2019

Effects of measurement timing on subgroup identification using growth mixture modeling: An empirical application to alcohol use

Anne M. Fairlie
University of Rhode Island

Michael Bernstein
University of Rhode Island

See next page for additional authors

Citation/Publisher Attribution
Fairlie, A. M., Bernstein, M., Walls, T. A., & Wood, M. D. (2019). Effects of measurement timing on subgroup identification using growth mixture modeling: An empirical application to alcohol use. *Psychol. Addict. Behav.*, *33*(3), 232-242. doi: 10.1037/adb0000446
Available at: http://dx.doi.org/10.1037/adb0000446

Authors
Anne M. Fairlie, Michael Bernstein, Theodore A. Walls, and Mark D. Wood

Effects of Measurement Timing on Subgroup Identification using Growth Mixture Modeling: An Empirical Application to Alcohol Use

Anne M. Fairlie, Michael Bernstein, Theodore A. Walls, and Mark D. Wood

University of Rhode Island, Kingston, Rhode Island

AUTHOR NOTE

At the time of the study, Anne M. Fairlie, Michael Bernstein, Theodore A. Walls, and Mark D. Wood were affiliated with the Psychology Department, University of Rhode Island.

Anne M. Fairlie is now at the Department of Psychiatry and Behavioral Sciences, University of Washington. Michael Bernstein is now at the Center for Alcohol and Addiction Studies, Brown University.

Drs. Wood and Walls served as mentors for Dr. Fairlie during her dissertation research at the University of Rhode Island. Dr. Fairlie's dissertation was completed in 2012, and Dr. Wood passed away in April 2015. Dr. Fairlie was supported by grant number F31AA020164 from the National Institute on Alcohol Abuse and Alcoholism while at the University of Rhode Island. Dr.

Bernstein was supported by F31AA024358 and T32DA016184. Dr. Walls was supported by an award from the American Cancer Society (No. MRSG-07-015-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Alcohol Abuse and Alcoholism or the National Institutes of Health.

This work is part of Dr. Fairlie's dissertation (2012), portions of which were presented at the 2012 meeting of the Research Society on Alcoholism (San Francisco, CA) and the 2012 meeting of the Modern Modeling Methods conference (Storrs, CT).

Correspondence concerning this article should be addressed to Anne Fairlie, Ph.D., Center for the Study of Health and Risk Behaviors, University of Washington, Box 354944, Seattle, WA 98195. Email: afairlie@uw.edu.

Abstract

Growth mixture modeling (GMM) identifies latent classes exhibiting distinct longitudinal patterns on an outcome. Subgroups identified by GMM may be artifactually influenced by measurement timing (e.g., timing of the initial assessment, length of the interval from the first to the last assessment, total number of assessments) as well as the theoretically posited developmental patterns of the behavior. The current study investigated this possibility using alcohol data from the 1997 National Longitudinal Survey of Youth (n = 2686; 49.44% female; 71.84% White/Caucasian). Three assessment configurations were examined: all 12 waves, first 6 waves, and last seven waves. Five subgroups were identified using all 12 waves: Normative (71.33%), Low-Increasing (8.45%), Low-Steady (8.97%), High-Slowly Decreasing (7.67%), Extreme-Sharply Decreasing (3.57%). When comparing participants' subgroup membership for all 12 waves to the first 6 waves, 14% of the sample was differentially classified. When comparing all 12 waves to the last seven waves, 62% of the sample was differentially classified. Alterations in the timing of the initial assessment had a substantial impact on latent class estimation, underscoring the importance of selecting the developmental window *a priori* based on theory and empirical knowledge. The time-bounded nature of mixture modeling solutions (i.e., a selected developmental window within the course of a phenomenon) suggests that the latent subgroups should not be interpreted as representing subgroups that are present in the population. Future directions and strategies for testing alternative interpretations are presented.

Keywords: measurement timing; growth mixture model; alcohol; adolescents

Effects of Measurement Timing on Subgroup Identification using Growth Mixture Modeling: An Empirical Application to Alcohol Use

Adolescence and young adulthood are important developmental periods with respect to the initiation, escalation, and maintenance of alcohol use (Grant et al., 2006; Schulenberg et al., 2018). A growing body of literature has sought to identify subgroups of individuals with different longitudinal courses of alcohol use in an effort to characterize the heterogeneity in these patterns of use and also identify which etiological factors predict the various longitudinal courses of alcohol use (Chassin, Flora, & King, 2004; Chassin, Pitts, & Prost, 2002; Colder, Campbell, Ruel, Richardson, & Flay, 2002; Goudriaan, Grekin, & Sher, 2007; Greenbaum & Dedrick, 2007; Greenbaum, Del Boca, Darkes, Wang, & Goldman, 2005; Hix-Small, Duncan, Duncan, & Okut, 2004; Li, Duncan, Duncan, & Acock, 2001; Sher, Jackson, & Steinley, 2011). The current study examines whether the identification of subgroups' longitudinal courses of alcohol use (e.g., steady increase over time, flat and stable over time) varies in accordance with measurement timing (i.e., timing and spacing of the assessments). We elucidate these considerations by drawing on methodological concerns that have important substantive implications. Longitudinal patterns of alcohol use are discussed in relation to the theoretically posited developmental patterns of the behavior.

Measurement Timing

The timing of survey assessments, or measurement timing, has long been recognized as a critical aspect of research design in longitudinal studies (Collins & Graham, 2002; Lerner, Schwartz, & Phelps, 2009; Nesselroade & Ghisletta, 2003; Wohlwill, 1973). Numerous researchers have emphasized the importance of the number and spacing of assessments to accurately measure how a behavior changes over time (Collins & Graham, 2002; Jackson &

Sher, 2006; Nesselroade & Ghisletta, 2003; Widaman, 1991; Willett, 1989). Notably, Wohlwill (1973) recognized that interindividual differences in a behavior may vary over time, as in the case of subgroups exhibiting different trajectories over time. For every longitudinal study, decisions must be made about the observation window, including the start of the observation period corresponding to the baseline assessment, the total length of the assessment period, the number of survey assessments, and the frequency and spacing of the assessments. It is critical that the timing of the assessments is determined by a conceptual understanding of how the phenomenon develops over time (Walls, Barta, Stawski, Collyer, & Hofer, 2012). Examining longitudinal trends of alcohol use is one area that would benefit from thoughtful consideration of these questions, given the developmental shifts that occur with respect to alcohol use during adolescence and young adulthood. Researchers could reasonably choose a variety of observation windows to reflect unique developmental transitions (e.g., college matriculation) that often coincide with marked increases or decreases in drinking, and the developmental period of interest may be wide (e.g., spanning adolescence into adulthood) or narrow (e.g., college semester). Under circumstances where the interval necessary to observe changes in the phenomenon is uncertain, it is advisable to use the smallest interval for which change may be evident (Cohen, 1991; Gollob & Reichardt, 1991; Smith & Walls, 2016). However, the total number of assessments may be subject to practical limitations due to limited resources, and decisions about how many assessments to include and over what time interval require careful balance of the practical limitations and the theoretical process of change. Understanding how these decisions may impact study findings is of critical importance.

