TPM®

PREDICTING TRAJECTORIES OF SUBSTANCE USE DURING EMERGING ADULTHOOD: EXPLORING THE BENEFITS OF GROUP-BASED TRAJECTORY MODELING FOR ZERO-INFLATED OUTCOMES

STEPHANIE HAZEL COOK
ERICA PAIGE WOOD
MARC SCOTT
KRISTYN AMBER PIERCE
FARZANA KAPADIA
NEW YORK UNIVERSITY

PERRY N. HALKITIS
RUTGERS UNIVERSITY

The application of group-based trajectory analysis with a zero-inflation specification to understand adaptations in alcohol and marijuana use among sexual minorities during the emerging adulthood period is underutilized. We propose a series of four steps in the application of group-based trajectory analysis for outcomes that follow a zero-inflated distribution. In addition, using a longitudinal cohort study of emerging adult sexual minority men (*n* = 597) we provide an example of how to obtain group-based trajectory profiles of alcohol and marijuana use among emerging adult sexual minority men with a zero-inflated specification. In addition, we examined racial/ethnic differences in trajectory profiles. Findings suggest that there were five distinct alcohol and marijuana use trajectories that were reliably predicted by racial/ethnic identity for alcohol but not marijuana utilizing the zero-inflated specification. A summary of findings and concluding remarks related to the utility of this modeling technique are presented.

Keywords: Group-based trajectory analysis; Zero-inflated modeling; Sexual minority youth; Racial/ethnic disparities; Substance use.

Correspondence concerning this article should be addressed to Stephanie Cook, School of Global Public Health, New York University, 715 Broadway, 10012 New York, United States of America. Email: sc5810@nyu.edu

Emerging adulthood, or the period encompassing ages 18 to 29, is a key developmental phase in which individuals experience rapid changes including entering new social environments (e.g., college) and generally experiencing a greater sense of independence (Arnett, 2004). While the emerging adulthood period is critical for accomplishing adulthood milestones such as financial independence, the socioenvironmental changes experienced during this period place emerging adults at particular risk for poor health behaviors including substance use and misuse (Stone, Becker, Huber, & Catalano, 2012). Indeed, much research documents that rates of substance use tends to increase and reach its peak during the emerging adulthood period, and then steadily declines as one gets closer to adulthood, that is, after age 29 (Stone et al., 2012). However, there are vast disparities in risky substance use behaviors across emerging adult populations (Stone et al., 2012). In particular, rates of problematic substance use continue to rise among young sexual minority men (YSMM) during the emerging adulthood period (Halkitis et al., 2014; Hatzenbuehler, McLaughlin, & Xuan,

TPM Vol. 27, No. 3, September 2020 – 291-311 – Special Issue – doi:10.4473/TPM27.3.1– © 2020 Cises

2015; Marshal et al., 2008; Marshal, Friedman, Stall, & Thompson, 2009; Newcomb, Ryan, Greene, Garofalo, & Mustanski, 2014). Nevertheless, the complex patterns of substance use within the emerging adulthood period among YSMM are currently underexplored.

To date, there have been a plethora of tutorials and demonstration articles concerning the different methods for understanding developmental dynamics in substance use behavior among adolescents and emerging adults (Greenwood et al., 2019). Substance use behavior sometimes follows a non-normal distribution (typically counts and often with an excess number of zero values in the number of substance use instances). Models and studies of patterns of substance use behavior and its correlates are more defensible when they more-closely capture this non-normality (Cook & Wood, 2018). Recent studies have handled the non-normal distribution by using zero-inflated Poisson (ZIP) growth models (Preacher, Wichman, MacCallum, & Briggs, 2008), which can account for an excess number of zero values commonly encountered. A popular methodology for data collected over time is latent growth curve (LCG) models. Linear versions of these models are used to incorporate individual-specific latent variables in terms of the intercept (initial values of the outcome variable) and the slope (the rate of change in the outcome variable over time; Muthén & Muthén, 2000). Thus, this estimation procedure is person-centered and can provide level differences, overall rate of change and the variation around level, and rate of change differences among subjects. The Poisson model can be incorporated into the LGC framework as a generalized linear mixed model (Box & Cox, 1964). Use of the ZIP model is more complicated, because it is a mixture model that utilizes a Poisson distribution while allowing for excess zeros (Cook & Wood, 2018; Liu, 2007). Although there has been considerable progress made in applying the usefulness of ZIP growth models across multiple disciplines, it remains relatively limited.

Group-based trajectory modeling (GBTM) was originally proposed by Nagin and Land (1993) to model the behavior of different groups on a specific outcome over a specified time period. This modeling procedure, and similar modeling procedures named growth mixture models (GMM) and growth curve models (GCM; Muthén, 2001), assumes that different groups have their own mean trajectory and that this specific trajectory is able to capture most of the overall variation of the group (Kreuter & Muthén, 2008). This model is similar to the more general form of latent growth curve modeling and there are detailed discussions of the similarities and differences reported in-depth elsewhere (Kreuter & Muthén, 2008). However, briefly, GBTMs assume that there are two or more subpopulations that follow a traditional GCM distribution. However, GBTM requires specifying only the form of the distributions related to the trajectories; the statistical procedure leverages structure embedded in the data to establish distinct groups of behavior with common structure. Thus, GBTM can be advantageous in the study of behaviors such as substance use in emerging adulthood due to its ability to discover unique patterns of behavior and its ability to handle the non-normality of such outcomes via choices in distributional form for each latent group. For the purposes of this paper, we primarily focus on the zero-inflated specification, which accounts for the overdispersion of zero values present in the outcome measurement.

In Poisson models, a rate term λ represents the mean usage (count) during the common exposure time in zero-inflated Poisson, as is implemented in GBTM (Kreuter & Muthén, 2008; Nagin, 2010; Nagin & Land, 1993). Inflation terms establish the probability of a zero, which we will call π, and then the remaining fraction $1–\pi$ follows a Poisson distribution. This mixing of "pure" zeros with a "regular" Poisson, establishes the inflation or overdispersion (when there are excess zeros, the variance of the distribution is more than one expects with Poisson, thus we label it overdispersed).

It is important to establish the mixture model portion of the GBTM. For this discussion, we assume that we know that there are K different "types" of behavior, and that the rate parameter is different for each type. Further, we allow that parameter to change over time in a manner prescribed by a polynomial, such as a line, a quadratic curve, or a cubic curve. This flexible form allows for a variety of behaviors, including growth and decline over the same time span. If we knew the group C_i for every subject *i*, then we would specify the rate portion of our model conditional on that value C_i, here given as a quadratic function of time *t*:

$$\ln(\lambda_{i|C_i=k}) = \eta_{0k} + \eta_{1k}t + \eta_{2k}t^2,$$

where each η refers to model estimates for a quadratic model over time, and *k* represents the trajectory group. In other words, η_{0k} is the estimated intercept for group *k* and η_{1k} and η_{2k} are the estimates for time and time squared for group *k* at time *t*. Note that in comparison to most classical growth modeling techniques, the η's in GBTMs vary across latent classes rather than across individuals (Kreuter & Muthén, 2008). It is additionally important to note that the above zero-inflated GBTM assumes the same growth factors *within* each class (i.e., individuals are assumed to follow the same trajectory over time within each class, with variation due only to the realization of the Poisson process). However, the growth factors (e.g., intercepts, slopes, quadratic terms) are designed to vary *between* classes (Kreuter & Muthén, 2008; Nagin, 2010; Nagin & Land, 1993). Thus, one of the key benefits to this particular methodology is its ability to identify distinct classes of developmental trajectories within a population while also accounting for the zero-inflated nature of the data in a computationally less intense manner (Nagin, 2010). For the social behavioral scientist, this technique is particularly useful as it has the ability to better specify intervention design. For instance, understanding the racial/ethnic differences in substance use trajectories could aid in the tailoring of intervention design and/or shifting of policy-based substance use prevention policies.

The goal of the present examination is two-fold. First, given the rather limited information regarding substance use dynamics across the emerging adulthood period among YSMM, we present data that define alcohol and marijuana use trajectories among a diverse sample of YSMM. Second, we provide a useful demonstration of the group-based trajectory modeling (GBTM) approach as a means to account for zero-inflated data and to address some of the complications that may arise from more computationally dense approaches like LCG modeling.

