The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	Serial Sexual Assaults: A Longitudinal Examination of Offending Patterns Using DNA Evidence
Author(s):	Rebecca Campbell, Ph.D., Steven J. Pierce, Ph.D., Dhruv Sharma, Ph.D., Hannah Feeney, M.A., Rachael Goodman-Williams, M.A., Wenjuan Ma, Ph.D.
Document Number:	252707
Date Received:	March 2019
Award Number:	2014-NE-BX-0006

SERIAL SEXUAL ASSAULTS: A LONGITUDINAL EXAMINATION OF OFFENDING PATTERNS USING DNA EVIDENCE

FINAL SUMMARY REPORT

2014-NE-BX-0006

Principal Investigator:

Rebecca Campbell, Ph.D.
Professor of Psychology
Michigan State University
Email: rmc@msu.edu
Phone: 517-432-8390

Co-Investigators:

Steven J. Pierce, Ph.D.
Acting Director and Associate Director, Center for Statistical Training & Consulting (CSTAT)
Michigan State University, East Lansing, MI 48824
Phone: (517) 353-1058, Email: pierces1@msu.edu.

Dhruv Sharma, Ph.D.
Senior Statistician, Center for Statistical Training & Consulting (CSTAT)
Michigan State University, East Lansing, MI 48824
Phone: (517) 353-1058, Email: sharmadh@msu.edu

Project Director:

Hannah Feeney, M.A.
Department of Psychology
Michigan State University, East Lansing, MI 48824
Phone: (517) 432-7082, Email: feeneyha@msu.edu.

Project Staff:

Rachael Goodman-Williams, M.A.
Department of Psychology
Michigan State University, East Lansing, MI 48824
Phone: (517) 432-7082, Email: goodm169@msu.edu.

Wenjuan Ma, Ph.D.
Statistician, Center for Statistical Training & Consulting (CSTAT)
Michigan State University, East Lansing, MI 48824
Phone: (517) 353-1058, Email: mawenjua@msu.edu

Grant Period:

January 1, 2015 - December 31, 2017
No-Cost Extension Approved to December 31, 2018

Introduction

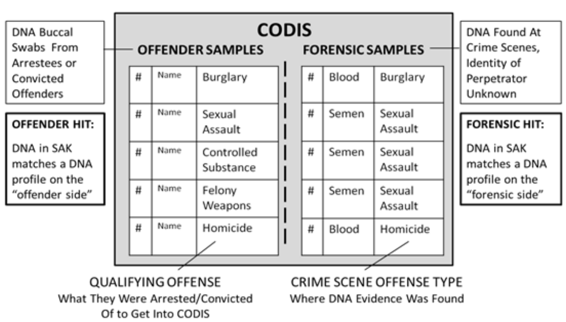
National epidemiological data indicate that 1-in-5 women will be sexually assaulted in her lifetime, most often by a male perpetrator (Black et al., 2011; Breiding et al., 2014). Given these data, does that mean that 1-in-5 men will sexually assault a woman in his lifetime? No, there are more victims than perpetrators because many offenders are serial sexual perpetrators who assault two or more individuals, in separate incidents over time (Edelstein, 2016). To date, there have been two primary ways social scientists have studied the problem of serial sexual offending. First, researchers have analyzed court records to determine how many individuals convicted of sexual offenses are re-arrested and/or re-convicted of subsequent sexual assaults. Meta-analyses suggest that 10-15% of adult sexual offenders recidivate after five years, with higher rates found in studies that track beyond five years

(Hanson & Morton-Bourgon, 2005; Lussier & Cale, 2013). Recidivism researchers note that these rates *substantially underestimate* serial sexual offending, given that most sexual assaults are not reported to the criminal justice system (Lonsway & Archambault, 2012), and of those that are, the vast majority are not adjudicated and therefore would not appear in arrest/conviction records (Campbell et al., 2014; Pattavina et al., 2016; Spohn & Tellis, 2012; Spohn et al., 2014). Thus, a second approach to studying serial sexual offending has been surveying individuals directly about acts they have committed. One-year re-offending incidence rates for male college students and military recruits range from 9-18% (Abbey et al., 2012; McWhorter et al., 2009), and longer-term re-offending rates are markedly higher, 63-78% (Abbey et al., 2001; Lisak & Miller, 2002; McWhorter et al., 2009; Zinzow & Thompson, 2015). However, self-report methods may also under-estimate the frequency of serial sexual offending, as individuals may be reluctant to disclose their behaviors, even in anonymous surveys. Recently, a third approach has emerged for studying serial sexual offending, one that complements and extends these other methods by relying on an entirely different data source: *biological evidence.*

In many sexual assaults, biological evidence is left behind by the perpetrator (semen, saliva), which can be collected by medical professionals in standardized sexual assault kits (SAKs). Hospital emergency departments began offering victims SAK collection in the late 1970s and this became routine in the 1990s (Martin, 2005; DOJ, 2013). Analysis of the evidence in the kit could yield a DNA profile of the suspect, and that profile could be uploaded to the federal criminal DNA database CODIS (Combined DNA Index System) and searched against other DNA profiles to identify offenders. Matching DNA profiles across cases can link multiple crimes to the same perpetrator (Butler, 2005). As such, examining DNA linkages through SAK evidence offers another approach for studying repeat perpetration, one that does not require that the perpetrator's arrest, conviction, or self-disclosure of the incident.

CODIS consists of two indexing systems (Figure 1), and the data in these systems can identify serial sexual assaults. First, the *offender index* contains known DNA profiles from arrestees/convicted offenders, obtained at their "qualifying offense" (i.e., a prior criminal offense that met federal requirements for CODIS entry). When a new DNA profile is entered into CODIS, there may be an "offender hit," such that the new profile matches the DNA of a known offender profile already in the system. If the DNA in a SAK matches an offender whose qualifying offense was a sexual assault, then the hit reveals a pattern of multiple sexual assaults. Second, the *forensic index* in CODIS contains unknown DNA profiles obtained at crime scenes; matches to these samples are termed "forensic hits." If the DNA in a SAK matches a forensic sample from a previous sexual assault, that hit establishes a pattern of multiple sexual assaults.

