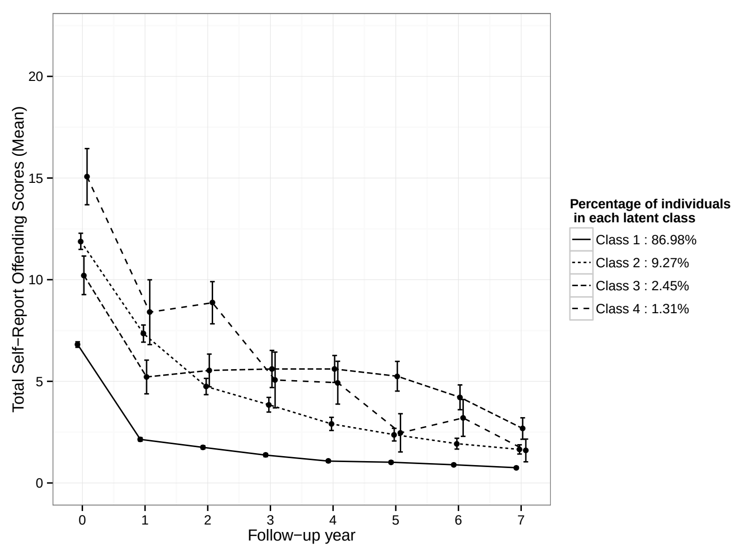

Figure 4.9: Classification results using LCGA.

The second and fourth groups of individuals (9.27% & 1.31%, respectively) identified with LCGA ($Class_{LCGA}$ 2 & $Class_{LCGA}$ 4) showed SRO change patterns similar to those of the "slow desisters" identified using GMM ($Class_{GMM}$ 2 & $Class_{GMM}$ 3). These participants reported having engaged in a relatively large number of offending acts (about 12 SRO) at the initial interviews (see Table 4.8 & Figure 4.9); over time, the amount of antisocial behavior gradually decreased. Although the reduction in offending behavior seemed to be less consistent for participants assigned in $Class_{LCGA}$ 3 than in $Class_{LCGA}$ 2, individuals in both $Class_{LCGA}$ 2 and $Class_{LCGA}$ 4 followed a general decrease in SRO involvement across time.

Table 4.9: Proportion of individuals assigned into varying latent classes using GMM and LCGA.

		LCGA classes			
		1	2	3	4
GMM classes	1	79.55 %	4.63 %	-	0.09 %
	2	6.47 %	3.32 %	0.09 %	0.61 %
	3	0.96 %	1.22 %	1.05 %	0.44 %
	4	-	0.09 %	1.31 %	0.17 %

Note. GMM classes referred to the latent classes identified by the best fitting solution using GMM (in Mplus). LCGA classes refer to the latent classes identified by the best fitting solution using LCGA.

The pattern of changes in SRO scores for individuals classified to be in $Class_{LCGA}$ 3 with LCGA were similar to that of participants in $Class_{GMM}$ 4. This small group of individuals (2.45%) reported being involved in a relatively large amount of offending at the initial assessments. Despite a reduction in offending behavior after the first year, they maintained a moderate amount of involvement in antisocial behavior over time. These individuals appeared to be relatively consistent with those identified as "continuous offenders" using GMM.

4.2.3 k-means Classification Results

The solution k-means found to be best-fitted to the "Pathways" data consisted of eight latent classes. The mean SRO total scores across time for individuals assigned into different classes were presented in Table 4.8 and Figure 4.10. The largest three groups of individuals (51.84%, 22.81%, & 9.62%) identified using k-means ($Class_{kmeans}$ 1, $Class_{kmeans}$ 6, & $Class_{kmeans}$ 7) were similar to the participants in $Class_{GMM}$ 1 and/or

$Class_{LGCA}$ 1 (Table 4.10). These individuals reported initial moderate amount of offending behavior, followed by a substantial reduction in illegal activities and continued to be involved in a minimum amount of antisocial acts in the subsequent years. SRO patterns reported by individuals in $Class_{kmeans}$ 1, 6, and 7 were consistent with those of "early desisters."