STEPS IN GROUP-BASED TRAJECTORY MODELING WITH ZERO-INFLATED OUTCOMES

A series of four steps are proposed to accommodate researchers who want to examine group-based trajectory modeling for developmental processes utilizing count outcomes that may follow a zero-inflated distribution (e.g., substance use, sexual risk behavior). This multistep procedure is important in order to ensure that the most appropriate modeling procedure is used to account for potential zero-inflation of the count outcome. The four steps are described in detail below.

Step 1: Data Exploration

The first step involves an exploration of the study variables. For longitudinal count outcomes, it is important that the data are collected in a parallel fashion; in other words, the "counts" should reflect the number of times engaging in a particular behavior of interest over the *same time interval* (e.g., over a 30-day period) so the exposure time is consistent across time points. The "data exploration" step includes an examination of the outcome at each time point (e.g., alcohol or marijuana use) and across time points to get a sense of average trends over time. Further, another important step in order to determine if the outcome follows a zero-inflated distribution is to examine the summary statistics and frequency distributions of your count outcome at each time point — including the minimum value, maximum value, mean, and variance. There may be many "zero" values present in the data (e.g., 60% reported zero substance use over the study period). Further, another sign that the outcome may be overdispersed is if the variance of the outcome is greater than the mean. This is an important step to take, as standard techniques such as Poisson modeling may not be able to accommodate the excess number of zeros (Cook & Wood, 2018). This is because Poisson models require that that the conditional variance of the outcome equals the conditional mean. When there is an excess number of zeros, however, this assumption is often violated (i.e., there is an *overdispersion* of zeros). Note that overdispersion identified this way suggests including zero inflation in the models, but the *exact form* of that inflation in the context of growth mixture models will need to be specified (and tested) further, which we will present in our model specification for our example described below. In addition, any time-varying covariates or risk factors to be included in the model should be explored in this step.

Step 2: Consider a Square Root Transformation of the Outcome

We recommend two methods to proceed with modeling a count outcome. The first is the established approach of attempting to "normalize" the outcome using a transform in the Box-Cox hierarchy (Box & Cox, 1964). It has been shown that the square root transform is often quite effective at minimizing high levels of skewedness. One then can proceed to model the GBTM as if it were normally distributed, and the $\ln(\lambda)$ term is simply replaced by the outcome itself (note that GBTMs are often specified using censored normal distributions, but for our purposes, we can describe them as normal). However, we acknowledge that this is not capturing the structure of the data generating process (DGP), which are discrete counts of use or nonuse over time. If we do not closely capture the "true" underlying process, we could have strongly model-specification dependent findings. Thus, we will use this approach as a robustness check on the example analysis. If approximately the same conclusions can be made with more than one approach (using different assumptions), one can have greater confidence in their findings. We, of course, will contrast this to the second method, which is to build GBTMs with a (potentially zero-inflated) Poisson distribution specification.

Step 3: Determining the Best Number of Groups and Trajectory Shapes

This step requires the determination of the number of groups and trajectory shapes that best describe the data. It is important to use a combination of theory, previous research, and conceptual frameworks as well as statistical diagnostics to determine the "best" number of groups and trajectory shapes. It is applied in almost the same manner for the censored normal and zero-inflated Poisson models, but the latter require specification of the inflation to be described. Typically, one considers a range of polynomials and number of groups. For example, one might start with linear, quadratic, and cubic specifications — identical for each trajectory group. Using the Bayesian information criterion (BIC; Schwarz, 1978), one quickly can determine the approximate "maximal" polynomial needed, which may be cubic for growth and decline in substance use studies. Note that the Akaike information criterion (AIC) is also used in model selection of growth mixture models, so we include that criterion in our analysis as well (Burnham & Anderson, 2004). If the number of groups is somewhat large, this may yield groups with mostly linear or even constant "growth" over time, so it is important to consider different degrees of polynomial for each class. The number of groups is also best determined by BIC (Fraley & Raftery, 1998; see also Lukočiene & Vermunt, 2009), but there is a complex interaction between this choice and the degree of the polynomial that fits the GBTM best. Importantly, at this stage, one is documenting the structure of the variation for the mixture of trajectory types without any additional predictors (see next step), so one must be prepared to alter these choices (using BIC as a guide) at every stage of model selection. While it is interesting which terms in a polynomial are statistically significant, the overarching choice of degree using BIC is more important. It would be surprising to find a "best" model in which most of the higher order polynomial terms are nonsignificant. The next step with Poisson models is to establish the degree of inflation. A straightforward approach used in GBTM is to model group k-specific inflation π_k (where k represents the class or trajectory group) as $\text{logit}(\pi_k) = \alpha_{0k} + \alpha_{1k}t + \alpha_{2k}t^2$ to establish a time-dependent level of inflation, where each α represents the quadratic model for each group specific $\text{logit}(\pi_k)$. In our analysis of substance use, as you will see below, it appears that zero-inflation diminishes as one ages, and thus the coefficient of linear terms in such a polynomial tend to be negative. One uses the same approach of search guided by BIC for the inflation, with the possibility of identifying homogeneous inflation across groups, or a lack of dependence on k (we did not find this in our analysis). Unfortunately, the choice between normal and Poisson models cannot be made using BIC due to an incompatibility in the likelihoods calculated in the available software.

In addition to using the AIC and BIC in selecting the "best" model fit, recommendations of Nagin and Odgers (2010) state that one should additionally examine posterior probabilities. Posterior probabilities are estimated by the GBTM and refer to the probability that each individual is assigned to each trajectory group. For example, in a 4-group trajectory model, an individual may be assigned a posterior probability of .20 for belonging to trajectory group one, .40 for trajectory group two, .10 for trajectory group three, and .30 for trajectory group four. An individual is then classified into the trajectory group in which they have the highest posterior probability (e.g., in the previous example, the individual would be assigned to trajectory Group 2). According to Nagin and Odgers (2010), "good" model fit are models with average within-group posterior probabilities of above .70 (i.e., the average posterior probabilities within each trajectory group should be high). In addition, there should be a close fit between predicted group and model fitted group proportions.

Step 4: Addition of Risk Factors and Time-Varying Covariates

While it might not be immediately apparent, the GBTM accomplishes two objectives at the same time. First, it establishes a set of trajectories that typify the developmental process. Second, it establishes the relative frequency of each type in a set of mixture parameters. Moreover, each subject may be assigned a *posteriori* probability of group membership and may be assigned to the group for which this value is largest. Thus, a widely used method is to export the trajectory assignments from a final unadjusted model, test these for associations, and then use a multinomial modeling procedure to examine individual-level characteristics that may predict group membership. The advantage to this is that the trajectory groups are "pure," having been established exclusively using the outcome and time. However, it is a "two-stage" approach, and thus the inference in the second model is conditional on the first (and ignores precision). The second approach adds risk factors and even time-varying covariates to the model itself. "Risk factors" (i.e., time-invariant individual-level characteristics such as racial/ethnic identity) establish person-specific probabilities of trajectory group membership (Nagin, 2005), and these are estimated simultaneously with the group characteristics. This can lead to slightly different group trajectories forming, as individuals are essentially "moved" from one group to another based on these risk factors. In fact, individuals are not so much re-assigned as the weight to which they contribute to the parameter estimation of each group is more dynamic. The other potential addition, time-varying covariates, allow for the examination of whether factors that change during the course of development (e.g., loss of employment) alter the trajectory shapes (Nagin, 2005). Prior to adding in “risk factors,” however, one could export the predicted trajectory group profiles of the unadjusted model and use a chi-square test of independence to examine if the “risk factors” (in this case, categorical) are associated with trajectory group membership. Once the “risk factor(s)” and/or time-varying covariate(s) are added, model fit should be reassessed using the aforementioned guidelines and should be compared to the unadjusted models in order to establish potential differences in the polynomial terms capturing change over time. As we will see in the example described later, polynomial terms may need to be dropped in order to account for the addition of risk factors to the model.