FIGURE 1: Overview of the Structure of CODIS



DNA testing of sexual assault kit biological evidence and analysis of CODIS hit patterns offers another methodological option for studying the problem of serial sexual assault. However, CODIS is massive—a national-scale repository—and it is not possible to mine it for associations and patterns;

there must be a starting query instigated by law enforcement personnel, as access to CODIS is legislatively limited (Butler, 2010). In this project, our starting query was a sample of approximately 7,000 previously-untested SAKs from Detroit, MI. First, we examined how many offenders may have committed more than one sexual assault, based on DNA matches in CODIS and other sexual assaults identified through a criminal history records check. Second, for the offenders who were suspected of committing multiple sexual assaults (based on CODIS and criminal history record data), we used growth mixture models to explore their patterns of serial sexual offending over time.

These models identified groups of serial sexual offenders (called latent classes) that differ with respect how often they engaged in sexual assault across nine age periods between ages 16 and 60. The resulting classes have sexual assault criminal career trajectories with different shapes because of differences in how many rapes were associated with offenders during their most prolific period of offending, when (i.e., in which age period) that occurred, plus the rates at which sexual offending rose to and declined from the levels estimated for that period.

Methods

Sample

The sample for this study was drawn from the population of 11,219 SAKs that were discovered in August, 2009 in a police property storage facility in Detroit, MI (Campbell et al., 2015). Some kits had been previously submitted to the police crime lab and were tested for DNA (*n* = 1,932), and some were tested later in grant projects (*n* = 400 by the Office of Violence Against Women, *n* = 1,600 by the National Institute of Justice). Our goal in this project was to examine the forensic testing outcomes of the *remaining* SAKs in that original inventory, *N* = 7,287 SAKs.

Procedures & Measures: Forensic Testing Outcome Data

All *N* = 7,287 SAKs were outsourced for testing to independent forensic laboratories. The forensic testing outcome data is at the *SAK-level of analysis* (i.e., how many SAKs progressed to DNA testing, yielded a CODIS-eligible profile, produced a CODIS hit, and matched to another sexual assault case). We also constructed variables at the *perpetrator-level of analysis* to capture how many unique perpetrators committed these acts (see Appendix A for an explanation of the transition from SAK-level data to perpetrator-level data). Table 1 (next page) lists all forensic testing outcome variables.

Procedures & Measures: Criminal History Record Data

We also obtained the criminal history records (CHR) for the unique & identifiable perpetrators who were revealed by the forensic testing outcome data. The Michigan State Police Forensic Science Division submitted a list of those individuals directly to the Michigan State Police Criminal History Record Division, which queried their adult criminal history records and assigned an anonymized research ID number. For each identified perpetrator, every criminal incident committed in Michigan, since age 16, is listed with corresponding arrest, charge, and judicial outcome data for that incident (see Table 2, next page, for list of variables obtained through the criminal history records and the variables we computed from those records).

TABLE 1: DNA Forensic Testing/CODIS Variables

	CODIS Variable	Description of CODIS Variable
SAK-Level	DNA Testing Rate	The probability that a kit will pass from Step 0, Screening (i.e., determining if male DNA was present in DNA sample) to Step 1, DNA Testing.
	CODIS Entry Rate	The probability that a kit will pass from Step 1, DNA Testing to Step 2, CODIS Entry (i.e., the process by which DNA profile was uploaded into the CODIS database).
	CODIS Hit Rate	The probability that a kit will pass from Step 2, CODIS Entry, to Step 3, CODIS Hit (i.e., the process by which the uploaded DNA profile matched an existing profile in the CODIS database).
	CODIS Serial Sexual Assault Hit Rate	The probability that a kit will pass from Step 3, CODIS Hit to Step 4, CODIS Serial Sexual Assault Hit (i.e., the process by which the CODIS hit was to another sexual assault case).
Perpetrator -Level	# of Unique Perpetrators	Number of perpetrators, accounting for DNA profiles yielding two or more CODIS Hits.
	# of Unique Serial Sexual Assault Perpetrators	Number of unique perpetrators when reviewing only those with CODIS Hits that were associated with other sexual assault cases.
	# of Unique & Identifiable Perpetrators	Number of unique perpetrators with identifying information in CODIS (i.e., identity is known). This is lower than the number of unique perpetrators as some hits are to the forensic index.
	# of Unique & Identifiable Serial Sexual Assault Perpetrators	Number of unique & identifiable perpetrators when reviewing only those with CODIS Hits that were associated with other sexual assault cases.

TABLE 2: Criminal History Records Variables		
	Criminal History Variable	Description of Criminal History Variable Creation
Criminal History Records (CHR)	Age at first arrest	Compared each offender's date of birth to the earliest arrest date with arrest offense record.
	Arrest diversity	Coded arrest offenses into 13 crime categories then computed the standardized diversity index for each offender from the offender-specific frequency distribution among those categories. Measure ranged from 0-1, with 0 indicating a single crime category (no diversity), and 1 indicating equal dispersion across all possible categories (maximum diversity).
	Arrest span	Number of years elapsed between the first and last arrest dates.
	Confinement	Sum of confinement sentences (in years) in each offender's adjudicated charge records, with vacated sentences=0 years and lifetime sentences=100 years. Variable was rescaled to decades in our statistical models.
	Escalation	Ranked 13 crime categories based on severity and measured escalation scale as an ordinal variable that sums the cumulative change in severity of arrest offenses across incident dates.
	# of incidents without sexual assault evidence	Count variable of incidents in each offender's criminal history that were not associated with any sexual assault evidence (arrest, prosecutor charges, or judicial charges for sexual assault).
CHR & CODIS	# of sexual assaults	Count variable of the total number of sexual assaults (from CODIS hits data & criminal history data) for each perpetrator during each of nine 5-year age periods from ages 16-60 .
	Total # of sexual assaults	Sum of the total number of sexual assaults (from CODIS hits data and criminal history data) associated with each perpetrator across each of nine 5-year age periods from ages 16-60 .
	Proportion of incidents with sexual assault	Proportion of incidents in each offender's criminal history that were associated with any evidence of sexual assault (from arrest offenses, prosecutor charges, or judicial charges).
GMM Variables	# of observed age period	Divided criminal histories into nine 5-year age periods (16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, and 56-60 years), with the last observed age period being that with the last known event date associated with that individual.
	# of observed age periods with sexual assaults	Count variable of the number of observed age periods where the offender had at least one incident with evidence of sexual assault.
	Proportion of observed age periods with sexual assault	Proportion of observed age periods where the offender had at least one incident with evidence of sexual assault.
	Number of peaks	A simple measure of intermittency in sexual assault offending where a consecutive set of age periods that all have at least one sexual assault is considered a single peak.
	Peak duration	Divided the number of peaks by the number of observed age periods with sexual assaults.
	# of sexual assaults per peak	Divided the total number of sexual assaults associated with a perpetrator by the number of peaks in the perpetrator's response pattern.