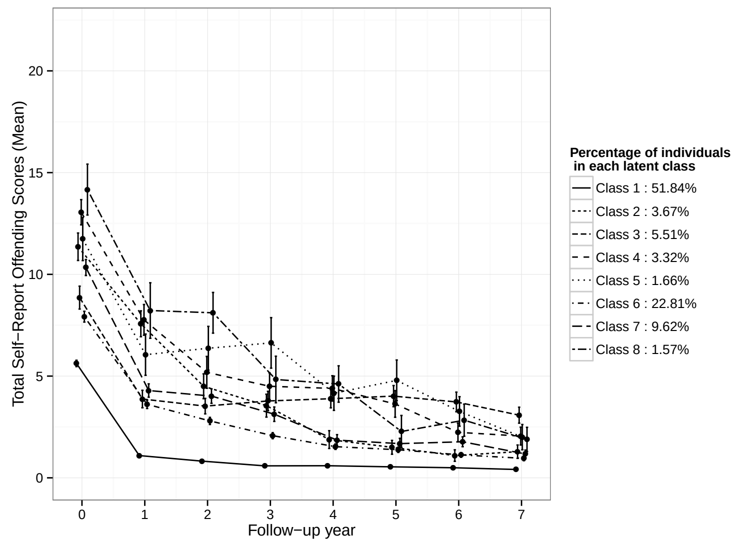

Figure 4.10: Classification results using k-means.

Individuals demonstrating similar SRO behavior across time as the "slow desisters" were identified in $Class_{kmeans}$ 2, $Class_{kmeans}$ 4, and $Class_{kmeans}$ 8 (Table 4.8 $\&$ Figure 4.10). Participants assigned to be in these groups follow a general trend of decreased involvement in offending behavior over time, despite reported having engaged in a moderate amount of antisocial acts at the initial assessments. Most of the individuals classified into these three classes were also assigned to the "slow desisters" group using GMM ($Class_{GMM}$ 2 $\&$ $Class_{GMM}$ 3) and LCGA ($Class_{LCGA}$ 2 $\&$ $Class_{LCGA}$ 4). Nonetheless, inconsistencies between methods were also observed; most of the participants in $Class_{kmeans}$ 2 were identified as "early desisters" ($Class_{GMM}$ 1) in GMM, whereas most were determined to be "slow desisters" ($Class_{LCGA}$ 2) in LCGA. It is possible that different techniques may have different classification criteria for the individuals, thus resulting in classification differences between methods.

The final two groups identified using k-means ($Class_{kmeans}$ 3 & $Class_{kmeans}$ 5) self-reported having been involved in a relatively large number of antisocial behavior initially (Table 4.8 & Figure 4.10). Although they reported reduced involvement in offending behavior after the first year, these individuals participated in a moderate amount of illegal activities over time. This pattern of offending behavior seemed to be consistent with those identified as "continuous offenders" with GMM ($Class_{GMM}$ 4) and LCGA ($Class_{LCGA}$ 3).

Table 4.10: Proportion of individuals assigned into varying latent classes using GMM and k-means.

		k-means classes							
		1	2	3	4	5	6	7	8
GMM classes	1	51.49 %	3.15 %	0.52 %	0.26 %	0.35 %	20.63 %	7.78 %	0.09 %
	2	0.35 %	0.44 %	2.80 %	2.19 %	-	2.19 %	1.84 %	0.70 %
	3	-	-	2.01 %	0.87 %	0.17 %	-	-	0.61 %
	4	-	0.09 %	0.17 %	-	1.14 %	-	-	0.17 %
LCGA classes	1	51.84 %	0.09 %	3.93 %	-	0.26 %	22.81 %	8.04 %	-
	2	-	3.58 %	0.44 %	3.32 %	0.09 %	-	1.57 %	0.26 %
	3	-	-	1.14 %	-	1.31 %	-	-	-
	4	-	-	-	-	-	-	-	1.31 %

Note. GMM classes refer to the latent classes identified by the best fitting solution using GMM (in Mplus). LCGA classes refer to the latent classes identified by the best fitting solution using LCGA. k-means classes refer to the latent classes identified by the best fitting solution using k-means.

5. Discussion

The primary objective of the present study was to evaluate how well several methods commonly used to model the heterogeneity of changes in repeatedly measured data, focusing specifically on the changes in self-report offending behavior as offending adolescents transition into young adults. Because the goal of modeling longitudinal data is to identify groups of individuals with similar patterns of change across time, the simulation study examined the performance of growth-mixture modeling (GMM), latent class growth analysis (LCGA), k-means clustering, and latent transition analysis (LTA; even though it was found to be infeasible for data of this size) across a variety of latent class numbers and distances between classes.