It is important to note that in the absence of risk factors or time-dependent predictors, GBTM is a form of clustering. It is a variant of model-based clustering for non-normal outcomes (Banfield & Raftery, 1993). However, once the researcher adds risk factors or predictors other than time to the model, the goal shifts from description to prediction, and one must contextualize the findings with this in mind. The noncovariate, nonrisk factor models of growth and decline establish typologies; the risk models attempt to quantify under which conditions will certain trajectories become more common and predictors allow dynamic alteration of the trajectory due to additional information/circumstances. In the latter scenario, there is no longer one homogeneous group, making it less a form of clustering in the traditional sense

GBTM Application: Understanding Trajectories of Alcohol and Marijuana Use among Emerging Adult Sexual Minority Men

Sexual minority stress (SMS) theory (Hatzenbuehler, 2009) posits that the negative social valuation of sexual minority identities (e.g., homophobia) leads to an excess burden of stress among sexual minority populations as compared to their heterosexual counterparts. This excess stress, in turn, can lead to the usage of substances as a coping mechanism (Hatzenbuehler, 2009; Talley, Sher, Steinley, Wood, & Littlefield, 2012). Research suggests that sexual minorities are at a particularly increased risk for experiencing discrimination and victimization within the emerging adulthood period in particular (Friedman et al., 2011; Morgan, 2013). Moreover, sexual minorities who report experiencing discrimination are more likely to abuse substances (e.g., alcohol) than sexual minorities who do not (McCabe, Bostwick, Hughes, West, & Boyd, 2010). In a longitudinal study examining college-age students, Talley et al. (2012) found that although YSMM reported lower alcohol usage than their heterosexual counterparts they had faster increasing trajectories of reported negative consequences of alcohol abuse than their heterosexual counterparts over a 4-year period. Further, these authors also found that YSMM were more likely than their heterosexual counterparts to report using alcohol as a coping mechanism. In another longitudinal study Hatzenbuehler, Corbin, and Fromme (2008) found that YSMM had faster increasing trajectories of alcohol use throughout college as compared to heterosexual males. Further, Halkitis et al. (2015) found evidence for increasing trajectories of alcohol, marijuana, and other drug use across an 18-month period among YSMM in New York City.

However, evidence also suggests heterogeneity exists within patterns of substance use behaviors among YSMM across the emerging adulthood period, such that those who identify as a racial/ethnic minority may use substances at lower rates than their White counterparts (Newcomb, Birkett, Corliss, & Mustanski, 2014; Slater, Godette, Huang, Ruan, & Kerridge, 2017). Indeed, a longitudinal study utilizing data from the P18 Cohort Study found that Black, Asian, and non-White identifying YSMM consistently used alcohol at lower rates over time as compared to their White counterparts (Halkitis et al., 2014). Further, another longitudinal study examining trajectories of substance use across the emerging adulthood period among YSMM found that Black YSMM were more likely than their White counterparts to be classified in trajectory groups that consisted of low alcohol, marijuana, and other drug use over time (Swann, Bettin, Clifford, Newcomb, & Mustanski, 2017). Thus, we often see flatter trajectories of substance use over the emerging adulthood period for non-White YSMM. However, there continues to be considerable variation with respect to substance use behaviors within populations of YSMM that warrants further exploration (Marshal et al., 2008, 2009; Talley, Sher, & Littlefield, 2010).

Hypotheses

Based on the extant research literature, we hypothesize that the largest trajectory group in the unadjusted alcohol and marijuana models will show an overall increase in alcohol and marijuana use during the early part of the emerging adulthood period and then a general decline in late emerging adulthood among young sexual minority men (YSMM; Halkitis et al., 2014; Hatzenbuehler et al., 2015). In addition, we hypothesize that there will be additional trajectory groups that will have lower membership and will show patterns of excessive usage and no usage. Lastly, based on the research literature examining racial/ethnic differences in trajectories of substance use during the emerging adulthood period among YSMM, we hypothesize that in comparison to White YSMM, Black YSMM, non-White Hispanic YSMM, and YSMM identifying as another non-White racial/ethnic identity will have lower trajectories (i.e., the curves will be "flatter") of alcohol and marijuana use over the emerging adulthood period.

DATA SOURCE

The P18 Cohort Study

Data came from the Project 18 (P18) Cohort Study, which was a 6-year cohort study examining the factors influencing the health and well-being of 600 gay and bisexual men across the emerging adulthood period (18-24). Many of the authors were part of the investigative team that implemented the study and collected the data. At baseline, participants were between the ages of 18 and 19 years old. One person was missing data on alcohol and marijuana use, and two people were missing data on racial/ethnic identity, thus these participants were dropped from the analysis and the final analytic sample consisted of 597 YSMM. Data were collected from 2010-2019 in the greater New York City metropolitan area. Participants were recruited utilizing active and passive recruitment strategies. To be eligible for the study, participants had to be between the ages 18 and 19 years old at baseline, identify as cisgender male, self-report having had sex with another man in the past six months, and self-report a negative HIV status at the time of baseline. After providing informed consent, participants completed an audio computer-assisted self-interview (ACASI) baseline survey in order to provide information pertaining to sociodemographics, mental health, and psychosocial factors. Thereafter, participants completed follow-up assessments every six months. At each study visit, participants reported on how many days they engaged in substance use behaviors (e.g., alcohol use, marijuana use) over the past 30 days through an interview-administered calendar-based methodology called the Timeline followback (TLFB; Robinson, Sobell, Sobell, & Leo, 2014). Full study details have been described elsewhere (Halkitis et al., 2015, 2013). For the current analyses, we utilized the survey data collected every six months for a total of 14 data collection points (i.e., "waves") encompassing the period between 18 and

26 years of age. We used the TFLB responses to assess substance usage over the last 30 days in order to examine alcohol and marijuana use trajectories. We utilized Stata v15 for all our analyses.

Plan of Analysis

Our analytical plan followed the four aforementioned steps. First, descriptive and bi-variable statistics were conducted for all study variables, including an examination of overdispersion for the study outcomes of alcohol and marijuana at each of the 14 waves. Second, we considered a square root transformation of our outcomes in order to examine if this minimized high levels of skewedness. Third, we used the *traj* package (Jones & Nagin, 2013) in Stata in order to understand the potentially different classes of substance use trajectories among YSMM using both censored normal and zero-inflated Poisson distributions. In this step, we examined several censored normal and zero-inflated Poisson GBTMs, separately, for both substances of interest (i.e., alcohol and marijuana) prior to choosing the final models based on BIC, AIC, and posterior probabilities. For each substance, group sizes of two to five were tested in order to have a range of models that may feasibly fit the data. Any tests greater than five were not considered feasible given the sample size. In each model, we initially allowed all groups to have "Order 3," indicating the use of cubic terms of time to model the outcomes of marijuana and alcohol, respectively, in each group. We conducted many models that included other orders (e.g., 0, 1), but based on the fit statistics, Order 3 and Order 2 provided the best model fit and aligned with our general conception of the substance use trajectories among YSMM for alcohol and marijuana use, respectively. Based on BIC (and sometimes reference to the AIC), along with overall visual fit and group size, four- and five-group models were chosen for additional fit analyses for both alcohol and marijuana use over time. This included running models with mixed polynomial terms for those models with polynomial term(s) that did not reach statistical significance to examine if the dropping of polynomials led to better model fit (Marshall et al., 2015). The models with the smallest absolute BIC and AIC were included in the final results (Nagin, 2005). Lastly, after selecting the GBTMs for alcohol and marijuana use among YSMM, we examined racial differences in the different trajectory profiles (Supplemental File 2[1] displays example syntax for GBTM in Stata).