Decision-making about measurement timing is particularly challenging in the context of subgroup estimation, because different subgroups exhibit unique trends in the outcome over time that correspond to different points of inflection for each subgroup. Timing the assessments to capture all the potential points of inflection could prove to be challenging. It is important to also recognize that researchers may investigate trajectories as secondary data analyses, without having considered trajectory estimation as thoroughly in the planning stage as would have been done if conducted as primary analyses; as such, concerns about measurement timing are potentially heightened. More generally, if assessments are improperly timed, the subgroups may be an artifact of the way the repeated-measurements were spaced rather than reflective of the subgroups that exist in the population. It is important to better understand how substantive findings drawn from methods that identify distinct subgroups may be artifactually influenced by variations in research designs, including the observation window, so that investigators can be better informed about the potential effects of assessment timing when making substantive conclusions (Cohen, 1991; Gollob & Reichardt, 1991; Jackson & Sher, 2006; Sher et al., 2011).

Findings on Measurement Timing from the Applied Literature

Few empirical studies in applied domains have examined how variations in the observation window may affect the identification of subgroups with distinct patterns of longitudinal change. Several empirical studies have analyzed prospective longitudinal data to examine how methodological factors impact subgroup identification (Eggleston, Laub, & Sampson, 2004; Hipp & Bauer, 2006; Jackson & Sher, 2005, 2006, 2008; Sher et al., 2011; Tan, Dierker, Rose, Li, & The Tobacco Etiology Research Network, 2011). Growth mixture modeling (GMM) and latent class growth analysis are often used to empirically identify subgroups with different longitudinal courses of a behavior. Jackson and Sher examined how subgroups varied according to different types of alcohol-related behaviors (Jackson & Sher, 2005) and the cutoff employed to denote heavy episodic drinking (Jackson & Sher, 2008). Four empirical studies have examined the effect of measurement timing in the context of GMM and latent class growth analysis: one study examined criminality among delinquent boys through age 70 (Eggleston et al., 2004), a second study examined smoking behavior among college freshmen over 35 consecutive weeks (Tan et al. 2011), and two studies examined alcohol use among college students up to seven years post-college (Jackson & Sher, 2006; Sher et al., 2011). Eggleston and colleagues underscored the findings that the latent class solutions (e.g., class size, average trajectory shape) were altered based on the inclusion of long-term follow-up data. Tan and colleagues found that even though the class-specific average trajectories were similar across the assessment configurations (35 waves, six waves, and four waves), individuals were assigned to different subgroups depending on the assessment configuration. This finding emphasizes the link between measurement timing and the points of inflection in individual-level trajectories, such that greater individual-level fluctuation (instability) over the study period may introduce inconsistent subgroup assignment.

Germane to the present study, Sher and colleagues (2011) evaluated the degree to which four prototypical classes (stable low, increasing, decreasing, stable high), which were referred to as the “cat’s cradle, were observed across eight different assessment configurations. Sher and colleagues examined 3058 incoming college students who completed a total of eight assessments with baseline occurring during the summer prior to matriculation and follow-up assessments occurring for seven consecutive college semesters. Latent class growth analysis was used to estimate models for a binary indicator of heavy episodic drinking in the past 30 days. Results showed that there was a general tendency for the four prototypical classes to emerge across the eight assessment configurations, in a “cat’s cradle” configuration. However, the shape of the average class trajectories varied to some degree. Specifically, there were differences depending upon whether all eight waves were used, only the first four waves were used, or only the last four waves were used. Notably, participants who exhibited an increasing or decreasing trajectory in either the Waves 1-4 or Waves 5-8 configurations were not classified into a comparable group using all eight waves, thus the individuals' trajectories were time-bounded.

The Current Study

The current study provides a focused investigation of the effects of observation window and measurement timing on subgroup (or class) identification through an empirical application with national data on adolescent and young adult alcohol use. The present study is a follow-up to Sher and colleagues (2011) using a different database (National Longitudinal Survey of Youth; NLYS) in order to determine the extent to which results would be comparable. First, the observation window for participants in NLSY corresponded to mid-adolescence through young adulthood. In doing so, we examined a developmental period with substantially more intra-individual change than the more narrow college sample (ranging from freshman year through senior year) used by Sher and colleagues. The extended age range of the sample enabled us to test the stability of classes within a more dynamic and shifting developmental period with respect to alcohol use behaviors. Second, we used a large-scale nationally representative sample whereas Sher et al. (2011) examined college students from a single university. Finally, Sher et al. used "heavy drinking" as the outcome. Our dependent variable (drinks consumed per month) captures participants at the lower end of the drinking spectrum (i.e., exhibiting non-use or limited use). We compared results from three subgroup analyses that included different assessment configurations: (1) all 12 waves, (2) the first 6 waves, and (3) the last seven waves. Comparisons were made on the basis of the shape of the class-specific average trajectories, class proportions, and consistency in individuals' class assignment. The aim was to determine the extent to which the subgroups identified in each analysis were similar on these key characteristics across the three assessment configurations.

Method

Participants

Publicly available data from the National Longitudinal Survey of Youth (NLSY) 1997 were analyzed. Participants were recruited in 1997 and were between the ages of 12 and 16 years at Wave 1. The first follow-up (Wave 2) occurred approximately 18 months later, and ten subsequent follow-ups (Waves 3 to 12) occurred annually for a total of 12 waves. Only individuals in the nationally representative sample who were 15 and 16 years old at baseline were included in the analyses, resulting in a sample of 2,686 participants. By selecting this restricted cohort, the sample excludes younger participants who predominantly reported no drinking in the past month at baseline (e.g., 97.83% of 12 year olds). Older participants were also excluded since there were proportionally fewer older participants in the sample compared to 15 and 16 year olds (i.e., 1404 participants were 15 years old; 1282 participants were 16 years old; 491 participants were 17 years old).

Approximately half (49.44%) of the participants in the final sample (n = 2686) were female. The mean age of the participants at Wave 1 was 15.94 years (*SD* = 0.57), and the mean age at Wave 12 was 27.50 years (*SD* = 0.63). The majority of the participants were White/Caucasian (71.84%), followed by Black/African American (16.41%), Asian/Pacific Islander (2.55%), American Indian (0.60%), and another race not specified (8.60%). Twenty-three participants did not report race. The majority of the participants were non-Hispanic (85.92%) with nine participants not reporting ethnicity. The participants' place of residence was distributed across the United States: Northeast (19.02%), North Central (26.28%), South (34.25%), and West (20.44%).

Measures

Demographics. Demographic information included sex, race, ethnicity, and place of residence according to region in the United States.

Alcohol use. Across the 12 waves, alcohol use was assessed with four items: two screening items, a quantity item, and a frequency item. The first screening item assessed whether or not participants had ever had a drink of an alcoholic beverage. Participants were instructed that a drink was considered "a can or bottle of beer, a glass of wine, a mixed drink, or a shot of liquor." The second screening item assessed whether or not participants had a drink of an alcoholic beverage since the date of the last interview. Participants who did not endorse drinking on the screening items were coded with zeroes on the quantity and frequency items.