Results from the simulation analyses yielded three main findings. First, the correct number of, or close enough, latent classes were much more likely to be obtained using GMM and LCGA than k-means, when the data consisted of few latent classes. On the other hand, when there existed a relatively large number of latent classes, k-means were much more likely to identify the correct number of, or close enough, latent classes. Similar patterns were observed relatively consistently across varying distances between latent classes. Such findings suggested that GMM and LCGA tended to choose fewer latent classes, whereas k-means were more likely to separate the data into more classes. Note that the sizes of the simulated latent classes varied with the Class conditions (n = 334, 200, and 143, for Class = 3, 5, and 7, respectively), whereas the total sample sizes remained relatively similar (N = 1,000 to 1,002). For data in which fewer latent classes existed, the similarities within classes, and/or differences between classes, may be more readily identified with GMM and LCGA, given the large sample sizes within each class. The presence of more latent classes in the data may be an indication of more heterogeneous data. Previous research (Peugh & Fan, 2012) suggested that even larger sample sizes may be required to obtain accurate parameter estimates for GMM; it is possible that a sample size of $N \sim$ 1,000 may not be sufficient for GMM and LCGA to identify the five or seven classes in the data. As such, GMM and LCGA may yield fewer, moderately-sized classes, whereas k-means obtains more, small-sized classes.

Second, the proportion of misclassification clearly increased as the number of latent classes increased in the simulated data. When there exists more latent classes in the data, the number of choices increases. For instance, with only three latent classes, individuals can only be assigned to one of the three classes, whereas individuals can be classified to one of the seven classes if there are seven classes. Accordingly, the probability of assigning an individual into an incorrect class increases with increased number of latent classes in the data. Although this pattern was similar for GMM, LCGA, and k-means, GMM was consistently out-performed by the other two methods. In particular, a drastic increase in the proportion of misclassification was observed when there were five or more classes in the simulated data; the probability of assigning an individual into their correct class was no better than random chance. These results demonstrated that, compared to k-means and LCGA, GMM may not be as accurate in classifying individuals into their corresponding classes. It is possible that, as larger sample sizes were suggested to obtain accurate parameter estimates for GMM (Kim, 2012; Peugh & Fan, 2012), larger class sizes may be required for GMM in order to achieve correct classification.

Third, the proportion of misclassification decreased with increased distance between latent classes. One can think about the distance between latent classes as the distinctiveness of each class. If the distance between two classes is small, one may be less certain there exist two distinct classes; when the distance between two classes is large, it may be easier to determine the presence of two distinct classes. As such, this finding was not surprising in and of itself. However, this trend of reduced misclassification with increased distance between classes was less prominent with increased number of latent classes. These results indicated that classification accuracy operates as a function of the number of latent classes and the distance between classes.

Overall, results from the simulation analyses suggested that GMM and LCGA tended to obtain fewer latent classes, whereas k-means were more likely to identify more latent classes. With respect to classification accuracy, individuals are more likely to be correctly classified into their corresponding classes when there exists a small number of relatively distinct latent classes. Furthermore, k-means was found to consistently perform better than, or at least as well as, the other two methods across varying simulation conditions. These results correspond to the suggestion that there were no "rule of thumb" for minimum sample sizes for k-means cluster analysis (Mooi & Sarstedt, 2011). It is also possible that, although the optimal number of latent classes determined by k-means might be incorrect, the k-means solution might be an adequate representation of the true solution. The Kullback-Leibler (KL) divergence was recently proposed as a metric, regardless of whether or not the optimum latent class solution was correct, to evaluate the solutions obtained from clustering algorithms (Martin & von Oertzen, 2015).

The KL divergence is a measure of relative entropy, which represents the amount of information lost when the distribution yielded from the the optimum solution was used to approximate the true distribution. KL divergence is defined as:

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{+\infty} \ln\left(\frac{p(x)}{q(x)}\right) p(x)dx, \tag{5.1}$$

where P and Q are distributions of a continuous random variable, $p(x)$ refers to the true distribution, and $q(x)$ refers to a model, or approximated distribution, of $q(x)$. KL divergence represents the amount of information lost when $q(x)$ is used to approximate $p(x)$, where lower values indicate $q(x)$ is a better approximation of $p(x)$ (Kullback & Leibler, 1951). In the context of the current simulation study, $p(x)$ would be the simulated data in which individuals were assigned into their correct latent class, and $q(x)$ would be the optimum solution obtained from the clustering methods. Unfortunately, no closed form solution is available for KL divergence for Gaussian mixtures, which is the typically assumed distribution in clustering approaches. To approximate the KL divergence, Martin and von Oertzen (2015) suggested a Monte Carlo simulation by:

1. Randomly select n independent samples from $p(x)$.
2. Compute $\widehat{D_{KL}}(P \parallel Q) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{p(x)}{q(x)}\right)$.