RESULTS

Descriptive Statistics (Step 1)

Table 1 displays the descriptive characteristics of the P18 Cohort Study, including the frequency of alcohol and marijuana use at each of the 14 waves. Thirty-eight percent of the participants identified as Hispanic, 29% as White, 15% as Black, and the remaining identified as another racial/ethnic identity. 86% of the participants were currently enrolled in school at the time of baseline. In terms of socioeconomic status (SES), the majority of the sample reported a familial SES of middle-class (37%) to upper-middle class (25%). Further, 89% of the study sample reported being born inside of the United States. In addition, most of the participants identified as exclusively homosexual (41%), while 29% identified as predominately homosexual and incidentally heterosexual, 13% identified as predominately homosexual but more than incidentally heterosexual, 12% identified as equally heterosexual and homosexual, and 2% identified as predominately

TABLE 1
P18 cohort study descriptives

	%/*M*(*SD*)	Minimum	Maximum
Racial/ethnic identity			
White	29	--	--
Black	15	--	--
Hispanic	38	--	--
Other	18	--	--
Student status			
Enrolled in school at initial visit	86	--	--
Not enrolled in school at initial visit	14	--	--
Familial socioeconomic status			
Upper-class	4	--	--
Upper-middle class	25	--	--
Middle-class	37	--	--
Lower-middle	24	--	--
Lower-class	10	--	--
Sexual orientation			
Exclusively homosexual	41	--	--
Predominately homosexual, incidentally heterosexual	29	--	--
Predominately homosexual, more than incidentally heterosexual	13	--	--
Equally heterosexual and homosexual	12	--	--
Predominately heterosexual, more than incidentally homosexual	3	--	--
Predominately heterosexual, incidentally homosexual	2	--	--
Alcohol use[a]			
Wave 1	4.08 (4.15)	0	27
Wave 2	2.16 (3.14)	0	26
Wave 3	2.26 (3.49)	0	30
Wave 4	2.61 (3.63)	0	30
Wave 5	2.98 (4.36)	0	30
Wave 6	3.29 (4.94)	0	29
Wave 7	3.85 (5.38)	0	30
Wave 8	4.88 (5.53)	0	30
Wave 9	4.91 (5.36)	0	30
Wave 10	4.62 (5.35)	0	30
Wave 11	4.41 (4.60)	0	24
Wave 12	4.41 (4.64)	0	24
Wave 13	4.06 (4.57)	0	30
Wave 14	4.48 (5.15)	0	30
Marijuana Use[a]			
Wave 1	4.79 (8.67)	0	30
Wave 2	5.12 (8.82)	0	30
Wave 3	5.58 (9.40)	0	30
Wave 4	6.84 (10.47)	0	30
Wave 5	6.86 (10.57)	0	30
Wave 6	7.73 (11.07)	0	30
Wave 7	8.16 (11.54)	0	30
Wave 8	9.33 (11.97)	0	30
Wave 9	8.49 (11.68)	0	30
Wave 10	8.58 (11.99)	0	30
Wave 11	8.67 (11.94)	0	30
Wave 12	8.54 (11.99)	0	30
Wave 13	8.57 (11.91)	0	30
Wave 14	9.25 (12.11)	0	30

Note. [a]Use over the preceding 30-days.

heterosexual and incidentally homosexual. None of the participants identified as exclusively heterosexual. In terms of alcohol use, the average number of days that participants reported using alcohol varied over time; however, at each wave the variance was more than twice the mean suggesting an overdispersion of zero values at each of the 14 waves. Moreover, on average, marijuana use increased over time, with variance also being greater than twice the mean suggesting an overdispersion of zero values at each wave.

Unadjusted Models (Steps 2 and 3)

Alcohol Use Trajectories

Censored normal model. The mathematically best-fitting model for alcohol use consisted of five trajectories over time that each followed a cubic pattern (see Supplemental Table 1[1]). Model fit statistics, including AIC, BIC (Supplemental Table 1[1]), along with group size (Table 2), suggested that the model was well-fitting for these data compared to others which were tested (group size should not be very small as a proportion of cases). Supplemental Figure 1[1] displays the five trajectories of alcohol use over time. We classified the five trajectories as follows: those who started out moderate on alcohol and decreased initially before sharply increasing over time (“moderate increase”; 7.1% of the study sample); those who did not drink but gradually increased in use over time (“low increase”; 10.6% of the study sample); those who stayed moderate drinkers over time (“moderate”; 53.0% of the study sample); those who started out moderate alcohol users and decreased over time (“moderate decrease”; 15.0% of the study sample); and lastly, those who started out with a high level of alcohol use and increased over time (“high increase”; 14.3% of the study sample). Table 2 displays average posterior probabilities for each of the five trajectory groups, the proportion of the sample assigned to each group, and the estimated proportion assigned to each group. In addition to BIC and AIC, the final model demonstrated good fit.

ZIP model. The mathematically best-fitting model for alcohol use displayed five distinct trajectories with four cubic polynomials and one quadratic polynomial (see Supplemental Table 2[1]). Model fit statistics, including AIC and BIC, along with group size (Table 2), suggested that the model was well-fitting for these data compared to others which were tested. Table 2 displays average posterior probabilities for each of the five trajectory groups, the proportion of the sample assigned to each group, and the estimated proportion assigned to each group. In conjunction with AIC and BIC estimates, this model indicated an overall good model fit. The final model displayed five unique patterns of alcohol use over time. Supplemental Figure 2[1] displays the five unique trajectories of alcohol use in emerging adulthood among YSMM: low gradual increase drinkers (“slow increase”, 27.7% of the study sample); consistently low drinkers (“low”; 30.2% of the study sample); moderate drinkers who waiver over time and eventually decrease (“moderate decrease”, 21.1% of the study sample); moderate drinkers who become heavy drinkers over time (“heavy” 4.6% of the study sample); and consistently moderate drinkers (“moderate increase”, 16.4% of the study sample). The terms listed as alpha0 and alpha1 represent the linear model for each group specific $\mathrm{logit}(\pi_k) = \alpha_{0k} + \alpha_{1k}t$. One notices that all alpha0 terms are negative, near –1, while all alpha1 terms are negative, around –0.10. This translates to an initial logit of –1.1, but then 14 waves later (i.e., 7 years), it would be –2.4. Using the inverse logit formula $\pi_k = \frac{e^{\alpha_{0k}+\alpha_{1k}t}}{1+e^{\alpha_{0k}+\alpha_{1k}t}}$, , this corresponds to about 25% dropping to less than 10% by age 25. These fractions represent the “never used” group (for that period) in excess of what would normally be expected in Poisson model. These excessive zeros clearly decline substantially over time.

TABLE 2
Diagnostic statistics of group-based trajectory models (N = 597)

	Average posterior probability within group	Proportion (#) assigned to each group	Estimated proportion using model	BIC (N = 4,695)	BIC (N = 597)	AIC
Alcohol trajectories, unadjusted model: Censored normal (Supplemental Figure 1[1])						
1. Moderate increase	.76	.06 (34)	.07	–7069.89	–7044.11	–6989.21
2. Low increase	.76	.13 (76)	.11			
3. Moderate	.81	.59 (355)	.53			
4. Moderate decrease	.77	.11 (63)	.15			
5. High increase	.86	.12 (69)	.14			
Alcohol trajectories, unadjusted model: Zero inflated poisson (Supplemental Figure 2[1])						
1. Slow increase	.76	.27 (164)	.28	–11763.73	–11729.70	–11657.23
2. Low	.81	.33 (197)	.30			
3. Moderate decrease	.84	.20 (122)	.21			
4. Heavy	.97	.04 (25)	.05			
5. Moderate increase	.92	.15 (89)	.16			
Marijuana use trajectories, unadjusted model: Censored normal (Supplemental Figure 3[1])						
1. Abstainers	.87	.41 (247)	.37	–6632.19	–6611.57	–6567.65
2. Sharp increase	.86	.11 (65)	.14			
3. Moderate increase	.80	.13 (76)	.14			
4. Moderate decrease	.77	.15 (89)	.14			
5. Frequent	.90	.20 (120)	.20			
Marijuana use trajectories, unadjusted model: Zero inflated poisson (Figure 1)						
1. Abstainers	.92	.47 (283)	.44	–12709.65	–12681.07	–12619.59
2. Gradual increase	.94	.11 (65)	.12			
3. Fast increase	.86	.10 (59)	.11			
4. Moderate decrease	.96	.12 (73)	.13			
5. Frequent	.98	.20 (117)	.20			
Alcohol use trajectories, adjusted model: Censored normal model with risk factors (Figure 2)						
1. Moderate increase	.75	.07 (41)	.08	–7090.99	–7052.84	–6971.59
2. Low increase	.81	.12 (73)	.11			
3. Moderate	.81	.58 (348)	.52			
4. Moderate decrease	.76	.12 (72)	.16			
5. High increase	.86	.11 (63)	.13			
Alcohol use trajectories, adjusted model: Zip model with risk factors (Figure 3)						
1. Moderate decrease	.84	.11 (63)	.11	–11764.35	–11717.95	–11619.13
2. Low	.84	.35 (206)	.32			
3. Moderate increase	.86	.36 (212)	.36			
4. Heavy increase	.96	.04 (25)	.05			
5. Moderate	.93	.15 (91)	.16			
Marijuana use trajectories, adjusted model: Censored normal model with risk factors (Figure 4)						
1. Abstainers	.87	.41 (246)	.37	–6668.81	–6635.82	–6565.55
2. Sharp increase	.86	.11 (63)	.14			
3. Moderate increase	.76	.14 (85)	.14			
4. Moderate decrease	.86	.13 (79)	.15			
5. Frequent	.90	.21 (124)	.20			