Descriptive Results

We counted how many SAKs reached each stage of forensic testing, computed unconditional and conditional progression rates, and obtained plausible ranges for those rates using continuation ratio modeling (Agresti, 2002; Hosmer et al., 2013) in R software, version 3.5.0 (R Development Core Team, 2017). All *N* = 7,287 SAKs were submitted for testing and screened, and *n* = 5,048 SAKs progressed to DNA testing (69.3% unconditional DNA Testing Rate, 95% CI .68-.70). A total of *n* = 2,938 SAKs had a DNA profile that met eligibility for upload into CODIS (40.3% unconditional and 58.2% conditional CODIS Entry Rate, 95% CI .57-.60), *n* = 1,675 SAKs yielded a CODIS Hit (23.0% unconditional and 57.0%

conditional CODIS Hit Rate, 95% CI .55-.59), and *n* = 775 SAKs produced a CODIS Serial Sexual Assault Hit (10.6% unconditional and 46.3% conditional CODIS Serial Sexual Assault Hit Rate, 95% CI .44-.49).

From the *n* = 1,675 SAKs with CODIS Hits, there were *n* = 1,424 unique perpetrators and *n* = 1,270 unique & identifiable perpetrators associated with those hits. From the *n* = 755 SAKs with CODIS Serial Sexual Assault Hits, there were *n* = 508 unique serial sexual assault perpetrators associated with those hits. Most of these perpetrators (*n* = 358 of 508) had two sexual assaults documented via DNA/CODIS data, but a sizable percentage had committed more than two: *M* = 2.64 sexual assaults via DNA/CODIS (*SD* = 1.37), range 2 to 12 sexual assaults. From the sample of 508 unique serial sexual assault perpetrators, there were *n* = 365 unique & identifiable serial sexual assault perpetrators: M = 2.67 sexual assaults via DNA/CODIS (*SD* = 1.47), range 2 to 12 sexual assaults.

Based only on the information available in CODIS, *n* = 365 of the *n* = 1270 unique & identifiable perpetrators were identified as unique & identifiable serial perpetrators. When the criminal history records of all unique & identifiable perpetrators were examined in addition to CODIS records, however, additional sexual assault incidents were identified, and the number of unique & identifiable serial sexual assault perpetrators increased from *n* = 365 to *n* = 504 These *n*= 504 perpetrators committed, on average, *M* = 3.27 sexual assaults (*SD* = 2.02), range 2 to 15 sexual assaults. Thus, based only on CODIS data, 28.7% of the subsample of unique & identifiable perpetrators were suspected serial sexual assault offenders, and when we included criminal history data, 39.7% were suspected serial sexual offenders.

Growth Mixture Model Results

We used zero-inflated Poisson GMMs (Kreuter & Muthén, 2008) to identify subgroups of serial sexual assault perpetrators (n = 392 for whom we had CHR data) with distinct longitudinal trajectories for the frequency of sexual assaults across nine 5-year age periods spanning the ages of 16 to 60 years. There were two submodels for each latent class: a Poisson count growth curve submodel that predicts the number of sexual assaults associated with the perpetrator at each time period and a logistic zero-inflation growth curve submodel that predicts whether the perpetrator can *only have a count of zero* at a given time point. The combined submodels gracefully handle count outcome data with many zeroes.

We estimated all GMMs via robust maximum likelihood (MLR) in Mplus 8 (Muthén & Muthén, 2017; Muthén et al., 2017) with numerical integration. We varied random number seeds and starting values for each model to ensure sufficient replication of the best log likelihood. We used the Bayesian Information Criterion (BIC), the Lo-Mendell-Rubin adjusted likelihood ratio test (LMRT; Lo et al., 2001 2001), and the bootstrapped likelihood ratio tests (BLRT; Nylund et al., 2007) to determine how many latent classes to extract. Both the LMRT and the BLRT test the null hypothesis that the data are drawn from a mixture of *k*-1 instead of *k* latent classes. We started modeling with *k* = 1 classes and incremented *k* until the LMRT result failed to reject *k*-1 classes.

The full set of parameters for the final 4-class model are reported in Appendix B. The tests of model fit yielded a Pearson $\chi^2(19655) = 1548.250$, p = 1.000, and a likelihood ratio $\chi^2(19655) = 616.386$, p = 1.000. The final model had low entropy (0.512) with classification accuracy of 87.8%, 61.4%, 48.2%, and 44.4% respectively for classes 3, 2, 1, and 4; the probability of misclassification from the other classes into class 3 ranged from 34.3 to 41.6%. The low entropy for the 4-class model indicates that it is less clear which perpetrators best fit in each class than we would like, but the BLRT nevertheless indicated that allowing multiple classes improved model fit. Descriptive characteristics of the final 4-class solution are presented in Table 3. Appendix C shows individual perpetrators' observed longitudinal trajectories for the number of sexual assaults by age period, divided into panels based on the perpetrator's most likely latent class and the number of peaks in the observed trajectory.