The use of the KL divergence would complement the current evaluation of clustering methods. Because the estimation of KL divergence does not depend on the number of latent classes, either in the simulated data or in the optimum solution, the evaluation of method performance may be extended beyond the choice of correct number of classes and proportion of misclassification. It would, thus, be possible to compare the extent of misfit between solutions of different number of latent classes (Martin & von Oertzen, 2015). For example, one may question whether the "incorrect" eight-class solution identified using k-means approximate the actual simulated data as well as the "correct" three-class solution chosen by GMM and LCGA. Based on the evaluation criteria utilized in the current simulation study, one would conclude that GMM and LCGA outperformed k-means. However, if the estimated KL divergence values were similar for the three solutions, one would reach the conclusion that k-means performed at least as well as GMM and LCGA. On the other hand, suppose the estimated KL divergence value was the smallest for the k-means solution, one would conclude that, although the correct latent class solution was selected with GMM and LCGA, classification results obtained using k-means was actually a better fit for the data.

The exploration of model misfit based on the KL divergence is beyond the scope of the current study, however, a second simulation study is suggested to compare solutions obtained by different clustering techniques using the KL divergence. Because the correct number of latent class and the classification of individuals are unknown in real data, the goal of statistical modeling is to identify the model(s) that best approximate the actual data. It would be reasonable to state that a solution that approximates the data well is preferred over one that does not. One may also argue that a many-class solution that approximates the data well may be more favorable than a few-class solution that does not. Accordingly, such a study would provide information as to how the optimum solution for latent class is related to classification misfit, when different clustering approaches are used.

In order to examine how GMM, LCGA, and k-means model the changes in real self-report offending behavior data, these three methods were applied to the SRO total scores in the "Pathways" data. Because the "Pathways" data was collected from real offending adolescents, neither the correct number of latent classes nor individuals' correct class assignments were known. Consistent with the simulation results, GMM and LCGA selected solutions with fewer latent classes, whereas k-means opted for those with more latent classes. GMM and LCGA determined the "Pathways" participants can be classified into three to four latent classes, based on their SRO scores. On the other hand, k-means identified there were as many as eight groups of individuals in the "Pathways" data.

Given that the true classifications were unknown in real data, classification results obtained by different methods were explored by examining the differences in SRO score changes in different latent classes. Based on the changes in SRO scores over time, individuals in the "Pathways" data may belong into one of the three broad categories–early desisters, slow desisters, and continuous offenders. Early desisters referred to those who started with a moderate amount of offending acts, and quickly reduced their involvement in illegal behavior to a minimum across time. Slow desisters follow a similar decline in offending behavior over the years; on average, however, they reported having participated in more antisocial behavior over the years than the early desisters. The continuous offenders, as the term suggests, remained involved in unlawful behavior over time.

It is worth noting that different GMM results were obtained using the "lcmm" package (Proust-Lima et al., 2014) in R (R Core Team, 2013) and Mplus (Muthen & Muthen, 1998-2012). When applied to the "Pathways" data, the two programs obtained different model fitness (as indicated by BICs) for varying cluster solutions. The number of latent classes determined to be the best fit for the data also differed. Nonetheless, the patterns of change in SRO scores were similar for the three major classes described above–early desisters, slow desisters, and continuous offenders. Such findings suggested that, despite differences in likelihood values obtained, both the "lcmm" package and Mplus were able to identify similar groups of individuals in the "Pathways" data.

The application results showed the large-sized latent classes identified using GMM and LCGA were separated into several small-sized latent classes by k-means. Perhaps due to the need for large sample sizes previously described, GMM and LCGA appeared to capture the overall changes in SRO pattern among large groups of individuals, thus creating relatively large-sized latent classes. k-means seemed to be able to capture small differences in SRO change patterns among groups of participants, therefore resulting in numerous latent classes, each consisted of a smaller number of individuals. For GMM and/or LCGA classes that were identified as separate classes using k-means, closer examination of the SRO scores showed the smaller classes actually follow similar patterns of change, but differ in the amount of self-reported offending behavior. Such observation suggested these classes may not be truly distinct groups of individuals; rather, these groups of participants differ in the severity of involvement in offending behavior.

At this point, one may recall the earlier discussion about evaluating model misfit using the KL divergence (Kullback & Leibler, 1951; Martin & von Oertzen, 2015). It is possible that, by choosing a larger number of latent classes as the optimum solution using k-means, the SRO changes over time were slightly different between classes. These slight differences may not be sufficiently distinct enough for GMM and LCGA; thus, fewer latent classes were selected using GMM and LCGA. As such, one may question whether the optimum solutions yielded by GMM, LCGA, and k-means approximate the "Pathways" data equally well? It is suggested that, following the second simulation study previously proposed, the extent to which each solution approximates the "Pathways" data using the KL divergence should be examined.