Marijuana Use Trajectories

Censored normal model. The best-fitting group-based trajectory censored normal model consisted of five groups with a mixture of polynomial terms, which was one linear term, one cubic term, and three quadratic terms (see Supplemental Table 3[1]). Compared to the other models, the five-group was superior in terms of AIC and BIC, along with relative group size. Diagnostic statistics are found in Table 2, which, overall, indicated good model fit. Supplemental Figure 3[1] displays the five unique trajectories of marijuana use over time. The five trajectories were classified as follows: abstainers (37.5% of the study sample); abstainers who sharply increased in use over time ("sharp increase"; 14.2% of the study sample); moderate users who gradually increased use over time ("moderate increase"; 13.9% of the study sample); moderate users who decreased use over time ("moderate decrease"; 14.5% of the study sample); and frequent users who increased over time ("frequent"; 19.9% of the study sample).

ZIP model. The mathematically best-fitting group-based trajectory model for marijuana consisted of five unique trajectories using a mixture of polynomials (one linear term and four quadratic terms; see Supplemental Table 4[1]). Compared to all other models, the five-group model was superior in terms of AIC and BIC along with relative group size. Table 2 displays the diagnostic statistics, which indicated good model fit. Figure 1 displays the five distinct patterns of marijuana use over time. The resulting trajectories display five patterns of use: abstainers (44.5%), moderate users who gradually increased use ("gradual increase"; 11.8%), low users who drastically increased use over time ("fast increase"; 11.2%), moderate users who decreased use over time ("moderate decrease"; 13.0%), and frequent users who increased to heavy users ("frequent"; 19.5%). The patterns of excessive zeros, captured in the alpha0 and alpha1 terms (not shown graphically; see Supplemental Table 4[1]) are much more complicated in this model, revealing heterogeneity between groups captured very well by the ZIP GBTM. Group 1 has alpha0 values near 1; with alpha1 value also near zero, this is a stable inflation over time of over 70%. Change over time in zero inflation for Group 2 is driven primarily by the linear term, with a change from near 50% initially, to nearly 0% by age 25. This does not indicate a lack of use; merely, a lack of excessive zeros. This is consistent with the growth in rate of marijuana use by the end of the study, since the overall rate is larger (arguably, this low to high trend is what suggests that these individuals are in the same group).

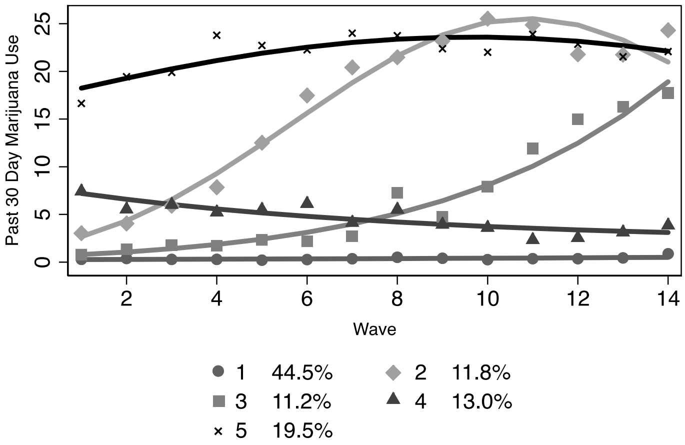

FIGURE 1
ZIP marijuana trajectories, unadjusted.

Note. Each "dot" represents the average number of days that an average individual in the trajectory group is expected to use marijuana over the preceding 30 days at each wave. Each line represents the average trajectory of marijuana use for each trajectory group.

EXPLORATION OF TRAJECTORY GROUP MEMBERSHIP AND RACIAL/ETHNIC IDENTITY AMONG EMERGING ADULT SEXUAL MINORITY MEN

Adjusted Group-Based Trajectory Models (Step 4)

Alcohol use trajectories and racial/ethnic identity

Censored normal model. Prior to adding racial/ethnic identity to the model as "risk factors" for group membership, we exported the predicted group membership profiles and conducted a chi-square test of independence to examine if trajectory group membership was associated with racial/ethnic identity. We found evidence to suggest that racial/ethnic identity was associated with trajectory group membership in our study sample, $\chi^2(12) = 45.64$, $p < .001$. The best-fitting model consisted of five trajectory groups with all cubic polynomials (see Supplemental Table 5[1]). Table 2 displays the diagnostic statistics, which indicated a good model fit.

The five trajectory groups were classified in a similar fashion to the unadjusted model (see Figure 2): those who started out moderate on alcohol and decreased initially before sharply increasing over time ("moderate increase"; 7.9% of the study sample); those who did not drink but gradually increased in use over time ("low increase"; 10.9% of the study sample); those who stayed moderate drinkers over time ("moderate"; 51.8% of the study sample); those who started out moderate alcohol users and decreased over time ("moderate decrease"; 16.0% of the study sample); and lastly, those who started out with a high level of alcohol use and increased over time ("high increase"; 13.4% of the study sample). With respect to group membership, we found that as compared to White YSMM, Hispanic YSMM (OR = 7.10, $p < .05$; Supplemental Table 5[1]) and those classified in the other racial/ethnic identity group (OR = 21.97, $p = .02$; Supplemental Table 5[1]) had higher relative odds of being classified in the "low increase" group relative to the "moderate increase" group. In addition, we found that relative to White YSMM, Black YSMM had lower relative odds of being classified in the "moderate" group relative to the "moderate increase" group (OR = .26, $p = .02$; Supplemental Table 5). Lastly, as compared to White YSMM, Black (OR = .09, $p < .01$; Supplemental Table 5[1]) and Hispanic (OR = .24, $p = .04$; Supplemental Table 5[1]) YSMM had lower relative odds of being classified in the "high increase" group relative to the "moderate increase" group.

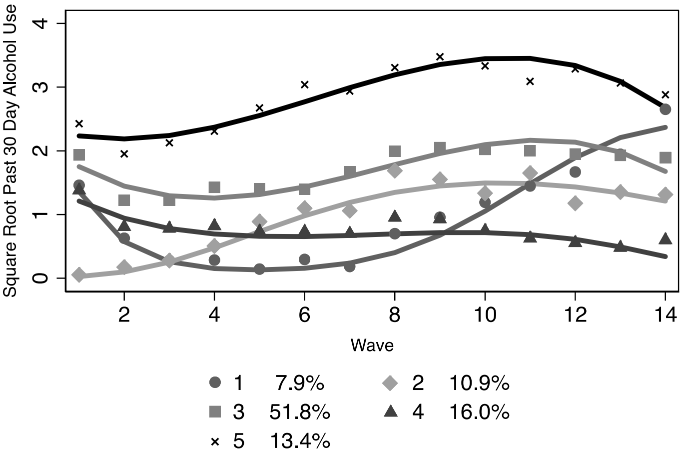

FIGURE 2

Censored normal alcohol trajectories, adjusted.