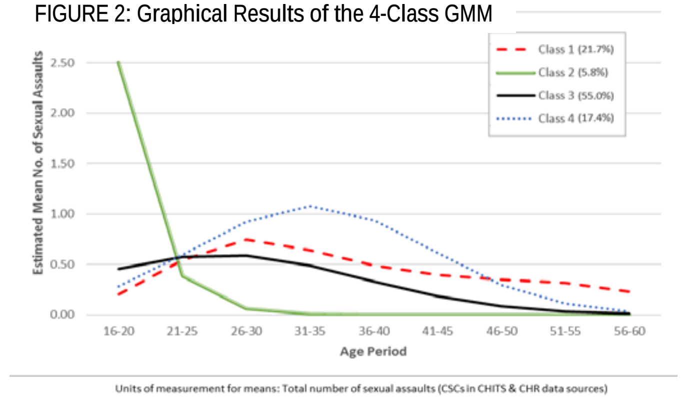
TABLE 3: Descriptive Characteristics of Final 4-Class Solution

Class	Size of Class *(%)*	Total Mean # of Sexual Assaults (*SD*)	Age Period of Class Peak (*M* # of Sexual Assaults during the peak)	Summary of Class
Class 1: Moderate-volume lifelong serial offenders	85.1 (*21.7*)	3.40 (*1.93*)	26-30 (0.75)	This class has a moderately high mean number of sexual assaults and a perpetration rate that remains stable across the lifespan.
Class 2: Apprehended early serial offenders	22.9 (*5.8*)	2.90 (*1.26*)	16-20 (2.50)	The smallest class, these prolific offenders perpetrate frequently early in their criminal career, but then decline abruptly.
Class 3: Low-volume lifelong serial offenders	215.7 (*55.1*)	2.33 (*0.76*)	26-30 (0.59)	The largest class, these offenders have a lower rate of perpetration when compared to other classes, which sustains across the lifespan.
Class 4: High-volume midlife serial offenders	68.3 (*17.4*)	5.26 (*2.82*)	26-30 (0.92) 31-35 (1.08) 36-40 (0.94)	This class has the highest mean number of sexual assaults over the longest peak. This peak sustains over the entirety of the midlife, but then declines quickly.

Moving forward, we interpret the classes in order of relative size to emphasize the most commonly occurring trajectories. The largest class, class 3, comprises more than half (55.1%) of the serial sexual offenders. Figure 1 shows that the mean number of sexual assaults for these *low-volume lifelong serial offenders* starts at 0.45 during the 16-20 age period, peaks at 0.59 during the 26-30 age period, then declines slowly to 0.01 by the 56-60 age period. These serial offenders have the lowest peak of the four classes and exhibit the most stable pattern of offending, with neither sharp increases nor decreases in offending from adolescence through the mid-30's, at which point they begin a decline that lasts through older age. The count submodel's linear slope indicates slow decline over time (α_S = -3.929, t = -7.393, $p < .001$) paired with deceleration (α_Q = -6.584, t = -3.399, $p < .001$) in the rate of change. Essentially, these offenders are suspected to have committed fewer sexual assaults when compared to the other latent classes, but their lower rates of offending are sustained across the lifespan. This class has no zero-inflation due to the constraints ($\alpha_{SI} = 0$, $\alpha_{QI} = 0$, and $\nu = -15$) added during model fitting.

Class 1, *moderate-volume lifelong serial offenders* starts at its minimum mean number of sexual assaults at 0.20 during the 16-20 age period and peaks at 0.75 during the 26-30 age period before declining to 0.23 by the 56-60 age period. This group's pattern is different from the previous class by less offending in adolescence but more in young adulthood and through the remainder of the lifespan.

FIGURE 2: Graphical Results of the 4-Class GMM



The count submodel's linear slope indicates slow decline over time ($\alpha_S = -2.442$, *SE* = 0.501, *t* = -4.873, *p* < .001) but its quadratic slope ($\alpha_Q = 4.767$, *SE* = 2.374, *t* = 2.008, *p* < .001) indicates that acceleration offsets the linear decline, effectively flattening out the trajectory. Meanwhile, the zero-inflation submodel threshold was small ($\nu = -1.784$, *SE* = 1.730, *t* = -1.031, *p* = .302) with a slowly decreasing propensity to have a count of zero sexual assaults ($\alpha_{SI} = -3.061$, *SE* = 1.392, *t* = -2.199, *p* < .028). The quadratic slope is not significant despite its large, positive value ($\alpha_{QI} = 14.695$, *SE* = 10.162, *t* = 1.446, *p* < .148), so it provides little evidence for a non-linear trend in that propensity. In sum, this class may be characterized by suspected serial sexual offending that remains moderately high and stable across the lifespan.

Regarding Class 4, the mean number of sexual assaults is 0.28 from 16-20 years old. It then rapidly rises to hover near 1 during the three consecutive age periods spanning the ages 26-40: those means are respectively 0.92, 1.08, and 0.94 sexual assaults. The mean then drops off more rapidly than in most of the other classes until it reaches 0.03 by the 56-60 age period. We have therefore labeled this category the *high-volume midlife serial offenders* to reflect both their high, sustained peak in middle adulthood and their lower rates of offending in adolescence and older adulthood. The count submodel's linear slope indicates slow decline over time ($\alpha_S = -2.261$, *SE* = 0.602, *t* = -3.755, *p* < .001) but its quadratic slope ($\alpha_Q = -9.320$, *SE* = 1.915, *t* = -4.868, *p* < .001) indicates more rapid deceleration than seen in the other classes. This class has some zero-inflation (ν = 0.561, *SE* = 0.121, *t* = 4.628, *p* < .001), but it is constant over time (α_{SI} and α_{QI} were fixed at 0) so there is no longitudinal trend in the propensity to have counts of zero. Overall, this class has the highest mean number of sexual assaults that sustains over the entirety of the midlife, but then declines quickly after age 40.