When modeling techniques are applied to real-world data, such as the "Pathways" data in the present study, it may not be feasible to choose the optimal number of latent classes purely based on model fit. The perfect model that best describes the raw data would, ideally, be one in which the every participant belongs to his/her own latent class. However, such a perfect model has almost no practical implications, as it is no different from the raw data. Rather, the goal of modeling longitudinal data is to determine the most parsimonious way that best approximates the raw data. In addition to model fit indices (e.g., BIC, KL divergence), it may be advisable to consider the proportion of individuals in each latent class in the determination of the optimal latent class solution. Depending on the sample size, researchers may want to decide upon a minimum threshold for the proportion of individuals to be included in a latent class. For example, one may determine that a latent class should include at least 2% of the individuals (i.e., 20) in a data set with 1,000 participants. It is suggested that this decision should be made in consideration with the substantive topic. For instance, the objective of modeling adolescents' offending behavior as they transition into young adults is to identify distinctive patterns of behavior between adolescent-limited (AL) and life-course persistent (LCP) offenders. Prior studies have shown there exists a small subgroup of LCP offenders (DeLisi et al., 2009; Moffitt, 1993); but how small can this group of offenders be until there are no practical implications? If a new policy was to be implemented targeting this small group of LCP offenders, an informed decision should be made by also taking into account the cost-effectiveness. Should the policy be implemented if there were 20 LCP offenders in every 1,000 delinquent youths? What if there were 10, or 5, in every 1,000 adolescent offenders?

In summary, researchers who are interested in modeling the heterogeneity of longitudinal data should take advantage of the varying latent class solutions selected by different methods. As k-means seem to be able to capture small differences in pattern changes across time, the optimal solution chosen by k-means may be used for exploratory purposes. The various number of latent classes described using k-means may be informative in describing many diverse change patterns across time. Results yielded with GMM and/or LCGA may offer a broad picture of the data, with a few general trends in the changes of measured behavior over time. Taken together, it would thus be possible for researchers to identify potential subgroups of individuals within larger latent classes. Because the true number of latent classes is unknown in real data, a more informative picture about the data may be obtained by taking full advantage of more than one clustering techniques.

5.1 Limitations and Future Directions

Several limitations in the current study should be noted. First, results regarding classification accuracy in terms of the proportion of misclassification were limited to solutions that yielded the correct number of latent classes. As described above, it is possible for solutions that selected an incorrect number of latent classes to approximate the true distribution as well as, or even better than, those that obtained the correct number of classes (Martin & von Oertzen, 2015). In order to obtain a comprehensive picture of how well different cluster solutions may, or may not, represent the true distribution, the next logical step will be to examine the differences using the estimated KL divergence (Kullback & Leibler, 1951) across varying simulation conditions. Taking into account the extent to which the KL divergence values vary across simulation conditions, it would be possible to explore how the optimum solution obtained using different approaches may be related to classification misfit. As such, a more comprehensive evaluation of the utility of different techniques in modeling the heterogeneity of longitudinal data may be obtained.

Second, the simulation work was conducted under the assumption that the data was a mixture of Gaussian distributions, in which the simulated SRO data can be completely summarized by their means and covariances (Bauer & Curran, 2003). In order to examine how well different techniques perform in the real data, the same assumption was made for the "Pathways" data. However, the self-report offending (SRO) measure is a count variable; a value of 0 represents no involvement in illegal activities, whereas a positive value indicates the participation in at least one type of offending behavior. Closer examination of the "Pathways" data further indicated that, even among this group of serious offending adolescents, the majority of youths indicated having no involvement in offending behavior, resulting in an excessive number of participants with a count of 0 SRO in the data. It is, therefore, posited that a zero-inflated Poisson model should be fitted to the "Pathways" data (Piquero et al., 2013). It is reasonable, then, to expand the current simulation study to explore how well the different methods perform when the data does not constitute a mixture of Gaussian distributions. Results from these analyses would further inform researchers as to what modeling approaches should be applied, based on the differences in data distribution.