Note. Each "dot" represents the square root of the average number of days that an average individual in the trajectory group is expected to use alcohol over the preceding 30 days at each wave. Each line represents the average trajectory of alcohol use for each trajectory group.

ZIP model. Before "risk factors," we exported the predicted group membership profiles and conducted a chi-square test of independence to examine if trajectory group membership was associated with racial/ethnic identity. We found evidence to suggest that racial/ethnic identity was associated with trajectory group membership, $\chi^2(12) = 62.95$, $p < .001$, and thus proceeded to add risk factors to the model. Similarly to the unadjusted ZIP model, we found evidence for five distinct trajectories of alcohol use over time with the addition of racial/ethnic identity as a "risk factor," four of which follow a cubic pattern and one which follows a quadratic pattern over time (see Supplemental Table 6[1]). Model fit statistics, including AIC and BIC, along with group size (see Table 2), suggested that the model was well-fitting for these data compared to others which were tested. Figure 3 displays the five trajectories of alcohol use over time among our sample of YSMM. We classified the five trajectories of alcohol use over time as follows: those who started out as moderate drinkers and who decrease over time ("moderate decrease", 11.2% of the study sample); consistently low drinkers ("low", 32.0% of the study sample); moderate drinkers who increase over time ("moderate increase", 35.7% of the study sample); those who start out as moderate drinkers and sharply increase use over time ("heavy increase"; 4.6% of the study sample); and consistently moderate drinkers ("moderate", 16.5% of the study sample).

With respect to racial/ethnic identity, we found that as compared to White YSMM, Black (OR = 6.49, $p < .01$) and Hispanic YSMM (OR = 5.70, $p < .001$), and YSMM of another racial/ethnic identity (OR = 58.56, $p < .001$) had higher relative odds of being classified in the "consistently low" group as compared to the "moderate decrease" group (Supplemental Table 6[1]). In addition, YSMM classified in the "other" racial/ethnic identity category had higher relative odds of being classified in the "heavy increase" group as compared to the "moderate decrease" group (OR = 15.96, $p = .03$; see Supplemental Table 6[1]). Lastly, YSMM classified in the "other" racial/ethnic identity category had higher relative odds of being classified in the "consistently moderate" group as compared to the "moderate decrease" group (OR = 11.59, $p = .04$; see Supplemental Table 6[1]).

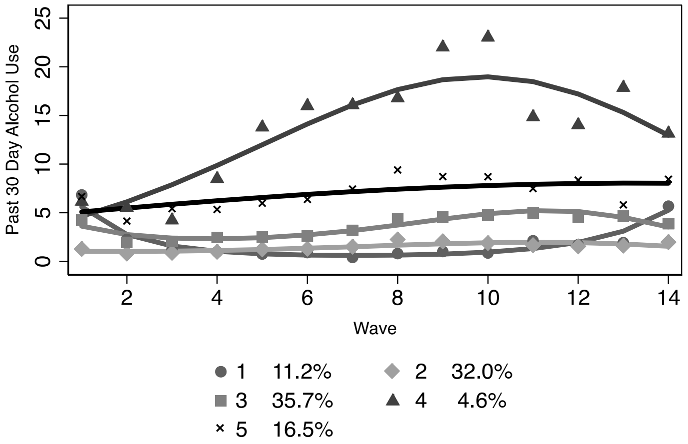

FIGURE 3
Zip alcohol trajectories, adjusted.

Note. Each "dot" represents the average number of days that an average individual in the trajectory group is expected to use alcohol over the preceding 30 days at each wave. Each line represents the average trajectory of alcohol use for each trajectory group.

Censored normal model. Before adding racial/ethnic identity to the model as “risk factors” for group membership, we exported the predicted group membership profiles and conducted a chi-square test of independence to examine if trajectory group membership was associated with racial/ethnic identity. We found evidence to suggest that racial/ethnic identity was associated with trajectory group membership in, $\chi^2(12) = 27.36$, $p < .01$. We found evidence for five distinct trajectories of use over time that followed a mixture of polynomial terms (one linear and four quadratic terms; see Supplemental Table 7[1]). Table 2 displays the diagnostic statistics, which indicated a good model fit relative to the other models tested. The five trajectories were classified similarly to the “unadjusted” model (see Figure 4): abstainers (37.4% of the study sample); abstainers who sharply increased in use over time (“sharp increase”; 13.8% of the study sample); moderate users who gradually increased use over time (“moderate increase”; 13.8% of the study sample); moderate users who decreased use over time (“moderate decrease”; 14.7% of the study sample); and frequent users who increased over time (“frequent”; 20.3% of the study sample). With respect to racial/ethnic identity, we found evidence to suggest that, in comparison to White YSMM, those who were classified in the “other” racial/ethnic identity category were had lower relative odds of being classified in the “moderate decrease” group relative to the “abstainers” group (OR = .28, $p = .02$; see Supplemental Table 7[1]). We did not find evidence to suggest differences in group membership by racial/ethnic identity among the other four trajectory groups (see Supplemental Table 7[1]).

ZIP model. A chi-square test of independence was run to determine if there were differences in trajectory group membership of the “unadjusted” marijuana group-based ZIP trajectory model prior to running the group-based trajectory model with the addition of racial/ethnic identity as a “risk factor.” We did not find enough evidence to suggest that there was an association between trajectory group membership and racial/ethnic identity in the unadjusted models, $\chi^2(12) = 17.44$, $p = .13$. Thus, we do not present this model.

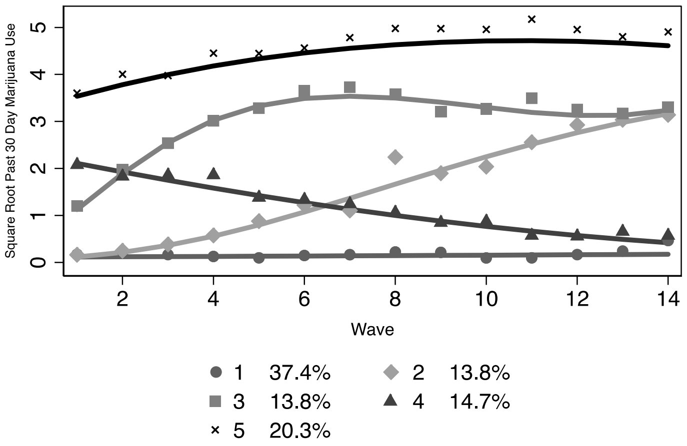

FIGURE 4
Censored normal marijuana trajectories, adjusted.
Note. Each “dot” represents the square root of the average number of days that an average individual in the trajectory group is expected to use marijuana over the preceding 30 days at each wave. Each line represents the average trajectory of marijuana use for each trajectory group.

DISCUSSION

The current example displayed an application of group-based trajectory modeling with zero-inflated count outcomes to examine trajectories of alcohol and marijuana use among YSMM over the emerging adulthood period. Results from the study demonstrate the complexity of group-based trajectory modeling. In particular, we found that transforming our outcomes with a square root transformation was insufficient in dealing with the zero-inflation of alcohol and marijuana use over time. Thus, the use of a zero-inflated specification was warranted for both alcohol and marijuana use among YSMM over time. For example, when comparing the unadjusted censored normal and ZIP models for alcohol use over time we found qualitative differences in the five trajectory curves such that the censored normal models were unable to capture the group of individuals who stayed near zero in terms of alcohol use over time. Further, in these same models, the alpha terms were significant and suggested that the number of zeros exceeded that which would have been predicted by a standard Poisson distribution, providing further evidence towards the use of a zero-inflated specification. Indeed, through the provided steps and example we have demonstrated why we prefer group-based trajectory models with a ZIP specification in comparison to the censored normal specification for zero-inflated outcomes. The ZIP specification allowed us to adjust for the extreme overdispersion of zero values and identify trajectories that we may not have been able to capture in a censored normal model. We further discuss the implications of our study findings below.