Class 2 has the most unexpected trajectory in the model. The smallest class (5.8%), these *apprehended early serial offenders* start out with a mean of 2.50 sexual assaults during the 16-20 age period, then their mean number of sexual assaults drops precipitously to 0.38 during the 21-26 age period and still further to 0.06 in the 26-30 age period. After that, their mean remains close to 0.00 sexual assaults for all remaining age periods. Though we lacked complete data regarding incarceration, we examined the criminal sentencing data we had and determined that many of these offenders' declines in offending could be reasonably attributed to incarceration. The only estimated count submodel parameter for this class was the intercept (α_I= -6.583, *SE* = 0.148, *t* = -44.462, *p* < .001) due to constraints imposed on the class during model fitting (α_S= 0, α_Q = 0). This class has no zero-inflation (fixed parameters: α_{QI} = 0, α_{QI} = 0, and ν = -15).

To validate the GMM results, we identified six antecedent variables that reflected characteristics of the offenders' criminal history to predict latent class membership. These included the number of incidents without sexual assault evidence, the offender's age at their first arrest, the number of years between their first and latest arrest, their diversity in offending, their escalation in offending, and their history of confinement (Appendix D). We also validated the model with one distal outcome, the proportion of each perpetrator's criminal incidents associated with sexual assault evidence. The antecedents model produced four sets of model parameters by varying which latent class was considered the reference group. For brevity, we report only the analyses using Class 3 (the largest class: low-volume lifelong serial offenders) as the reference. Relative to class 3, one antecedent was a significant predictor of being in class 1: each additional year of age at first arrest was associated with 9%

increase in the odds of being in class 1 (*OR* = 1.095, 95% CI = [1.033, 1.162]). Similarly, only one of the antecedents was a significant predictor of membership in class 2 instead of class 3. Each additional year of age at first arrest was associated with a 35% decrease in the odds of being in class 2 (*OR* = 0.653, 95% CI = [0.450, 0.948]). One antecedent significantly predicted membership in class 4 instead of class 3: each additional decade of sentenced confinement was associated with a 6% increase in the odds of being in class 4 (*OR* = 1.059, 95% CI = [1.010, 1.110]). In other words, offenders who were older at first arrest were more likely to be in class 1 and less likely to be in class 2 compared to class 3, and offenders who had a longer period of total confinement were more likely to be in class 4 compared to class 3. Regarding the distal outcome of the proportion of each perpetrator's criminal incidents associated with sexual assault evidence, we found that the high-volume, midlife serial offenders in class 4 (M = 0.311) were more specialized in sexual assault offending than the low-volume, lifelong offenders in class 3 (M = 0.224).

We acknowledge the following limitations and caveats regarding the GMM results. The diverse observed trajectories made identifying a small number of meaningful latent classes harder, as did censoring at later age periods. Analyzing a broader sample (all sexual offenders) may improve entropy because non-serial sexual offenders' observed trajectories would be simpler than those of serial sexual offenders. Subgroups of serial offenders may be more difficult to differentiate from one another than from non-serial offenders. Observed trajectories with multiple peaks (i.e., intermittency) are challenging to model via quadratic GMMs, which assume a single peak in each latent trajectory. Models that do not assume a simple quadratic growth curve shape should be explored in the future.

Discussion of Findings & Implications

In this study, we used DNA evidence as a data source for studying serial sexual assault perpetration, which is a novel methodological approach in this literature. It is not as stringent a threshold as repeat sexual assault criminal convictions, which is used in recidivism studies, as DNA evidence is evidence of a *reported* crime, rather than a fully adjudicated crime. However, the historic under-prosecution of sexual assault means that recidivism studies likely underestimate the scope of serial sexual perpetration. Indeed, the rates of repeat sexual offending documented in this study (39.7% of unique and identifiable perpetrators) are higher than recidivism studies (typically 10-15%). However, these data sources may also underestimate actual offending because many sexual assault victims do not have a SAK collected and/or report their assaults to the criminal justice system. The data sources used in this study (CODIS hits and criminal history data) are only able to account for sexual assaults in which victims sought post-assault health care and reported to the police. With these caveats in mind, we contend that using CODIS data and criminal history records are still a useful and credible data source for studying sexual assault perpetration.

Our identification of four latent trajectories of sexual assault perpetration supports literature suggesting that heterogeneous patterns underlie trajectories of sexual offending (Lussier et al., 2010; Swartout et al., 2015). Similar to Lussier et al. (2010), we found that offending frequency changes throughout the adult lifespan. This finding highlights the importance of extended longitudinal data when drawing conclusions related to offending trajectories, as shorter-term data could mistake a break in offending for desistance and misrepresent patterns of perpetration.

Our findings indicate that SAKs are uniquely capable of identifying serial sexual offending, particularly when criminal history records of suspects are included in the analysis. The sizeable number of serial offenders identified in our study should encourage law enforcement officers to approach investigations as suspected serial, rather than isolated, incidents.

References

Abbey, A., McAuslan, P., Zawacki, T., Clinton, M., & Buck, P. (2001). Attitudinal, experimental, and situational predictors of sexual assault perpetration. *Journal of Interpersonal Violence, 16,* 784-807.

Abbey, A., Wegner, R., Pierce, J., & Jacques-Tiura, A. J. (2012). Patterns of sexual aggression in a community sample of young men: Risk factors associated with persistence, desistance, and initiation over a 1-year interval. *Psychology of Violence*, *2*, 1-15.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Black, M. C., Basile, K. C., Breiding, M. J., Smith, S. G., Walters, M. L., Merrick, M. T., Chen, J., & Stevens, M. R. (2011). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Breiding, M. J., Smith, S. G., Basile, K. C., Walters, M. L., Chen, J., & Merrick, M. T. (2014). *Prevalence and characteristics of sexual violence, stalking, and intimate partner violence victimization--national intimate partner and sexual violence survey, united states, 2011* (1546-0738). Retrieved from https://www.cdc.gov/violenceprevention/pdf/NISVS-StateReportBook.pdf

Butler, J.M. (2005). *Forensic DNA typing: Biology, technology, and genetics of STR markers,* 2nd Edition. Boston, MA: Elsevier Academic Press.