Given the amount of computation required for latent transition analysis (LTA), an evaluation of the LTA approach in modeling the changes in repeated measured data was not feasible in the present study. Unlike GMM, LCGA, or k-means that model the total SRO scores, the changes in each SRO item over time are modeled using LTA. By taking into account item-level data, it is possible that LTA may yield different solutions from those obtained from GMM, LCGA, or k-means, across varying simulation conditions. To evaluate LTA's performance under the current simulation conditions, one may consider reducing the size of the contingency table representing the changes in item responses across time by *a*) reducing the number of simulated SRO items in the data, and/or *b*) reducing the number of measurement time points. Rather than simulating all 22 SRO items, it may be more feasible to simulate summarized SRO scores for a few types of offenses (e.g., income offense, drug offense, aggressive offense). It may also be advisable to reduce the number of time points by selecting the time points that were critical to changes in behavior, such as the first (initial interview), third, fifth, and the eighth (last assessment) time points. By reducing the size of the contingency table, it may then be possible to evaluate how well

LTA performs under varying simulation conditions, in comparison with the other modeling approaches.

5.2 Conclusion

The current study evaluated the modeling results obtained using GMM, LCGA, and k-means, under a variety of simulation conditions. Results showed GMM and LCGA were likely to select solutions with fewer latent classes, whereas $k-$means tended to opt for many-class solutions. With respect to classification accuracy, $k-$means demonstrated the most stable performance across all simulation conditions, with performance at least as well as, if not better than, GMM and LCGA. When the self-report offending behavior data was modeled using these three methods, the larger-sized classes identified with GMM and/or LCGA were found made up of two or more smaller-sized classes identified using k-means. The present findings suggested that researchers should take advantage of multiple modeling techniques to obtain a comprehensive picture of their repeatedly measured data.

References

Aldenderfer, M. A., & Blashfield, R. K. (1984). *Cluster analysis*. Beverly Hills, CA: Sage.

Auerbach, A. J., & Collins, L. M. (2006). A multidimensional developmental model of alcohol use during emerging adulthood. *Journal of Studies on Alcohol*, *67*, 917–925.

Azzalini, A., & Scarpa, B. (2012). *Data analysis and data mining.* New York, NY: Oxford University Press.

Barnett, A., & Lofaso, A. J. (1985). Selective incapacitation and the Philadelphia cohort data. *Journal of Quantitative Criminology*, *1*, 3–36.

Bartusch, D. R., Lynam, D. R., Moffitt, T. E., & Silva, P. A. (1997). Is age important? Testing a general versus a developmental theory of antisocial behavior. *Criminology*, *35*, 13–48.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338-363.

Benghezal, M. O. (2013). kmlcov: Clustering longitudinal data using the likelihood as a metric of distance [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=kmlcov (R package version 1.0.1)

Boers, K., Reinecke, J., Seddig, D., & Mariotti, L. (2010). Explaining the development of adolescent violent delinquency. *European Journal of Criminology*, *7*(6), 499-520. doi: 10.1177/1477370810376572

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models. a structural equation perspective.* New York, NY: Wiley.

Caulkins, J. P. (2001). How large should the strike zone be in “three strikes and you’re out” sentencing laws? *Journal of Quantitative Criminology*, *17*, 227–246.

Collins, L. M. (2002). Using latent transition analysis to examine the gateway hypothesis. In D. B. Kandel & M. Chase (Eds.), *Stages and pathways of drug involvement: Examining the gateway hypothesis* (pp. 254–269). Cambridge, UK: Cambridge University Press.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: John Wiley & Sons,.

Decker, S. H., & Salert, B. (1988). Predicting the career criminal: An empirical test of the greenwood scale. *Journal of Criminal Law and Criminology*, *77*, 215–236.

DeLisi, M., Barnes, J. C., Beaver, K. M., & Gibson, C. L. (2009). Delinquent gangs and adolescent victimization revisited: A propensity score matching approach. *Criminal Justice and Behavior*, *36*(8), 808-823. doi: 10.1177/0093854809337703

Donovan, K. L., & Brassard, M. R. (2011). Trajectories of maternal verbal aggression across the middle school years: Associations with negative view of self and social problems. *Child abuse & neglect*, *35*(10), 814-830. doi: 10.1016/j.chiabu.2011.06.001

D’Unger, A., Land, K., McCall, P., & Nagin, D. (1998). How many latent classes of delinquent/criminal careers? Results from mixed poisson regression analyses. *American Journal of Sociology*, *103*, 1593–1630.