The quantity and frequency items were open-ended and assessed alcohol use in the past month. The frequency item assessed the number of days in the last 30 days that participants had one or more drinks of an alcoholic beverage. The quantity item assessed the number of drinks participants usually consumed on the days that they drank. The quantity and frequency items were multiplied and log transformed (see "Data Analyses" section); these log transformed scores were used as the outcome for all analyses.

Procedure

The survey was conducted by the Bureau of Labor Statistics, U.S. Department of Labor, and data were made available for public use. The initial sample was randomly selected from households to be nationally representative (n = 6,748), plus a supplemental sample that over-sampled for Hispanic or Latino and African American individuals (n = 2,236). Participants had to be 12 to 16 years old as of December 31, 1996 in order to be eligible to take part in the survey. The sampling frame included all household residents in this age range. Therefore, a subset of the participants (*n* = 154) resided in the same household (e.g., siblings); given that the aims of the study were methodological in nature combined with our focus on model estimation without the inclusion of covariates, we retained these 154 participants in the analyses. Interviewers administered the 1-hour youth questionnaire via a computer-assisted personal interview system. Questions on substance use were administered with audio computer-assisted self-interview technology.

In general, retention decreased over the course of the study with 92.07% retention at Wave 2 and 80.57% retention at Wave 12, which occurred approximately 11.5 years post-baseline. Wave 9 had the lowest retention rate (77.74%). The retention rates reflect both participant dropout and intermittent non-response, where participants may have been unavailable at a given wave but completed the interview at a later follow-up. The majority (79.52%) of the participants completed nine or more of the 12 waves, and 56.48% of the participants completed all 12 waves. Comparisons were made between those who did and did not complete Wave 7 as well as those who did and did not complete Wave 12. Participants who did not complete Wave 7 (*M* = 15.50, *SD* = .55) were older than those who did complete Wave 7 (*M* = 15.43, *SD* = .53), t(672.95) = -2.46, p = .01. No other significant differences were found between those who did and did not complete Wave 7 or Wave 12 on age, sex, race, or alcohol use at baseline.

Assessment Configurations

The following three assessment configurations were tested based on our goal of making comparisons across different observation windows with implications for developmental timing: 1) *all 12 waves*; 2) *first 6 waves* (Waves 1-6); and 3) *last seven waves* (Waves 6-12). These assessment configurations allowed comparisons across different timeframes; specifically, targeted comparisons were made between the configuration with all 12 waves and the two alternative configurations. The average age of the participants at Waves 1, 6 and 12 were 15.94 years, 21.59 years, and 27.50 years, respectively.

Data Analyses

The current study uses GMM, which is a mixture modeling technique that may be used to identify latent subgroups of individuals who exhibit distinct patterns in an outcome over time, namely a mixture of subgroups with different trajectories (Arminger, Stein, & Wittenberg, 1999; Jones & Nagin, 2007; Muthén, 2004; Muthén & Shedden, 1999; Nagin, 1999, 2005; Pearson, 1894). GMM summarize the heterogeneity associated with interindividual differences in intraindividual change by identifying a set of commonly occurring patterns of longitudinal change on an outcome. GMM evaluates individuals' trajectories, which are estimated from the repeated-measurements of an outcome. GMM summarizes individuals' trajectories into multiple commonly occurring longitudinal patterns using a categorical latent class variable to denote class membership. Mixture models assume that individuals exhibit a similar pattern on the outcome within each latent class and that these patterns differ across the classes (class-specific intercepts and slopes).

Descriptive statistics were calculated in SAS Version 9.2 (SAS Institute Inc., 2008). All latent growth modeling and growth mixture modeling analyses were performed in Mplus Version 6 (Muthén & Muthén, 1998-2010). The distributional properties of the quantity item (drinks per drinking day) and frequency item (days drinking per month) were examined first. The responses on the quantity item ranged from zero to 99 drinks per drinking day. Responses greater than 30 drinks per drinking day were considered extreme (e.g., implausible) and were re-coded using ranking. Nineteen or fewer participants provided these extreme responses at any given wave, thus representing a very small portion of the data. Next, the outcome, number of drinks consumed in the last month, was created based on the product of the quantity and frequency items (see “Measures”). To reduce skew and kurtosis, a natural log transformation was applied to the outcome. The log-transformed outcome was modeled as a continuous measure with a normal distribution in all of the primary analyses.

Specifically, based on previous procedures and recommendations, four major procedural steps were taken to analyze the NLSY1997 alcohol data (Jackson & Sher, 2005; Jung & Wickrama, 2008; Ram & Grimm, 2009). First, latent growth modeling was used to determine the most appropriate set of slope parameters (e.g., linear, quadratic, and cubic slope parameters) for estimating the functional form of the outcome based on all 12 waves. Second, GMMs with two to seven classes were estimated to determine the appropriate number of latent classes using all 12 waves. Third, GMMs were estimated in accordance with the two alternative assessment configurations. To allow for direct comparisons, the number of latent classes estimated in the best fitting GMM using all 12 waves was also used in the estimation of the GMM for each alternative assessment configuration. As done in previous studies (Jackson & Sher, 2005, 2006, 2008), the indicators for comparing GMM results across the assessment configurations included: inspection of the plots for the class-specific average trajectories, class proportions based on the estimated model, average posterior probabilities, and Cohen’s kappa κ (Cohen, 1960). GMM results were also evaluated by creating contingency tables to compare the degree to which participants were classified into a comparable alcohol use subgroup (i.e., similar trajectory shape) across the assessment configurations. Fourth, latent growth models and GMMs were re-fit to determine whether or not the best fitting GMM for each assessment configuration coincided with the five-class GMM using all 12 waves (Sher et al., 2011). For all Mplus analyses, missing data were handled using maximum likelihood estimation with robust standard errors and an accelerated EM algorithm (Muthén & Muthén, 1998-2010; Schafer & Graham, 2002). Missing data were assumed to be missing at random, such that the probability of missingness is not related to the missing values of the outcome.

Results

Descriptive Analysis

Almost three-quarters (71%) of the participants reported no past-month drinking at Wave 1, and this decreased to 32% at Wave 12. The median number of drinks consumed per month was zero at Wave 1 and increased to six drinks per month by Wave 6 then remained constant through Wave 12. The correlations among the 12 log transformed outcome measures ranged from .15 to .68. As expected, correlations were generally higher for adjacent waves, and correlations decreased as the time lag increased (e.g., Waves 1 and 3 versus Waves 1 and 10).

Model-Building Using All 12 Waves

As the first step latent growth modeling was conducted to determine the functional form of the outcome over time using all 12 waves (N = 2,686). Latent growth models included up to three slope factors (linear, quadratic, cubic) as the model with a quartic slope factor did not converge. Chi-square difference tests showed significant improvements in model fit when a linear slope factor was added to the intercept-only model [$\Delta\chi^2$ (3) = 2765.23, p < .001], followed by the addition of a quadratic slope factor [$\Delta\chi^2$ (4) = 1886.46, p < .001], and a cubic slope factor [$\Delta\chi^2$ (5) = 384.82, p < .001] (Table 1). All three slope factor variances were significantly different from zero, suggesting the presence of heterogeneity in the latent factors. Model fit improved when slopes factors were added as indicated by: 1) increasing values for CFI and TLI, and 2) decreasing values for RMSEA, SRMR, AIC, BIC, and aBIC. Therefore, an intercept factor and linear, quadratic, and cubic slope factors were included in all subsequent growth mixture models.