Butler, J. M. (2010). *Fundamentals of forensic DNA typing*. Boston, MA: Academic Press/Elsevier.

Campbell, R., Bybee, D., Shaw, J.L., Townsend, S.M., & Karim, N. (2014). The impact of sexual assault nurse examiner (SANE) programs on criminal justice case outcomes: A multi-site replication study. *Violence Against Women, 20,* 607-625.

Campbell, R., Fehler-Cabral, G., Pierce, S. J., Sharma, D., Bybee, D., Shaw, J., Horsford, S., & Feeney, H. (2015). *The Detroit sexual assault kit (SAK) action research project (ARP)* (NIJ 2011-DN-BX-0001). Washington, DC: National Institute of Justice.

Department of Justice (DOJ). 2013. *A national protocol for sexual assault medical forensic examinations: Adults & adolescents,* 2nd Edition. Washington, DC: Author.

Edelstein, A. (2016). Rethinking conceptual definitions of the criminal career and serial criminality. *Trauma, Violence, & Abuse, 17*, 62-71.

Hanson, R.K., & Morton-Bourgon, K.E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-1163.

Hosmer, D.W., Lemeshow, S. & Sturdivant, R.X. (2013). *Applied logistic regression* (3rd ed.). Hoboken: Wiley.

Kreuter, F., & Muthén, B. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology, 24*, 1-31.

Lisak, D., & Miller, P.M. (2002). Repeat rape and multiple offending among undetected rapists. *Violence & Victims, 17*, 73-84.

Lonsway, K.A., & Archambault, J. (2012). The "justice gap" for sexual assault cases: Future directions of research and reform. *Violence Against Women, 18*, 145-169.

Lussier, P., & Cale, J. (2013). Beyond sexual recidivism: A review of the sexual criminal career parameters of adult sex offenders. *Aggression and Violent Behavior, 18*, 445-457.

Martin, P.Y. (2005). *Rape work: Victims, gender, and emotions in organization and community context*. New York, Routledge.

McWhorter, S. K., Stander, V. A., Merrill, L. L., Thomsen, C. J., & Milner, J. S. (2009). Reports of rape reperpetration by newly enlisted Navy personnel. *Violence & Victims, 24*, 204-218.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, B. O., Muthén, L. K., Asparouhov, T., & Nguyen, T. (2017). Mplus (Version 8.0) [Computer program]. Los Angeles, CA: Muthén & Muthén. www.statmodel.com.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14, 535-569.

Pattavina, A., Morabito, M., & Williams, L.M. (2016). Examining connections between the police and prosecution in sexual assault case processing: Does the use of exceptional clearance facilitate a downstream orientation? *Victims & Offenders, 11*, 315-319.

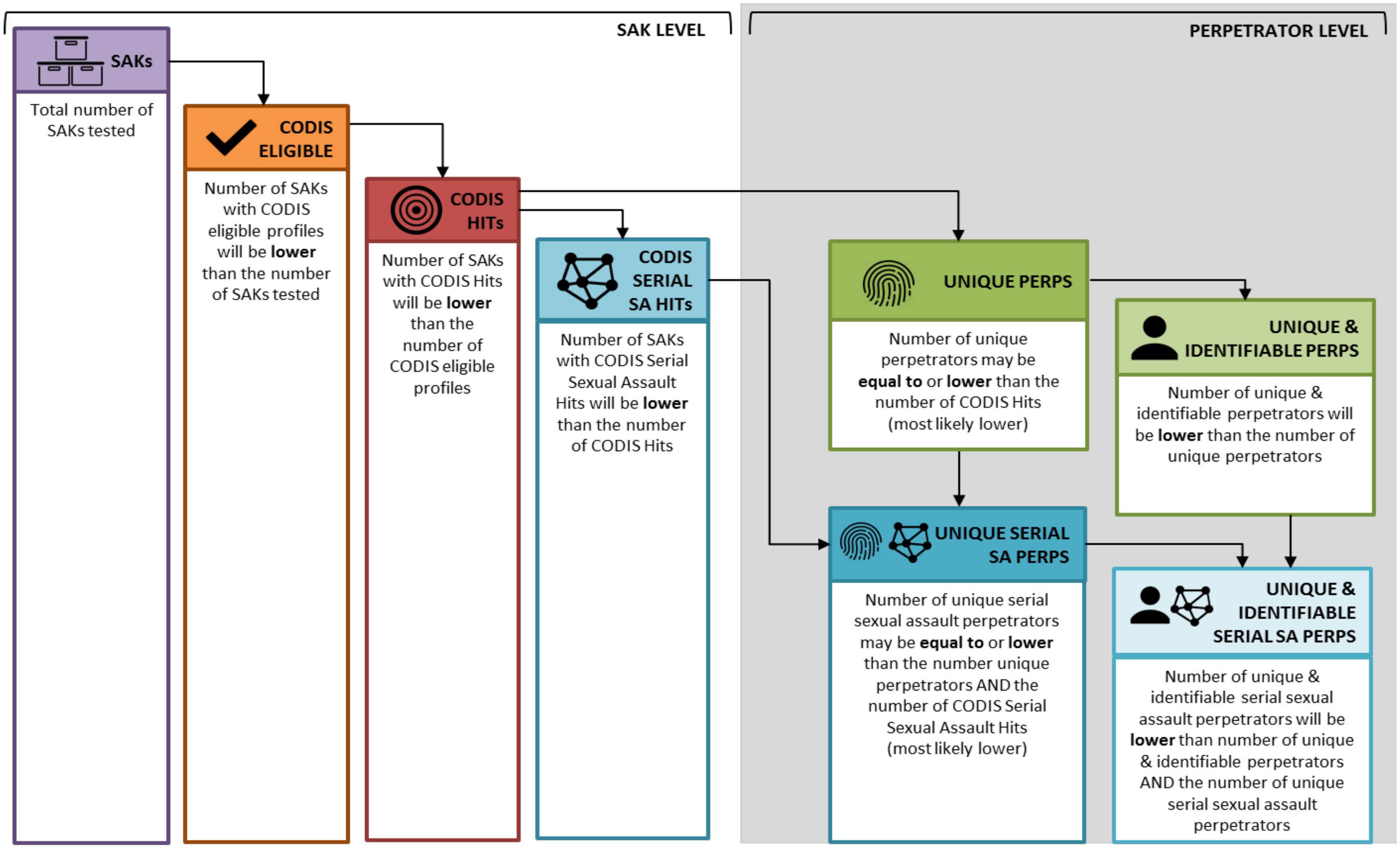
R Development Core Team. (2017). R: A language and environment for statistical computing (Version 3.5.0) [Computer Program]. Vienna, Austria: R foundation for Statistical Computing. Retrieved from http://www.R-project.org