Eggleston, E. P., Laub, J. H., & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology*, *20*(1), 1-26. doi: 10.1023/B:JOQC.0000016696.02763.ce

Espelage, D. L., Cauffman, E., Broidy, L., Piquero, A. R., Mazerolle, P., & Steiner, H. (2003). A cluster-analytic investigation of MMPI profiles of serious male and female juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(7), 770-777. doi: 10.1097/01.CHI.0000046877.27264.F6

Everitt, B. S. (2006). *The Cambridge dictionary of statistics.* Cambridge, UK: Cambridge University Press.

Ezell, M. E., & Cohen, L. E. (1997). Age, crime, and crime control policies: A longitudinal analysis of youthful, serious, chronic offenders with implications for "three strikes" legislation. *Studies on Crime and Crime Prevention*, *6*, 169–199.

Farrington, D. P. (1996). The explanation and prevention of youthful offending. In J. Hawkins (Ed.), *Delinquency and crime: Current theories.* Cambridge, UK: Cambridge University Press.

Greenwood, P. W. (1983). *Selective incapacitation*. Santa Monica, CA: Rand.

Hair, J. F., & Black, W. C. (2000). Cluster analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 147–205). Washington, DC: American Psychological Association.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, *34*, 93-101.

Henry, B., Capsi, A., Moffitt, T. E., & Silva, P. (1996). Temperament and familial predictors of violent and non-violent criminal convictions: From age 3 to age 18. *Developmental Psychology*, *32*, 614–623.

Hoeve, M., Blokland, A., Dubas, J. S., Loeber, R., Gerris, J. R. M., & van der Lann, P. H. (2008). Trajectories of delinquency and parenting styles. *Journal of abnormal child psychology*, *36*(2), 223-235. doi: 10.1007/s10802-007-9172-x

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302-317.

Kass, R. E., & Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association*, *90*, 775–795.

Kim, S.-Y. (2012). Sample size requirements in single- and multiphase growth analysis and growth mixture models: A Monte Carlo simulation study. *Structural Equation Modeling*, *19*, 457-479.

Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*, 79-86.

Lacourse, E., Dupere, V., & Loeber, R. (2008). Developmental trajectories of violence and theft. In R. Loeber, D. P. Farrington, M. Stouthamer-Loeber, & H. R. White (Eds.), *Violence and serious theft.* New York/London: Rouledge.

Lauritsen, J. L. (1998). The age-crime debate: Assessing the limits of longitudinal self-report data. *Social Forces*, *77*, 127–155.

Males, M. A., & Brown, E. A. (2014a). Authors' rejoinder to respondents (Shulman, Steinberg, & Piquero, 2013). *Journal of Adolescent Research*, *29*(1), 35-44.

Males, M. A., & Brown, E. A. (2014b). Teenagers' high arrest rates: Features of young age or youth poverty? *Journal of Adolescent Research*, *29*(1), 3-24.

Martin, D. P., & von Oertzen, T. (2015). Growth mixture models outperform simpler clustering algorithms when detecting longitudinal heterogeneity, even with small sample sizes. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 264-275.

Meredith, M., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychological Review*, *100*(4), 674–701.

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

Moffitt, T. E., Lynam, D., & Silva, P. A. (1994). Neuropsychological test predicting persistent male delinquency. *Criminology*, *32*, 277–300.

Moffitt, T. E., & Walsh, A. (2003). The adolescence-limited/life-course persistent theory of antisocial behavior: What have we learned? In A. Walsh & L. Ellis (Eds.), *Biosocial criminology: Challenging environmentalism's supremacy.* Hauppauge, NY: Nova Science Publishers.

Monahan, K. C., Steinberg, L., Cauffman, E., & Mulvey, E. P. (2009). Trajectories of antisocial behavior and psychosocial maturity from adolescence to young adulthood. *Developmental Psychology*, *45*(6), 1654–1668.

Monahan, K. C., Steinberg, L., Cauffman, E., & Mulvey, E. P. (2013). Psychosocial (im)maturity from adolescence to early adulthood: Distinguishing between adolescence-limited and persisting antisocial behavior. *Developmental Psychopathology*, *25*, 1093-1105.

Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research*. Berlin, Germany: Springer-Verlag.

Mulvey, E. P., Steinberg, L., Fagan, J., Cauffman, E., Piquero, A. R., Chassin, L., ... Losoya, S. H. (2004). Theory and research on desistance from antisocial activity among serious adolescent offenders. *Youth Violence and Juvenile Justice*, *2*(3), 213–236. doi: 10.1177/1541204004265864

Mulvey, E. P., Steinberg, L., Piquero, A. R., Besana, M., Fagan, J., Schubert, C., & Cauffman, B. (2010). Trajectories of desistance and continuity of antisocial behavior following court adjudication among serious adolescent offenders. *Development and Psychopathology*, *22*, 453–475.