Growth Mixture Modeling Using 12 Waves

Model specifications. A series of models was estimated to determine the number of latent classes that best represented the data. Using all 12 waves, 2- through 7-class GMMs were estimated. The factor means were class-specific, while the factor variances and covariances and residual variances of the observed variables were constrained to be equal across classes.

GMMs were estimated with 500 initial stage random sets of starting values and 50 final stage optimizations to avoid local solutions (Hipp & Bauer, 2006). When the best likelihood value was not replicated, then the random starting values and final stage optimizations were increased (e.g., 1500 and 150, respectively). Due to inadmissible or inappropriate solutions (e.g., negative variances), the residual variance of the Wave 1 measure of the outcome was fixed at zero for the six-class and seven-class GMMs using 12 waves. The variance of the cubic factor was also fixed at zero for the seven-class GMM using 12 waves.

Model selection. Based on the model selection indices (e.g, AIC, BIC, aBIC), average posterior probabilities, class-specific average trajectories, and class interpretability, the five-class GMM was considered the best fitting model using all 12 waves (Table 2). The six-class and seven-class GMMs were not considered viable because both of those models had a class that included less than 1% of the sample (Jung & Wickrama, 2008). The values of the estimated parameters for the five-class GMM are shown in Table 3. Of note, the within-class variance of the intercept factor was not significantly different from zero. In contrast, within-class variances of the three slope factors (linear, quadratic, and cubic) were significantly different from zero.

Thus, there was a high degree of variability in the shapes of the individual-level trajectories, even among participants who belonged to the same latent class.

Class descriptions. Individuals in Class 1 (*Normative Class*; *n* = 1916, 71.33% of the sample), the largest class, generally did not drink at Wave 1 with steadily increasing use through Wave 6, followed by a period of relatively stable use (Figure 1, solid lines). Individuals in Class 2 (*Low-Increasing Class*; *n* = 227, 8.45%) consumed a relatively low number of drinks per month at Wave 1 and consumption gradually increased before leveling off. Individuals in Class 3 (*Low-Steady Class*; *n* = 241, 8.97%) consumed a moderate number of drinks per month at Wave 1 and consumption increased before leveling off. Individuals in Class 4 (*High-Slowly Decreasing Class*; *n* = 206, 7.67%) consumed a high number of drinks per month at Wave 1 and consumption slowly decreased over time. Individuals in Class 5 (*Extreme-Sharply Decreasing Class*; *n* = 96, 3.57%) consumed an extremely high number of drinks per month at Wave 1 and decreased over time. As shown in Table 4, the *High-Slowly Decreasing Class* and *Extreme-Sharply Decreasing Class* consisted of mostly men (57.77% and 63.54%, respectively) as well as the largest proportions of White participants (89.22% and 87.50%, respectively).

Comparisons across the Assessment Configurations

Results from the five-class GMM using 12 waves were compared to the five-class GMM results for each of the two alternative assessment configurations. Model specifications were imposed (e.g., fix variance of intercept factor at zero) as necessary to obtain model convergence. Comparisons were based on estimated class trajectories, class proportions, average posterior probabilities, parameter estimates, Cohen's kappa κ, and contingency tables.

Estimated class trajectories. The class-specific average trajectories were relatively similar when comparing the 12 wave configuration to the configuration with the first 6 waves

(Figure 1). Minor differences were observed in the shapes of the trajectories, particularly for Class 5. The 12 wave solution exhibited lower mean levels of drinks per months at the earlier waves. When using only the last seven waves, the class-specific trajectories no longer showed convergence with those found in the 12 wave solution, but instead showed trajectories with markedly different shapes (Figure 1).

Class proportions. Based on most likely class membership, class proportions were very similar when comparing results using 12 waves to results using the first 6 waves. Large differences in class proportions were observed for the comparison between all 12 waves and the last seven waves (.35, .18, .25, .18, and .04 for Classes 1 to 5, respectively).

Average posterior probabilities. Average posterior probabilities for the five latent classes were similar between the 12-wave configuration and the first 6 waves ($\geq$ 0.950 on diagonals and $\leq$ 0.037 on off-diagonals), but differed for the last seven wave configuration. The average posterior probabilities from the last seven wave solution exhibited less precision in class assignments ($\geq$ 0.820 on diagonal and $\leq$ 0.103 on off-diagonal).

Parameter estimates. The estimated means of the class-specific intercept factors were relatively similar across the three assessment configurations. There was greater variability among the assessment configurations for the estimated means of the three slope factors, which was reflected in the class-specific average trajectories. There was substantial variability in the estimated variance and covariance parameters when comparing the 12-wave configuration to the configurations with the first 6 waves and the last seven waves.

Cohen's kappa. Cohen's kappa, κ, assessed the degree of similarity in participants' class assignments across the assessment configurations (Cohen, 1960). According to Landis and Koch (1977), κ indicates slight agreement (0.00 to 0.20), fair agreement (0.21 to 0.40), moderate agreement (0.41 to 0.60), substantial agreement (0.61 to 0.80), or almost perfect agreement (0.81 to 1.00). Based on participants' most likely class membership, agreement in class assignment was high between the 12 wave configuration and first 6 waves (κ's = .69). There was low agreement between the 12-wave configuration and last seven wave configuration (κ = .10).

Contingency tables. Contingency tables examined the extent to which participants were assigned to a similar drinking class in the 12 wave configuration compared to the four alternative configurations based on participants' most likely class. Table 5 shows the degree of misclassification across the assessment configurations. Using the second row of the table as an example, 50.66% of the participants assigned to Class 2 using 12 waves were also assigned to Class 2 using the first 6 waves, while the remaining participants were assigned to Class 1 using the first 6 waves. When comparing the 12 wave configuration to the first 6 waves, approximately 14% of the total sample was misclassified, and misclassification was almost always to the adjacent class that had lower drinking at Wave 1. When comparing the 12 wave and last seven wave configurations, 62% of the total sample was misclassified with participants being assigned to adjacent and non-adjacent classes, thus representing a substantial degree of misclassification among the five classes. For example, only 42.82% of the participants classified into Class 1 using 12 waves were also classified into Class 1 using the last seven waves. This misclassification likely reflects substantial variation in the individual-level trajectories when trajectories were estimated over 12 waves (spanning adolescence into young adulthood) versus the last seven waves (spanning young adulthood only).

Model re-fitting. After fitting a series of latent growth models for each alternative assessment configuration, the best fitting trajectory shape for each of the four alternative configurations was found to include the same set of latent factors as that used for all 12 waves (i.e., intercept plus linear, quadratic, and cubic slopes). Similarly, after evaluating the GMMs, the five-class model remained the best fitting model for each of the two alternative configurations.