Spohn, C., & Tellis, K. (2012). The criminal justice system's response to sexual violence. *Violence Against Women, 18*, 169-192.

Spohn, C., White, C., & Tellis, K. (2014). Unfounding sexual assault: Examining the decision to unfound and identifying false reports. *Law &Society Review, 48*, 161-191.

Zinzow, H.M., & Thompson, M. (2015). A longitudinal study of risk factors for repeated sexual coercion and assault in US college men. *Archives of Sexual Behavior, 44,* 213-222.

APPENDIX A — Transitioning Forensic Outcome Data from the SAK-Level to Perpetrator-Level of Analysis



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

APPENDIX B — Final 4-Class GMM Solution

Zero-Inflated Poisson GMM Fit Statistics.

No. Classes (*k*)	*FP*	*LL*	*BIC*	Δ*BIC*	Entropy	*LMRT* *p*-value	*BLRT* *p*-value	Conclusion
1	6	-2160	4356		1.000			N/A
2	8	-2112	4272	-84	0.457	< .0001	< .0001	Reject 1 class
3	10	-2105	4269	-4	0.570	.0074	< .0001	Reject 2 classes
4	17	-2093	4288	19	0.512	.0624	< .0001	Reject 3 classes

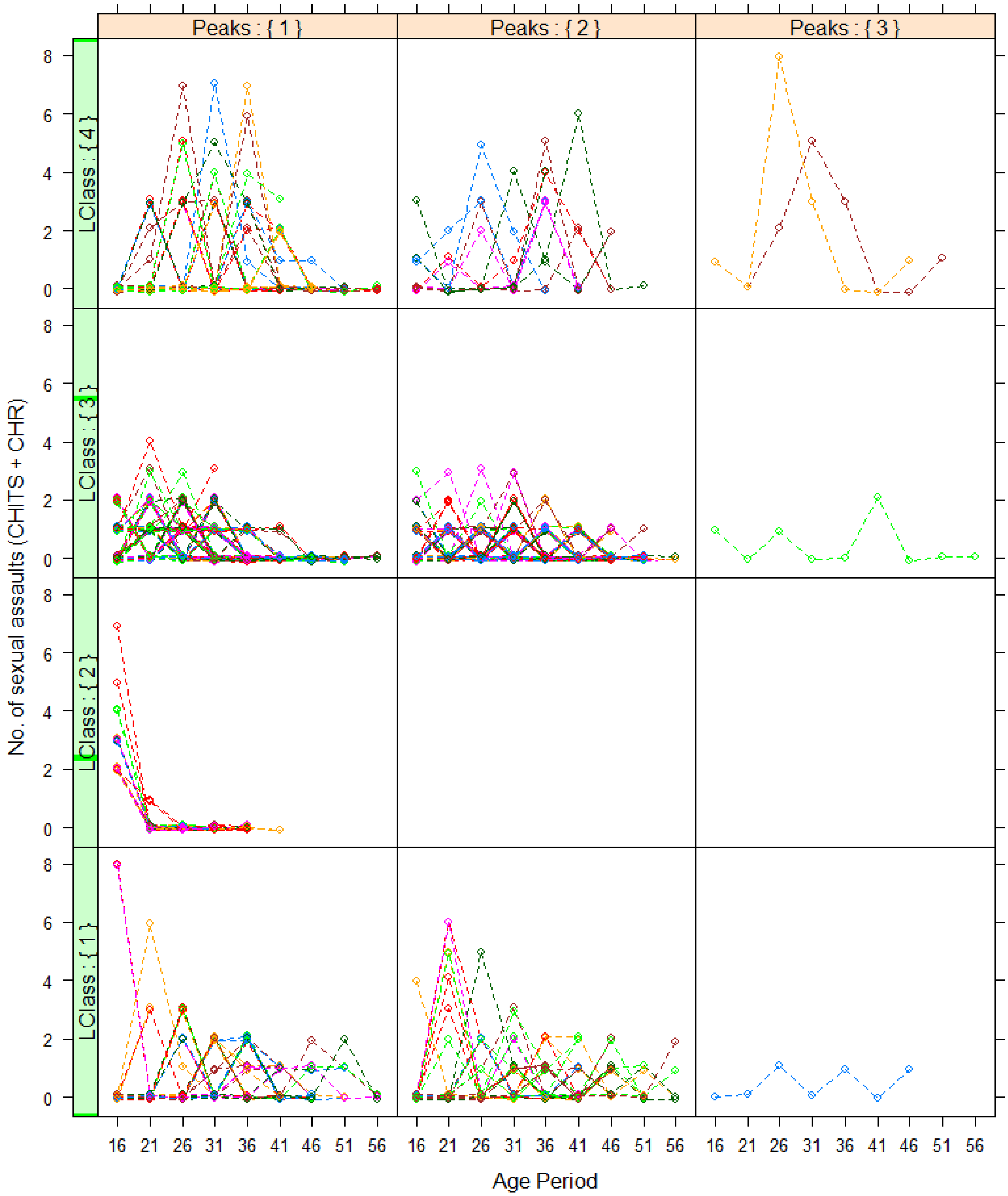
Note: Models with class-specific means and class-invariant covariances fixed at 0 were fit via robust maximum likelihood (MLR) in Mplus 8 with numerical integration (15 points). *BIC*, Bayesian information criterion; *BLRT*, bootstrapped likelihood ratio test; *FP*, number of free parameters estimated; LL, loglikelihood; *LMRT*, Lo-Mendell-Rubin adjusted likelihood ratio test.