Muthen, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469.

Muthen, L. K., & Muthen, B. O. (1998-2012). *Mplus user's guide. Seventh Edition*. Los Angeles, CA: Muthen & Muthen.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child development*, *70*(5), 1181-1196. doi: 10.1111/1467-8624.00086

Odgers, C. L., Moretti, M. M., Burnette, M. L., Chauhan, P., Waite, D., & Reppucci, N. D. (2007). A latent variable modeling approach to identifying subtypes of serious and violent female juvenile offenders. *Aggressive Behavior*, *33*(4), 339-352. doi: 10.1002/ab.20190

Osgood, D. W., O'Malley, P., Bachman, J., & Johnston, L. (1989). Time trends and age trends in arrests and self-reported illegal behavior. *Criminology*, *27*, 389–418.

Patrick, M. E., Collins, L. M., Smith, E., Caldwell, L., Flisher, A., & Wegner, L. (2009). A prospective longitudinal model of substance use onset among south african adolescents. *Substance Use & Misuse*, *44*, 647–662.

Peugh, J., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? A simulation study examining GMM's performance characteristics. *Structural Equation Modeling*, *19*, 204-266.

Piquero, A. R. (2000). Testing Moffitt's hypothesis for life-course-persistent offending. *Psychology, Crime, and Law*, *7*, 193–216.

Piquero, A. R. (2008). Taking stock of developmental trajectories of criminal activity over the life course. In A. Liberman (Ed.), *The long view of crime: A synthesis of longitudinal research* (pp. 23–78). New York, NY: Springer.

Piquero, A. R., & Brezina, T. (2001). Testing Moffitt's account of adolescence-limited delinquency. *Criminology*, *39*, 353–370.

Piquero, A. R., Monahan, K. C., Glasheen, C., Schubert, C. A., & Mulvey, E. (2013). Does time matter? Comparing trajectory concordance and covariate association using time-based and age-based assessments. *Crime & Delinquency*, *59*, 738–763.

Proust-Lima, C., Philipps, V., Diakite, A., & Liquet, B. (2014). lcmm: Estimation of extended mixed models using latent classes and latent processes [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=lcmm` (R package version 1.6.6)

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rafter, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–164.

Roeder, K., Lynch, K. G., & Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association*, *94*, 766–776.

Schwarz, G. (1987). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Shulman, E. P., Steinberg, L., & Piquero, A. R. (2013). The age-crime curve in adolescence and early adulthood is not due to age differences in economic status. *Journal of Youth and Adolescence*, *42*, 848-860.

Shulman, E. P., Steinberg, L., & Piquero, A. R. (2014). A mistaken account of the age-crime curve: Response to Males and Brown (2013). *Journal of Adolescent Research*, *29*, 25-34.

Steinberg, L., & Cauffman, E. (1996). Maturity of judgment in adolescence: Psychosocial factors in adolescent decision making. *Law and Human Behavior*, *20*, 249-272.

Steinberg, L., & Cauffman, E. (2000). (Im)maturity of judgment in adolescence: Why adolescents may be less culpable than adults. *Behavioral Sciences and the Law*, *18*, 741-760.

Thornberry, T. P. (2005). Explaining multiple patterns of offending across the life course and across generations. In R. J. Sampson & J. H. Laub (Eds.), *Developmental criminology and its discontents. the annals of the american academy of political and social science.* Thousand Oaks, CA: Sage.

Tibbetts, S. G., & Piquero, A. R. (1999). The influence of gender, low birth weight, and disadvantaged environment in predicting early onset of offending: A test of Moffitt's interactional hypothesis. *Criminology*, *37*, 843–877.

Tittle, C. (2000). Theoretical developments in criminology. In *National institute of justice 2000, vol. 1. the nature of crime: continuity and change.* Washington, DC: National Institute of Justice.

Walsh, A. (2002). *Biosocial criminology: Introduction and integration.* Cincinnati, OH: Anderson.

Walsh, A. (2009). Crazy by design: A biosocial approach to the age-crime curve. In A. Walsh & K. M. Beaver (Eds.), *Biosocial criminology: New directions in theory and research* (pp. 154–175). New York, NY: Routledge.

White, H. R., Bates, M. E., & Buyske, S. (2001). Adolescence-limited versus persistent delinquency: Extending Moffitt's hypothesis into adulthood. *Journal of abnormal psychology*, *110*(4), 600-609. doi: 10.1037/0021-843X.110.4.600