Discussion

The current study examined the effects of observation window (i.e., timing of survey assessments) in the context of GMM using an empirical application with prospective longitudinal data on alcohol use. Key findings were that: 1) the five-class GMM results exhibited a low degree of discrepancy when comparing all 12 waves to the configuration with the first 6 waves; and 2) the GMM results for the 12 wave and last seven wave configurations exhibited a large discrepancy based on all six indicators of class agreement. These findings suggest that alterations in the timing of the initial wave (i.e., all 12 waves where participants' mean age was 15.94 years at the initial assessment versus the last seven waves where participants' mean age was 21.59 years at the initial assessment) had the most substantial impact on the comparability of the latent classes, thus underscoring the need to carefully consider the most appropriate observation window in relation to the study aims and the developmental window of the participants.

The largest discrepancy in subgroup assignment (62% misclassification) was observed when comparing the five-class GMM results from the 12 wave configuration to those from the last seven wave configuration, and a portion of the misclassified participants belonged to non-adjacent classes (i.e., less similar class trajectories) rather than to adjacent classes (i.e., more similar class trajectories). This finding is not surprising given that these two configurations produced substantially different parameter estimates in the empirical application. Our findings are partially consistent with Sher et al (2011), who also demonstrated a relatively high degree of discrepancy among classes observed in a wave 1-8 and 5-8 configuration. For instance, while most participants identified as "chronic" or "nonbinger" in waves 5-8 were also identified in that manner in waves 1-8 (83.6% and 83.4%, respectively), there was little overlap between the "increase" and "decrease" classes (34.2% and 27.0% overlap, respectively). It is important to consider that the marked changes in development that occur between adolescence and young adulthood may have contributed to the high degree of misclassification in the current work. In fact, it may be that studies looking at a similar timespan in mid-life would not observe such marked differences, under the assumption that developmental changes in alcohol use are more constant during mid-life.

In the current study, participants' average ages ranged from 16 to 28 years across the 12 waves. For the last seven waves, participants' average ages ranged from 22 to 28 years, which is associated with the post-college years, career identification, and early adulthood (Brown et al., 2008). Observed differences among the assessment configurations with respect to individuals' class assignment is consistent with findings from Jackson and Sher (2006) and Tan and colleagues (2011). The high degree of misclassification suggests that individuals exhibited markedly different drinking patterns at different developmental windows across their lifetime, especially during and subsequent to emerging adulthood when heavy alcohol use may be incompatible with newly acquired social roles (e.g., marriage, parenthood) (Littlefield & Sher, 2010; Sher, Grekin, & Williams, 2005). Future research is needed to determine the extent to which individual-level drinking patterns vary within a shorter timeframe that is developmentally or theoretically meaningful (e.g., from semester to semester among college students). Furthermore, researchers may be interested in understanding a known developmental shift or inflection point that would likely be reflected in the class-specific trajectories, such as changes in alcohol use upon transitioning out of college (Sher et al., 2011). In this situation, the spacing of the assessments would need to adequately capture the shape of the individual-level trajectories both before and after the shift.

Several observations are warranted with respect to the patterns of the class-specific average trajectories and consistency in individuals' class assignment from the empirical application. In all three assessment configurations, the class-specific average trajectories exhibited divergence at the first wave and relative convergence by the last wave, similar to a reverse "fanning effect." Other empirical studies have also found similar class-specific average trajectories despite alterations to the measurement timing (Jackson & Sher, 2006; Sher et al., 2011). Notably, the current study did not identify a "cat's cradle" with respect to the class-specific average trajectories whereby four prototypical classes (stable low, increasing, decreasing, stable high) are identified, and this cat's cradle pattern of class-specific average trajectories was a major focus of Sher et al. (2011).

Study Strengths and Limitations

The importance of proper selection of the timing and length of the observation window to adequately capture longitudinal changes in a given phenomenon has been well recognized (Collins & Graham, 2002; Widaman, 1991; Wohlwill, 1973). The current study highlighted the effects that altering the observation window and timing of the assessments can have on GMM results, particularly in the case of all 12 waves versus the last seven waves, thus underscoring how measurement timing can affect subgroup identification. The current study contributes to the small body of research focusing on how observation window affects the performance of GMM (Eggleston et al., 2004; Jackson & Sher, 2006; Sher et al., 2011; Tan et al., 2011). The subsample used in the current analyses was drawn from a nationally representative sample of adolescents, thus extending the empirical applications on alcohol use to non-college samples.

Despite notable study strengths, several limitations should be acknowledged. First, the outcome here was modeled as a continuous measure (i.e., drinks per month). However, other distributions may be suitable for these data, including modeling the outcome as count data with a zero-inflated Poisson or negative binomial distribution (Reinecke & Seddig, 2011). Second, the current study was limited to comparisons among three assessment configurations, even though the data lend itself to the potential for examining alternative configurations (see Fairlie, 2012 that considers two-year intervals and uneven intervals). Third, we were unable to account for the clustering of siblings within families because it was limited to 5.7% of the sample, thus presenting issues if accounting for clustering where the majority of the clusters (i.e., families) would have only one observation. Despite these limitations, the current findings provide a useful contribution to the small body of literature examining measurement timing and the performance of GMM, given the large sample size and the ability to directly examine the effects of shifting the observation window.

Future Research

The current findings suggest several promising avenues for future research. First, in order to determine the generalizability of the current findings, empirical work should examine the variations in the observation window using databases with a variety of characteristics: (1) phenomena with different patterns of change over time (e.g., heavy episodic drinking, marijuana use), (2) samples with different characteristics (e.g., average age at Wave 1), (3) studies with different research designs (e.g., time lag between assessments, total length of study period), and (4) trajectories over different developmental periods. Notably, regarding this final point, it may be beneficial to examine how variations in measurement timing may impact our understanding of alcohol use trajectories in later life among older adults (Kuerbis, Sacco, Blazer, & Moore, 2014).

The notion of measurement timing is also pertinent to clinical studies with respect to being able to capture when change occurred. For example, studies may fail to identify reductions in alcohol use if intervention effects dissipate quickly and/or the assessments are too far apart.

Future empirical work should determine which factors are associated with misclassification across different assessment configurations. Jackson and Sher (2006) found participants' probability of class membership and amount of missing data were associated with misclassification. Other relevant predictors of misclassification may also be identified. For instance, individuals who drank heavily, but infrequently, or individuals who reported marked fluctuations in their drinking over time may have been more likely to be misclassified. Similarly, individuals who were classified into different classes for the 12 wave versus last seven wave configurations may have exhibited markedly different patterns of alcohol use during these two timeframes. It may be useful to apply piecewise GMMs in order to examine this type of misclassification, since piecewise GMMs would allow for different trajectory shapes, for example, over the first 6 and last seven waves (Li, Duncan, & Hops, 2001). A piecewise GMM using all 12 waves could then be compared to the first 6 and last seven wave configurations to determine whether the piecewise GMM reduced the proportion of participants who were misclassified. In this way, piecewise modeling could be used to test a viable alternative explanation of the current findings; it may be that consideration of a turning point that denotes the next stage of a growth process in the 12-wave model may partially account for what appears to be misclassification across assessment configurations. Furthermore, applications of latent transition analyses could also be used to better understand transitions among class membership, for instance, using the first 6 waves and the last seven waves (Davis, Ingram, Merrin, & Espelage (2018); Muthén & Muthén, 2000). This or related approaches could provide a unique examination of various factors that predict transitions among the classes (e.g., movement to a higher or lower use class from adolescence to young adulthood).