We have shown that GBTM can identify underlying typologies of use over time and that the zero-inflation itself can change, as more individuals in a group move from “never use” to using sometimes. These typologies are a useful first stage in an analysis, but one often wishes to go beyond clustering to predictive models of likely trajectory types (i.e., by identifying factors that predict group membership). We have shown that these are instructive when they reveal risk factors for behavior, but that the added complexity comes at a small price of potential adjustments to the typologies themselves. Further, the research literature examining trajectories of substance use across time usually shows an overall increase in alcohol and marijuana use during the early part of the emerging adulthood period and then a general decline in late emerging adulthood among young sexual minority men (YSMM; Halkitis et al., 2014; Hatzenbuehler et al., 2015). Our findings are largely supported by the extant research literature, with some deviations which we discuss below.

Trajectories of Alcohol Use

With respect to the censored normal models, the BIC indicated that the model *without* the inclusion of risk factors was the “best” model fit; however, on the other hand, the AIC indicated that this model was the “best” model fit. In addition, many of the coefficients for race were significant in this model in addition to the chi-square test of independence that was conducted utilizing the exported trajectory profiles of the unadjusted model. However, when looking at the results of the ZIP models, we found that the BIC was consistently telling us that the “best” model fit was the ZIP model with the inclusion of risk factors.

Further, the final selected ZIP models are supported by previous research and thus support our original hypotheses. Our results suggest that there are differing clusters of individuals with respect to alcohol use over time. One cluster seems to have substantial increases in alcohol usage over time, and then decreases to maintain an average of about 12 drinks in the last 30 days. However, this is only one trajectory group. The majority of YSMM seem to exhibit low to moderate drinking patterns over the emerging adulthood period, which is inconsistent with current theory or conceptual understanding of drinking behaviors among YSMM (Halkitis et al., 2014; Hatzenbuehler et al., 2008; McCabe et al., 2010).

Trajectories of Marijuana Use

The results of marijuana usage tell slightly different story, however. Similar to alcohol, we found that the BIC indicated that the model *without* the inclusion of risk factors was the "best" model fit. Nevertheless, the chi-square test of independence that was conducted utilizing the unadjusted model *and* the AIC from the adjusted model point to the inclusion of racial/ethnic identity in the "final" model. On the other hand, however, model fit statistics (including BIC and AIC) consistently suggested that the group-based trajectory model of marijuana over time among YSMM was best *without* the inclusion of racial/ethnic identity as a "risk factor." Thus, when conducting group-based trajectory models with zero-inflated outcomes one needs to examine several factors outside of BIC when determining the "final" models. First, examine the effects of the risk factors across models. In other words, are the effects of the "risk factors" the same (i.e., in terms of significance, magnitude, and direction) across functional forms (e.g., censored normal and ZIP). In addition, are the plots for each trajectory the same across different functional forms? Lastly, when utilizing ZIP models, are we seeing a different story?

Substantively, there is some literature to support our findings while some of our findings are more novel. In the final ZIP models, the majority of YSMM (44.5%) were in the little to no ("abstainer") marijuana usage trajectory group, which is consistent with the extant research literature. The consistently high marijuana use trajectory group ("frequent") had 19.5% group membership. In this trajectory group, YSMM averaged between 18 to around 24 days of marijuana usage in the last 30 days between the ages of 18 and 26 years. This finding is consistent with some research that identified marijuana users at high risk for substance use disorder (McCabe et al., 2010); however, much of the research literature in this area specifies a precipitous decline in marijuana use closer to the end of the emerging adulthood period among YSMM. There were also clusters of individuals who started with very little marijuana use but ended with between 18-25 days of marijuana usage in the last 30 days (11.8% of the study sample; the "low increase" group). The last two clusters showed that there was a group of young men who increased gradually and a group that decreased gradually in their marijuana usage, which was consistent with our study hypotheses. However, inconsistent with our study hypotheses, racial/ethnic identity did not predict group membership in our "final" model for marijuana usage. These mixed findings suggest that there are unique trajectory groups in terms of marijuana use among YSMM that should be explored further.

CONCLUSION

There are many tutorials on the different statistical tools to analyze developmental trajectories. However, far fewer of these tutorials consider specific models that can be practically used by researchers. It is also helpful that GBTM is able to account for the zero-inflated characteristics of the data often used to understand behaviors, such as substance use, during the emerging adulthood period.

Group-based trajectory analyses are most useful when guided by a substantive theoretical framework that can aid in the specifications of the size and number of trajectory groups (Nagin & Odger, 2010). However, as exemplified within this paper, this technique is beneficial when attempting to explore trajectory patterns that may not yet be conceptualized or defined. This is not to say that theory should be not be an initial starting point (indeed it many times should be); however, this analytic technique can help further develop theory and identify important trajectory patterns to be considered for further review.

Group-based trajectory analyses have valuable extensions that increase the flexibility and utility of this method. As noted, one of the positive components of this statistical tool is its flexibility in handling different data types (e.g., zero-inflated, binary, etc.). In addition, coupled with structural equation modeling techniques, random effects can be included in the model, covariates that predict the trajectory groups can be modeled, and maximum likelihood estimation procedures can be utilized to account for missing data (if missing at random; Nagin, 2005). However, as with all analyses, there are limitations in terms of the number of covariates that can be put in the model. In addition, researchers must be thoughtful about selecting the best model. Theory or conceptual considerations as well as statistical model fit must be considered. We consider this a strength when this process is well-documented and explored. However, it can take considerably more time and collaboration among colleagues to select the best model as compared to other methods where statistical fit statistics alone are relied upon.

Overall, GBTM is a useful and computationally lighter alternative to other methods when examining developmental trajectories with zero-inflated outcomes. We contend that this strategy is especially useful when attempting to assess subpopulations within a highly marginalized group, such as YSMM, where there may be key groupings that should be flagged for further exploration. Further, GBTM is a useful tool for examining outcomes that may follow non-normal distributions outside of substance use behaviors. For example, sexual risk behaviors such as condomless intercourse usually follow a zero-inflated distribution (Cook & Wood, 2018) and would lend itself well to a group-based trajectory approach. For example, one study utilizing data from the P18 Cohort Study found that relationship cognitions (i.e., fear of not being in a relationship and control over relationship functioning) were associated with condomless anal intercourse behaviors among YSMM over a two-year period using zero-inflated growth curve models (Cook, Halkitis, & Kapadia, 2018). However, this sort of question may be useful to look at through the lens of group-based trajectory modeling; for example, are there distinct “groups” of sexual risk behaviors over this study period? In addition, do relationship cognitions inform trajectory group membership? Such questions and multilevel analyses are critical for understanding dynamic adaptations and change during the emerging adulthood period, and utilizing GBTM provides an additional means (in addition to other procedures such as latent class growth modeling) of modeling these adaptations and changes. Further, utilizing GBTM can also identify key covariates that may influence trajectory groups in a meaningful way. For instance, in the example presented above we focused specifically on racial/ethnic identity, but in such GBTM models other important factors that influence substance use among emerging adult YSMM could be considered in the future (e.g., education, socioeconomic status, peer relationships, etc.).

In sum, group-based trajectory modeling techniques can be a useful tool for informing culturally relevant and tailored intervention designs. First, it can help researchers to uncover distinct classes of developmental trajectories underlying outcomes that may follow non-normal distributions such as substance use (e.g., when there is an overdispersion of zeros). Further examination of the classes can then determine the characteristics associated with class membership (e.g., race/ethnicity, stress exposure). This information, in turn, can help to inform intervention design through the uncovering of factors associated with patterns of both risk and resilience.

NOTE

1. Supplemental material is available at https://github.com/hazel275/TPM-Supplement

FUNDING

This research was funded by grants from the National Institute on Drug Abuse (1R01DA0225537; 2R01DA025537).