Final 4-Class GMM Parameter Estimates.

Parameter	Label	Estimate	*SE*	*t*	*p*-value	
Categorical Latent Variable						
Mean 1	C#1	0.218	0.443	0.492	0.623	
Mean 2	C#2	-1.093	0.458	-2.387	0.017	*
Mean 2	C#3	1.150	0.393	2.930	0.003	**
Latent Class 1						
Means						
Count intercept	α_I	-0.557	0.442	-1.259	0.208	
Count linear slope	α_S	-2.442	0.501	-4.873	0.000	***
Count quadratic slope	α_Q	4.767	2.374	2.008	0.045	*
Inflation linear slope	α_{SI}	-3.061	1.392	-2.199	0.028	*
Inflation quadratic slope	α_{QI}	14.695	10.162	1.446	0.148	
Inflation Thresholds	ν	-1.784	1.730	-1.031	0.302	
Latent Class 2						
Means						
Count intercept	α_I	-6.583	0.148	-44.462	0.000	***
Count linear slope [a]	α_S	-15.000				
Count quadratic slope [a]	α_Q	0.000				
Inflation linear slope [a]	α_{SI}	0.000				
Inflation quadratic slope [a]	α_{QI}	0.000				
Inflation Thresholds [a]	ν	-15.000				
Latent Class 3						
Means						
Count intercept	α_I	-1.107	0.150	-7.393	0.000	***
Count linear slope	α_S	-3.929	0.907	-4.332	0.000	***
Count quadratic slope	α_Q	-6.584	1.937	-3.399	0.001	***
Inflation linear slope [a]	α_{SI}	0.000				
Inflation quadratic slope [a]	α_{QI}	0.000				
Inflation Thresholds [a]	ν	-15.000				
Latent Class 4						
Means						
Count intercept	α_I	0.951	0.127	7.466	0.000	***
Count linear slope	α_S	-2.261	0.602	-3.755	0.000	***
Count quadratic slope	α_Q	-9.320	1.915	-4.868	0.000	***
Inflation linear slope [a]	α_{SI}	0.000				
Inflation quadratic slope [a]	α_{QI}	0.000				
Inflation Thresholds	ν	0.561	0.121	4.628	0.000	***

Note: Linear factor loadings were class-invariant and fixed (ranging from -0.500 to 0.500 in increments of 0.125), quadratic factor loadings were fixed to the squares of the linear loadings. Latent variances were fixed to 0 for identification, as were inflation latent intercepts (α_{II}). Inflation thresholds (ν) were class-specific but constrained equal across age periods. [a] Parameter was fixed to achieve identification. * $p < .05$, ** $p < .01$, *** $p < .001$.

APPENDIX C — Observed Trajectories by Most Likely Latent Class and Number of Peaks



Note: Line color varies across offenders. Due to limited numbers of colors, there are panels where multiple offenders have the same color. Number of sexual assaults is the total number of unique sexual assault incidents in a given age period identified by counting both SAKs submitted for forensic testing and criminal history incidents with any evidence of sexual assault (arrests, prosecutor charges, or adjudicated charges). Incidents represented in both sources were counted only once. CHITS, CODIS hits data set; CHR, criminal history records; LClass, most likely latent class.

APPENDIX D — GMM With Antecedents Predicting Latent Class Membership, Relative to Class 3

Parameter	Estimate	SE	OR [95% CI]	t	p-value	
Predictors of Class 1 (N = 85)						
No. incidents w/o sexual assault	0.019	0.043	1.019[0.937, 1.109]	0.439	0.661	
Age at first arrest (years)	0.091	0.030	1.095[1.033, 1.162]	3.010	0.003	**
Arrest span (years)	0.045	0.023	1.046[1.000, 1.094]	1.930	0.054	
Arrest diversity	-0.639	1.002	0.528[0.074, 3.762]	-0.638	0.524	
Escalation scale	0.058	0.035	1.060[0.989, 1.135]	1.687	0.092	
Confinement (decades)	0.047	0.029	1.048[0.990, 1.109]	1.620	0.105	
Predictors of Class 2 (N = 23)						
No. incidents w/o sexual assault	-0.141	0.118	0.868[0.689, 1.094]	-1.195	0.232	
Age at first arrest (years)	-0.426	0.190	0.653[0.450, 0.948]	-2.244	0.025	*
Arrest span (years)	-0.051	0.043	0.950[0.873, 1.034]	-1.166	0.244	
Arrest diversity	-1.270	1.557	0.281[0.013, 5.940]	-0.816	0.414	
Escalation scale	-0.105	0.067	0.900[0.790, 1.027]	-1.572	0.116	
Confinement (decades)	0.036	0.042	1.037[0.955, 1.126]	0.843	0.399	
Predictors of Class 4 (N = 68)						
No. incidents w/o sexual assault	-0.029	0.050	0.971[0.881, 1.071]	-0.575	0.565	
Age at first arrest (years)	0.045	0.037	1.046[0.973, 1.125]	1.242	0.214	
Arrest span (years)	0.043	0.025	1.044[0.994, 1.096]	1.696	0.090	
Arrest diversity	-0.617	0.993	0.540[0.077, 3.778]	-0.621	0.534	
Escalation scale	0.069	0.039	1.071[0.993, 1.157]	1.769	0.077	
Confinement (decades)	0.057	0.024	1.0590[1.010, 1.110]	2.361	0.018	*
Intercepts						
Class 1	-3.688	1.175		-3.140	0.002	**
Class 2	7.998	3.678		2.174	0.030	*
Class 4	-2.571	1.256		-2.048	0.041	*

Note: Class 3 (*N* = 216) is the reference group. * $p < .05$, ** $p < .01$, *** $p < .001$.