Conclusions

The current findings showed that differences in the observation window had a low impact on the performance of GMM when the first wave remained constant (i.e., the first assessment was consistent for the configurations with all 12 waves and the first 6 waves). In contrast, altering the timing of the first wave included in the analysis (i.e., all 12 waves versus the last seven waves) greatly impacted the comparability of the latent class solutions, thus underscoring the need for careful selection of the observation window, especially during developmental periods where an individual’s behavior may shift considerably over time (e.g., adolescence versus young adulthood). Design decisions about the timing of the assessments are particularly challenging in the context of GMM, because the latent subgroups exhibit different trends in the outcome. The time-bounded nature of mixture modeling solutions (i.e., a selected developmental window within the course of a phenomenon) suggests that the latent subgroups should not be interpreted as representing subgroups that are present in the population. Instead, subgroup identification is sensitive to variations in research design, which include, but may not be limited to, the observation window and corresponding developmental period under investigation. This research has important implications for the utility of subgroup analyses and the associated developmental trajectories of alcohol use into young adulthood. Findings may apply more broadly to identifying trajectory subgroups for other health and risk behaviors. Investigators should carefully consider the timing of their survey assessments and the circumstances under which the average trajectories of each subgroup may be altered by the observation window and timing of the assessments.

References

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika, 64*, 475-494.

Brown, S. A., McGue, M., Maggs, J., Schulenberg, J., Hingson, R., Swartzwelder, S., … Murphy, S. (2008). A developmental perspective on alcohol and youths 16 to 20 years of age. *Pediatrics, 121 Suppl. 4,* S290-S310.

Chassin, L., Flora, D. B., & King, K. M. (2004). Trajectories of alcohol and drug use and dependence from adolescence to adulthood: The effects of familial alcoholism and personality. *Journal of Abnormal Psychology, 113*, 483-498.

Chassin, L., Pitts, S. C., & Prost, J. (2002). Binge drinking trajectories from adolescence to emerging adulthood in a high-risk sample: Predictors and substance abuse outcomes. *Journal of Consulting and Clinical Psychology, 70,* 67-78.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational Psychological Measurement, 20,* 37-46.

Cohen, P. (1991). A source of bias in longitudinal investigations of change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 18-25). Washington, DC: American Psychological Association.

Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology, 70,* 976-985.

Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: temporal design considerations. *Drug and Alcohol Dependence, 68,* S85-S96.

Davis, J. P., Ingram, K. M., Merrin, G. J, & Espelage, D. L. (2018). Exposure to parental and community violence and the relationship to bullying perpetration and victimization among early adolescents: A parallel process growth mixture latent transition analysis. *Scandinavian Journal of Psychology*. doi: 10.1111/sjop.12493. [Epub ahead of print].

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

Fairlie, A. M. (2012). *Measurement timing in growth mixture modeling of alcohol trajectories*. (Doctoral dissertation, University of Rhode Island).

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 243-259). Washington, DC: American Psychological Association.

Goudriaan, A. E., Grekin, E. R., & Sher, K. J. (2007). Decision making and binge drinking: A longitudinal study. *Alcoholism: Clinical and Experimental Research, 31(6),* 928-938.

Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., Dufour, M. C., & Pickering, R. P. (2006). The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. *Alcohol Research & Health, 29(2),* 79-91.

Greenbaum, P. E., & Dedrick, R. F. (2007). Changes in use of alcohol, marijuana, and services by adolescents with serious emotional disturbance: A parallel-process growth mixture model. *Journal of Emotional and Behavioral Disorders, 15(1),* 21-32.

Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C.-P., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of Consulting and Clinical Psychology, 73,* 229-238.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11,* 36-53.

Hix-Small, H., Duncan, T. E., Duncan, S. C., & Okut, H. (2004). A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. *Journal of Psychopathology and Behavioral Assessment, 26(4),* 255-270.

Jackson, K. M., & Sher, K. J. (2005). Similarities and differences of longitudinal phenotypes across alternate indices of alcohol involvement: A methodologic comparison of trajectory approaches. *Psychology of Addictive Behaviors, 19,* 339-351.

Jackson, K. M., & Sher, K. J. (2006). Comparison of longitudinal phenotypes based on number and timing of assessments: A systematic comparison of trajectory approaches II. *Psychology of Addictive Behaviors, 20,* 373-384.

Jackson, K. M., & Sher, K. J. (2008). Comparison of longitudinal phenotypes based on alternate heavy drinking cut scores: A systematic comparison of trajectory approaches III. *Psychology of Addictive Behaviors, 22,* 198-209.

Jones, B. L., & Nagin, D. S. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods & Research, 35(4),* 542-571.

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

Kuerbis, A., Sacco, P., Blazer, D. G., & Moore, A. A. (2014). Substance abuse among older adults. *Clinics in Geriatric Medicine*, 30(3), 629-654.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33(1),* 159-174.

Lerner, R. M., Schwartz, S. J., & Phelps, E. (2009). Problematics of time and timing in the longitudinal study of human development: Theoretical and methodological issues. *Human Development, 52,* 44-68.

Li, F., Duncan, T. E., Duncan, S. C., & Acock, A. (2001). Latent growth modeling of longitudinal data: A finite growth mixture modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal, 8,* 493–530.

Li, F., Duncan, T. E., & Hops, H. (2001). Examining developmental trajectories in adolescent alcohol use using piecewise growth mixture modeling analysis. *Journal of Studies on Alcohol, 62*, 199-210.

Littlefield, A. K., & Sher, K. J. (2010). Alcohol use disorders in young adulthood. In J. E. Grant, & M. N. Potenza (Eds.), *Young adult mental health* (pp. 292-310). New York, NY: Oxford University Press.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882-891.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55(2),* 463-469.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4,* 139-157.

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

National Longitudinal Survey of Youth (NLSY). (1997). Accessed at https://www.bls.gov/nls/nlsdoc.htm

Nesselroade, J. R., & Ghisletta, P. (2003). Structuring and measuring change over the life span. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 317-337). Boston: Kluwer Academic Publishers.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A, 185,* 71–110.

Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying difference in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33,* 565-576.

Reinecke, J., & Seddig, D. (2011). Growth mixture models in longitudinal research. *Advances in Statistical Analysis, 95,* 415-434.

SAS Institute Inc. (2008). *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7(2),* 147-177.

Schulenberg, J. E., Johnston, L. D., O'Malley, P. M., Bachman, J. G., Miech, R. A. & Patrick, M. E. (2018). Monitoring the Future national survey results on drug use, 1975–2017:

Volume II, College students and adults ages 19–55. Ann Arbor: Institute for Social Research, The University of Michigan. Available at http://monitoringthefuture.org/pubs.html#monographs.

Sher, K. J., Grekin, E. R., & Williams, N. A. (2005). The development of alcohol use disorders. *Annual Review of Clinical Psychology, 1(1),* 493-523.

Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology, 120(2),* 322-335.

Smith, D., & Walls, T. A. (2016). Multiple time-scale models in sport and exercise science. *Measurement in Physical Education and Exercise Science, 20*, 1-15.

Tan, X., Dierker, L., Rose, J., Li, R., & The Tobacco Etiology Research Network. (2011). How spacing of data collection may impact estimates of substance use trajectories. *Substance Use and Misuse, 46,* 758-768.