REFERENCES

Arnett, J. J. (2004). *Emerging adulthood: The winding road from the late teens through the twenties*. Oxford University Press.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics, 49*(3), 803-821. doi:10.2307/2532201

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243. doi:10.1111/j.2517-6161.1964.tb00553.x

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261-304. doi:10.1177/0049124104268644

Cook, S. H., Halkitis, P. N., & Kapadia, F. (2018). Relationship cognitions and longitudinal trajectories of sexual risk behavior among young gay and bisexual men: The P18 cohort study. *Journal of Health Psychology*, *23*(14), 1884-1894. doi:10.1177/1359105316671187

Cook, S. H., & Wood, E. P. (2018). *Growth Curve Models for Zero-Inflated Count Data: An Application to Sexual Risk Behavior*. SAGE Publications Ltd.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, *41*(8), 578-588. doi:10.1093/comjnl/41.8.578

Friedman, M. S., Marshal, M. P., Guadamuz, T. E., Wei, C., Wong, C. F., Saewyc, E. M., & Stall, R. (2011). A meta-analysis of disparities in childhood sexual abuse, parental physical abuse, and peer victimization among sexual minority and sexual nonminority individuals. *American Journal of Public Health, 101*(8), 1481-1494. doi:10.2105/AJPH.2009.190009

Greenwood, C. J., Youssef, G. J., Betts, K., Letcher, P., Mcintosh, J., Spry, E., . . . Sanson, A. (2019). A comparison of longitudinal modelling approaches: alcohol and cannabis use from adolescence to young adulthood. *Drug and Alcohol Dependence*, *201*, 58-64. doi:10.1016/j.drugalcdep.2019.05.001

Halkitis, P. N., Kapadia, F., Bub, K. L., Barton, S., Moreira, A. D., & Stults, C. B. (2015). A longitudinal investigation of syndemic conditions among young gay, bisexual, and other MSM: the P18 Cohort Study. *AIDS and Behavior*, *19*(6), 970-980. doi:10.1007/s10461-014-0892-y

Halkitis, P. N., Kapadia, F., Siconolfi, D. E., Moeller, R. W., Figueroa, R. P., Barton, S. C., & Blachman-Forshay, J. (2013). Individual, psychosocial, and social correlates of unprotected anal intercourse in a new generation of young men who have sex with men in New York City. *American Journal of Public Health*, *103*, 889-895. doi:10.2105/AJPH.2012.300963

Halkitis, P. N., Siconolfi, D. E., Stults, C. B., Barton, S. C., Bub, K. L., & Kapadia, F. (2014). Modeling substance use in emerging adult gay, bisexual, and other YMSM across time: The P18 cohort study. *Drug and Alcohol Dependence*, *145*, 209-216. doi:10.1016/j.drugalcdep.2014.10.016

Hatzenbuehler, M. L. (2009). How does sexual minority stigma “get under the skin”? A psychological mediation framework. *Psychological Bulletin*, *135*(5), 707-730. doi:10.1037/a0016441

Hatzenbuehler, M. L., Corbin, W. R., & Fromme, K. (2008). Trajectories and determinants of alcohol use among LGB young adults and their heterosexual peers: Results from a prospective study. *Developmental Psychology*, *44*(1), 81-90. doi:10.1037/0012-1649.44.1.81

Hatzenbuehler, M. L., McLaughlin, K. A., & Xuan, Z. (2015). Social networks and sexual orientation disparities in tobacco and alcohol use. *Journal of Studies on Alcohol and Drugs*, *76*, 117-126. doi:10.15288/jsad.2015.76.117

Jones, B. L., & Nagin, D. S. (2013). A note on a Stata plugin for estimating group-based trajectory models. *Sociological Methods & Research*, *42*(4), 608-613. doi:10.1177/0049124113503141

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), 1-31. doi:10.1007/s10940-007-9036-0

Liu, H. (2007). Growth curve models for zero-inflated count data: An application to smoking behavior. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 247-279. doi:10.1080/10705510709336746

Lukočiene, O., & Vermunt, J. (2009). Determining the number of components in mixture models for hierarchical data. In A. Fink, B. Lausen, W. Seidel, & A. Ultsch (Eds.), *Advances in Data Analysis, Data Handling and Business Intelligence. Studies in Classification, Data Analysis, and Knowledge Organization*. Heidelberg, Berlin: Springer.

Marshal, M. P., Friedman, M. S., Stall, R., King, K. M., Miles, J., Gold, M. A., . . . Morse, J. Q. (2008). Sexual orientation and adolescent substance use: a meta-analysis and methodological review. *Addiction*, *103*, 546-556. doi:10.1111/j.1360-0443.2008.02149.x.

Marshal, M. P., Friedman, M. S., Stall, R., & Thompson, A. L. (2009). Individual trajectories of substance use in lesbian, gay and bisexual youth and heterosexual youth. *Addiction*, *104*, 974-981. doi:10.1111/j.1360-0443.2009.02531.x

Marshall, B. D. L., Operario, D., Bryant, K. J., Cook, R. L., Edelman, E. J., Gaither, J. R., . . . Fiellin, D. A. (2015). Drinking trajectories among HIV-infected men who have sex with men: A cohort study of United States veterans. *Drug and Alcohol Dependence*, *148*, 69-76. doi:10.1016/j.drugalcdep.2014.12.023

McCabe, S. E., Bostwick, W. B., Hughes, T. L., West, B. T., & Boyd, C. J. (2010). The relationship between discrimination and substance use disorders among lesbian, gay, and bisexual adults in the United States. *American Journal of Public Health*, *100*(10), 1946-1952. doi:10.2105/AJPH.2009.163147

Morgan, E. M. (2013). Contemporary issues in sexual orientation and identity development in emerging adulthood. *Emerging Adulthood*, *1*(1), 52-66. doi:10.1177/2167696812469187

Muthén, B. O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change: Decade of behavior* (pp. 291-322). American Psychological Association.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*(6), 882-891. doi:10.1111/j.1530-0277.2000.tb02070.x

Nagin, D. S. (2005). *Group-based modeling of development*. Harvard University Press.

Nagin, D. S. (2010). Group-based trajectory modeling: an overview. In A. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 53-67). Springer.

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, *31*(3), 327-362. doi:10.1111/j.1745-9125.1993.tb01133.x

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*, 109-138. doi:10.1146/annurev.clinpsy.121208.131413

Newcomb, M. E., Birkett, M., Corliss, H. L., & Mustanski, B. (2014). Sexual orientation, gender, and racial differences in illicit drug use in a sample of US high school students. *American Journal of Public Health*, *104*, 304-310. doi:10.2105/AJPH.2013.301702

Newcomb, M. E., Ryan, D. T., Greene, G. J., Garofalo, R., & Mustanski, B. (2014). Prevalence and patterns of smoking, alcohol use, and illicit drug use in young men who have sex with men. *Drug and Alcohol Dependence*, *141*, 65-71. doi:10.1016/j.drugalcdep.2014.05.005

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Sage.

Robinson, S. M., Sobell, L. C., Sobell, M. B., & Leo, G. I. (2014). Reliability of the Timeline Followback for cocaine, cannabis, and cigarette use. *Psychology of Addictive Behaviors*, *28*(1), 154-162. doi:10.1037/a0030992

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461-464. doi:10.1214/aos/1176344136

Slater, M. E., Godette, D., Huang, B., Ruan, W. J., & Kerridge, B. T. (2017). Sexual orientation-based discrimination, excessive alcohol use, and substance use disorders among sexual minority adults. *LGBT Health*, *4*(5), 337-344. doi:10.1089/lgbt.2016.0117

Stone, A. L., Becker, L. G., Huber, A. M., & Catalano, R. F. (2012). Review of risk and protective factors of substance use and problem use in emerging adulthood. *Addictive Behaviors*, *37*(7), 747-775. doi:10.1016/j.addbeh.2012.02.014

Swann, G., Bettin, E., Clifford, A., Newcomb, M. E., & Mustanski, B. (2017). Trajectories of alcohol, marijuana, and illicit drug use in a diverse sample of young men who have sex with men. *Drug and Alcohol Dependence*, *178*, 231-242. doi:10.1016/j.drugalcdep.2017.05.015

Talley, A. E., Sher, K. J., & Littlefield, A. K. (2010). Sexual orientation and substance use trajectories in emerging adulthood. *Addiction*, *105*(7), 1235-1245. doi:10.1111/j.1360-0443.2010.02953.x

Talley, A. E., Sher, K. J., Steinley, D., Wood, P. K., & Littlefield, A. K. (2012). Patterns of alcohol use and consequences among empirically derived sexual minority subgroups. *Journal of Studies on Alcohol and Drugs*, *73*(2), 290-302. doi:10.15288/jsad.2012.73.290