Walls, T. A., Barta, W. D., Stawski, R. S., Collyer, C. E., & Hofer, S. M (2012). Time-scale-dependent longitudinal designs. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 46-64). New York, NY: Guilford Press.

Widaman, K. F. (1991). Qualitative transitions amid quantitative development: A challenge for measuring and representing change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 204-217). Washington, DC: American Psychological Association.

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49,* 587-602.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.

Table 1

Model Selection Indices to Compare Latent Growth Models that Vary by Latent Growth Parameters

Model parameters	Chi-square	df	CFI	TLI	AIC	BIC	aBIC	RMSEA	SRMR	Significant factor means and variances?
Intercept only	5330.26	76	0.58	0.64	95231.2	95313.7	95269.2	0.16	0.2	Yes, both
Intercept, linear slope	2565.03	73	0.8	0.82	92471.9	92572.2	92518.2	0.11	0.1	Yes, all
Intercept, linear and quadratic slopes	678.57	69	0.95	0.95	90593.5	90717.3	90650.6	0.06	0.04	Yes, all
Intercept; linear, quadratic, and cubic slopes	293.75	64	0.98	0.98	90218.7	90371.9	90289.3	0.04	0.03	Yes, all

Note. All models included 12 waves of data. df = degrees of freedom. CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. AIC = Akaike information criterion. BIC = Bayesian information criterion. aBIC = sample size adjusted BIC. RMSEA = root mean square error of approximation. SRMR = standardized root mean square residual.

Table 2

Model Selection Indices to Compare GMMs that Vary in the number of Classes

Classes	AIC	BIC	aBIC	Entropy	VLMR LRT (p-value)	LMR adj. LRT (p-value)
1 class	90218.65	90371.94	90289.33	n/a	n/a	n/a
2 classes	88299.71	88482.48	88383.98	0.953	$p < .0001$	$p < .0001$
3 classes	87106.27	87318.52	87204.14	0.968	$p < .01$	$p < .01$
4 classes	86154.61	86396.34	86266.07	0.980	0.04	0.04
5 classes	85625.39	85896.60	85750.44	0.979	0.10	0.10
6 classes	84792.23	85087.02	84928.15	0.987	0.04	0.04
7 classes	84604.49	84905.17	84743.13	0.991	0.17	0.18

Note. All models included 12 waves of data. In the 6- and 7-class GMMs, the residual variance of the Wave 1 measure of the outcome was fixed at zero. In addition, for the 7-class GMM, the variance of the cubic factor was fixed at zero. AIC = Akaike information criterion. BIC = Bayesian information criterion. aBIC = sample size adjusted BIC. VLMR LRT = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. LMR adj. LRT = Lo-Mendell-Rubin adjusted Likelihood Ratio Test.

Table 3

Estimated Parameters for the 5-class GMM using all 12 Waves

Latent factor	*b*	*SE*	95% CI for *b*	β
Class 1: Estimated mean parameters				
Intercept	0.00	0.00	[0.00, 0.01]	0.03
Linear	5.45***	0.19	[5.08, 5.83]	0.82
Quadratic	-5.40***	0.42	[-6.23, -4.57]	-0.41
Cubic	1.70***	0.25	[1.22, 2.18]	0.23
Class 2: Estimated mean parameters				
Intercept	0.96	0.03	[0.91, 1.01]	9.68
Linear	4.46***	0.62	[3.25, 5.67]	0.67
Quadratic	-4.67***	1.41	[-7.43, -1.91]	-0.35
Cubic	1.52***	0.81	[-0.08, 3.12]	0.21
Class 3: Estimated mean parameters				
Intercept	2.02***	0.03	[1.96, 2.09]	20.37
Linear	1.90**	0.65	[0.63, 3.17]	0.29
Quadratic	-1.56	1.47	[-4.44, 1.33]	-0.12
Cubic	0.04	0.87	[-1.66, 1.74]	0.01
Class 4: Estimated mean parameters				
Intercept	3.27***	0.05	[3.18, 3.36]	32.89
Linear	-2.45***	0.73	[-3.88, -1.03]	-0.37
Quadratic	3.68*	1.56	[0.63, 6.74]	0.28
Cubic	-1.97*	0.88	[-3.69, -0.24]	-0.27
Class 5: Estimated mean parameters				
Intercept	4.79***	0.09	[4.61, 4.97]	48.21
Linear	-9.91***	1.19	[-12.24, -7.59]	-1.49
Quadratic	14.65***	2.54	[9.67, 19.62]	1.10
Cubic	-6.95***	1.46	[-9.82, -4.08]	-0.95
Estimated variance parameters				
Intercept	0.01	0.02	[-0.03, 0.04]	1.00
Linear	44.40***	2.31	[39.87, 48.93]	1.00
Quadratic	177.44***	12.03	[153.87, 201.01]	1.00
Cubic	53.25***	4.11	[45.19, 61.31]	1.00
Estimated covariance parameters				
Linear-Intercept	-0.01	0.10	[-0.21, 0.19]	-0.02
Quadratic-Intercept	0.02	0.17	[-0.31, 0.34]	0.01
Quadratic-Linear	-83.01***	5.13	[-93.07, -72.96]	-0.94
Cubic-Intercept	-0.01	0.08	[-0.17, 0.15]	-0.01
Cubic-Linear	41.43***	2.90	[35.75, 47.11]	0.85
Cubic-Quadratic	-95.17***	6.97	[-108.83, -81.51]	-0.98

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4

Demographic Information by Class for the 5-Class GMM using 12 Waves

Demographics	Class 1, Normative % (n)	Class 2, Low-increasing % (n)	Class 3, Low-steady % (n)	Class 4, High-slowly decreasing % (n)	Class 5, Extreme-sharply decreasing % (n)
Sex					
Male	50.05 (959)	44.49 (101)	48.96 (118)	57.77 (119)	63.54 (61)
Female	49.95 (957)	55.51 (126)	51.04 (123)	42.23 (87)	36.46 (35)
Race					
White	67.18 (1275)	78.76 (178)	81.17 (194)	89.22 (182)	87.50 (84)
Black	19.60 (372)	13.72 (31)	7.95 (19)	4.90 (10)	5.21 (5)
Amer. Indian	0.63 (12)	0.0 (0)	0.84 (2)	0.49 (1)	1.04 (1)
Asian	3.16 (60)	2.21 (5)	0.42 (1)	0.49 (1)	1.04 (1)
Other	9.43 (179)	5.31 (12)	9.62 (23)	4.90 (10)	5.21 (5)
Hispanic					
No	85.33 (1629)	88.55 (201)	83.33 (200)	91.22 (187)	86.46 (83)
Yes	14.67 (280)	11.45 (26)	16.67 (40)	8.78 (18)	13.54 (13)
Region					
Northeast	18.63 (357)	20.26 (46)	17.84 (43)	21.84 (45)	20.83 (20)
North Central	26.62 (510)	24.67 (56)	24.90 (60)	25.24 (52)	29.17 (28)
South	35.23 (675)	37.00 (84)	30.29 (73)	27.67 (57)	32.29 (31)
West	19.52 (374)	18.06 (41)	26.97 (65)	25.24 (52)	17.71 (17)
Age at Wave 1					
Mean	15.91	15.93	16.00	16.05	16.17

SD	0.57	0.56	0.55	0.56	0.59

Note. Class 1 (Normative, *n* = 1916), Class 2 (Low-Increasing, *n* = 227), Class 3 (Low-Steady, *n* = 241), Class 4 (High-Slowly Decreasing, *n* = 206), and Class 5 (Extreme-Sharply Decreasing, *n* = 96). Class membership was based on assignment to the participants' most likely class. Amer. Indian = American Indian.

Table 5

Percentage of Participants Assigned to Classes 1 to 5 in the Alternative Assessment Configurations (Column) Based on Class Membership using the 12-Wave Solution (Row)

Most likely class using 12-waves	Most likely class in alternative configurations					
	Class 1	Class 2	Class 3	Class 4	Class 5	Marginal *n*
	First 6 waves, χ^2 (16, N = 2685) = 5872.04, p < .0001					
Class 1	100	0	0	0	0	1915
Class 2	49.34	50.66	0	0	0	227
Class 3	0	58.09	41.91	0	0	241
Class 4	0.49	0	42.23	57.28	0	206
Class 5	0	0	0	27.08	72.92	96
	Last seven waves, χ^2 (16, N = 2505) = 181.03, p < .0001					
Class 1	42.82	17.04	23.61	14.20	2.34	1796
Class 2	23.70	21.80	30.33	20.38	3.79	211
Class 3	18.55	12.22	34.84	27.60	6.79	221
Class 4	20.86	11.76	26.74	32.09	8.56	187
Class 5	24.44	10.00	24.44	31.11	10.00	90

Note. Percentages are row percentages, such that the values in each row sum to 100%.

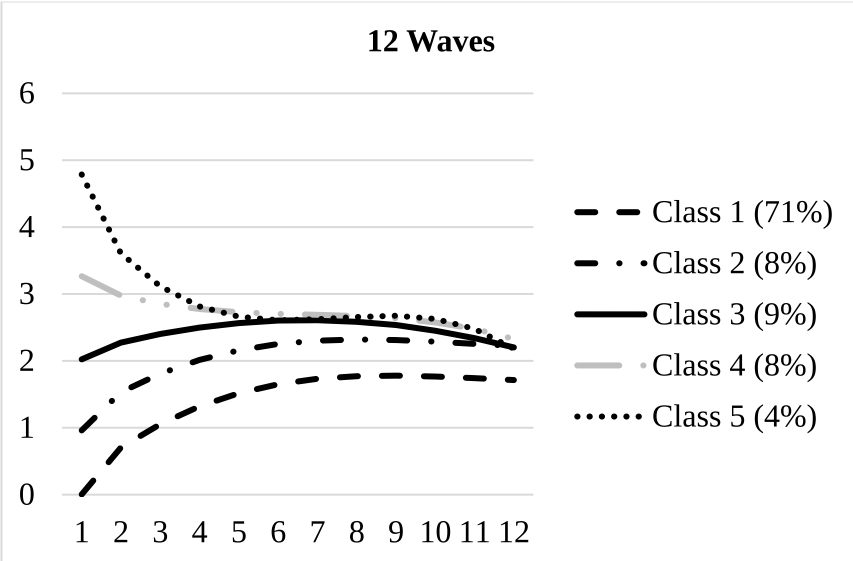

Figure 1, Panel A

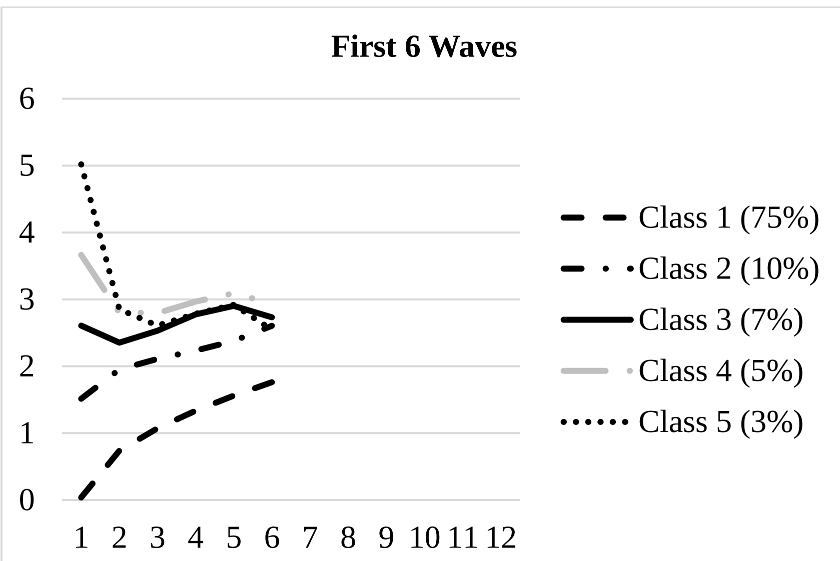

Figure 1, Panel B

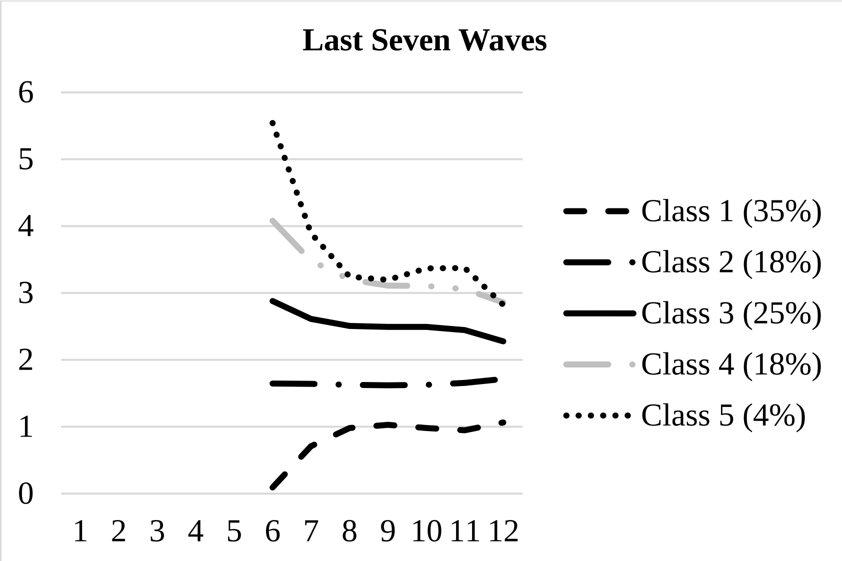

Figure 1, Panel C

Figure 1. Comparing class-specific average trajectories from the five-class GMMs using 12 waves (*N* = 2686; Panel A), the first 6 waves (*N* = 2685; Panel B), and the last seven waves (*N* = 2505; Panel C). Estimated means are plotted on the y-axis. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month, 3 = 19.09 drinks per month, 4 = 53.60 drinks per month, 5 = 147.41 drinks per month). To aid interpretation, the average age of the participants at Waves 1, 6 and 12 were 15.94 years, 21.59 years, and 27.50 years, respectively.